GOODNESS OF FIT OF PREDICTION MODELS AND TWO STEP PREDICTION

A thesis submitted to the University of Nottingham for the Degree of Doctor of Philosophy

G. JANACEK

May 1973

GOODNESS OF FIT OF PREDICTION MODELS

Abstract

Given a second order stationary time series X_t it can be shown that there exists an optimum linear predictor of X_k, say X_k^*, which is constructed from $\{X_t, t=0,-1,-2\ldots\}$ the mean square error of prediction being given by

$$e_k = E\left[|X_k - X_k^*|^2\right] .$$

In some cases however a series can be considered to have started at a point in the past and an attempt is made to see how well the optimum linear form of the predictor behaves in this case.

Using the fundamental result due to Kolmogorov relating the prediction error e_1 to the power spectrum $f(w)$

$$e_1 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(w)dw \right\}$$

estimates of e_1 are constructed using the estimated periodogram and power spectrum estimates. As is argued in some detail the quantity e_1 is a natural one to look at when considering prediction and estimation problems and the estimates obtained are non-parametric.

The characteristic functions of these estimates are obtained and it is shown that asymptotically they have distributions which are approximately normal. The rate of convergence to normality is also investigated.

A previous author has used a similar estimate as the basis of a test of white noise and the published results are extended and in the light of the simulation results obtained some modifications are suggested.

To increase the value of the estimates e_1 their small sample distribution is approximated and extensive tables of percentage points are provided. Using these approximations one can construct a more powerful and versatile test for white noice and simulation results confirm that the theoretical results work well.

The same approximation technique is used to derive the small sample distribution of some new estimates of the coefficients in the model generating $\{X_t\}$. These estimates are also based on the power spectrum. While it is shown small sample theory is limited in this situation the asymptotic results are very interesting and useful.

Several suggestions are made as to further fields of investigation in both the univariate and multivariate cases.

PAGE NUMBERING AS ORIGINAL

ACKNOWLEDGMENTS

My sincere thanks are due to a number of people who assisted in the production of this thesis. In particular I owe a great deal to my supervisor Prof. C. Granger whose kindness and guidence has been invaluable.

I would like to acknowledge the help of Mr. A. Jack who helped in the writing of the Fast Fourier Transform routines and to the members of the Mathematics department who discussed my problems with me. My wife Pamela has been an unfailing source of great encouragement and help.

This research was supported in part by the Science Research Council.

PREFACE

A random or stochastic process may be defined formally as set of random variables $\{X(t)\}$ indexed by the parameter t, which is an element of some index set T.

In many physically useful processes the parameter t represents time; this is not always so however as may be seen by considering $\{X(t)\}$ as a barometric pressure reading and t the distance east from a fixed reference point.

Thus if we observe, for values of some parameter t, a quantity $\{X(t)\}$ subject to random fluctuation, the "random function" X(t) obeys probabalistic rather than mathematical laws and we are observing a random process. Instances of such processes are numerous.

e.g. (i) the observation of a seismograph trace

(ii) "Shot noise" across a vacuum diode

(iii) Woolworth's stock prices 1946-1950

(iv) Monthly figures for the production of pig iron in the U.S.A. 1877-1956

In what follows we shall be mainly concerned with second order stationary processed. These are processes $\{X(t)\}$ whose covariance structure is invariant under a shift of origin in the scale of measurement of t. This is not as severe a restriction as might be thought at first glance since one can often remove non-stationary trends and sometimes a non-stationary process can be transformed into a stationary form.

Clearly the prediction of such processes is of considerable interest and there is an extensive literature dealing with the prediction of stochastic processes. In addition to the pure prediction problem there are processes where one is interested in the probabalistic structure of the process.

If one has a vector stochastic process $\{ X_k(t); t = 1, \ldots n \}$ then the determination of the structure of the process and the relationships between the component series is of great interest. For example $\{ X_1(t) \}$ might be disposable personal income and $\{ X_2(t) \}$ personal consumption expenditure (Fishman (1969)).

At the present time a good deal of effort is being spent on the development of forecasting methods and the construction of forecasts. It would seem essential that attempts be made to assess the value of individual forecasts and to provide usual diagnotic tools which could enable one to compare forecasts with some objective criterion. This is the main raison d'etre for what follows.

Chapter 1 outlines the basic theory required for later use. In particular the concept of the spectrum is considered in detail. This is essentially the harmonic decomposition of the variance of the stochastic process and has a very natural connection with the optimum linear predictor of the process. Indeed one can consider the spectrum as being at the heart of the prediction problem.

In chapter 2 we look at a particular optimum predictor and consider its accuracy in a finite sample case. The minimum prediction error is also discussed in some detail.

Chapter 3 outlines the estimation of the minimum mean square error and considers the distribution of spectral estimates.

In chapter 4 we look in detail at a proposed test for "white noise", that is a test to determine whether a process is composed only of independent random variables.

Chapter 5 contains work on the approximation of the minimum mean square error and its distribution. The theoretically derived results are compared with those of some simulation experiments, the agreement is generally good. Of necessity this chapter contains an outline of approximation methods.

Chapter 6 develops the methods of Chapter 4 further and removes most of the rather more limiting restrictions required. An exact method is developed for hypothesis testing but a suitable approximation isseen to work well.

Chapter 7 describes some alternatives to the methods developed and demonstrates the drawbacks of using traditional spectral estimates. The role of the band width of the smoothing window is emphasised.

In Chapter 8 we look at the factored spectrum and show how one can introduce some distribution theory into forecasting methods proposed by other authors. We show that one can estimate moving average and autoregressive parameter from the factored spectrum and how these estimates can be used.

The last chapter provides some practicle work and discusses extension of the methodology developed. We also consider briefly the multivariate problems. Suggestions for the extension of the methods presented in previous chapters are given in some detail.

CHAPTER I

Introduction

A random or stochastic process may be defined as a set of random variables $\{X_t\}$ indexed by the parameter t which is an element of some index set T. Usually the ordering is in time, thus giving rise to the term Time Series, however there are many cases where the ordering is spatial. The ordering of the random variables is crucial as in essence one is interested in the relationship between the random variables over time.

More formally we define a stochastic process as follows:

Consider a probability space Ω on which is defined a σ algebra $\mathcal{F}$ of sets S. A real valued function X(w) $(w \in \Omega)$ defined on Ω is called a stochastic variable if it is measurable with respect to $\mathcal{F}$ i.e. $\{X(w) < a\} \in \mathcal{F}$.

Let T = $\{t\}$ be a set of indices, then if for each the function $X_t(w)$ is a stochastic variable (considered as a function of w), $X_t(w)$ is a stochastic process when t,w vary over T x Ω

The set T can be finite, countable or uncountable. We shall discuss sets of the form

(i) $T = 0, \pm 1, \pm 2, \ldots\ldots$

(ii) T is the real line

or their subsets.

In common with the accepted practice we shall often suppress the argument w.

Two interpretations of $X_t(w)$ are possible, $\{ X_t(w) \}$ can be regarded as a set of time functions indexed on w in which case each time function is a sample value or "realisation" of the process. From the alternative viewpoint $X_t(w)$ is a collection of random variables indexed by $\{t\}$.

The question arises whether it is possible to determine a probability measure on Ω in such a way that the joint distribution of $X_{t_1}(w)$, $X_{t_2}(w) \ldots\ldots X_{t_n}(w)$ will be the same as that of X_{t_1} $X_{t_2} \ldots\ldots$ X_{t_n} Where X_{t_j} is defined as a function on Ω. That such a measure can be determined has been shown by Kolmogorov (1950), the condition being that the initial joint distributions of the $\{ X_t \}$ should be compatible. That is the marginal distribution of any subset of $X_{t_1} \ldots\ldots$ X_{t_n} should be the same as the prescribed distribution of this subset. For further details see Hannan (1960), Rosenblatt (1962) or Anderson (1971).

Dropping the w we see that we now have in effect a multivariate problem since $\{X_t\}$ can be regarded as a random vector. Thus far we have made no assumptions about the relationships between the $\{ X_t \}$, these being determined by the phenomenon generating the series. For most purposes it is convenient to split stochastic processes into two major clauses, stationary processes which do not exhibit major change over a change in origin of the time scale and evolutionary processes.

To be more precise we define a strictly stationary process $\{ X_t \}$ as a process for which, for all n, and h, $t_1 \ldots$ tn (these latter being understood to belong to the index set T) the distributions of

$$X_{t_1}, X_{t_2} \ldots\ldots X_{t_n} \quad \text{and} \quad X_{t_1+h}, X_{t_2+h}, \ldots\ldots X_{t_n+h}$$

are identical. For more detail see Pitt (1963).

This is a rather severe restriction and can be weakened. We define stationarity of order m as follows:

A process is staionary to order m if for any admissable subset $\{t_1 \cdots t_n\} \in T$ all the joint moments of $\{X_{t_1} \cdots X_{t_n}\}$ up to and including those of order m exist and are equal to the corresponding joint moments of $\{X_{t_1+k} \quad X_{t_2+k} \cdots \quad X_{t_n+k}\}$ for all values of k.

All processes considered subsequently will be assumed to be second order stationary unless it is explicitly mentioned otherwise. In this case m = 2 and inconsequence

$$E[X_t] = \mu < \infty \qquad \text{a constant}$$

$$E[X_t X_s] = E[X_0 X_{t-s}] \qquad \text{a function}$$

of (s-t) only. We shall define

$$E[X_t X_s] = R_{t-s} < \infty$$

as the autocovaniance function. We may also define the autocorrelation function ρ_τ

$$\rho_\tau = R_\tau / R_0 \qquad 1.1$$

We can immediately establish for real valued processes

$$R_0 = \text{var}(X_t) \qquad \rho_0 = 1$$
$$R_\tau \leq R_0 \qquad |\rho_\tau| \leq 1.$$
$$R_{-\tau} \leq R_0 \qquad \rho_{-\tau} = \rho_\tau$$

Second order stationarity has been referred to as "Generalised Stationarity" by Wold (1938), Weakly Stationary by Doob (1953) and Covariance stationary by Parzen (1961). Whittle (1963) has also used the expression "time homogeneous". The concept of stationarity appears to have been orginated by Khinchine (1934).

We point out that for a Normal process second order stationarity implies strict stationarity since the normal distribution is completely specified by its variance covariance matrix.

For out purposes we shall consider the index set to be

$$T = \{ 0, \pm 1, \pm 2, \ldots \} \qquad 1.2$$

or some subset. In this case the process is called a discrete stochastic processes or a stochastic sequence, Yaglom (1962).

Most of the processes we shall consider are discrete and by a suitable transformation can be considered to have an index set of the form 1.2 above. However even if T is constructed from the real line one finds in practise that one is forced to consider sampling values of T in which case we again have in effect an index set T of the form given by 1.2

Linear Models

Basic to the study of discrete processes are several models. The first we consider being the autoregressive model of $\{X_t\}$ which may be written

$$\sum_{k=0}^{p} a_k X_{t-k} = \varepsilon_t \qquad 1.3$$

or

$$A(B) X_t = \varepsilon_t \qquad 1.4$$

where $$A(z) = \sum_{k=0}^{p} a_k z^k$$

and B denotes the backward shift operator defined on the index where

$$B X_t = X_{t-1}$$

The process $\{\varepsilon_t\}$ is considered to be a sequence of independent random variables, the so called white noise process.

The equation

$$A(z) = 0 \qquad 1.5$$

is sometimes called the associated equation.

It can be proved that if all the roots of the associated equation 1.5 are lessthan one in absolute value then X_t can be written as an infinite linear combination of $\varepsilon_t, \varepsilon_{t-1}, \ldots$ and in addition X_t is independent of $\varepsilon_{t+1}, \varepsilon_{t+2}, \ldots$ For details see Anderson (1971)

Multiplying 3 by X_{t-s} we have

$$\sum_{k=0}^{p} a_k X_{t-k} X_{t-s} = \varepsilon_t X_{t-s} \qquad s = 1, 2 \ldots$$

and taking expectations

$$\sum_{k=0}^{p} a_k R_{s-k} = 0$$

These are often called the Yule-Walker equations, and have often been used to find the $a_j, j=0 \cdots p$ when the R_s are known.

Autoregressive processes appear to have been first suggested and used by Yule (1927), however they have been much used since, perhaps since they are easily handled and provide simple models. Box & Jenkins (1970) have used the expression 'partial autocorrelations' for the $\{a_k\}$.

The model

$$X_t = \sum_{k=0}^{q} b_k \varepsilon_{t-k} \qquad 1.6$$

or

$$X_t = \theta(B) \varepsilon_t$$

where

$$\theta(z) = \sum_{k=0}^{q} b_k z^k$$

is called the moving average model of order q when $\{\varepsilon_t\}$ is a white noise process.

The autoregressive and moving average models can be combined to give in the terminology of Box & Jenkins (1972) an autoregressive moving average model

$$A(B)\, Y_t = \theta(B)\, \varepsilon_t \qquad 1.7$$

The Spectrum

These models form the basic theoretical models used in the study of stationary stochastic processes. We now consider the use of another entity, the power spectrum. We shall outline the main results here but for more detail and proofs see Grenander & Szego (1958), Grenander & Rosenblatt (1957) or Anderson (1971).

We state our main result as a theorem

The covanance function R_τ of a stationary stochastic process with $E[Y_t] = 0$ can be represented as a Fourier-Stieltjes integral

$$R_\tau = \int_{-\pi}^{\pi} e^{i\omega\tau}\, dF(\omega) \qquad 1.8$$

with a bounded and non-decreasing function F(w).

The function F(w) is uniquly determined at all points of continuity and is called the spectrum of the process.

It is well known that every bounded and non-decreasing function F(w) can be written as the sum of three components

$$F(w) = F_1(w) + F_2(w) + F_3(w) \qquad 1.9$$

where

(a) $F_1(w)$ has a derivatve which exists almost everywhere

(b) $F_2(w)$ is a step function which is constant except at the discontinuities of F(w)

(c) $F_3(w)$ is a singular component with zero derivative almost everywhere.

In practical situations we usually consider processes having an absolutely continuous spectrum, or at most an absolutely continuous spectrum with a few discontinuities superimposed.

It is worth noting that for a continuous process i.e. $T =]-\infty, \infty[$ one can prove a similar theorem with

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} \, dF(\omega) \qquad 1.10$$

We can also connect the spectrum to the process in an alternative manner to obtain the spectral representation of the process by the following theorem due to Cramer.

Any discrete stationary process which is continuous in the mean can be represented in the form

$$X_t = \int_{-\pi}^{\pi} e^{it\omega} \, dZ(\omega) \qquad 1.11$$

where $Z(\omega)$ is an orthogonal process i.e. its increments are uncorrelated and

$$E\left[|dZ(\omega)|^2 \right] = dF(\omega)$$

$$E\left[dZ(\omega_1) \, d\bar{Z}(\omega_2) \right] = 0 \quad \omega_1 \neq \omega_2 \qquad 1.12$$

In addition to dividing stochastic processes into stationary and evolutionary types we can also split stationary processes into two distinct groups, those that are deterministic and those that are non-deterministic.

Consider the problem of predicting X_{t+h} given $\{ X_n, n = t, t-1, \ldots \}$. We construct our predictor of X_{t+h} from some linear combination of $X_t, X_{t-1}, \ldots$ and denote it by X^*_{t+h}. We shall look at the mean-square error of prediction e_h where

$$e_h = E\left[|X^*_{t+h} - X_{t+h}|^2 \right] \qquad 1.13$$

If $e_n = o$ we clearly have rather an odd process in which there is an exceptionally strong dependence between X_{t+h} and $X_t, X_{t-1}, \ldots$. Such processes are said to be deterministic.

Kolmogorov (1939) proved the remarkable theorem that for a non-deterministic process with an absolutly continuous spectrum

$$e_1 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e 2\pi f(\omega)\, d\omega \right\} \qquad 1.14$$

for the optimum predictor and further that a necessary and sufficient condition for a process to be deterministic is that

$$\int_{-\infty}^{\infty} \log_e f(\omega)\, d\omega = -\infty \qquad 1.15$$

We consider 1.14 in more detail at a later stage but for the moment present another remarkable result due to Wold (1938).

If $e_1 > 0$ we may decompose the process $\{X_t\}$ as follows

$$X_t = U_t + V_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j} + V_t$$

where

(a) $E[\varepsilon_t \varepsilon_s] = \sigma^2 \delta_{t,s}$

(b) $\alpha_0 = 1, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty$

(c) $E[\varepsilon_s V_t] = 0$ for all s, t

(d) The V_t process is deterministic.

In addition the U_t process has an absolutely continuous spectral density function with spectral density f(w).

$$f(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \alpha_j e^{i\omega j} \right|^2$$

while the V_t process has a spectral distribution fuction

$$F_V(\omega) = F_2(\omega) + F_3(\omega)$$

where these are defined by 1.9.

When V_t is absent then $\{X_t\}$ is a purely non-deterministic process. One has an intuitive feeling that for processes met in practice $e_1 \neq 0$ and one may hope to remove the deterministic component by regression techniques, Hannan (1960).

Before turning to estimation we consider the effect of linear filters.

If $\{X_t\}$ and $\{Y_t\}$ are two discrete parameter processes and

$$X_t = \sum_{u=-\infty}^{\infty} g_u Y_{t-u}$$

Then $f_x(w) = |\Gamma(\omega)|^2 f_y(\omega)$ 1.16

where

$$\Gamma(\omega) = \sum_{u=-\infty}^{\infty} g_u e^{-i\omega u}$$

$\Gamma(\omega)$ transfer function and $|\Gamma(\omega)|^2$ the filter and $f_x(w)$, $f_y(w)$ denote the spectra of the $\{X_t\}$ and $\{Y_t\}$ processes.

Filtering is an important and useful tool in spectral analysis and many problems can be reduced to constructing the appropriate filter. Indeed even the problem of finding the optimum linear predictor can be reduced to a filtering problem.

All the above theory can be extended to multivariate random processes. For example $\{X_1(t), X_2(t)\}$ is called a bivariate stationary process (of second order)

if

(i) $X_1(t)$ and $X_2(t)$ are stationary to order 2 with autocorrelation and spectral density functions given by $R_{11}(\tau)$, $f_{11}(\omega)$, $R_{22}(\tau)$, $f_{22}(\omega)$ respectively.

(ii) $$E\left[\{X_1(t)-\mu_1\}\{X_2(t+\tau)-\mu_2\}\right] = R_{12}(\tau)$$

is a function only of τ

We may then define the cross spectrum $f_{12}(w)$ by

$$f_{12}(w) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R_{12}(\tau)\, d\tau$$

and the spectral density matrix by

$$\underline{f(\omega)} = \begin{pmatrix} f_{11}(\omega) & f_{12}(\omega) \\ f_{21}(\omega) & f_{22}(\omega) \end{pmatrix}$$

Two components of the cross spectrum are of interest, the coherence

$$C(\omega) = \frac{|f_{12}(\omega)|}{\sqrt{f_{11}(\omega) f_{22}(\omega)}}$$

and the phase which is the argument of $f_{12}(\omega)$ written in complex form. Much greater discussion of bivariate and multivariate spectral analysis can be found in Granger (1964) and Hannan (1970).

From the theoretical outline we now turn our attention to the important problem of estimation. We have however a problem in that given a sequence $\{X_t\}$ we have only one realisation of a whole possible ensemble, and moreover we will have in practical situations only a finite set T. In fact one is forced to consider the realisation $\{x_t(w)\}$ as being in some sense typical of the entire ensemble. We are forced into a position where one must estimate autocorrelation co-efficients for a predetermined set of $\{w\}$.

Naturally much work has gone into investigating the conditions for which time averages produce consistent estimates of the population parameters of the ensemble and many delicate and powerful theorems have been proved. In essence our time averages provide reasonable estimates if the ergodic property holds for our process. Sufficient

conditions for this property to hold are

(a) $\lim_{\tau \to \infty} R_\tau \to 0$

or (b) $\{X_t\}$ is fourth order stationary

or (c) The absence of jumps in F(w)

Condition (c) is also a necessary condition.

We refer interested readers to Doob (1953), Billingsley (1965) and Rozanov (1967).

Estimation

As autoregressive, moving average and mixed processes have been proposed as models for stationary processes our next concern is the estimation of parameters.

Mann & Wold (1943) developed the maximum likelihood estimates and asymptotic theory for the model

$$\sum_{r=0}^{\infty} \alpha_r X_{t-r} = \varepsilon_t \qquad 1.17$$

for $\{\varepsilon_t\}$ having zero mean Normal distributions and showed that maximum likelihood estimates are asymptotically efficient.

The derivation of the estimates is as follows:

since $\varepsilon_1 \ldots \varepsilon_N$ are independent of $X_0, X_{-1} \ldots X_{-p+1}$ the conditional p.d.f. of $\varepsilon_1 \cdots \varepsilon_N$ is the same as the unconditional density and is given by

$$\frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{N} \varepsilon_t^2 \right\} \qquad 1.18$$

When X_0 X_{-1}... are fixed we may use 1.17 to give the likelihood of $X_1 \cdots X_N$ as

$$\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{N} \left[\sum_{r=0}^{p} \alpha_t X_{t-r} \right]^2 \right\}$$

thus the maximum likelihood estimates are obtained by minimising

$$Q = \sum_{t=1}^{N} \left\{ \sum_{r=1}^{P} \alpha_r X_{t-r} \right\}^2$$

which is equivalent to the usual least squares problem.

Writing and setting $\alpha_0 = 1$

$$C_h = \frac{1}{N} \sum_{t=1}^{N-h} X_t X_{t+h}$$

$$r_h = C_h / C_0 \qquad h = 1, 2, \ldots N-1$$

we have for large N

$$\underline{R}\,\underline{\alpha} = -\underline{r}$$

$$\underline{R} = \begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & \\ \vdots & & & & \\ r_{p-1} & & \cdots & & 1 \end{pmatrix}$$

and

$$\underline{r}' = (r_1, r_2 \ldots r_p), \quad \underline{\alpha}' = (\alpha_1, \alpha_2 \ldots \alpha_p)$$

thus giving

$$\underline{\alpha} = -(\underline{R})^{-1} \underline{r} \qquad 1.19$$

The structure of $\underline{R}$ enables one to use simple iterative methods of evaluating the α details of which were developed by Durbin (1960).

There is a considerable body of theory on the asymptotic behavior of these estimates, perhaps the best account being given by Anderson (1971).

While there is much asympotic theory the small sample theory is less well developed. Perhaps the simplest autoregressive process is the AR1 process, or first order Markov process given by

$$X_t - \alpha X_{t-1} = \varepsilon_t \qquad 1.20$$

where $|\alpha| < 1$ for stationarity.

Hurwicz (1950) studied 1.20 under the condition that X_1 was Normal with zero mean and variance

$$\frac{var(\varepsilon_t)}{1-\alpha^2}$$

and derived the distribution of the maximum likelihood estimate of for N=3 and 4 with $X_0=0$. He showed that the estimate was biased but the arguments he used became too complex to pursue for larger N.

Marriot & Pope (1954) again discovered bias in the estimator and suggested a modification in view of their results and those of Kendall (1954).

White (1961) gave asymptotic expansions for the mean and variance of the serial correlation co-efficient for both cases

$X_0 = 0$ and X_0 being $N\left(0, \frac{var(\varepsilon_t)}{1-\alpha^2}\right)$

where $\sigma^2 = var(\varepsilon_t)$ He also showed White (1958), (1959) that for $|\alpha| > 1$ the asymptotic distribution of $\hat{\alpha}$ is Cauchy

Copas (1966) compared various estimates of α in some simulation experiments and found little difference. He found some evidence that a mean-likelihood method worked rather well compared to a least-squares estimate over a limited range. Thornbur (1967) gave some interesting results from a Baysian viewpoint and pointed out that the fact α is constrained did affect the estimation procedure.

Thus it would seem that while the large sample theory is well known there is still much to be learned in the small sample situation, see Orcutt et al (1969).

The estimation of moving average and mixed models is rather more complex. Perhaps because the autoregressive has proved of rather more interest there appears to have been less examination of small sample situations.

Let $$X_t = \sum_{j=0}^{q} \beta_j \varepsilon_{t-j} \qquad 1.21$$

if the $\{\varepsilon_t\}$ are normal then the X_t are normal and we can consider estimating the $q+1$ parameters from $\{ X_t, t = 1,2 \ldots N \}$. Unfortunately the maximum likelihood estimates cannot be solved directly, see Anderson (1971).

Walker (1961) used maximum likelihood applied to the sample correlations and used his estimates $\hat{\rho}$ to calculate $\underline{\beta}$. Durbin (1959) suggested that the autoregressive representations of 1.21

$$\varepsilon_t = \sum_{r=0}^{\infty} \gamma_r X_{t-r}$$

could be approximated by the finite sum

$$\bar{\varepsilon}_t = \sum_{r=0}^{n} \gamma_r X_{t-r} \qquad 1.22$$

for large n. For suitably large n it can be argued that the $\bar{\varepsilon}_t$ are nearly uncorrelated which suggests estimating $\underline{\gamma}' = (\gamma_1, \gamma_2 \ldots \gamma_n)$

by

$$\underline{R}\ \underline{\gamma} = -\underline{r}$$

For details see Durbin (1959) or Anderson (1971).

In the case of mixed model both Walker (1962) and Durbin (1960b) have proposed modifications of their methods for dealing with moving average models.

Clearly we do not have space to go into greater detail in the estimation of linear model parameters but mention must be made of Box &

Jenkins (1972) who make extensive use of such models in their forecasting techniques. We defer fuller discussion of their methods until we examine forecasting in greater detail.

It is often the case that our models can be considered to contain a deterministic component that may be described as some smooth function such as a polynomial in t, or even a trigonometric polynanial. Grenander (1954) and Grenander & Rosenblatt (1957) obtain the presant result that a straight forward least squares procedure is efficient under quite general conditions. Some additional results may be found in Hannan (1957).

Spectral Estimation

It is beyond our scope to review the extensive development of the theory of statistical spectral analysis over the past few years and we restrict our scope to providing an outline of results. For more detail there are a number of review papers, see Jenkins (1961), Parzen (1961) and Priestley (1962) while the books of Granger (1964) and Fishman (1968) provide much greater detail.

In early work on time series the series was often envisaged as satisfying the model

$$X_t = \sum_{j=1}^{n} a_j \cos(\omega_j t + \phi_j) + \varepsilon_t \qquad 1.23$$

in which the $\{\varepsilon_t\}$ are white noise and the ω_j, a_j, ϕ_j are to be estimated. A technique known as the periodogram was introduced by Schuster consisting of the function

$$I_N(\omega) = \frac{2}{N} \left| \sum_{t=1}^{N} X_t e^{-i\omega t} \right|^2 \qquad 1.24$$

It is easily shown that $I_N(\omega)$ will have a peak at $\omega = \omega_0$ if the model contains a periodic term at $\omega = \omega_0$ and there will be subsidary peaks at $\omega = \omega_0 + \frac{2\omega_0}{N}$ Significance tests for such peaks have been proposed and a survey can be found in Jenkins & Priestley (1957). Examples of estimated periodograms can be found in Davis (1941) and Beveridge (1922).

This technique became somewhat discredited when it became apparent that an excessive number of cycles were being found. This prompted workers in the field to consider other possible models.

However Wiener's (1930) paper enabled the theory to be put into a more coherent framework by emphasising the link between the auto-correlations and the spectrum (see 10). The modern spectral estimation approach was developed from this paper.

Before considering the distributional problems in more detail we consider the phenomenon of "aliasing". Let $\{X(t)\}$ be a sample record of length T $(0 < t < T)$ which has been read from a continuous record. Let $\{X_t\}$ be a discrete set of readings obtained by sampling at a fixed interval Δt from the continuous trace. Thus

$X_t = X(t.\Delta t)$ and $\Delta t . N = T$ The sampling of the discrete sequence has two consequences.

(i) The spectrum of the discrete series is now defined only for frequencies up to $\omega_0 = \frac{\pi}{\Delta t}$ being called the Nyquist or Folding frequency.

(ii) The power at frequencies higher than ω_0 is superimposed on the portion of the spectrum lying between $\pm \omega_0$

It is convenient to work in the "standard" frequency range $\pm \pi$ which corresponds to sampling the record at unit intervals i.e. $\Delta t = 1$ The spectrum $f^*(\omega)$ in terms of the original frequency scale being

$$f^*(\omega) = \Delta t \, f(\omega . \Delta t) \qquad |\omega| \leq \frac{\pi}{\Delta t} \qquad 1.25$$

Now for a continuous record the autocorrelations are given by

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} \, dQ(\omega) \qquad \tau = 0, \pm 1, \pm 2, \ldots$$

$$= \sum_{s=-\infty}^{\infty} \int_{(2s-1)\pi}^{(2s+1)\pi} e^{i\omega\tau} \, dQ(\omega)$$

$$= \sum_{s=-\infty}^{\infty} \int_{-\pi}^{\pi} e^{i(\omega+2\pi s)\tau} \, dQ(\omega+2\pi s) \qquad \tau = 0, \pm 1, \ldots$$

Then writing $dF(\omega) = \sum_{s=-\infty}^{\infty} dQ(\omega+2\pi s)$ we have

$$f_s(\omega) = \sum_{-\infty}^{\infty} f(\omega + 2\pi s) \qquad 1.26$$

where $f_s(\omega)$ is the spectrum of the sampled sequence.

To minimise the effect of aliasing it is usual to try and choose the sampling interval Δt in such a way that for frequencies higher than the Nyquist $f(\omega)$ is small.

From now on we shall consider only discrete series in our discussion of estimation.

As we have seen in 1.8

$$R_\tau = \int_{-\pi}^{\pi} e^{i\omega\tau} \, dF(\omega)$$

and since we shall assume F(w) is uniformly continuous with $f(\omega) = F'(\omega)$

we can write

$$f(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R_s \cos s\omega \qquad |\omega| \le \pi \qquad 1.27$$

since R_τ and hence $f(\omega)$ is symmetric.

From 1.27 it would seem that the natural way to estimate $f(\omega)$ is

$$\hat{f}_s(\omega) = \frac{1}{2\pi} \sum_{s=-N+1}^{N-1} \hat{R}(s) \cos s\omega \qquad 1.28$$

using a suitable estimate of $\hat{R}_s$ Many estimates of R_s have been suggested the most popular being

$$\hat{R}_s = \frac{1}{N-|s|} \sum_{t=1}^{N-|s|} X_t X_{t+|s|} \qquad 1.29$$

and

$$\hat{\hat{R}}_s = \frac{1}{N} \sum_{t=1}^{N-|s|} X_t X_{t+|s|} \qquad 1.30$$

The form given in 1.29 can be shown to be (asymptotically) unbiased, however Parzen (1961) has pointed out that the form given in 1.30 has smaller mean-square error in that

$$E\left[\hat{\hat{R}}_s - R_s\right]^2 \le E\left[\hat{R}_s - R_s\right]^2 \qquad 1.31$$

while

$$E\left[\hat{\hat{R}}_s\right] \sim \left(1 - \frac{|s|}{N}\right) R_s \qquad 1.32$$

For this reason we shall use the second form almost exclusively.

Thus one form of estimate of $f(\omega)$ is

$$\hat{f}_s(\omega) = \frac{1}{2\pi} \sum_{s=-N+1}^{N-1} \hat{R}_s \cos s\omega$$

which can be shown to be intimately connected with $I_N(\omega)$ in that

$$I_N(\omega) \simeq 4\pi \hat{f}_s(\omega)$$

and in fact $I_N(\omega) = 4\pi \hat{f}_s(\omega)$ when R_s (1.29) is used. The periodogram is very nice to handle from the mathematical viewpoint but has some drawbacks as we shall see in Chapter 3 and one can prove some very elegent theorems. Good reviews are given by Olsen (1967) and Walker (1965).

Fisher (1929) devised a test for jumps in F(w) based on the periodogram.

However although $\hat{f}_{s(w)}$ is an asymptotically unbiased estimator of $f(w)$ it is well known that is not a consistent estimator, Priestley (1965). More generally it is usual to consider estimates for $f(w)$ of the form

$$\hat{f}(w) = \frac{1}{2\pi} \sum_{s=-N+1}^{N-1} \lambda_s \hat{R}_s \cos sw \qquad 1.33$$

where $\{\lambda_s\}$ is a suitably chosen weighting sequence, sometimes called the covariance window.

Bartlett (1950) suggested

$$\lambda_s = \begin{cases} 1 - \frac{|s|}{m} & |s| \leq m \\ 0 & |s| > m \end{cases} \qquad 1.34$$

where m is some integer less than N and $m/N \to 0$ as both m and N tend to infinity

Tukey (1949) suggested

$$\lambda_s = \begin{cases} 1 - 2a + 2a\cos\frac{\pi s}{m} & |s| \leq m \\ \frac{1}{2} - 2a & |s| = m \\ 0 & |s| > m \end{cases} \qquad 1.35$$

with $a = 0.23$ while Daniel (1946) proposed

$$\lambda_s = \frac{\sin s\pi/m}{s\pi/m} \quad \text{for all } s \qquad 1.36$$

A somewhat more restricted form of estimate has been discussed in detail by Parzen (1957a 1957b) is

$$\hat{f}(w) = \frac{1}{2\pi} \sum_{s=-N+1}^{N-1} k(B_N s) \hat{R}_s \cos sw \qquad 1.37$$

where $k(u)$ is a continuous function of u and B_N is a sequence of constants such that $\lim_{N\to\infty} B_N = 0$ In addition Parzen has shown (1957a)

that for estimates for the form 1.37

$$E\left[\hat{f}(\omega) - f(\omega)\right] \sim B_N^r \left| k^{(r)} \right| f(\omega) / \left\{ \lambda_r(\omega) \right\}^r \qquad 1.38$$

where r is the largest positive integar such that

$$k^{(r)} = \lim_{u \to 0} \left\{ \frac{1 - k(u)}{|u|^r} \right\} \quad \text{exists and is non-zero}$$

and

$$\lambda_r(\omega) = \left| \frac{f(\omega)}{f^{(r)}(\omega)} \right|^{1/r} \qquad 1.39$$

where

$$f^{(r)}(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |s|^r R_s \cos s\omega \qquad 1.40$$

Parzen has also suggested the following form of $k(u)$

$$k(u) = \begin{cases} 1 - 6u^2 + 6|u|^3 & |u| < 0.5 \\ 2(1 - |u|)^3 & 0.5 < |u| < 1.0 \\ 0 & |u| > 1 \end{cases} \qquad 1.41$$

the attraction of which is that it ensures positive estimates of the spectrum, Granger (1964).

Grenander and Rosenblatt (1957) have shown that for normal series the variance of $\hat{f}(\omega)$ is given by

$$V(\omega) \simeq \frac{2\pi f^2(\omega)}{N} \int_{-\pi}^{\pi} W^2(\theta) d\theta \qquad 1.42$$

where $W(\theta) = \sum_{-\infty}^{\infty} \lambda_s e^{is\theta}$. This result has been modified by Neave (1970) to:

$$V(\omega) = \int_{-\gamma^{-1}}^{\gamma^{-1}} \frac{2\pi}{N} f^2(\omega) (1 - \gamma|\theta|) W^2(\theta) d\theta \qquad 1.43$$

Here $\gamma = \frac{B_N}{N}$ whereas in the traditional case $\gamma = 0$

The introduction of $W(\theta)$ the spectral window enables us to write $\hat{f}(\omega)$ in an alternative manner as

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} I_N(\theta) W(\theta - \omega) d\theta$$

in effect smoothing the periodogram by using a suitable kernal.

To choose the value of the parameter m is a rather vexed question. Various procedures have been mentioned, the problem being

that ones "resolution" of the spectrum depends on the "bandwidth" of the spectral window employed. For our purposes we define the bandwidth of the width of a rectangle whose height and area correspond to those of The problems involved are too detailed to be discussed here but we feel that the approach of Priestley (1965), (1962) is the most appropriate.

It is perhaps worth mentioning that we have all along implicitly assumed that $\{X_t\}$ has a zero mean. If this is not true we can of course modify our autocovariance estimates to

$$\hat{R}_s = \frac{1}{N} \sum_{t=1}^{N-|s|} (X_t - \bar{X})(X_{t+|s|} - \bar{X})$$

however this does introduce some bias into our estimates of the spectrum. Usually it is assumed to be negligable however Fishman (1969) has shown that in this case

$$E[\hat{f}(\omega)] = f(\omega) - \frac{2\pi}{N} f(0) \int_{-\pi}^{\pi} W(\theta) d\theta \qquad 1.44$$

This can be compensated for by noting that

$$E[\hat{f}(0)] \sim f(0)\left[1 - \frac{z}{N}\right] \qquad 1.45$$

where

$$z = 2\pi \int_{-\pi}^{\pi} W(\theta) d\theta$$

and using as our estimate

$$\hat{f}_u(\omega) = \hat{f}(\omega) + \hat{f}(0)\left(1 - \frac{z}{N}\right)\frac{z}{N} \qquad 1.46$$

As it may be imagined estimates of the form 1.37 involve considerable computational effort and often for long series may prove too much for a computer. Cooley and Tukey (1965) rediscovered an algorithm for computing the Fourier co-efficients

$$J(\omega) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} X_t e^{-i\omega t}$$

in $N \log_2 N$ operations instead of N^2 for the usual methods. This Fast Fourier Transform (FFT) technique enables one to reduce computational problems considerably by enabling the periodogram to be estimated very

easily and hence the spectrum by using

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} I_N(\theta)\, W(\theta-\omega)\, d\theta \qquad 1.47$$

In practice most weighting of the periodogram is carried out by averaging adjacent periodogram ordinates. The main drawback of this technique apart from some rather bad machine implementations is that one does not compute the autocorrelations which can be very informative. This however can be overcome by performing the inverse transformation on the estimated spectrum using the FFT once again.

Forecasting Time Series

Kolmogorov (1939, 1941) appears to have been the first to completely solve the problem of finding the optimum linear predictor of discrete time series although Wiener (1949) independently solved the problem for continuous processes and later considered the multivariate case Wiener and Masani (1957,58). We shall be considering the mathematical theory in detail in Chapter II but before looking at optional methods we consider some standard techniques that have been proposed.

One of the earliest methods of forecasting a time series consisted of a simple decomposition of the series into three parts, these being a trend, a cycle (usually 12 monthly) and a residual. The basic idea being that if one can estimate fairly accurately the trend and cyclical components then one may obtain reasonable forecasts, since one would expect the contribution to the total variance by the residual term to be minor.

The usual procedure is to fit various functions to the data to estimate the trend using regression techniques and then to use the

residuals from the trend to estimate the seasonal cycle. One method of determining the cycle being to use a regression technique with dummy variables. Other techniques are described by Granger (1967).

Another approach mainly investigated by Granger & Newbold (1971) is to use a stepwise regression method. To generate a forecasting model one attempts to explain X_{n+1} in terms of $\{ X_{n-j} ; j = 0, 1 \ldots \}$ and possibly $\{ Y_{i,n-j} ; j \geq 0 \quad i = 1 \ldots m \}$ where these are a set of explanatory series. The technique used is a stepwise-regression method. While generally good results are reported one must bear in mind that there is no attempt to explain the structure of the process being forecast and its relation with any explanatory series. In consequence there is some danger of developing a spurious model and care must be taken to ensure that one's regression model is not some artifact arising from some local behaviour of the process.

Perhaps the most interesting methods, apart from the optimum ones to be considered later are the so-called adaptive methods which range from a simple experimental smoothing technique to the more detailed work of Box and Jenkins (1972).

Adaptive techniques produce forecasting methods which are, above all, quick and easy to apply with a minimum of data storage. However they do not make use of a great deal of structural underlying the process being forecast, which makes them easy to apply but does restrict the amount of information available to construct forecasts.

The simplest method is the "exponential smoothing" or "exponentially weighted moving averages" technique proposed by Holt (1957) which

has been discussed by many authors notably Winters (1959) (1962) Kalman (1960) Brown (1959,1963) and Harrison (1965).

Given a stationary process $\{X_t\}$ we define $\{X_t^*\}$ by

$$X_t^* = a X_{t-1} + (1-a) X_{t-1}^* \qquad 0<a<1$$
$$= X_{t-1}^* + a(X_{t-1} - X_{t-1}^*) \qquad 1.48$$

which can be written

$$X_t^* = a \sum_{j=1}^{\infty} (1-a)^{j-1} X_{t-j} \qquad 1.49$$

i.e. $\{X_t^*\}$ is a weighted average of past values of $\{X_t\}$

Examination of 1.49 shows that X_t^* is unbiased. Thus to forecast X_{t+1} one uses X_{t+1}^* where

$$X_{t+1}^* = a X_t + (1-a) X_t^*$$

$$= a \sum_{k=0}^{\infty} (1-a)^k X_{t-k} \qquad 1.50$$

This method is adaptive in the sense that previous forecasts are updated by a proportion of previous forecast errors.

If we have a non-stationary process with for example a trending mean the above procedure must be modified. One modification is to try and incorporate an estimate of the trend.

$$X_t^* = a X_t + (1-a)(X_{t-1}^* + T_t)$$

where the variable T is our estimate of trend where

$$T_t = b(X_t^* - X_{t-1}^*) + (1-b) T_{t-1} \qquad 0<b<1.$$

This method of adjusting for a trending mean appears to work well as long as the trend is locally linear.

As well as adjusting the adaptive procedure for trend seasonal cycles have also been incorporated. If one has a cycle of length n units then a seasonal factor F_t may be introduced

$$F_t = c(X_t - X^*_t) + (1-c)F_{t-n} \qquad 0<c<1$$

giving forecasts of the form

$$X^*_t = a(X_{t-1} + F_{t-1}) + (1-a)(X^*_{t-1} + T_{t-1})$$

One should notice however that these additional factors do complicate the forecasting method and detract from its simplicity. In addition one needs considerable amounts of information to estimate the seasonal factors accurately.

Crucial to the adaptive methods outlined above is the estimation of the a, b and c. As one is not fitting a structural model to the series there are no opimium values of these parameters and one needs must proceed in an empirical fashion. If the values of a, b, c are close to zero then the constructed series $\{X^*_t\}$ has a long memory in that the contributions from part $\{X_t\}$ contribute significantly even for distant values, see 1.50. In this case the final forecasts are stable and do not respond quickly to fluctuations in $\{X_t\}$. When the parameter values tend to one the series has a correspondingly short memory and reacts to changing circumstances very quickly.

It would appear that the most useful method of evaluating a, b, c is to use a search procedure in which forecasts are made and errors e_j determined from

$$e_j = X^*_{t+j} - X_{t+j} \qquad j > m$$

using selected values of a,b, c over a sample of $\{X_t\}$. One then chooses the parameter combination that provides the smallest sum of squares

$$Q_s = \sum_{j=m+1}^{n} e_j^2$$

One would hope that the sum of squares surface is relatively flat near its minimum to ensure that slight deviations from the "best" parameter values do not entail large forecast errors.

The initial values of the predictor series X_0^*, T_0, F_0 may be computed in the same manner if one splits up the sample $\{X_t\}$ and estimates the set of parameters in differing segments.

Brown (1963) has demonstrated that one can produce very much the same results using only one parameter. A generalisation of this approach is general exponential smoothing, in which it is assumed that the process being forecast can be represented by

$$X_t = a_1 f_1(t) + a_2 f_2(t) + \ldots + a_n f_n(t) + \varepsilon_t$$
$$= \underline{a}'(t)\,\underline{f}(t) + \varepsilon_t \qquad 1.51$$

where the $f_i(t)$ are known functions of time and the a_i are co-efficients to be updated. To estimate $\underline{a}(t)$ we use discounted least squares, minimising

$$Q_t = \sum_{j=0}^{t} \beta^j \left(X_{t-j} - \underline{a}'(t)\,\underline{f}(-j) \right)^2$$

the solution being

$$\underline{a}'(t) = \underline{g}'(t)\,\underline{F}^{-1}(t) \qquad 1.52$$

where

$$\underline{F}(t) = \beta^t \underline{f}(t)\,\underline{f}'(t) + \underline{F}(t-1) \qquad 1.53$$

Since 1.52 implies that everytime one needs to update the $\underline{a}'(t)$ a matrix inverse is needed a more restricted set of $\underline{f}_i(t)$ has been suggested in order that there exists a non-singular $\underline{L}$ such that

$$\underline{f}(t) = \underline{L}\,\underline{f}(t-1)$$

In this case we can obtain $\underline{g}(t)$

$$\underline{g}(t) = X_t \underline{f}(0) + \beta L^{-1} \underline{g}(t-1) \qquad 1.54$$

and $\underline{a}'(t) = \underline{g}'(t)\,\underline{F}^{-1}(t)$ can be written

$$\underline{a}'(t) = \underline{L}^{-1}\,\underline{a}(t-1) + \underline{F}(t)^{-1}\underline{f}(t)\left(X_t - X^{*}_{t-1}\right) \qquad 1.55$$

One may assume from 1.54 for $|\beta| < 1$ that $\underline{F}(t)$ will converge to a stable form independent of t in which case 1.55 is easily updated as new information becomes available, and in addition only one inversion is needed.

It does appear however that if one is using a computer for producing forecasts that such restrictions are quite unecessary since there are now fast and accurate matrix inversion routines available.

From the theoretical viewpoint Muth (1960) looked at a model where

$$X_t = Y_t + \eta_t$$

and

$$Y_t = Y_{t-1} + \varepsilon_t$$

where η_t was a transitory component introduced at each time period and considered to be normal with zero mean and variance σ_η^2. If the ε_t are similarly defined with variance σ_ε^2 Muth showed that a simple exponential forecast is optimal with

$$\alpha = -\frac{\sigma_\eta^2}{2\sigma_\varepsilon^2} + \frac{\sigma_\eta}{\sigma_\varepsilon}\left(1 + \frac{\sigma_\eta^2}{4\sigma_\varepsilon^2}\right)^{1/2}$$

The problem is however to estimate these variances and no solution has yet been proposed.

Kalman (1960) considered a much more general model of the form

$$\underline{X}_t = \underline{H}_t \underline{Y}_t + \underline{U}_t \qquad 1.56$$

$$\underline{Y}_t = \underline{\Phi}_t \underline{Y}_{t-1} + \underline{V}_t \qquad 1.57$$

and produced optimium forecasts on the assumption that the covariance matrices of $\underline{U}_t$ and $\underline{V}_t$ were known.

A rather more interesting and instructive forecasting procedure is that due to Box and Jenkins (1970) (1967). Using their notation they consider fitting mixed models of the form

$$\Phi(B) X_t = \theta(B) \varepsilon_t \qquad 1.58$$

where

$$\Phi(B) = 1 + \sum_{j=1}^{p} \Phi_j B^j$$

and

$$\theta(B) = 1 + \sum_{j=1}^{q} \theta_k B^k$$

Where B is the "backshift operator" $B X_t = X_{t-1}$ c.f the use of E and E^{-1} in Hartree (1958). In addition it is assumed that the $\{\varepsilon_t\}$ are a sequence of zero mean independent normal variates with common variance σ^2

Clearly 1.58 does not represent a great variety of series met in practice and Box & Jenkin suggest that by differencing a series $\{X_t\}$ a sufficient number of times one can obtain a stationary series. Thus if we introduce the differencing operator

$$\nabla X_t = X_t - X_{t-1}$$

a more general model is

$$\Phi(B) \nabla^d X_t = \theta(B) \varepsilon_t \qquad 1.59$$

This model they term an autoregressive integrated moving average model (ARIMA) of order (p, d, q) They also impose the natural condition that the roots of

$$\theta(z) = 0$$

must lie outside the unit circle.

The deriving of forecasts from 1.59 falls naturally into three parts, the first being the identification. The identification process is inexact and later diagnostic checks may require one to modify the model chosen. To identify a particular model we note

(i) when a series has been differenced until a stationary process if achieved the high order autocorrelations tend to zero.

(ii) For autoregressive models of order p say the partial autocorrelations damp down to . $\Phi(B)\rho_k = 0$

(iii) For a moving average process of order q the autocorrelations of order higher than q are zero, while the partial correlation tend to zero

(iv) For a mixed process

$$\Phi(B)\ \rho_k = 0 \qquad k > q$$

Using the sample autocorrelation and partial correlation functions one attempts to identify the model and then estimates the parameters using the low order sample autocorrelations.

Thus for example in a first order autoregressive

$$(1 - \phi B) X_t = \varepsilon_t$$

$$\hat{\phi} \simeq r_1$$

while for a model

$$X_t = (1 - \theta_1 B - \theta_2 B^2)\varepsilon_t$$

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} \qquad 1.60$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \qquad 1.61$$

Having identified the model one may then turn to estimation. The parameters of

$$\phi(B)\, X_t = \theta(B)\varepsilon_t$$

are evaluated by minimising the sum of squares using non-linear least squares techniques or a minimisation procedure of the form used by Flecher & Powell (1963)

Having estimated the parameter one must then determine the adequacy of the model obtained. If $\{\hat{a}_t\}$ denotes the residuals from the fitted model let

$$\rho_k = \text{corr}(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+k}) .$$

Following Bartlett (1946) and Anderson (1942) we can show that if the $\hat{\varepsilon}_t$ are independent normal

then $$\text{var}(\rho_k) \simeq \frac{1}{N}$$

and $$\text{corr}(\rho_k, \rho_{k+s}) \simeq 0$$

while the distribution of ρ_k is asympotically Normal. This is an approximation however since even if the fitted model is appropriate the $\hat{\varepsilon}_t$ will be autocorrelated. However for large samples Box and Pierce (1970) and Durbin (1971) have shown that

$$n \sum_{k=1}^{m} \rho_k^2 \qquad 1.62$$

is approximately χ^2 with m-s degrees of freedom, where s in the total number of parameters to be estimated. The choice of m appears to be quite arbitary.

Box Jenkins and Bacon (1967) have further extended their class of models to

$$\Phi(B)\,\Phi^*(B^s)\,\nabla^d(1-B^s)^{d_s}\,X_t = \theta(B)\,\theta^*(B^s)\,\varepsilon_t \qquad \ldots\ldots 1.63$$

where

$$\Phi^*(z^s) = 1 - \sum_{k=1}^{P} \Phi^*_k z^{kps}$$

$$\theta^*(z^s) = 1 - \sum_{k=1}^{q} \theta^*_k z^{kqs} \qquad 1.64$$

where s is the length of the seasonal.

The methods outlined above have proved very effective and since we seek to discover the underlying structure of the process there is less arbitary data minding as in the exponential smoothing approach. However, one does need substantial amounts of data, especially to use the anymptotic results in the diagnostic stages, otherwise the sampling behaviour of the sample autocorrelations becomes too dominant. In addition the identification of various models requires considerable skill and is not really suitable for automatic forecasting, especially as it is assumed the underlying model does not change over the series . The IBM data quoted by Box and Jenkins (1970) appears to have a changing model.

It might be added that these procedures also require that one must perform a considerable amount of computation.

The only extensive trials of forecasting procedures appear to be by Reid (1969) and Granger and Newbold (1972). From Reid's work it would appear that the Box-Jenkins method appears to be best under most

circumstances provided one has sufficient data to obtain reliable estimates of the model parameters. For short series the Brown method appears to work rather better. Granger and Newbold also showed that Box Jenkins techniques worked very well but for long lead times the advantage, particularly over a stepwise procedure of fitting autoregressives was very small.

Often one is in a position where one has two or more forecasts of the same quantity available. Rather than choose one of them Bates & Granger(1969) have suggested taking a weighted sum of these forecasts to produce a combined forecast. Thus if one has forecasts $\{F_{1,t}\}$ $\{F_{2,t}\}$ a combined forecast would be

$$C_t = k_t F_{1t} + (1 - k_t) F_{2t} \qquad 0 \leq k \leq 1 \qquad \ldots\ldots 1.65$$

Using this form we have unbiased forecasts if F_{1t} and F_{2t} are unbiased. If one wishes to minimise the variance of the combined forecast error then

$$k_t = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

where $\sigma_j^2 = \text{var}(X_t - F_{jt}) \qquad j = 1, 2$

and ρ denotes the correlation between the error series produced by $\{F_{1t}\}$ and $\{F_{2t}\}$

Unfortunately the parameters determining k_t are unknown. Bates and Granger (1969) have suggested some practical procedures and Granger and Newbold (1972) have shown that combining statistical and economic forecasts can be very effective, especially if one of the components forecasts is based on a persons judgement e.g. a brand managers forecasts of sales.

The Evaluation of Forecasts

While we have considered forecasting methods we have not paid a great deal of attention to the objectives in mind when forecasts are made. Clearly one would wish to make the "best" possible forecast, the question being what does one mean by best. Given a process $\{X_t\}$ and a corresponding sequence of forecasts $\{X_t^*\}$ one could in principle assign a cost to the magnitude of the errors

$$d_j = X_j - X_j^* \qquad 1.66$$

If the cost of an error of size z is $C(z)$ then the best forecast will be that which minimises $C(z)$

The idea of using a cost function is a natural one and gives us a very reasonable method of ranking forecasts. It would appear reasonable to suppose that $C(0) = 0$ and that $C(z)$ increases monotonically with $z > 0$ and $C(z)$ increases monotonically as z tends to minus infinity. However one needs must bear in mind that $C(z)$ need not be symmetric about $z = 0$ Indeed there are many cases one can visualise where $C(z)$ is markedly non-symmetric.

Another complication is that $C(z)$ may well change shape over time. If $C_t(z) = A(t)C(z)$ for some positive $A(t)$ then one can proceed as if the cost function is $C(z)$ however if $C_t(z)$ changes shape over time one is in the position of having to forecast the form of the future cost function. In this case the cost function must be forecast before one can even decide on what forecast criteria to use.

Granger (1968, 1973) has suggested that a reasonable procedure is to forecast by least-squares and then to allow for possible non-

symmetry of $C(z)$ by adding a bias. Gilchrist (1968) has suggested a similar technique.

In what follows we shall assume that one wishes to minimise costs at a particular time rather than the sum of costs over a period and in addition we shall assume a quadratic cost function which arises from the consideration of a minimum mean square error criteria.

If one uses a mean-square error criteria the obvious measure of a forecasting method is

$$D_n^2 = \frac{1}{n} \sum_{t=1}^{n} (X_t - X_t^*)^2 \qquad 1.67$$

In fact Newbold and Granger (1973) show that the use of any statistic that is not a monotonic function of D_n may give rise to misleading conclusions. In particular they show that the statistic proposed by Theil (1958)

$$U_1 = \frac{D_n}{\left\{\frac{1}{n}\sum X_t^2\right\}^{1/2} + \left\{\frac{1}{n}\sum X_t^{*2}\right\}^{1/2}} \qquad 1.68$$

and used by Kumichika (1971) and Lounguist (1971) is not a minimum when one uses the optimal linear predictor of the first order autoregressive model.

In addition to U_1 above, Theil (1958) noted that

$$D_n^2 = (\bar{X}^* - \bar{X})^2 + (S^* - S)^2 + 2(1-r)S^*S \qquad 1.69$$

and

$$D_n^2 = (\bar{X}^* - \bar{X})^2 + (S^* - Sr)^2 + (1-r^2)S^2 \qquad 1.70$$

where $\bar{X}^*$ and $\bar{X}$ denote the sample means of the original process and the forecast sequence and S^* and S their respective sample standard deviations while r is the sample correlation.

1.69 and 1.70 lead Theil to the definition of

$$U^M = (\bar{X} - \bar{X}^*)^2 / D_n^2 \qquad 1.71$$

$$U^S = (S^* - S)^2 / D_n^2 \qquad 1.72$$

$$U^C = 2(1-r)S^*S / D_n^2 \qquad 1.73$$

where $U^M + U^S + U^C = 1$

Granger and Newbold (1972) throw some doubt on the usefulness of these quantities and suggest from 1.66 the use of

$$U^M = (\bar{X}^* - \bar{X})^2 / D_n^2 \qquad 1.74$$

$$U^R = (S^* - rS)^2 / D_n^2 \qquad 1.75$$

$$U^D = (1 - r^2) S^2 / D_n^2 \qquad 1.76$$

as diagnostic checks on a predictors performance. They show that U^M and U^R tend to zero for the optimum predictor while U^D tends to one and suggest these quantities be viewed together with a "prediction realisation" diagram. This latter is a diagram of predicted values plotted against actual values, or predicted change against actual change.

We feel however that ideally one should examine the whole forecast error sequence $\{d_t\}$ and try to obtain some idea of its closeness to a "white noise" sequence. Ideally one would wish to perform a full spectral or correlogram analysis but it seems inevitable that the $\{d_n\}$ sequences obtained in practice are too short. Granger and Newbold (1972) suggest a test based on the Von Neuman ratio

$$Q = \frac{\frac{1}{n-1} \sum_{t=2}^{n} (d_t - d_{t-1})}{\frac{1}{n} \sum_{t=1}^{n} (d_t - \bar{d})^2} \qquad 1.77$$

using Harts (1942) tables. We suggest in Chapter 6 an alternative measure of "randomness" which would seem applicable here.

For an overall criteria however it would appear the most attractive approach would be to compare D_n^2 given by 1.67 against the theoretical minimum value D^2_{min}

Then the quantity

$$G_n = |D^2_{min} - D_n^2| \quad \text{or} \quad F_n = \frac{D^2_{min}}{D_n^2}$$

gives an intuitively reasonable idea of goodness of fit. A further attraction is that as we shall show D^2_{min} can be estimated directly from the $\{Y_t\}$ sequence without any model building. It would thus seem that one has an objective measure for ranking forecast procedures, especially when this is used in conjunction with an examination of the $\{d_n\}$ sequence.

It must be stressed however that even if one has constructed an adequate prediction process the forecast errors must be monitored to enable one to pick up changes in the deep structure of the underlying model which might invalidate ones forecasting approach.

We now turn to consider the optimum linear predictors (in a mean square sense) and its attractive mathematical behaviour.

CHAPTER 2

Optimum Linear Prediction

Given a stationary (discrete) process $\{X_t\}$ with zero mean and variance σ_x^2 we now consider the problem of predicting X_{n+k} given $\{X_t : t = n, n-1, n-2 \cdots\}$. We shall derive the optimum linear predictor and consider its convergence to an exact non-asymptotic solution. In addition we attempt to bring out the central role of the spectrum, particularly in view of Kolmogorov's theorems on minimum mean square error.

To obtain an estimate X^*_{k+n} of X_{t+n} we need a definition of closeness of fit of our estimate to the true value. In most published work and in the original papers of Weiner (1949) and Kolmogorov(1939, 1941) the criterion has been taken to be the minimisation of the mean square error e_k where

$$e_k = E\left[|X^*_{n+k} - X_{n+k}|^2 \right] \qquad 2.1$$

This has perhaps been used as much for its attractive mathematical properties as for being intuitively reasonable. A further compelling reason can be demonstrated. Clearly the solution of 2.1 is given by

$$X^*_{n+k} = E\left[X_{n+k} \mid X_s \;\; s = n, n-1 \cdots \right] \qquad 2.2$$

which is rather attractive. However if $\{X_t\}$ is a Gaussian process then 2.2 implies that X^*_{n+k} is a linear function of past values of the series. Thus if one assumes, as is usual, linear predictors, in the Normal case one is assured that the optimum predictor is liner. Doob (1953 sect.II) considers the geometric viewpoint.

Consider the linear form

$$X^*_{n+k} = \sum_{t=0}^{r} a_t X_{n-t} \qquad 2.3$$

where the $\{ a_i ,\ i = 0, 1 \ldots r \}$ are a set of unknown co-efficients which must be estimated from our realisation. Then the expected mean square error of prediction is

$$e_k = E\left[\left(X^*_{n+k} - X_{n+k} \right)^2 \right] \qquad 2.4$$

$$= \sigma_x^2 - 2 \sum_{j=0}^{r} a_j E\left[X_{n+k} X_{n-j} \right] + \sum_j \sum_k a_j a_k E\left[X_{n+k-j} X_{n-k} \right] \qquad 2.5$$

giving

$$e_k = \sigma_x^2 - 2 \sum_{j=0}^{r} a_j R_{k+j} + \sum_j \sum_k a_j a_k R_{j-k} \qquad 2.6$$

where $\sigma_x^2 = R_0$ denotes the variance of $\{ X_t \}$ To minimise the mean square error we choose a set of $\{ a_j \}$ which satisfy

$$\frac{\partial e_k}{\partial a_j} = 0 \qquad j = 0, 1, \ldots r \qquad 2.7$$

The equations 2.7 can be written in matrix form as

$$\begin{pmatrix} R_0 & R_1 & R_2 & \cdots & R_r \\ R_1 & R_0 & R_1 & & \vdots \\ \vdots & & & & \vdots \\ R_r & R_{r-1} & & & R_0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \\ \\ a_r \end{pmatrix} = \begin{pmatrix} R_k \\ R_{k+1} \\ \\ \\ R_{k+r} \end{pmatrix} \qquad 2.8$$

or $$\underline{M} \cdot \underline{a} = \underline{R} \qquad 2.9$$

Thus taking an autoregressive process of first order

$$X_t = \alpha X_{t-1} + \varepsilon_t$$

where the autocorrelation function is given by

$$\rho_{rr} = \alpha^{|r|} \qquad r \geq 0$$

then 2.9 becomes

$$\begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^r \\ \alpha & 1 & \alpha & \dots & \alpha^{r-1} \\ \vdots & & & & \vdots \\ \alpha^r & \alpha^{r-1} & & \alpha & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_r \end{pmatrix} = \begin{pmatrix} \alpha^k \\ \alpha^{k+1} \\ \vdots \\ \alpha^{k+r} \end{pmatrix} \qquad 2.10$$

The solution in this case being

$$a_0 = \alpha^r$$
$$a_j = 0 \qquad j \neq 0$$

and the optimum predictor of X_{n+k} is just

$$X^*_{n+k} = \alpha^k X_n \qquad 2.11$$

Thus for a first order autoregressive process the solution is simple and easily derived and all one now requires is an estimate of α^k. As we have seen in Chapter 1 the estimation of α and indeed α^k is not entirely straight forward. For the moment we shall only consider the form of the predictor, and will assume that the underlying model is known.

In general the equations 2.9 are rather more complex and analytical solutions are rather difficult to find except for auto-regressive processes. The one other case we have found quoted in the literature is for a model of the form

$$X_t = \frac{1}{\sqrt{m}} \sum_{j=0}^{m-1} \varepsilon_{t-j} \qquad 2.12$$

by Kozulyaew (1941) who derives the form of the optimum predictor.

It must be pointed out however that the previous derivation does assume that the entire part of the process is known, as may be seen from 2.2. As one usually only has a finite sample it would be inter-

esting to know what kind of error is introduced by using the "semi-infinite" form of the predictor in a finite sample case. Unforetunately in the time domain it does appear that considerable problems are introduced. However by considering the first order autoregressive case we have managed to obtain some idea of the approximations involved. We now examine this case.

Finite sample approximations

Consider $X_t = \alpha X_{t-1} + \varepsilon_t \qquad |\alpha| < 1$

which "began" at X_0 and we shall attempt to predict X_{n+k} given $\{X_t, t=0..n\}$

Then clearly

$$X_t = \varepsilon_t + \alpha\varepsilon_{t-1} + \cdots + \alpha^{t-1}\varepsilon_1 + \alpha^t X_0 \qquad 2.13$$

and we shall assume

$$E[X_0] = 0$$

$$\text{var}(X_0) = \sigma^2(X)$$

Naturally we may wish to consider a fixed starting value X_0 and this may easily be done by setting $\sigma^2(X)$ to zero

From 2.13

$$E[X_t X_{t-k}] = \alpha^k\sigma^2 + \alpha^{k+2}\sigma^2 \cdots + \alpha^{2k-k-2}\sigma^2 + \alpha^{2t-k}\sigma^2(X)$$

$$= \frac{\alpha^k\sigma^2(1-\alpha^{2(t-k)})}{1-\alpha^2} + \alpha^{2t-k}\sigma^2(X) \qquad 2.14$$

where $E[\varepsilon_t^2] = \sigma^2$ and we have assumed that X_0 is independent of all the ε_t. From 2.14 it is clear that if t is sufficiently large for

$$\alpha^{2t-k}\sigma^2(X) \to 0 \quad \text{and} \quad \alpha^{2(t-k)} \to 0$$

then

$$E[X_t X_{t-k}] = \frac{\alpha^k\sigma^2}{1-\alpha^2}$$

which gives the "semi-infinite case autocovariances

Writing 2.14 in the form

$$R_s(k) = \alpha^k X + \alpha^{-k} Y \qquad 2.15$$

where $X = \dfrac{\sigma^2}{1-\alpha^2}$ $\qquad Y = \alpha^{2k}\left(\sigma^2(X) - \dfrac{\sigma^2}{1-\alpha^2}\right)$

the equation corresponding to 2.8 became

$$\begin{pmatrix} X+Y & \alpha X + \alpha^{-1}Y & \dots & \alpha^r X + \alpha^{-r}Y \\ \alpha X + \alpha^{-1}Y & & & \\ \vdots & & & \\ & & & \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_r \end{pmatrix} = \begin{pmatrix} \alpha^k X + \alpha^{-k}Y \\ \vdots \\ \alpha^{k+r} X + \alpha^{-k-r}Y \end{pmatrix} \qquad 2.16$$

that is $\underline{M}\,\underline{a} = \underline{R}$

Thus the vector $\underline{a}$ is given by $\underline{a} = \underline{M}^{-1}\underline{R}$

where

$$\underline{M} = X\begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^r \\ \alpha & & & & \\ \vdots & & & & \\ \alpha^r & \dots & \dots & & 1 \end{pmatrix} + Y\begin{pmatrix} 1 & 1/\alpha & 1/\alpha^2 & \dots & 1/\alpha^r \\ 1/\alpha & 1 & & & \\ \vdots & & & & \\ \alpha^{-r} & \dots & \dots & & 1 \end{pmatrix}$$

$$= X\underline{A} + Y\underline{B} \qquad 2.17$$

Thus $M^{-1} = (X\underline{A} + Y\underline{B})^{-1} = \left[X\underline{A}\left(\underline{I} + \dfrac{Y}{X}\underline{A}^{-1}\underline{B}\right)\right]^{-1}$ 2.18

$$= \left(\underline{I} + \frac{Y}{X}\underline{A}^{-1}\underline{B}\right)^{-1}(XA)^{-1} \qquad 2.19$$

where $\underline{I}$ denotes the unit matrix.

Then expanding 2.19, assuming for the moment that the expansion converges we have

$$\underline{M}^{-1} = \left[\underline{I} - \frac{Y}{X}\underline{A}^{-1}\underline{B} + \left(\frac{Y}{X}\right)^2 \underline{A}^{-1}\underline{B}\underline{A}^{-1}\underline{B} + \cdots \right](X\underline{A}^{-1})^{-1}$$

$$= \frac{1}{X}\underline{A}^{-1} - \frac{Y}{X^2}\underline{A}^{-1}\underline{B}\underline{A}^{-1} + \frac{Y^2}{X^3}(\underline{A}^{-1}\underline{B})^2\underline{A}^{-1} - \cdots \qquad 2.20$$

Setting Y = 0 gives us the asymptotic case in 2.10, the remaining terms thus provide the finite sample correction

The inverse of $\underline{A}$ is easily obtained by noticing that

$$\underline{P}\underline{A}\underline{P}' = \begin{pmatrix} 1 & -\alpha & 0 & \cdots & 0 \\ 0 & 1 & -\alpha & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & \cdots & & 1 \end{pmatrix} \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^r \\ \vdots & & & & \vdots \\ \alpha^r & \cdots & & & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\alpha & 1 & \cdots & 0 \\ 0 & -\alpha & \cdots & 0 \\ & & & \\ 0 & \cdots & & -\alpha \end{pmatrix}$$

$$= \begin{pmatrix} 1-\alpha^2 & 0 & \cdots & 0 \\ 0 & 1-\alpha^2 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & & 1 \end{pmatrix} = \underline{D} \qquad 2.21$$

Thus $\underline{A}^{-1} = \underline{P}'\underline{D}^{-1}\underline{P}$

$$= \frac{1}{1-\alpha^2}\begin{pmatrix} 1 & -\alpha & 0 & \cdots & 0 \\ -\alpha & 1+\alpha^2 & \cdots & \cdots & \\ 0 & -\alpha & & & \\ \vdots & & & & \\ 0 & \cdots & & & 1 \end{pmatrix} \qquad 2.22$$

and further we can obtain $\underline{A}^{-1}\underline{B}$ in the form

$$\underline{A}^{-1}\underline{B} = \frac{1}{1-\alpha^2}\begin{pmatrix} 0 & \frac{1}{\alpha}-\alpha & \frac{1}{\alpha^2}-1 & \cdots & \frac{1}{\alpha^{r-1}}-\frac{1}{\alpha^r} \\ 0 & \alpha^2-1 & 0 & 0 & \vdots \\ \vdots & 0 & \alpha^2-1 & 0 & \vdots \\ \vdots & & & & 0 \\ \frac{1}{\alpha^{r-1}}-\frac{1}{\alpha^r} & \cdots & \cdots & & 1 \end{pmatrix}$$

and the second term in the matrix expansion 2.20 may be seen to be

$$\underline{A}^{-1}\underline{B}\underline{A}^{-1} = \frac{1}{(1-\alpha^2)^2}\begin{pmatrix} \alpha^2-1 & 0 & 0 & \cdots & 0 \\ -\alpha(\alpha^2-1) & -(1-\alpha^4) & -\alpha(\alpha^2-1) & \cdots & 0 \\ 0 & -\alpha(\alpha^2-1) & \cdots & & 0 \\ & & & & \\ -\frac{1}{\alpha^{r-3}} & +\frac{1}{\alpha^{r-1}} & 0 & \cdots & \alpha^k \end{pmatrix} \qquad 2.23$$

It is apparent that the other terms in expansion may be evaluated, at the cost of some algebra.

Thus $\underline{a}$ is given by

$$\underline{a} = \underline{M}^{-1}\underline{R}$$

from 2.20

$$\underline{a} = \frac{\underline{A}^{-1}\underline{R}}{X} + \frac{Y}{X^2}\,\underline{A}^{-1}\underline{B}^{-1}\underline{A}^{-1}\underline{R} + \cdots \qquad 2.24$$

Taking only the first term of 2.24 we have

$$\begin{aligned} a_0 &= \alpha^k \\ a_j &= 0 \qquad j \neq 0 \quad j \neq r \\ a_r &= Y\left(\alpha^{-k+r-2} - \alpha^{(k+r)}\right) \end{aligned} \qquad 2.25$$

that is

$$\begin{aligned} a_0 &= \alpha^k \\ a_r &= \frac{\alpha^{2k}}{1-\alpha^2}\left(1+(1-\alpha^2)\frac{\sigma^2(x)}{\sigma^2}\right)\left(\frac{1}{\alpha^2}-1\right)\alpha^{-k-r} \end{aligned} \qquad 2.26$$

$$a_j = 0 \quad \text{otherwise}$$

Thus taking only the first term we see that a "pertubation error" is introduced in the finite sample case.

Taking a further term we have

$$\underline{a} = \frac{\underline{A}^{-1}\underline{R}}{X} + \frac{Y}{X^2}\underline{A}^{-1}\underline{B}\underline{A}^{-1}\underline{R} \qquad 2.27$$

giving

$$a_0 = \alpha^k + \frac{1}{X}(\alpha^2-1)\left(\alpha^k + \alpha^{-k}\frac{Y}{X}\right) \qquad 2.28$$

$$a_j = 0 \qquad j \neq 0, j \neq r$$

$$a_r = \left\{ Y\alpha^{-k-r}\left(\tfrac{1}{\alpha^2}-1\right) + \left(\alpha X + \tfrac{Y}{\alpha}\right)\left(1-\tfrac{1}{\alpha^2}\right)\tfrac{1}{\alpha^{r-1}} + \alpha^2\left(X\alpha^r + \tfrac{Y}{\alpha^r}\right)\right\}\tfrac{1}{X^2} \qquad 2.29$$

the dominant term in the expansion of a_r being the order of $\alpha^{2k-k-r-2}$

We would judge that for $\alpha < 1$ two terms of the matrix expansion would provide a sufficiently accurate approximation to the finite sample predictor especially for small k and t larger than 10.

From the above it would seem apparent that the reasonable sample sizes the "semi-infinite" case predictor works well as an approximation when one has a finite sample. However in situations with large α values i.e. $1 > |\alpha| > 0.8$ and for long term forecasts or indeed value of X_0 with a large variance then it may well be worth checking the order of magnitude of the error involved.

All that now remains is to prove that the expansion of $\underline{M}^{-1}$ given in 2.20 is convergent. We do this as follows:

Let $\|A\|$ denote a norm of a matrix $\underline{A}$

i.e. $$\|A\| = \left(\sum_i \sum_j |\alpha_{ij}|^2\right)^{1/2} \qquad 2.30$$

where $\underline{A} = (\alpha_{ij})$

Then a sufficient condition for 2.20 and hence 2.24to converge is for $\| \frac{Y}{X} \underline{A}^{-1} \underline{B} \| < 1$ Wilkinson (1965) shows further that a sufficient condition for convergence is that any subordinate norm of the component matrices in the expansion should have modulus less than 1.

Consider the norm of Z defined by

$$\| Z \|_2 = \sqrt{\text{maximal eigenvalue of } \underline{Z}'\underline{Z}}$$

which is subordinate to 2.30

Then $$\| \frac{Y}{X} A^{-1} B \|_2 \leq \frac{Y}{X} \| A^{-1} \|_2 \| B \|_2 \qquad 2.31$$

and if we donate the maximal eigenvalues by λ_A^{-1} and λ_B respectively

$$\| A^{-1} B \|_2 \leq \lambda_A^{-1} \lambda_B \qquad 2.32$$

since $\underline{A}^{-1}$ and $\underline{B}$ are symmetric.

From 2.21 the eigenvalues of A are 1 and $1 - \alpha^2$ thus $\lambda_A^{-1} = \frac{1}{1-\alpha^2}$

Similarly the eigenvalues of B are 1 and $1 - 1/\alpha^2$ thus

$$\frac{Y}{X} \lambda_A^{-1} \lambda_B = \alpha^{2t-2} \left(1 + \frac{\sigma^2(X)(1-\alpha^2)}{\sigma^2} \right) \qquad 2.33$$

and in consequence $\| \frac{Y}{X} \underline{A}^{-1} \underline{B} \|_2 < 1$ for suitable values of Thus our matrix expansion converges. The dependence on $\sigma^2(X)$ would appear reasonable for a short series while for $t \to \infty$ convergence is certain.

From the brief outline given above it does seem that the analytic solution of the prediction problem as presented above is intractable. One must also bear in mind that we have assumed that the correlation structure is known and we have used exact values of the auto-

correlations, while in practice we can at most expect only sample values.

As an alternative to obtaining an analytic solution we could estimate the covaniance matrix in 2.8 and obtain a solution of the equation using a numerical equation solving algorithm While it is quite feasible to handle large matrices in this fashion the errors in estimation of high order autocorrelation introduce insuperable accuracy and stability problems.

Spectral Formulation

It is quite possible and as we shall see very convenieht to reformulate the prediction problem into spectral terms. In what follows we outline some main results and methods available. It should be noticed that we assume the spectral density function exists and that it is a rational function. These restrictions can be lifted if generality is sought but we confine an outline of the more general methods to Appendix I.

For algebraic simplicity we reformulate our problem slightly to consider the prediction of X_{n+k} given $\{ X_{n-j} , j = 1, 2 \ldots \}$ Then the minimisation equations corresponding to 2.8 become

$$R_{k+m} = \sum_{s=1}^{r} a_s R_{m-s} \qquad m = 1, 2 \ldots r \qquad 2.34$$

and using the spectral representation (2.11) we have

$$\int_{-\pi}^{\pi} \left\{ e^{i(k+m)\omega} - \sum_{s=1}^{r} a_s e^{i(m-s)\omega} \right\} f(\omega)\, d\omega \qquad m = 1, 2 \ldots r \qquad 2.35$$

Equivalently

$$\int_{-\pi}^{\pi} e^{im\omega} \left\{ e^{ik\omega} - \Phi_{k,r}(\omega) \right\} f(\omega)\, d\omega = 0 \qquad m = 1, 2 \ldots r \qquad 2.36$$

where $$\Phi_{k,r}(\omega) = \sum_{s=1}^{r} a_s e^{-is\omega} \qquad 2.37$$

Our problem is now that of finding the function $\underline{\Phi}_{k,r}(\omega)$ In addition to 2.36 we can also express the mean-square error in terms of the $\underline{\Phi}_{k,r}(\omega)$ as

$$\sigma_{k,r} = \int_{-\pi}^{\pi} |e^{ik\omega} - \Phi_{k,r}(\omega)|^2 f(\omega)\, d\omega \qquad 2.38$$

where $$e_{k+1} = \min(\sigma_{k,r})$$

Following Yaglom (1962) we let $n \to \infty$ and consider the problem of finding the function $\Phi_k(\omega)$ where

$$\Phi_k(\omega) = \sum_{j=1}^{\infty} a_j e^{-i\omega j} \qquad 2.39$$

where $$\lim_{n\to\infty} \int_{-\pi}^{\pi} |\Phi_k(\omega) - \sum_{s=1}^{r} a_s e^{-is\omega}|^2 f(\omega) d\omega = 0 \qquad 2.40$$

As is apparent we are assuming the entire part of the sequence is known and that it may be used to predict future values. The corresponding mean-square error, at the minimum is given by

$$e^2_{k+1} = \int_{-\pi}^{\pi} |e^{ik\omega} - \Phi_k(\omega)|^2 f(\omega) d\omega \qquad 2.41$$

We have thus reduced our problem to finding the function $\underline{\Phi}_k(\omega)$ if it exists. In general this is a very difficult problem, first solved by Kolmogorov (1941a). The outlines presented in appendix one give some idea of the difficulties. However if one places restrictions on $\Phi_k(\omega)$ and assumes $f(\omega)$ is a rational function of exponentials then Yaglom (1962) shows how one may find $\Phi_k(\omega)$ given the spectral density. For more general results see Grenander and Szego (1958), Doob (1953) or Grenander & Rosenblatt (1957).

This approach while mathematically appealing does present formidable difficulties when the spectrum is unknown. However allied to this approach is a result of considerable value upon which much of our later work is based. This is Kolmogorovs (1939) result that

$$e_1 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(\omega) d\omega \right\} \qquad 2.42$$

If the process is detminstic and can be predicted perfectly by a linear combination of past values then

$$e_1 = \cdots 0$$

As can be seen from Appendix I that a necessary and sufficient condition for a stationary process with an absolutely continuous spectrum to be representable as a one sided (backwards in time) moving average is that it is non-deterministic. See also Anderson (1971).

It may be helpful to consider $\{X_t\}$ written as

$$X_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \theta(B)\varepsilon_t \qquad 2.43$$

For simplicity let $b_0 = 1$, this does not involve any less of generality.

Then

$$f_x(\omega) = \frac{\sigma^2}{2\pi} \theta(e^{i\omega}) \theta(e^{-i\omega}) . \qquad 2.44$$

Then

$$\int_{-\pi}^{\pi} \log 2\pi f_x(\omega) d\omega = \int_{-\pi}^{\pi} \log \sigma^2 d\omega + \int_{-\pi}^{\pi} \log \theta(e^{-i\omega}) d\omega + \int_{-\pi}^{\pi} \log \theta(e^{i\omega}) d\omega . \qquad 2.45$$

Let the polynomials $\theta(z)$ have zeros at β_j, $j = 1, \cdots$ etc.

Then $$\int_{-\pi}^{\pi} \log \theta(e^{i\omega}) d\omega = \sum_{r=1}^{\infty} \int_{-\pi}^{\pi} \log(1 - \beta_r e^{i\omega}) d\omega \qquad 2.46$$

If $|\beta_j| < 1$ for all j then $\log \theta(z)$ is analytic in $|z| < 1$

and

$$\int_{-\pi}^{\pi} \log(1 - b_s e^{i\omega}) d\omega = \int_{-\pi}^{\pi} \left\{ -b_s e^{i\omega} - \left(\frac{b_s e^{i\omega}}{2}\right)^2 \cdots \right\} d\omega \qquad 2.47$$

Now all the terms of 2.47 involving exponentials vanish giving

$$\int_{-\pi}^{\pi} \theta(e^{i\omega}) d\omega = 0 \qquad 2.48$$

and similarly $$\int_{-\pi}^{\pi} \theta(e^{-i\omega}) d\omega = 0$$

Thus $$\int_{-\pi}^{\pi} \log 2\pi f(\omega) d\omega = 2\pi \log \sigma^2 \qquad 2.49$$

and thus $$\sigma^2 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) 2\pi \, d\omega \right\} \qquad 2.50$$

or equivalently

$$\sigma^2 = 2\pi \exp\left\{ \int_{-\pi}^{\pi} \log f(\omega) d\omega \right\} \qquad 2.51$$

Clearly the minimum mean square error of prediction is given in this case by the 'innovation variance' σ^2. It is clear that if $b_0 \neq 1$ then

$$b_0^2 \sigma^2 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(\omega) d\omega \right\} \qquad 2.52$$

The moving average representation does enable one to obtain the optimium linear predictor very easily.

Consider $\{ Y_t \}$ being non-deterministic, then

$$X_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \qquad 2.53$$

and we assume a linear form of the predictor

$$X^*_{t+m} = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u}$$

This is reasonable since we may write

$$X^{*}_{t+m} = \sum_{j=0}^{r} a_j X_{t-j}$$

$$= \sum_{j=0}^{r} a_j \sum_{k=0}^{\infty} b_k \varepsilon_{t-j-k}$$

$$= \sum_{u=0}^{\infty} g_u \varepsilon_{t-u} \qquad 2.54$$

Then 2.55

$$E\left[\left(X^{*}_{t+m} - X_{t+m}\right)^2\right] = E\left[\left\{\sum_{u=0}^{\infty} g_u \varepsilon_{t-u} - \sum_{u=0}^{\infty} b_u \varepsilon_{t+m-u}\right\}^2\right]$$

$$= E\left[\left\{\sum_{u=0}^{\infty} (g_u - b_{u+m})\varepsilon_{t-u} - \sum_{u=0}^{m-1} b_u \varepsilon_{t+m-u}\right\}^2\right]$$

$$= \sum_{u=0}^{\infty} (g_u - b_{u+m})^2 \sigma^2 + \sum_{u=0}^{m-1} b_u^2 \sigma^2 \qquad 2.56$$

Clearly 2.56 is minimised by setting

$$g_u = b_{u+m} \qquad u = 0, 1, 2$$

and the mean square error of prediction is given by the last term of 2.56

$$e_m = \sigma^2 \sum_{u=0}^{m-1} b_u^2 \qquad 2.57$$

i.e. $e_1 = \sigma^2 b_0$

Further if we consider the Fourier expression of $\log f(\omega)$ ($\log f(\omega)$ being bounded)

$$\log f(\omega) = a_0 + a_1 \cos\omega + a_2 \cos 2\omega + \cdots \qquad 2.58$$

then it may be shown that

$$e^{(a_1 z + a_2 z^2 + \cdots)/2} = b_0 + b_1 z + \cdots$$

which is the result quoted by Kolmogorov (1939), see also Doob (1953, p.577).

From 2.58

$$a_r = \frac{1}{\pi}\int_0^{\pi} \log f(\omega)\cos r\omega\, d\omega \qquad r = 0, 1\ldots$$

then after some manipulation we can find the moving average co-efficients b_j.

While much has been written on the subject of linear prediction when the entire past of a process is known little has been done in the case when only a finite sample of part values is available. This is especially so when the minimum mean square error is considered.

The sharpest large sample results appear to be those of Grenander & Szego (1958), and we summarise these here.

Let e_1 denote the minimum mean square error using the whole part of the series and let $e_{1,N}$ that when only $X_1 \ldots X_N$ are used. Then

(a) A necessary and sufficient condition that $\delta_N = e_{1,N} - e_1$ decreases at least exponentially to zero as n tends to infinity is that f(w) coincides in $[-\pi, \pi]$ almost everywhere with a function which is analytic for real ω and has no real zeros.

(b) Let f(w) be defined a.e. by the expression

$$f(\omega) = p(\omega)\prod_{\nu=1}^{s} |e^{i\omega} - e^{i\omega_\nu}|^{2\alpha_\nu}$$

where $p(\omega) > 0$ with $k \geq 4$ integrable derivatives, α_ν are positive integers and ω_ν are distinct points in $[-\pi, \pi]$.

Then if $S = 0$ and f(w) has no zeros

$$\delta_N = o(N^{S/2-k})$$

and if $S > 0$

$$\delta_N = O(N^{-1})$$

In the case of the first order moving average process they show that

$$\delta_N = \frac{1}{N+1}$$

Ibragimov (1964) has shown that δ_N decreases to zero slower than $N^{-(1+\varepsilon)}$ for any $\varepsilon > 0$ when f(w) has zeros or is unbounded. In addition Ibragimov & Solev (1967) have shown that for a spectral density function of the form

$$f(\omega) = p_1(\omega) \prod_{\nu=1}^{M} |e^{i\omega_\nu} - e^{i\omega}|^{2\alpha_\nu}$$

where $p_1(\omega)$ satisfies a Lipschitz condition of order $\alpha \geq 1/2$ and is strictly positive

then

$$0 < \underset{N\to\infty}{\underline{\lim}} \; \delta_N / 1/N \leq \underset{N\to\infty}{\overline{\lim}} \; \frac{\delta_N}{1/N} < \infty$$

Neither of these last two papers are of much practical value.

Further results were presented by Davidson (1965) who considered linear predictors of the form

$$X_0^* = \sum_{r=1}^{N} \alpha_r X_{-r}$$

to predict X_0 where $f(\omega) = |g(e^{i\omega})|^2$. He showed that

$$e_{1N} = -\rho_{-k} a_k$$

where f(w) is expressed as

$$f(\omega) = \sum_{-k}^{k} \rho_k e^{i\omega k} \qquad k < \infty$$

Thus for

(i) $f(\omega) = |e^{i\omega} - 1|^{2M}$

$$e_{1,N} = 1 + O(1/N)$$

and

(ii) $f(\omega) = |(e^{i\omega} - e^{i\theta})(e^{i\omega} - e^{-i\theta})|^2 \quad -\pi < \theta < \pi$

$$e_{1,N} = 1 + \frac{2}{N+1}$$

(iii) $f(\omega) = |1 - be^{i\omega}|^4 \quad |b| < 1$

$$e_{1N} = 1 + b^{2(N+1)}(1 - b^4)$$

Of rather more interest is his derivation of an upper bound for δ_N. If $f(\omega) \neq 0$ and if there exists a $g(z) \neq 0$ for $1 < |z| \leq R$ then

$$\delta_N \leq \sum_{m=0}^{N} \sum_{n=N+1}^{\infty} \max \left| \frac{g(0)}{g(z)} \right|^2 \frac{|b_{m-n}|}{R^{m+n}}$$

and $b_k = R_k$

Thus for $f(\omega) = |(1 - 0.5e^{i\omega})(1 - 0.25e^{i\omega})|^2$

we have

$$\delta_N \leq 120(1/3)^N$$

These results do have the drawback that f(w) must be completely specified. However it does appear that for reasonable length series we can fairly assume that

$$\delta_N = O(N^{-1})$$

We have not touched on the prediction of deterministic processes so far in our account. In this field little theoretical work has been done but there are some published results e.g. Rosenblatt (1957) on prediction when the spectrum is identically zero over an interval. Our interest is in non-deterministic processes and it is on these we shall concentrate.

CHAPTER III

The Estimation of the Mean Square Error

Time series models are often constructed for the purpose of making forecasts and indeed it can be argued that the only true assessment of the fit of a particular model is its predictive power. To assess the predictive power of a model is difficult problem and we feel that perhaps the only rational criteria is to compare the mean square error of our fitted model with that of the optimum linear predictor e_1. Indeed one can imagine situations when e_1 is too large to make prediction worth while at all.

If we wish to proceed on these lines it is clear that we need some method of producing reliable estimates of e_1 for a non-deterministic process and the construction of such estimates is the next step we consider.

As we have seen, the minimum mean-square error of prediction using a linear function of part values of a time series $\{Y_t\}$ can be expressed as

$$e_1 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi f(\omega)\, d\omega\right\} \qquad 3.1$$

under suitable regularity conditions. We shall now attempt to exploit this link between the spectrum and e_1 by estimating e_1 using suitable estimates for the spectrum. As we shall see we can obtain quite satisfactory estimates of the e_1. For simplicity however we shall concentrate on the estimation of

$$\Delta = \log e_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi f(\omega)\, d\omega \qquad 3.2$$

using a sample of a realisation from a second order stationary process $\{X_t : t=1,2..N\}$ Clearly we must first consider the estimation of f(w).

As we have seen in Chapter I all the spectral estimates which have been used to date may be written in the form

$$\hat{f}(w) = \frac{1}{2\pi} \sum_{s=-m}^{m} \lambda_s \hat{R}_s \cos sw \qquad 3.3$$

where m is a parameter and $\{\lambda_s\}$ a suitable weighting sequence. The form of the approximation admits at most m independent estimates of f(w) (Whittaker(1935)) and it is convenient for our purposes to consider these to be at the $w_i = \frac{2\pi i}{m}$ $i = 0, \pm 1, \pm 2 \ldots \pm [\frac{m}{2}]$ where $[x]$ denotes the largest integer smaller than x . As a consequence we need to evaluate the integral in 3.2 using a finite set of values approximating f(w). Thus we must consider a finite sum approximating the integral. This can be done in many ways as can be seen from the literature on numerical analysis, for example Hildebrand (1956), HMSO (1961) and Clenshaw and Curtis (1960). The obvious candidates amongst numerical integration methods are Simpsons Rule and one of the Gaussian quaderature methods. The drawback of Simpsons rule would appear to be the introduction of rather a complex correlation struture. The Gaussian expansions appear to be unpredictable for high order approximations and involve evaluating the spectrum at points which are not easily calculated.

We must also bear in mind that spectral estimates are themselves wildly inaccurate by numerical analysis standards and we feel that there is much to be gained in simplicity in using the simple Reimann sum approximation.

$$\hat{\Delta} \simeq \frac{2}{M} \sum_{p=1}^{\lfloor M/2 \rfloor} \log \hat{f}(\omega_p) \qquad \omega_p = \frac{2\pi p}{M} \qquad p = 0 \cdots \left[\frac{M}{2}\right] \quad 3.4$$

This can be modified to the "extended trapizodal rule"

$$\hat{\Delta} = \frac{2}{M} \left\{ \sum_{p=1}^{\left[\frac{M}{2}\right]+1} \log \hat{f}(\omega_p) + \frac{1}{2} \log \hat{f}(0) + \frac{1}{2} \log \hat{f}(\pi) \right\} \quad 3.5$$

The advantage of this latter form is that it can provide an explicit error bound by application of the Euler-Maclaren summation formula as shown by Whittaker and Watson (1965 p.128). Naturally this bound is obtained on the assumption that $\hat{f}(\omega)$ approximates to f(w) with an arbitarily high degree of accuracy but nevertheless it does provide a useful guideline.

In addition to the questions of quaderature we also require estimates of f(w). In addition to the standard estimates using Parzen or Tukey weights we have the raw periodogram and as we shall see it can provide consistent estimates. This might be expected since integration is a smoothing operation while in addition it is generally held that logarithmic transformations stabilise variances. It might be added that direct methods of smoothing the periodogram using the fast Fourier transform of Cooley & Tukey (1965) based on the algorithm of Singleton (1968) are also candidates, although their attraction in their ease of computation rather than any theoretical advantage.

We shall consider first the properties of the periodogram $I_N(\omega, x)$ given by

$$I_N(\omega, x) = \frac{2}{N} \left| \sum_{t=1}^{N} X_t e^{i\omega t} \right|^2 \qquad 3.6$$

and we consider

$$\hat{\Delta} = \frac{2}{M} \sum_{p=0}^{\left[\frac{M}{2}\right]} \log I_N(\omega_p, x) \qquad \omega_p = \frac{2\pi p}{M}$$

where $m = N$ and the prime denotes that the first and last terms have been divided by 2.

Our investigation of the behaviour of the periodogram starts with Kolmogorov's theorem (Appendix I) from which we know that any non-deterministic process can be represented as a moving average viz:

$$X_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j} \qquad 3.7$$

where the $\{\varepsilon_t\}$ process is one of uncorrelated random variables with zero mean, variance σ^2 and finite fourth cumulant K_4. Then if the periodogram of the $\{\varepsilon_t\}$ is written

$$I_N(\omega, \varepsilon) = \frac{2}{N} \left| \sum_t \varepsilon_t e^{-i\omega t} \right|^2 \qquad 3.8$$

we have Bartletts (1955) result

$$I_N(\omega, \varepsilon) = 2\pi f(\omega) I_N(\omega, \varepsilon) + O(N^{-1/2}) \qquad 3.9$$

see Hannan (1960 p 55)

Further Hannan (1960) shows that

$$E\left[|R_N|^{2k}\right] = E\left[\left| I_N(\omega, x) - 2\pi f(\omega) I_N(\omega, \varepsilon)\right|^{2k}\right] = O(N^{-k}) \qquad 3.10$$

provided K_4 is finite and

$$\sum_{j=0}^{\infty} |\beta_j| j^{1/2} < \infty \qquad 3.11$$

Olshen (1967) in a very illuminating paper on the periodogram derives similar expressions for $J_N(\omega, x)$ and $J_N(\omega, \varepsilon)$ where

$$\left| J_N(\omega, x) \right|^2 = I_N(\omega, x) \qquad 3.12$$

The consequence of 3.10 is that one can in effect work in terms of the residual series $\{\varepsilon_t\}$ which makes derivations considerably easier

Now

$$E[I_N(\omega,\varepsilon)] = \frac{2}{N}\left|\sum_{t=1}^{N}\varepsilon_t e^{i\omega t}\right|^2$$

$$= \frac{2}{N}\sum_{t=1}^{N}\sum_{s=1}^{N} e^{i\omega(t-s)}E[\varepsilon_t\varepsilon_s] = 2\sigma^2 \qquad 3.13$$

and

$$E[I_N(\omega_p,\varepsilon)\,I_N(\omega_q,\varepsilon)]$$

$$= \frac{4K_4}{N} + 4\sigma^2 + \frac{4\sigma^4}{N}\left\{\frac{\sin N(\omega_p+\omega_q)/2}{\sin(\omega_p+\omega_q)/2}\right\}^2$$

$$+ \frac{4\sigma^4}{N}\left\{\frac{\sin N(\omega_p-\omega_q)/2}{\sin(\omega_p-\omega_q)/2}\right\}^2 \qquad 3.14$$

and hence if ω_j of the form $\omega_j = \frac{2\pi j}{N}$

$$\operatorname{cov}[I_N(\omega_p,\varepsilon)\,I_N(\omega_q,\varepsilon)] = \frac{4K_4}{N} \qquad \omega_p \neq \omega_q \qquad 3.15$$

$$\operatorname{var}[I_N(\omega_p,\varepsilon)] = 4\sigma^2 + \frac{4K_4}{N} \qquad \omega \neq 0,\pi$$

$$= 8\sigma^2 + \frac{4K_4}{N} \qquad \omega = 0,\pi \; . \qquad 3.16$$

Thus we see that at the points $\omega = 0,\pi$ the variance of $I_N(\omega,\varepsilon)$ doubles and so does that of $I_N(\omega,x)$ producing what might be thought of as an edge effect since the spectrum is undefined outside $[-\pi,\pi]$ It is also clear that when $K_4 = 0$ the periodogram estimates at intervals of $\frac{2\pi}{N}$ are uncorrelated.

Further if we define

$$A(\omega) = \sqrt{\frac{2}{N}} \sum_{t=1}^{N} \varepsilon_t \cos \omega t \qquad 3.17$$

$$B(\omega) = \sqrt{\frac{2}{N}} \sum_{t=1}^{N} \varepsilon_t \sin \omega t \qquad 3.18$$

then if $E[\varepsilon_t] = 0 \quad E[\varepsilon_t \varepsilon_s] = \sigma^2 \delta_{t,s}$

$$E[A(\omega)] = E[B(\omega)] = 0$$

and $$E[A^2(\omega)] = E[B^2(\omega)] = \sigma^2 \qquad 3.21$$

since

$$\frac{1}{n} \sum_{k=1}^{n} \sin^2 kx = \frac{n}{2} + O\left(\frac{1}{n}\right)$$

$$\frac{1}{n} \sum_{k=1}^{n} \cos^2 kx = \frac{n}{2} + O\left(\frac{1}{n}\right)$$

in addition

$$E[A(\omega_p) B(\omega_q)] = \frac{2\sigma^2}{N} \sum_{t=1}^{N} \sin \omega_p t \cos \omega_q t$$

$$= 0$$

for $\omega_p = \frac{2\pi p}{N}$, $\omega_q = \frac{2\pi q}{N}$

since $$\sum_{k=1}^{n} \sin kx = \sin\frac{(n+1)x}{2} \sin\frac{nx}{2} \operatorname{cosec}\frac{x}{2}$$

Thus when the $\{\varepsilon_t\}$ are normally distributed each of the A(w) and B(w) have independent $N(0,\sigma^2)$ distributions and in consequence

$$I_N(\omega, \varepsilon) = A^2(\omega) + B^2(\omega) \qquad 3.22$$

has a $\sigma^2 \chi^2_2$ distribution. From this and 3.9 we see that has a distribution which approximates $2\pi f(\omega)\sigma^2 \chi^2_2$. In addition we see from 3.21 that the $I_N(\omega_p, x)$ are independent.

Similar results can be obtained under weaker conditions as to the behaviour of the $\{\varepsilon_t\}$ We state a theorem a proof of which is given by Olshen (1967).

If $\{X_t\}$ is a second-order stationary process which is non-deterministic and where represented in the form 3.7 the $\{\varepsilon_t\}$ obey the central limit theorem then the joint distribution of

$$I_N(\omega_j, x) \quad , \quad \omega = \frac{2\pi}{N} j \qquad j = 1 \ldots k$$

tends to that of k independent random variables each having a $2\pi f(\omega)\sigma^2 \chi^2_2$ distribution if $\omega \neq 0, \pm\pi$ and to $4\pi f(\omega)\chi^2_1$ in the special cases $\omega = 0, \pm\pi$.

Having obtained the distribution of $I_N(\omega, x)$ we are now in a position to look more closely at

$$\hat{\Delta} = \frac{1}{n} \sum_{k=0}^{n}{}' \log I_N(\omega_k, x) .$$

It should be recalled that

$$\sum_{k=0}^{p}{}' a_k = \frac{a_0}{2} + \frac{a_p}{2} + \sum_{k=1}^{p-1} a_k .$$

Then from 3.9

$$\hat{\Delta} \simeq \frac{1}{n} \sum_{k=0}^{n}{}' \left[\log 2\pi f(\omega) + \log I_N(\omega_k, \varepsilon) \right]$$

$$= \frac{1}{n} \sum_{k=0}^{n}{}' \log 2\pi f(\omega) + \frac{1}{n} \sum_{k=0}^{n}{}' \log I_N(\omega_k, \varepsilon) \qquad 3.23$$

To consider the distribution of $Y_k = \log I_N(\omega_k, x)$ we require the standard transformation theorem outlined below.

Let Y = g(X) where X is a one dimensional continuous random variable with a probability density function f(x) and g(x) is a strictly monotonic function such that $g'(x)$ exists and is non-zero everywhere except possibly at a finite number of points.

Further let

$$\lim_{x \to -\infty} g(x) = a$$

$$\lim_{x \to \infty} g(x) = b$$

Then Y = g(X) is a continuous random variable with a probability density function given by

$$h(y) = \frac{f(x)}{\left|\frac{dy}{dx}\right|} \qquad a < y < b$$

$$0 \qquad \text{elsewhere}$$

Using the above when X is χ^2_2 then

$$f(x) = \frac{e^{-x/2}}{2} \qquad x > 0$$

and Y = $\log_e X$ has a probability density function given by

$$h(y) = \frac{e^y}{2} \exp\left[-\frac{1}{2}e^y\right] \qquad -\infty < y < \infty \quad \ldots\ldots 3.24$$

Since it involves little extra labour we can consider a chi-squared distribution with ν degrees of freedom, where

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{(\nu/2-1)x} e^{-x/2} \qquad x > 0 \qquad 3.25$$

In this case Y = $\log_e$ X has a distribution given by

$$h(y) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{\nu y/2} \exp\left[-\frac{1}{2}e^y\right] \qquad 3.26$$

This degree of generality will prove useful later.

Since we are dealing with sums of independent random variables we find it useful to obtain the characteristic function of 3.26 as

$$\phi_y(t) = E[e^{itY}] = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_{-\infty}^{\infty} e^{\nu y/2} e^{ity} \exp[-\tfrac{1}{2}e^y]\,dy$$

$$= \frac{2^t}{\Gamma(\nu/2)} \int_0^{\infty} z^{(it+\nu/2-1)} e^{-z}\,dz$$

TEXT CUT OFF IN THE ORIGINAL

Using the substitution $x = z/2$ this can be seen to be the Euler integral of the second kind defining the Gamma function (Whittaker and Watson (1965)).

Thus we have

$$\phi(t) = \frac{2^{it}}{\Gamma(\nu/2)} \Gamma(it + \nu/2) \qquad 3.27$$

and the characteristic function of the sum of M random variables having this distribution is

$$\phi(t) = \frac{2^{iMt}}{\Gamma(\nu/2)^M} \left[\Gamma(it + \nu/2)\right]^M \qquad 3.28$$

Neglecting for the moment the end points of $\hat{\Delta}$ and considering

$$L_S = \frac{1}{M} \sum_{k=1}^{M} \log I_N(\omega_k, c)$$

then the characteristic function of L is given by

$$\phi_{L_S}(t) = \frac{2^{it}}{\Gamma(\nu/2)^M} \left[\Gamma\left(\frac{it}{M} + \nu/2\right) \right]^M \qquad 3.29$$

If one considers the end points where the degrees of freedom are halved one has for

$$L = \frac{1}{M} \sum_{k=0}^{M}{}' \log I_N(\omega_k, c) \qquad 3.30$$

$$\phi_L(t) = \frac{2^{it(m-1)/m}}{\Gamma(\nu/2)^{m-1}} \left[\Gamma\left(\frac{it}{M} + \nu/2\right)\right]^{M-1} \left\{ \frac{2^{it/2m}}{\Gamma(\nu/2)} \Gamma\left(\frac{it}{2m} + \nu/4\right) \right\}^P \qquad \ldots 3.31$$

Where p = 2 if both end points are included and p = 1 if only one is included.

To find the distribution of L or L' all we require is the Fourier transform of either 3.29 or 3.31. We have been unable to

evaluate these transformations and are of the opinion that no analytic form of the inverse can be found. Since we cannot use a direct approach we are forced to consider some other technique to find the appropriate distribution function. One possibility is the numerical inversion of the characterstic function for fixed ν and m using a suitable quaderature technique. In principle this would seem quite feasible, however for any theoretical treatment this would seem quite unsatisfactory, and in addition one is left with the problem of determining the accuracy of one's solution. We shall pursue alternative approaches in later chapters, for the moment we shall examine the moments of the distribution of L_S and L'.

Using the well known expansion

$$\phi(t) = 1 + i\mu t + \cdots \qquad 3.32$$

this is quite straightforward, however we shall find it convenient to work in terms of the cumulant generating function

$$K(t) = \log \phi(t)$$

and the corresponding cumulants $\kappa_1, \kappa_2 \cdots$ etc.

In the case of L_s we have

$$K_{L_S}(t) = it \log 2 + m \log\left[\Gamma\left(\frac{it}{m} + \nu/2\right)\right] - \log \Gamma(\nu/2) \qquad \ldots 3.33$$

Since the gamma function is continuous and possesses continuous derivatives of all orders

$$\frac{\partial K}{\partial t} = i \log 2 + \frac{i\,\Gamma'\left(\frac{it}{m} + \nu/2\right)}{\Gamma\left(\frac{it}{m} + \nu/2\right)} \qquad 3.34$$

and hence

$$\kappa_1 = \log 2 + \frac{\Gamma'(\nu/2)}{\Gamma(\nu/2)} \qquad 3.35$$

$$\kappa_2 = \frac{1}{m}\left\{\frac{\Gamma''(\nu/2)}{\Gamma(\nu/2)} - \left[\frac{\Gamma'(\nu/2)}{\Gamma(\nu/2)}\right]^2\right\} \qquad 3.36$$

and

$$K_r = \frac{1}{(i)^r}\left[\frac{\partial^r K}{\partial t^r}\right]_{t=0} \qquad 3.37$$

To evaluate expressions of the form 3.34 one can use the Weirstraus definition of Γ as an infinite product.

$$\log \Gamma(z) = -\log z - \gamma z - \sum_{r=1}^{\infty}\left[\log\left(1+\frac{z}{r}\right)-\frac{z}{r}\right] \qquad 3.38$$

where γ is Euler's constant and obtain

$$K_s = \frac{1}{m^{s-1}} \sum_{r=0}^{\infty} (-1)^s \frac{(s-1)!}{(r+\nu_2)^s} \qquad s \geq 2 \qquad 3.39$$

since $\log \Gamma(z)$ is represented by an absolutely and uniformly convergent sequence of analytic functions.

The K_2 can be easily computed from 3.39 to any reasonable accuracy since for large s the series converges quickly. For small s we can use the Zeta function of Reimann

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \qquad 3.40$$

which has been tabulated by Abramowitz and Segun (1968).

Alternatively we can usefully use the Psi-function

$$\psi(z) = \frac{d}{dz}\left[\log \Gamma(z)\right] \qquad 3.41$$

and it derivatives, which are tabulated by Abramowitz and Segun (1968). These polygamma functions give us the very neat relations

$$K_1 = \log 2 + \psi(\nu_2) \qquad 3.42$$

$$K_2 = \frac{1}{m}\psi'(\nu_2) \qquad 3.43$$

$$K_3 = \frac{1}{m^2}\psi''(\nu_2) \qquad 3.44$$

$$-(m-1)\log\Gamma(\upsilon/2) - p\log\Gamma(\upsilon/4)+\{\frac{it(m-1)}{m}+\frac{itp}{2m}\}\log 2$$

$$+(m-1)\log\Gamma(\frac{it}{m}+\frac{\upsilon}{2}) + p\log\Gamma(\frac{it}{2m}+\frac{\upsilon}{4})$$

$$=\kappa_{L_S}(t) + \frac{it}{2m}(p-2)\log 2 + \log\Gamma(\upsilon/2) - p\log\Gamma(\upsilon/4)$$

$$-\log\Gamma(\frac{it}{m}+\frac{\upsilon}{2}) + p\log\Gamma(\frac{it}{2m}+\frac{\upsilon}{4}) ,$$

$$K_r = \frac{1}{m^{r-1}} \psi^{(r-1)}(v/2) \qquad 3.45$$

which determine the cumulants of L_S.

For L' we have a similar set of relations. The cumulant generating function is

As given on the opposite page 3.46

giving as cumulants

$$\kappa_1 = \left(\frac{2m+p-2}{2m}\right) \log 2 + \left(\frac{m-1}{m}\right)\psi\left(\frac{v}{2}\right) + \frac{p}{2m}\psi\left(\frac{v}{4}\right) \qquad 3.47$$

$$\kappa_2 = \left(\frac{m-1}{m^2}\right)\psi'\left(\frac{v}{2}\right) + \frac{p}{(2m)^2}\psi'\left(\frac{v}{4}\right) \qquad 3.48$$

$$\kappa_3 = \left(\frac{m-1}{m^3}\right)\psi''\left(\frac{v}{2}\right) + \frac{p}{(2m)^3}\psi''\left(\frac{v}{4}\right) \qquad 3.49$$

$$\kappa_4 = \left(\frac{m-1}{m^4}\right)\psi'''\left(\frac{v}{2}\right) + \frac{p}{(2m)^4}\psi'''\left(\frac{v}{4}\right) \qquad 3.50$$

and $$\kappa_s = \frac{m-1}{m^s}\psi^{(s-1)}\left(\frac{v}{2}\right) + \frac{p}{(2m)^s}\psi^{(s-1)}\left(\frac{v}{4}\right) \qquad 3.51$$

Having obtained these cumulants we can if we wish obtain the raw moments or central moments by using the relation

$$K(t) = \log \phi(t) \qquad 3.52$$

Kendall and Stuart (1963) provide tables from the expansion of 3.52 giving

$$\mu_r' = \sum_{m=0}^{r} \sum \left(\frac{K_{p_1}}{p_1!}\right)^{\pi_1}\left(\frac{K_{p_2}}{p_2!}\right)^{\pi_2}\cdots\left(\frac{K_{p_m}}{p_m!}\right)^{\pi_m}\frac{r!}{\pi_1!\cdots\pi_m!} \qquad \text{..... } 3.53$$

where the second summation extends over all non-negative values of the π such that $\sum_{j=1}^{m} \pi_j p_j = r$

or alternatively

$$\frac{\partial \mu_r'}{\partial \kappa_j} = {}^{r}C_j \, \mu'_{r-j} \qquad 3.54$$

For convenience we provide some explicit relations for low order cumulants and moments

For the raw moments

$$\begin{aligned}
\mu_1' &= \kappa_1 \\
\mu_2' &= \kappa_2 + \kappa_1^2 \\
\mu_3' &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3 \\
\mu_4' &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4 \\
\mu_5' &= \kappa_4 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 10\kappa_3\kappa_1^2 \\
&\quad + 15\kappa_1^2\kappa_1 + 10\kappa_2\kappa_1^3 + \kappa_1^5
\end{aligned} \qquad 3.55$$

while for the moments about the mean

$$\begin{aligned}
\mu_2 &= \kappa_2 \\
\mu_3 &= \kappa_3 \\
\mu_4 &= \kappa_4 + 3\kappa_2^2 \\
\mu_5 &= \kappa_5 + 10\kappa_3\kappa_2
\end{aligned} \qquad 3.56$$

A property of the cumulants we shall use later is that, apart from the first they are invariant under a change in origin. In fact for any linear transformation of Z say

$$W = aZ + b$$

the cumulants of W are given by

$$\kappa_r(w) = a^r \kappa_r(z) \qquad r > 1 \qquad 3.57$$

Having obtained the cumulants of the L_s or L' we can now obtain the cumulants

of $$\hat{\Delta} = \frac{1}{M} \sum_{k=0}^{M} \log 2\pi f(\omega) + L \qquad 3.58$$

or $$\hat{\Delta}_s = \frac{1}{m}\sum_{k=1}^{m} \log 2\pi f(\omega) + L_s \qquad 3.59$$

Since the cumulants are invariant under a change of origin we need consider only the first cumulant

For $\hat{\Delta}$ we have

$$K_1 = E[\hat{\Delta}] = \frac{1}{m}\sum_{k=0}^{m}{}' \log 2\pi f(\omega) + E[L] \qquad 3.60$$

and from 3.47 for L

$$K_1 = \left(1 + \frac{P}{2m}\right)\log 2 + \psi(\tfrac{1}{2}) + \frac{P}{m}\psi(\tfrac{1}{4}) \qquad 3.61$$

Now we know that $I_N(\omega, \varepsilon)$ is chi-squared with two degrees of freedom, that is $\nu = 2$

Then for L

$$K_1 = \left(1 + \frac{P}{2m}\right)\log 2 + \psi(1) + \frac{P}{2m}\psi(\tfrac{1}{4}) \qquad 3.62$$

Letting $m \to \infty$ we have

$$E[\hat{\Delta}] = K_1 \simeq \log 2 + \psi(1) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi f(\omega)\, d\omega \qquad 3.63$$

and

$$\mathrm{var}(\hat{\Delta}) = K_2 = \frac{1}{m}\psi'(1) + \frac{P}{4m^2}\psi'(\tfrac{1}{2})$$

which in the limit tends to zero.

Thus $\hat{\Delta}$ is a biased estimator of e_1 and can be seen to be inconsistent. However, it can be easily modified to

$$\hat{\Delta}_c = \hat{\Delta} - \log 2 - \psi(1) \qquad 3.64$$

$\hat{\Delta}_c$ is clearly asympotically unbiased and a consistent estimator of e_1.

In fact we shall prefer to use

$$\hat{\Delta}_c = \hat{\Delta} - \left(1 + \frac{P}{2m}\right)\log 2 - \psi(1) - \frac{P}{2m}\psi(\tfrac{1}{2}) \qquad 3.65$$

and we shall in future refer to this estimate as $\hat{\Delta}_I$ The evaluation of the constant terms in 3.65 involves no problems since

$$\psi(1) = -\gamma = -0.577216$$

and

$$\psi(\tfrac{1}{2}) = -\gamma - 2\log 2 = -1.963510$$

where γ denotes Euler's constant which may be defined as

$$\gamma = \lim_{n\to\infty}\left\{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right\}.$$

When m is sufficiently large then the terms in $\frac{1}{m}$ corresponding to the "end corrections" of the trapizodal estimate will tend to zero and as an approximation we may use the expression given in 3.64 and

$$\operatorname{var}(\hat{\Delta}_I) \simeq \frac{\psi'(1)}{m} = \frac{1.644934}{m} \qquad 3.66$$

It is of interest to note that we may write

$$\psi'(1) = \pi^2/6$$

since

$$\psi^{(n)}(1) = (-1)^{n+1} n!\, \zeta(n+1) \qquad n = 1, 2, \ldots$$

where $\zeta(z)$ is the Reimann Zeta function and

$$\zeta(2n) = \frac{(2\pi)^{2n}}{2(2n)!}\,|B_{2n}| \qquad 3.67$$

B_{2n} being one of the Bernoulli numbers. For details see Abramowitz & Segun (1968).

If we let $m \to \infty$ then our estimate 3.65 is in effect the sum of independent identically distributed random variables and from the central limit theorem we may deduce that the distribution of $\hat{\Delta}_I$ tends to that of a normal variate with mean $\log e_1$ and variance given by

$$\frac{\psi'(1)}{m} = \pi^2/6m \qquad 3.68$$

Thus $\hat{e}_1 = e^{\hat{\Delta}_I}$ would be expected to be log-normal.

While asympotic results are useful one really needs some information as to how quickly the distribution of $\hat{\Delta}_I$ tends to normality. Indeed one needs to know if the asympotic results can be used at all. For $\hat{\Delta}_I$ we can obtain some idea of the rates of convergence by considering the moments of its distribution, which as we have seen are readily available.

Since the odd moments of the normal distribution vanish it would appear sensible to consider the third moment of our estimate.

Now since $\mu_3 = \kappa_3$ we have

$$\kappa_3 = \frac{\psi''(1)}{m^2} \qquad 3.69$$

neglecting the terms in m^3 in

$$\kappa_3 = \frac{\psi''(1)}{m^2} + \frac{P}{8m^3}\psi''(1/2) \qquad 3.70$$

we have from 3.69

$$\kappa_3 = -\frac{2.40411}{m^2}$$

referring to Abromowitz and Segun (1968).

For m of the order of 100 κ_3 is very small however there does seem to be a criteria for the "smallness" of κ_3. Perhaps a more sensible idea is to consider the Pearson system of probability distributions which are solutions of

$$\frac{1}{y}\frac{dy}{dx} = \frac{-(x+c_1)}{c_0 + c_1 x + c_2 x^2} \qquad 3.71$$

The form of the solutions depends upon the values of the parameters c_0, c_1, c_2 and we can relate these to the moments of the solution as follows:

Let $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ $\beta_2 = \frac{\mu_4}{\mu_2^2}$ 3.72

then $$c_0 = \frac{\sigma^2(4\beta_2 - 3\beta_1)}{2(5\beta_2 - 6\beta_1 - 9)}$$

$$c_1 = \frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

$$c_2 = \frac{\sigma(2\beta_2 - 3\beta_1 - 6)}{2(5\beta_2 - 6\beta_1 - 9)}$$

where σ^2 denotes the variance of the distribution.

The solutions of 3.71 include a wide variety of interesting distributions amongst which is the Normal distribution. This being the solution when $\beta_1 = 0$ and $\beta_2 = 3$. We can compare our distribution with the normal by looking at

$$\beta_1 = [\psi''(1)]^2 / [\psi'(1)]^3_m \qquad 3.73$$

$$\beta_2 = [\psi'''(1)] / [\psi'(1)]^2_m + 3 \qquad 3.74$$

Directly in terms of the cumulants

$$\beta_1 = \frac{\kappa_3}{\kappa_2^{3/2}} \qquad \beta_2 = \kappa_4 / \kappa_2^2 \qquad 3.75$$

Table I contains tables of values of β_1 and β_2 for a range of values of m and figure 1 provides a diagrammatic representation. The coefficient β_1 is generally regarded as being a measure of skewness, see Kendall & Stuart (1963) and as we can see in table 1 the distribution of $\hat{\Delta}_I$ is skewed but as $m \to \infty$ tends to a symmetrical distribution.

For a more precise comparison we can consider the published values of the percentage points of Pearson curves given in Johnson et al (1963). These tables give the standardised percentage points of a number of Pearson curves corresponding to a range of values of $\sqrt{\beta_1}$ and β_2.

TABLE 1

m	$-\beta_1$	β_2
20	0.2548	3.1200
50	0.1612	3.0480
60	0.1417	3.0400
70	0.1362	3.0344
80	0.1274	3.0300
90	0.1201	3.0264
100	0.1140	3.0240
120	0.0963	3.0172
160	0.0901	3.0150
180	0.0849	3.0132
200	0.0806	3.0120
250	0.0721	3.0096
300	0.0658	3.0080
400	0.0570	3.0060
500	0.0510	3.0048
600	0.0465	3.0040
700	0.0431	3.0034
800	0.0403	3.0030
900	0.0380	3.0026
1000	0.0360	3.0024

For $\beta_2 = 3$ we have given a small set of values in table 2.

TABLE 2

Percentage points of Pearson curves for

	Lower				Upper			
$\sqrt{\beta_1}$	0.5	1.0	2.5	5	0.5	1	2.5	5
0	-2.58	-2.33	-1.96	-1.64	2.58	2.33	1.96	1.64
0.1	-2.82	-2.53	-2.10	-1.74	2.21	2.05	1.79	1.55
0.05	-2.76	-2.48	-2.06	-1.71	2.33	2.14	1.84	1.58
0.03	-2.72	-2.44	-2.04	-1.69	2.39	2.19	1.87	1.59
0.01	-2.66	-2.40	-2.01	-1.67	2.48	2.25	1.91	1.62

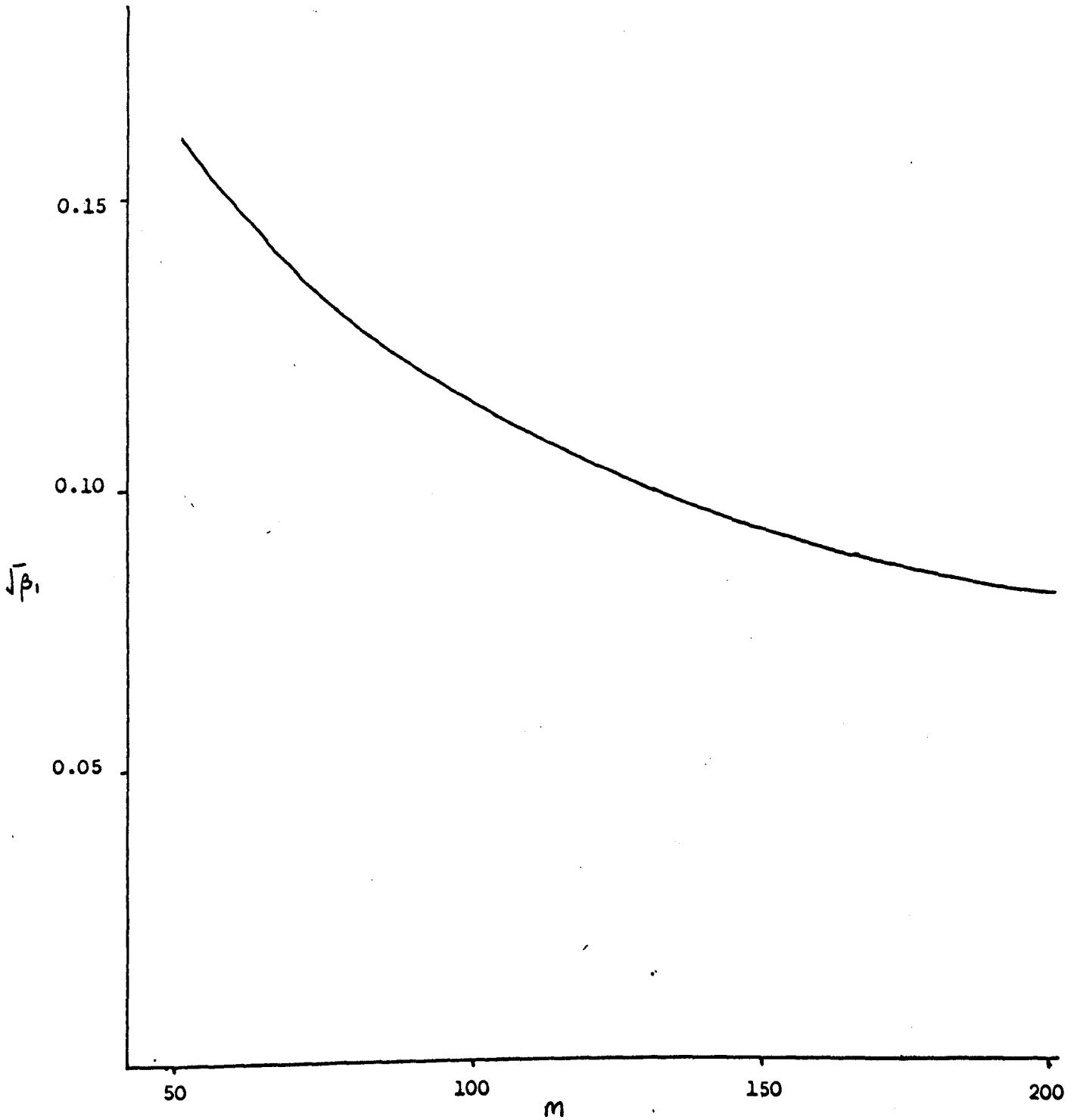
As can be seen from Table 2 even for small non-zero values of $\sqrt{\beta_1}$ there is a considerable difference in the tails of distributions when compared with the normal, while for $\sqrt{\beta_1} > 0.05$ there is a marked difference in the 5% point. It should also be born in mind that we have not considered differences in the β_2 parameter.

We conclude that while $\hat{\Delta}_I$ tends to normality for small values of m, say $m < 300$ which corresponds to realisation of less than 600 in length we cannot expect very accurate approximations using the normal distribution. Thus it would appear that for realisations of less than 200 in length we require the exact distribution.

For large samples however, we can estimate $\log e_1$ and e_1 and we know the distribution of our estimate. Using $\hat{\Delta}_I$ as we shall see it is easy to construct a test for white noise against any specific alternative. In addition estimates of e_1 are of great value

FIG. 1

Values of $\sqrt{\beta_1}$ for values of m



if one is interested in the estimation of models of the form

$$A(B^{-1})X_t = \theta(B) \qquad 3.76$$

and in comparison of the performance of predictors.

Estimates Using the Spectrum

Before looking further into such problems it is worth considering alternative estimators of Δ. We can if we wish estimate $\log_e$, by using the same numerical form as 3.5 but using a different estimator of f(w). Since the usual spectral estimator is a smooth periodogram (cf.1.28) we shall consider estimates of the form

$$\hat{f}(\omega) = \frac{1}{2\pi}\sum_{s=-m}^{m} \lambda_s \hat{R}_s \cos s\omega \qquad 3.77$$

where the $\{\lambda_s\}$ are a suitably chosen set of weights where

$$W(\theta) = \sum_{s=-\infty}^{\infty} \lambda_s e^{-is\theta} \qquad 3.78$$

Of the many suggested sequences we shall concentrate on that suggested by Parzen (1961)

$$\lambda_s = \begin{cases} 1 - 6\left(\frac{s}{m}\right)^2 + 6\left|\frac{s}{m}\right|^3 & \left|\frac{s}{m}\right| < 0.5 \\ 2\left(1 - \left|\frac{s}{m}\right|\right)^3 & 0.5 < \left|\frac{s}{m}\right| < 1 \\ 0 & \left|\frac{s}{m}\right| > 1 \end{cases}$$

and adapted from the Jackson-de la Vallee Poussin Kernal used in approximation theory, see Parzen (1963). Our main reason for this is that while it is one of the most used kernals when combined with autocovariance estimates of the form

$$\hat{R}_s = \frac{1}{N}\sum_{t=1}^{N-|s|} X_t X_{t+|s|}$$

the resulting estimates of f(w) are non-negative, Granger (1964). Since we need to consider $\log f(w)$ positive definateness seems a necessary

requirement. Of course one may adapt ones procedure to use a Tukey or Bartlett weight function but it is then necessary to impose an arbitary value on any negative estimates and gives rise to many complications. For simplicity we shall choose the Parzen weighting sequence although in principle the same derivations can be carried through for any sequence which gives rise to positive definate estimates.

We now look at the distributional properties of such spectral estimates. Now f(w) must be a function of $X_1 \ldots X_n$ and we can write the function as a quadratic form

$$\hat{f}(w) = \underline{X}' W \underline{X}$$

where $\underline{X}' = (X_1 X_2 \ldots X_n)$ and the matrix $\underline{W}$ is non-negative. To attempt to obtain explicit form of the distribution function we shall derive the characteristic function. Assuming the $\{X_t\}$ are normal variates then the characteristic function of

$$Q = \sum_{i,j} \omega_{ij} X_i X_j \qquad 3.79$$

where $W = (\omega_{ij})$ is given by

$$\phi_n(z) = \det | \underline{I} - 2iz\underline{R} W |^{-1/2} \qquad 3.80$$

as may be seen in Cramer (1946). Here $\underline{R}$ denotes the autocovariance matrix. If we denote the eigenvalues of $\underline{R}\,\underline{W}$ by $\lambda_1, \ldots \lambda_n$ we can write 3.80 as

$$\phi_n(z) = \prod_{\nu=1}^{n} | 1 - 2iz\lambda_\nu |^{-1/2} \qquad 3.81$$

or

$$\log \phi_n(z) = -\frac{1}{2} \sum_{\nu=1}^{n} \log(1 - 2iz\lambda_\nu) \qquad 3.82$$

This is not very useful since the λ_ν are in general not known since $\underline{R}$ is not known. We can however obtain the first moments

of f(w) from 3.82 as

$$\mu = E[\hat{f}(\omega)] = \sum_{\nu=1}^{n} \lambda_\nu \qquad 3.83$$

$$\sigma^2 = \mathrm{var}(\hat{f}(\omega)) = 2\sum_{\nu=1}^{n} \lambda_\nu^2 \qquad 3.84$$

The evaluation of these moments still entails knowledge of the eigenvalues or alternatively the trace of $\underline{R}\,\underline{W}$ and $(\underline{R}\,\underline{W})^2$

However the above has been used to derive some asymptotic results. Using the fact that

$$\underline{R} = \left(R_{\nu-\mu} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(\nu-\mu)\omega} f(\omega)\, d\omega \;;\; \nu,\mu = 1,2..n \right)$$

is a Toeplitz matrix Grenander and Szegö (1958) show that if

$$Y = \frac{\hat{f}(\omega) - \mu}{\sigma}$$

then the characteristic function of Y is given by

$$\psi(z) = \prod_{\nu=1}^{n} \left| 1 - 2iz\frac{\lambda_\nu}{\sigma} \right|^{-\frac{1}{2}} \left(\frac{izm}{\sigma} \right) \qquad 3.85$$

or

$$\log \psi(z) = -\frac{1}{2}\sum_{\nu=1}^{n} \log\left| 1 - 2iz\frac{\lambda_\nu}{\sigma} \right| + \frac{izm}{\sigma} \qquad 3.86$$

Thus

$$\log \psi(z) = +\frac{1}{2}z^2 + d_n \qquad 3.87$$

and

$$d_n \le A|z^3|\sigma^{-3}\sum_{\nu=1}^{n} |\lambda_\nu|^3 \qquad 3.88$$

and using the analytic theory of Toeplitz forms they show that when

$$0 < c_1 < f(\omega) < c_2$$

$$0 \le \omega_{ij} \le c_3$$

for some constants c_1, c_2, c_3 then $d_n = O(n^{-\frac{1}{2}})$

and

$$\lim_{n\to\infty} \psi(z) = e^{-\frac{1}{2}z^2}$$

Thus the normalised spectral estimate is asymtotically normal. This has also been suggested by Grenander and Rosenblatt (1957) and Lomnicki & Zaremba (1959) amongst others. Grenander and Szego try sharpening their approximation and to some extent do achieve this aim, however their later development cannot be put into a practical framework. Some closely related work was done by Kac (1954) whose results are closely paralled by the others mentioned above.

Several authors have looked at the problem of inverting the characteristic function given by 3.82 that is evaluating the integral.

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i.2x}\prod_{\nu=1}^{n}(1-2iz\lambda_\nu)^{-1/2}\,dz \qquad 3.89$$

Since the integrand of 3.89 is an analytic function of z except at the branch points the integration contour can be deformed into a set of circles enclosing pairs of branch points if n is even. By collapsing the circles one can obtain an expression for g(x) as a sum of finite integrals, viz:

$$g(x) = \frac{1}{\pi}\sum_{r=1}^{[\frac{n}{2}]}(-1)^{r+1}\int_{(2\lambda_{2\nu-1})^{-1}}^{(2\lambda_{2\nu})^{-1}} e^{-yx}\prod_{j=1}^{n}(1-2y\lambda_j)^{-1/2}dy. \qquad \ldots\ldots\ 3.90$$

This technique has been used to facilitate numerical evaluation by Slepian (1958). Robbins (1948) has suggested a power series expansion for $g(z)$ as well as a representation as a mixture of χ^2 distributions while Gurland (1955) has developed an expression in terms of Laguerre polynomials. All of these expansions however appear to converge slowly. Grenander, Pollack and Slepian (1959) discuss an integral equation method and some further Toeplitz approximations, however in practise one feels that their results while interesting, are not applicable, in general,

especially when one wishes to use standard spectral windows.

A common approximation, Blackman & Tukey (1958) or Granger (1964), is to assume that the f(w) have a distribution that is approximately χ^2. The rational being that since

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} I_N(\omega_p, x)\, W(\theta-\omega)\, d\theta \qquad 3.91$$

and hence

$$\hat{f}(\omega) \simeq 2\pi f(\omega) \int_{-\pi}^{\pi} I_N(\omega, \epsilon)\, W(\theta-\omega)\, d\omega \qquad 3.92$$

Now the integral in 3.92 is essentially a weighted sum of χ^2 variates when the ε_t are normal or at least follow the central limit theorem. Clearly the integral can, by following the arguments of Welch (1947) can have its distribution approximated by a χ^2 distribution.

Since the number of degrees of freedom ν of a χ^2_ν variate is given by

$$\nu = \frac{2\mu^2}{\sigma^2}$$

we can define the "equivalent degrees of freedom of f(w) as

$$\nu = \frac{2E\left[\hat{f}(\omega)\right]^2}{\mathrm{var}\left[\hat{f}(\omega)\right]} \sim 2N\left[2\pi \int_{-\pi}^{\pi} W^2(\theta-\omega)\, d\omega\right]^{-1} \quad \ldots 3.9$$

This quantity is also sometimes known as the "stability" of our estimate.

It is of some interest that Freiberger and Genander (1959) in a paper which considers the use of Toeplitz matrices in investigating distributions of quadratic forms also use a χ^2 approximation for a rectangular window $W(\theta)$ of the type proposed by Daniel (1946).

TEXT CUT OFF IN THE ORIGINAL

We shall in our future consideration of spectral estimates use the χ^2 approximation for its distribution. One further problem we must consider is the correlation that exists between spectral estimates at differing frequencies, unlike the case of the periodogram

We may see

$$\operatorname{cov}\left[\hat{f}(\omega_1)\hat{f}(\omega_2)\right] = \frac{1}{4\pi^2}\sum\sum \lambda_r \lambda_s \operatorname{cov}(\hat{R}_r \hat{R}_s)\cos r\omega_1 \cos s\omega_2 \quad \ldots 3.94$$

$$= \frac{\pi}{N}\int_{-\pi}^{\pi} f^2(\theta)\left[W(\theta-\omega_1)+W(\theta+\omega_1)\right]\left[W(\theta-\omega_2)+W(\theta+\omega_2)\right]d\theta + K_4 f(\omega_1) f(\omega_2) \quad \ldots 3.95$$

and asympotically the covariance tends to zero. However for sample sizes considered in practice one must choose estimates whose arguments are sufficiently different to minimise correlation. Table 3 gives the correlation between spectral estimates when the underlying series is white noise. This table is easily computed since

$$\operatorname{cov}\left[\hat{f}(\omega_1)\hat{f}(\omega_2)\right] = \left(\sum_{-m}^{m} \lambda_r^2 \cos r\omega_1 \cos r\omega_2\right) \Big/ \left(\sum_{r=-m}^{m} \lambda_s^2 \cos^2 s\omega\right) \quad \ldots 3.96$$

Clearly the correlation will depend on the "bandwidth" or how spiked the function $W(\theta)$ happens to be as may be seen from 3.95 From table 4 which gives the bandwidth for three differing spectral windows it is clear that while the Parzen window has smaller variance, this has to be paid for by accepting a larger bandwidth. Hence one has a greater degree of correlation and in consequence some loss of "focusing power" of the spectral window. This is easily seen in the values presented in table 3. One might expect that estimates of f(w) separated by at least one bandwidth would be only very slightly correlated. As can be seen for the Tukey window with bandwidth $\pi/5$ the correlation is 0.1667 while at this value the Parzen estimate has correlation of 0.4020. Looking at the bandwidth value for the Parzen

TABLE 3

This table gives the correlation between spectral estimates for a Normal white noise process using 3.92 with M = 10. Source Fishman (1969)

Separation	Correlation		Separation		
	Tukey	Parzen		Tukey	Parzen
0.0	1.000	1.000			
0.1	0.9961	0.9978	2.6	0.0292	0.2100
0.2	0.9843	0.9911	2.7	0.0184	0.1850
0.3	0.9650	0.9802	2.8	0.0101	0.1620
0.4	0.9385	0.9650	3.0	0.0000	0.1410
0.5	0.9054	0.9458	3.1	- 0.0026	0.1219
0.6	0.8663	0.9228	3.2	- 0.0040	0.1048
0.7	0.8220	0.8964	3.3	- 0.0045	0.0894
0.8	0.7734	0.8667	3.4	- 0.0044	0.0758
0.9	0.7213	0.8342	3.5	- 0.0038	0.0637
1.0	0.6667	0.7993	3.6	- 0.0030	0.0533
1.1	0.6105	0.7623	3.7	- 0.0021	0.0442
1.2	0.5537	0.7236	3.8	- 0.0013	0.0365
1.3	0.4972	0.6873	3.9	- 0.0006	0.0300
1.4	0.4417	0.6429	4.0	0.0000	0.0202
1.5	0.3881	0.6017	4.1	0.0004	0.0166
1.6	0.3369	0.5605	4.2	0.0006	0.0137
1.7	0.2888	0.5196	4.3	0.0006	0.0114
1.8	0.2442	0.4794	4.4	0.0006	0.0096
1.9	0.2034	0.4401	4.5	0.0005	0.0083
2.0	0.1667	0.4020	4.6	0.0003	0.0013
2.1	0.1340	0.3655	4.7	0.0002	0.0065
2.2	0.1054	0.3306	4.8	0.0001	0.0060
2.3	0.0809	0.2975	4.9	0.0000	0.0057
2.4	0.0602	0.2663	5.0	0.0000	0.0057
2.5	0.0431	0.2371			

FIG. 2

CORRELATIONS FROM TABLE 3

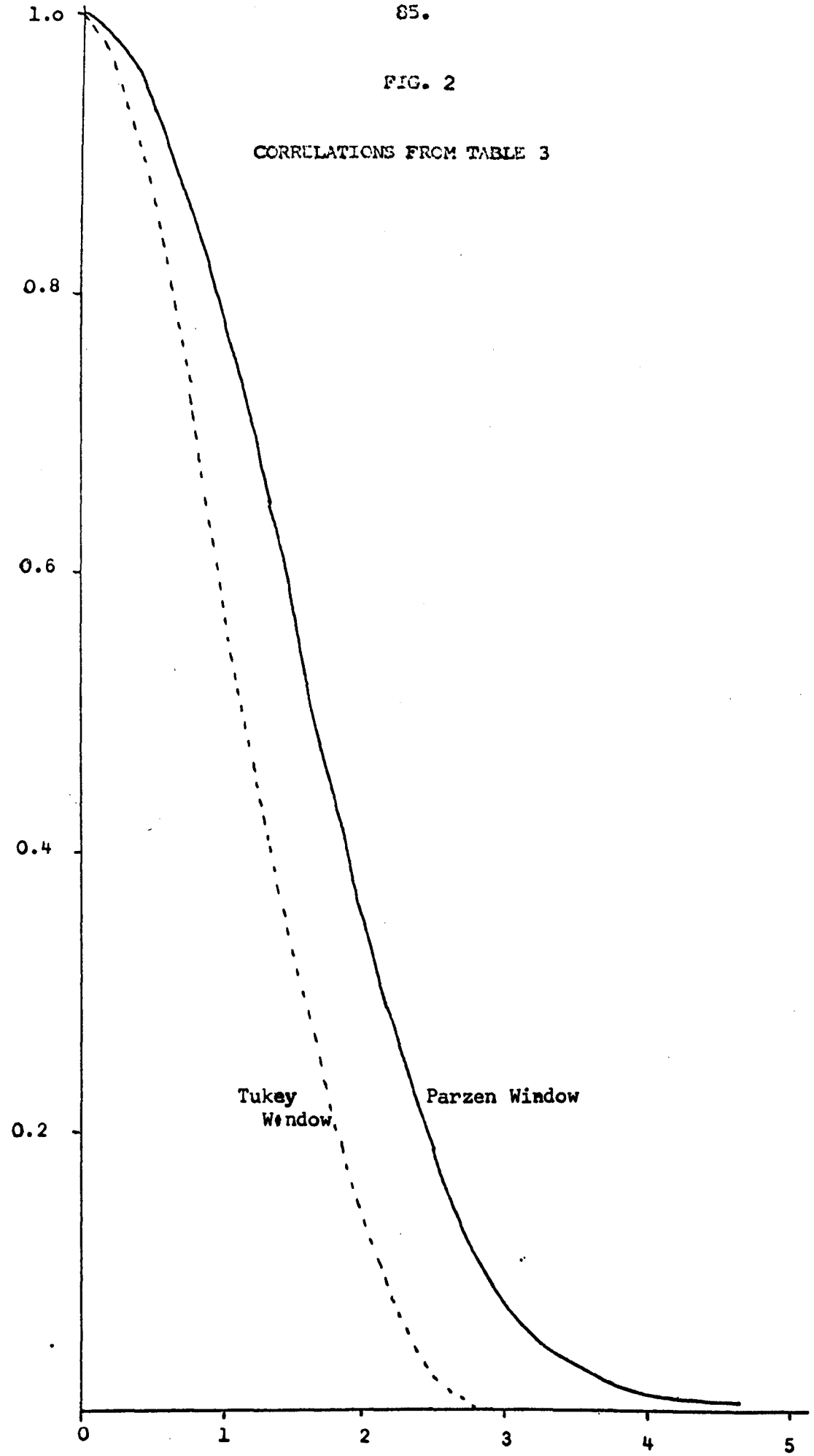


TABLE 4

Bandwidth, Variance and Equivalent degrees of freedom for same windows $W(\theta)$. N denotes the length of the realisation and M the truncation point.

W(O)	Bandwidth	Variance/$f(\omega)^2$	E.D.F.
Unit	$\pi/(M+\frac{1}{2})$	$2.000\ M/N$	N/M
Tukey	$2\pi/M$	$0.750\ M/N$	$2.7\ N/M$
Parzen	$8\pi/3M$	$0.539\ M/N$	$3.7\ N/M$

window the correlation drops to 0.1830.

From the above one may conclude that if one has spectral estimates separated by at least one bandwidth then one can assume that the estimates are approximately uncorrelated.

Having established the results above we can now persue our estimation of $\log e_1$

Let $$\hat{\Delta}_f = \frac{2}{M} \sum_{p=0}^{M}{}' \log 2\pi \hat{f}(\omega_p) \qquad 3.97$$

where $$\omega_p = \frac{2\pi p}{M} \qquad p = 1 \ldots \left[\frac{M}{2}\right]$$

where M is chosen such that the correlation between $f(\omega_p), f(\omega_{p+1})$ is minimised, that is $\frac{2\pi}{M}$ is at least, the bandwidth of the Parzen window. Then from 3.91 we have

$$\hat{f}(\omega) = f(\omega) Y(\omega)$$

where Y is a random variable with χ^2_ν distribution, ν being the equivalent degrees of freedom of our estimate. Thus from 3.97

$$\hat{\Delta}_f \simeq \frac{2}{M} \sum_{p}{}' \log 2\pi f(\omega_p) + \frac{2}{M} \sum_{p}{}' \log Y(\omega_p) \qquad 3.98$$

Notice that as in the case of the periodogram we have a doubling of the variance at $\omega = 0, \pm\pi$

If we assume the $Y(\omega_p)$ $p = 0, 1 \ldots$ are independent χ^2 variates we can proceed in exactly the same way as for $\hat{\Delta}_I$

Thus if one has a sufficiently large value of M that the end points of the sum may be ignored then

$$H_s = \frac{2}{M} \sum_{p=0}^{M}{}' \log Y(\omega_p) \qquad 3.99$$

has a characteristic function given by

$$\phi_{H_S}(t) = \frac{2^{it}}{\Gamma(\nu/2)} \left[\Gamma\left(\frac{it}{m} + \nu/2\right) \right]^m \qquad 3.100$$

while using $H = \frac{2}{m} \sum_{p=0}^{\lfloor m/2 \rfloor}{}' \log Y(\omega_p)$ the characteristic function becomes

$$\phi_H(t) = \frac{2^{it}}{\Gamma(\nu/2)} \left[\Gamma\left(\frac{it}{m} + \nu/2\right)^{m-1} \right] \left\{ \frac{2^{it/2m}}{\Gamma(\nu/2)} \Gamma\left(\frac{2it}{m} + \nu/2\right) \right\}^p \qquad 3.101$$

where $p = 1, 2$ depending if one or two end points are included.

As in the case of the periodogram it is clear that the characteristic function cannot be inverted directly and one needs must find some indirect approach. The cumulants and consequently the moments can easily be found as before

For H_S

$$\begin{aligned} K_1 &= \log 2 + \psi(\nu/2) \\ K_2 &= \frac{1}{m} \psi'(\nu/2) \\ &\vdots \\ K_r &= \frac{1}{m^{r-1}} \psi^{(r-1)}(\nu/2) \end{aligned} \qquad 3.102$$

while in the case of H we have the rather more complex formulation

$$\begin{aligned} K_1 &= \left(1 + \frac{p}{2m}\right) \log 2 + \psi(\nu/2) + \frac{p}{2m} \psi(\nu/4) \\ K_2 &= \frac{1}{m} \psi'(\nu/2) + \frac{p}{4m^2} \psi'(\nu/4) \\ &\vdots \\ K_r &= \frac{1}{m^{r-1}} \psi^{(r-1)}(\nu/2) + \frac{p}{2^r m^r} \psi^{(r-1)}(\nu/4) \end{aligned} \qquad \cdots 3.103$$

The computation of the moments is thus quite routine, and given the equivalent degrees of freedom one can find tabulated values of the Polygamma functions in Abromowitz and Segun (1968).

As is evident from the above our estimate is biased but as in the case of our previous estimate $\hat{\Delta}_I$ we can make the following modification

$$\hat{\Delta}_f = \hat{\Delta}_f - K_1 \qquad 3.104$$

where K_1 is defined as in 3.103 above. From now on we shall use only the unbiased form in 3.104 and for convenience shall denote it by $\hat{\Delta}_f$

From 3.104 the variance of $\hat{\Delta}_f$ is given by

$$K_2 = \frac{1}{m}\psi'(\nu/2) + \frac{2p}{4m^2}\psi'(\nu/4)$$

and as $m \to \infty$ $K_2 \to 0$ implying that $\hat{\Delta}_f$ is a consistent estimator with mean

$$K_1 = \frac{2}{m}\sum_{p=0}^{m/2} \log 2\pi f(\omega_p) \simeq \frac{1}{\pi}\int_0^{\pi} \log 2\pi f(\omega)\, d\omega$$

since $\psi'(z)$ is a bounded function.

For large values of m $\hat{\Delta}_f$ can be considered as a set of independent, identically distributed random variables. This follows from the arguments stated above that $\hat{\Delta}_f$ is approximately normally distributed and from the results of Grenander and Rosenblatt (1957) that

$$\lim_{m \to \infty} \operatorname{cov}\left[\hat{f}(\omega_1)\hat{f}(\omega_2)\right] \to 0 \qquad \omega_1 \neq \omega_2$$

Thus we can by appealing to the central limit theorem assume that has an asympotic normal distribution.

This results is not of great interest as it stands for we really need to know how fast the time distribution converges to a normal form when $m \to \infty$. As before by considering the third and fourth moments, or more usefully β_1 and β_2 where

$$\beta_1 = k_3^2 / k_2^3$$

$$\beta_2 = k_4 / k_2^2$$

we have, for large m

$$\beta_1 \simeq \frac{\psi''(\nu/2)^2}{m\psi'(\nu/2)^{3/2}} \qquad 3.105$$

$$\beta_2 = 3 + \frac{\psi'''(\nu/2)}{m\psi'(\nu/2)} \qquad 3.106$$

where we recall m is the number of lags used in the estimation of

For a Parzen window the equivalent degrees of freedom are given by

$$\nu = 3.7\, N/M \qquad 3.107$$

using the classical variance estimate. However if one uses the modification due to Neave (1966) discussed in Chapter 1 then this needs to be slightly modified. Since we are using approximations we shall use the classical result of 3.107 and table 5 gives the values of $\sqrt{\beta_1}$ and β_2 for three values of m for convenient values of the polygamma functions.

From the table we can see that $\sqrt{\beta_1}$ tends to zero and β to 3 as m increases. The other comparison we may wish to make is with our

estimate $\hat{\Delta}_F$ to do so we must bear in mind that m is the number of values of $\log f(\omega)$ included in $\hat{\Delta}_F$. Using a $M = \frac{N}{3}$ lags to estimate the spectrum with a separation of say $\frac{2\pi}{M}$ then $m = N/6$ while for the corresponding periodogram estimate $\hat{\Delta}_I$ $m = N/2$ and reference to the table 5 shows that $\hat{\Delta}_S$ is marginally better. It is more realistic to consider a separation of $\frac{3\pi}{M}$ in which case for $\hat{\Delta}_S$, $m = N/9$ and the difference is very small. However if one wishes to choose $M = N/4$ or less then the position may change completely. We must also remember that we have not considered the correlation between our estimates.

It is of interest to consider the asympotic variance of $\hat{\Delta}_F$ For a Parzen window, from table 4 $\nu = 3.7\frac{N}{M}$ thus since for large z

$$\psi'(z) = \frac{1}{z}$$

$$\text{var}(\hat{\Delta}_F) = \frac{M}{2NM} = \frac{1}{2KM} \qquad 3.108$$

while

$$\text{var}(\hat{\Delta}_I) = 1.6/N \qquad KM = N \qquad 3.109$$

Thus if we choose a minimum separation of $3\pi/M$ for our smoothed spectral estimates $M = M/3$

and $\text{var}(\hat{\Delta}_F) = 1.5/N$

the implication being that there is little to choose in the efficiency of the two estimators.

We must however admit some preference towards the estimate $\hat{\Delta}_I$ It would appear to offer great simplicity and in addition the estimates of the periodogram are at least independent. For small sample sizes we have to remember that the smoothed spectral estimates are correlated and as we shall see the effects of this correlation can be quite marked.

TABLE 5

Polygamma functions for 3 arguments

N/M	ν	$\psi'(\nu/2)$	$\psi''(\nu/2)$	$\psi'''(\nu/2)$
4	14.8	0.14468	-0.02089	0.00603
3	11.1	0.191196	-0.038846	0.01524
2	7.4	0.31004	-0.09540	0.05828

Table of values of $\sqrt{\beta_1}$ and β_2

		$\sqrt{\beta_1}$			β_2	
m	ν 14.8	11.1	7.4	14.8	11.1	7.4
20	0.08495	0.10387	0.17411	3.0144	3.02085	3.03032
50	0.05373	0.06569	0.11012	3.0058	3.00834	3.01213
80	0.04273	0.05193	0.08706	3.0036	3.00521	3.00758
100	0.03799	0.04645	0.07787	3.0029	3.00417	3.00606
200	0.02687	0.03285	0.05505	3.0014	3.00209	3.00303
250	0.02402	0.02938	0.04924	3.0015	3.00167	3.00243
300	0.02193	0.02681	0.04496	3.0010	3.00139	3.00202
350	0.02029	0.02482	0.04162	3.0008	3.00119	3.00173
400	0.01900	0.032322	0.03894	3.0007	3.00104	3.00152
500	0.01700	0.02078	0.03483	3.0006	3.00083	3.00121
600	0.01552	0.01897	0.03178	3.0005	3.00070	3.00101
700	0.01435	0.01755	0.02943	3.0004	3.00060	3.00087
800	0.01342	0.01643	0.02753	3.0004	3.00052	3.00076
900	0.01265	0.01549	0.02596	3.0003	3.00046	3.00067
1000	0.01200	0.01470	0.02462	3.0003	3.00042	3.00061

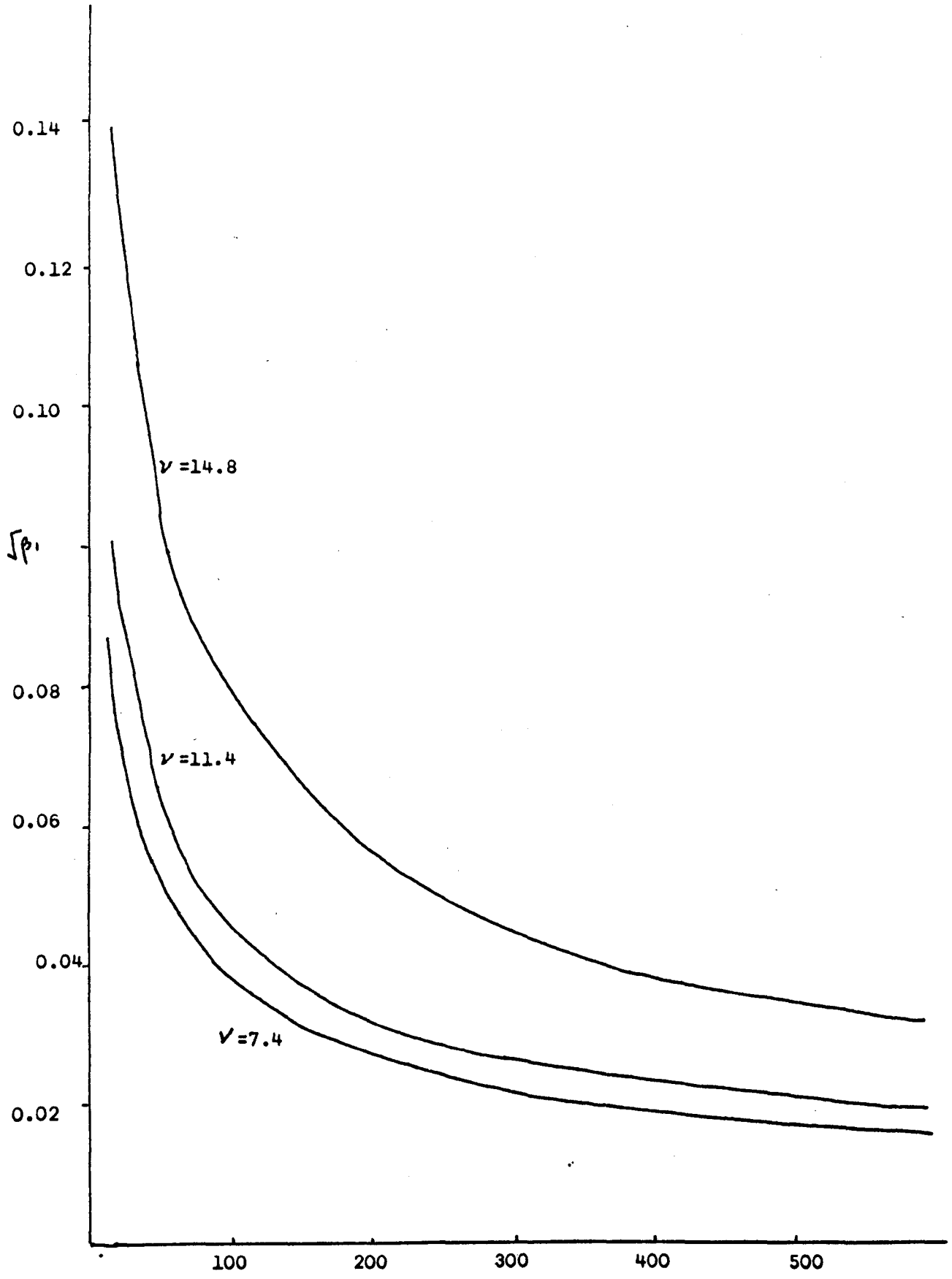
N denotes the sample size

M the number of lags used to evaluate the spectrum

m is the number of values of $\hat{f}(\omega)$ in the summation Δ_f

Fig 3

VALUES OF $\sqrt{\beta_1}$ FROM TABLE 5



CHAPTER 4

A TEST FOR WHITE NOISE

As we have seen in Chapter 3 we may estimate the logarithm of the minimum mean square error of prediction by $\hat{\Delta}_I$ or $\hat{\Delta}_S$ where

$$\hat{\Delta}_I = \frac{1}{n} \sum_{r=1}^{n} \log I_N(\omega, x) - \log 2 - \psi'(1) \qquad 4.1$$

and

$$\hat{\Delta}_S = \frac{1}{m} \sum_{r=1}^{m} \log 2\pi \hat{f}(\omega) - \log 2 - \psi'(\tfrac{1}{2}) \qquad 4.2$$

From the forms chosen above it is clear we shall be considering asympotic results and we shall look at a test for white noise described by Davis and Jones (1968) based on the asympotic properties of 4.1 . In addition we point out that one can base an equivalent test on 4.2, and consider some drawbacks of their procedure.

Suppose we have a non-deterministic time series $\{X_t\}$ which has the following moving average representation

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} \beta_j \varepsilon_{t-j} \qquad 4.3$$

where

$$E[\varepsilon_t] = 0$$
$$E[\varepsilon_t \varepsilon_s] = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases}$$

Then the variance of the series

$$\sigma_x^2 = \operatorname{var}(X_t)$$

can be written in the form

$$\sigma_x^2 = \sigma^2 \left(1 + \sum_{j=1}^{\infty} |\beta_j|^2\right). \qquad 4.4$$

A consequence of 4.4 is that $\sigma_x^2 = \sigma^2$ implies $\beta_k = 0$ for all $k \geq 1$

Using this fact Davis and Jones (1968) attempt to test for white noise by deciding between the two hypotheses

$$H_0' : \sigma^2 = \sigma_x^2 \qquad H_1 : \sigma^2 \leq \sigma_x^2$$

or equivalently

$$H_0 : \log\sigma^2 = \log\sigma_x^2 \qquad H_1 : \log\sigma^2 < \log\sigma_x^2$$

A convenient estimate of σ_x^2 can be obtained by noting that

$$\sigma_x^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)\,d\omega \qquad 4.5$$

since

$$R_L = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\omega t} f(\omega)\,d\omega \qquad 4.6$$

and we use as our estimate of σ_x^2

$$S_p^2 = \frac{1}{n}\sum_{j=1}^{n} I_N(\omega_j, x) \qquad 4.7$$

For large values of n it is fairly easy to show that S_p^2 is approximately equivalent to the sample variance since

$$I_N(\omega, x) = 2\sum_{s=-n+1}^{n-1}\frac{1}{N}\sum_{t=1}^{N-|s|} X_t X_{t+|s|}\cos s\omega \qquad 4.8$$

$$= 2\sum_{s=-N+1}^{N-1}\left(1 - \frac{|s|}{N}\right)\hat{R}_s\cos s\omega \qquad 4.9$$

This implies that if $\frac{|s|}{N} \to 0$

$$S_p^2 = \frac{1}{n}\sum_{\nu=1}^{n} I_N(\omega_\nu, x) \qquad 4.10$$

$$= \frac{1}{n}\sum_{\nu=1}^{n} 2\sum_{s=-N+1}^{N-1}\hat{R}_s\cos s\omega_\nu$$

However

$$\sum_{k=1}^{n} \cos kx = \frac{\cos(n+1)x \sin nx/2}{\sin x/2} \qquad 4.11$$

and hence

$$\sum_{r=1}^{n} \cos s\omega_r = 0 \quad s \neq 0$$

$$= n \quad s = 0$$

giving

$$s_p^2 = s_x^2 \qquad 4.12$$

where s_x^2 denotes the sample variance.

Davis and Jones (1968) use s_x^2 as their estimate of σ_x^2 however 4.12 enables us to consider s_x^2 as being approximately the sum of periodogram ordinates, which proves to be useful as we shall see later.

Using the sample variance we can assume that under H_0, s_x^2 has a $\chi^2 \frac{\sigma_x^2}{N-1}$ distribution which shows that $\log s_x^2$ is a biased estimate of $\log \sigma_x^2$ since

$$E\left[\log (N-1)\frac{s_x^2}{\sigma_x^2}\right] = \log 2 + \psi\left(\frac{N-1}{2}\right)$$

Thus

$$E[\log s_x^2] = \log \sigma_x^2 - \log\left(\frac{N-1}{2}\right) + \psi\left(\frac{N-1}{2}\right) \qquad 4.13$$

and

$$\operatorname{var}[\log s_x^2] = \psi'\left(\frac{N-1}{2}\right) \qquad 4.14$$

From the above we can see that

if

$$b = -\log\left(\frac{N-1}{2}\right) + \psi\left(\frac{N-1}{2}\right) \qquad 4.15$$

then $\log s_x^2 + b$ is an unbiased estimator of $\log \sigma_x^2$

Writing

$$\ell = \log s_x^2 - \hat{\Delta}_I + b \qquad 4.16$$

then under H_0 ℓ has a zero mean and for large n has a distribution which is approximately normal. To find the variance of ℓ we need to look at the correlation structure between $\log s_x^2$ and $\hat{\Delta}_I$ which we now proceed to do.

Now

$$\begin{aligned} &\operatorname{cov}\left[\log s_x^2, \hat{\Delta}_I\right] \\ &= \operatorname{cov}\left[\log \frac{1}{n}\sum_{\nu=1}^{n} I_N(\omega_\nu, x), \frac{1}{n}\sum_{\nu=1}^{n} \log I_N(\omega_\nu, x)\right] \end{aligned} \qquad 4.17$$

from 4.12 and under H_0 the spectrum is a constant function and the periodogram estimates are identically distributed thus

$$\operatorname{cov}\left[\log s_x^2 \hat{\Delta}_I\right] = \operatorname{cov}\left[\log \sum_{\nu} I_N(\omega_\nu, x), \log I(\omega_p, x)\right] \qquad 4.18$$

where ω_p is one of the values of ω at which the periodogram is evaluated.

We can simplify 4.18 by noticing that

$$\begin{aligned} &\operatorname{var}\left[\log \sum_{\nu} I_N(\omega_\nu, x) - \log I_N(\omega_p, x)\right] \\ &= \operatorname{var}\left[\log\left\{\frac{I_N(\omega_p, x)}{\sum_{\nu} I_N(\omega_\nu, x)}\right\}\right] \end{aligned} \qquad 4.19$$

$$\begin{aligned} = \operatorname{var}\left[\log \sum_{\nu} I_N(\omega_\nu, x)\right] + \operatorname{var}\left[\log I_N(\omega_p, x)\right] \\ - 2\operatorname{cov}\left[\log \sum I_N(\omega_\nu, x), I_N(\omega_p, x)\right] \end{aligned} \qquad 4.20$$

and since $\dfrac{I_N(\omega_p,x)}{\sum_v I_N(\omega_v,x)}$ is independent of $\sum I_N(\omega_j,x)$

see Hannan (1960) we have

$$\operatorname{var}\left[\log I_N(\omega_p,x)\right] = \operatorname{var}\left[\log\left\{\frac{I_N(\omega_p,x)}{\sum I_N(\omega_v,x)}\right\}\right] + \operatorname{var}\left[\log\sum I_N(\omega_v,x)\right] \quad 4.21$$

From 4.20, 4.21 we see that

$$\operatorname{var}\left[\log\sum I(\omega_p,x)\right] = \operatorname{cov}\left[\log I(\omega_p,x)\log\left[I_N(\omega_v,x)\right]\right] \quad 4.22$$

or equivalently

$$\operatorname{cov}\left[\log s_x^2, \hat{\Delta}_I\right] \simeq \operatorname{var}\left[\log s_x^2\right] \quad 4.23$$

giving

$$\operatorname{var}\left[\log s_x^2 + b - \hat{\Delta}_I\right] = \operatorname{var}(\hat{\Delta}_I) - \operatorname{var}\left[\log s_x^2\right] \quad 4.24$$

$$= \frac{\psi'(1)}{n} - \psi'\left(\frac{N-1}{2}\right) . \quad 4.25$$

Consequently as a result of our manipulations we can see that under H_0 the statistic

$$z = \frac{\log s_x^2 + b - \hat{\Delta}_I}{\sqrt{\dfrac{\psi'(1)}{n} - \psi'\left(\dfrac{N-1}{2}\right)}} \quad 4.26$$

has a distribution which tends to that of a standard Normal variate as N (or equivalently n) tends to infinity. It has been pointed out by

Davis and Jones (1968) that this is in fact equivalent to Barlett's test for the homogeneity of variances applied to periodogram ordinates.

If one wishes to use $\hat{\Delta}_f$ then much the same approach can be used.

Defining S_f^2 by

$$S_f^2 = \frac{1}{m} \sum_{\nu=1}^{m} \hat{f}(\omega_\nu) \qquad 4.28$$

for a suitable choice of m then as $m \to \infty$ we can show that

$$S_f^2 \to S_x^2 \qquad \text{in mean square}$$

where as before S_x^2 denotes the sample variance.

Proceeding as above we can write

$$b = -\log\left(\frac{N-1}{2}\right) + \psi\left(\frac{N-1}{2}\right) \qquad 4.29$$

and

$$L = \log S_x^2 + b - \hat{\Delta}_f \qquad 4.30$$

can be used as the basis for a test of white noise. Clearly we need to evaluate

$$\operatorname{cov}\left[\log S_x^2, \hat{\Delta}_f\right] \simeq \operatorname{cov}\left[\frac{1}{m}\sum \log \hat{f}(\omega_\nu), \log \frac{1}{m}\sum \hat{f}(\omega_\nu)\right] \qquad 4.31$$

which following our previous derivation can be written

$$\operatorname{cov}\left[\log S_x^2, \hat{\Delta}_f\right] = \operatorname{cov}\left[\log \sum_{\nu} \hat{f}(\omega_\nu), \log \hat{f}(\omega_p)\right] \qquad 4.32$$

where ω_p is one of the points at which $\hat{f}(\omega)$ is estimated. The derivation is identical to the periodogram case if we can show that $\dfrac{\hat{f}(\omega_p)}{\sum_{\nu} \hat{f}(\omega_\nu)}$ is independent of $\sum \hat{f}(\omega_\nu)$

The independence follows from a theorem due to Pitman (1937) which states that:- if $X_1 \ldots X_k$ are indepdendent random variables from a $\Gamma(\mu)$ distribution, then if $h(X_1 .. X_k)$ is a random variable such that for any $c > 0$ the function obeys

$$h(x_1, x_2 \ldots x_k) = h(c, x_1, cx_2, \ldots cx_k)$$

then $X_1 + X_2 + \cdots + X_k$ and $h(X_1 X_2 \ldots X_k)$ are independent.

Using this theorem we may show that under H_0 $\frac{\hat{f}(\omega_p)}{\sum \hat{f}(\omega_\nu)}$ and $\sum \hat{f}(\omega_p)$ are independent and in consequence if we proceed in exactly the same way as for the periodogram we have

$$var[\log s_x^2] \simeq var[\log s_x^2, L] \qquad 4.33$$

Using 4.33 and assuming asymptotic normality then

$$z' = \frac{L}{\sqrt{\frac{\psi'(\nu/2)}{m} - \psi'\left(\frac{N-1}{2}\right)}} \qquad 4.34$$

has a distribution which is approximately standard normal where ν denotes the degrees of freedom of the distribution of $\hat{f}(\omega)$

While the above statistics are of interest we really need some idea of their distributions under H_1, Davis and Jones consider the asymptotic distribution of z under the alternative hypothesis that $\{X_t\}$ is generated by the first order model

$$X_t = \alpha X_t + \varepsilon_t \qquad |\alpha| < 1 \qquad 4.35$$

however they do state that the distribution of z can be discovered

under any alternative which uses a model giving rise to a uniformly bounded spectrum.

Assuming a model of the form 4.35 we have

$$\sigma_x^2 = \mathrm{var}(X_t) = \frac{\sigma^2}{1-\alpha^2} \qquad 4.36$$

where $E[\varepsilon_t^2] = \sigma^2$. Then if we define V_n as

$$V_n = \frac{1}{n} \sum_{r=1}^{n} f^2(\omega_p) \qquad 4.37$$

it is easily seen from Parceval's theorem that

$$\lim_{n \to \infty} V_n = \int_{-\pi}^{\pi} f^2(\omega) d\omega = \sum_{t=-\infty}^{\infty} R_t^2 = V \qquad 4.38$$

where R_t denotes the t^{th} autocovariance. Using a Taylor expansion for the moments of $\log s_x^2$ we then find that for $n \to \infty$

$$E[\log s_x^2] = \log \frac{\sigma^2}{1-\alpha^2} \qquad 4.39$$

and $$\mathrm{var}[\log s_x^2] = \frac{1}{N} \sum_{-\infty}^{\infty} (\alpha^{|t|})^2 = \frac{1+\alpha^2}{(1-\alpha^2)N} \qquad 4.40$$

By appealing to the corollary to the Lindberg-Feller condition suggested by Eicker (1963) Davis and Jones show that the distribution of $\log\left[\frac{1}{n} \sum I(\omega_r, x)\right]$ is approximately normal with mean and variance given by 4.39, 4.40

Then if we assume that the correlation between s_x^2 and $\hat{\Delta}_I$ changes very little the covariance may be expressed in the same way as for the white noise case and the mean and variance of z defined by 4.26 are

$$E[z] = -\frac{\log(1-\alpha^2)}{\sqrt{\frac{\psi'(1)}{n} - \psi'\left(\frac{N-1}{2}\right)}} \qquad 4.41$$

$$\text{var}(z) = \frac{\frac{\psi'(1)}{n} + \frac{(1+\alpha^2)}{N(1-\alpha^2)} - 2\psi'\left(\frac{N-1}{2}\right)}{\frac{\psi'(1)}{n} - \psi'\left(\frac{N-1}{2}\right)} \qquad 4.42$$

under H_1

Naturally if one wishes to use any other model as an alternative this is possible provided one can evaluate V and in addition one is prepared to assume that the correlation between $\hat{\Delta}_I$ and $\log s_x^2$ does not change appreciably.

In the case of z' where one uses a smoothed spectral estimate one obtains the same result since we are using the same estimator $\log s_x^2$ Again if we assume that the correlation is unchanged we have the result that under H_1, z' is asymptotically normal with mean and variance given by

$$E[z'] = \frac{-\log(1-\alpha^2)}{\sqrt{\psi'(1/2)/m - \psi'\left(\frac{N-1}{2}\right)}} \qquad 4.43$$

$$\text{var}(z') = \frac{\psi'(1/2)/m - 2\psi'\left(\frac{N-1}{2}\right) + \frac{1+\alpha^2}{(1-\alpha^2)m}}{\frac{\psi'(1/2)}{m} - \psi'\left(\frac{N-1}{2}\right)} \qquad 4.44$$

Thus it may be seen that one has two alternative methods of testing for white noise based on z or z'. Since from Chapter 3 the rate of convergence to normality of $\hat{\Delta}_I$ or $\hat{\Delta}_P$ is approximately the same the choice seems fairly arbitary. If one has computed the spectrum then it is easier to use $\hat{\Delta}_P$ however if one has no intention of using the spectrum for other purposes it would appear that $\hat{\Delta}_I$ is simpler to use.

This is particularly so if one has very large values of N for one may then use a fast Fourier transform algorithem such as that due to Singleton (1968).

We feel that the two statistics have two considerable disadvantages.

(a) One requires large samples to achieve asymptotic normality, from Chapter 2 the indications are that a value of 200 would be required for $\hat{\Delta}_I$ or $\hat{\Delta}_L$ to be approximately normal in the tails of their distribution. It is just this tail area one would wish to use for any test of significance.

(b) The correlation structure between $\hat{\Delta}_I$ and $\log s_x^2$ is not clear and this is especially so when one is looking at the distribution under the alternative hypothesis.

One alternative procedure which overcomes at least objection (b) and has considerable intuitive appeal is to split the series into two parts

i.e. $X_1 \ldots X_m$ and $X_{m+1} \ldots X_{2m}$

From $X_1 \ldots X_m$ we estimate $\log e_1$ while from $X_{m+1} \ldots X_{2m}$ we estimate $\log \sigma_x^2$ using the sample variance.

Consider the two periodogram estimates

$$I_m^1(\omega, x) = \frac{2}{m} \left| \sum_{t=1}^{m} e^{-i\omega t} X_t \right|^2 \qquad 4.45$$

and

$$I_m^2(\omega, x) = \frac{2}{m} \left| \sum_{t=1}^{m} e^{-i\omega t} X_{t+m} \right|^2 \qquad 4.46$$

Then

$$E\left[I_m^1(\omega, x)\, I_m^2(\theta, x) \right] \simeq 4\pi f(\omega) f(\theta) + E\left[I_m^1(\omega, \varepsilon)\, I_m^2(\theta, \varepsilon) \right] \qquad 4.47$$

and

$$E\left[I_m^1(\omega, \varepsilon) I_m^2(\theta, \varepsilon) \right] = E\left[\frac{4}{m^2} \sum_{t,s} \varepsilon_t \varepsilon_s e^{i\omega(t-s)} \sum_{u,v} \varepsilon_{u+m} \varepsilon_{v+m} e^{i\theta(u-v)} \right]$$

$$= E\left[\frac{4}{m^2} \sum_t \sum_s \sum_u \sum_v \varepsilon_t \varepsilon_s \varepsilon_{u+m} \varepsilon_{v+m} e^{i(t-s)\omega + i(u-v)\theta} \right] \qquad 4.48$$

Then using the result due to Isserlis (1918) for random variables W, X, Y, Z

$$cov\,[WXYZ] = E[WX]E[YZ] + E[WY]E[XZ] + E[WZ]E[XY] + K_4 \quad \ldots\ 4.49$$

where K_4 denotes the fourth cumulant we have

$$cov\,[\,I_m^1(\omega,\varepsilon)I_m^2(\theta,x)] = \frac{4K_4}{m} \qquad \omega = \frac{2\pi j}{m} \quad \theta = \frac{2\pi k}{m} \qquad 4.50$$

where here K_4 is the fourth cumulant of the $\{\varepsilon_t\}$ process. Moreover if $\{Y_t\}$ is a general linear process and the $\{\varepsilon_t\}$ is a sequence of independent identically distributed random variables $K_4 = 0$

The drawback of the above procedure is that one is halving the available realisation and in consequence the convergence to normality is much slower. However the problems of correlation between $\hat{\Delta}_I$ and $\log s_x^L$ do vanish. When we look at small sample approximations to distributions we shall find this approach very useful.

While discussing large sample tests on split realisations we mention a further interesting procedure. If one were to split a realisation then one can estimate $\log e_1$ from the first segment while on the second segment one can fit a time series model, say

$$A(B)Y_t = \varepsilon_t$$

From the fitted model one can obtain residuals $\{a_t\}$ where

$$a_t = Y_t - \hat{Y}_t$$

and Mann and Wald (1943) and Anderson (1971) have shown that

$$\hat{\sigma}_R^2 = \frac{1}{n-1}\sum_{t=1}^{n}(a_t - \bar{a})^2 \qquad 4.51$$

has a distribution which is approximately normally distributed. By comparing $\log \hat{\sigma}_R^L$ with $\log e_1$ we can provide a "goodness of prediction" criteria for the fit of our time series model.

Davis and Jones outline some brief simulations using z as a test statistic and report good results using 50 series of 100 standard normal variates and 50 series of the form

$$X_t = 0.5\, X_{t-1} + \epsilon_t$$

using the original normal deviates as input for the "innovation process" $\{\epsilon_t\}$

We attempted to duplicate their results using the random number generator described in Appendix 2.

To obtain some idea of the approach to normality figures 3 to 6 show sample values of $\hat{\Delta}_I$ and z for some simple autoregressive processes. In each case every other sample value has been plotted. As may be seen from these plots on normal probability paper the discrepencies in the tails of the distributions can be quite large especially when the autoregressive parameter moves away from zero.

More detail of each simulation is given in tables 6 - 10. Looking at these tables we have for $\hat{\Delta}_I$

TABLES 6 AND 7

	Mean		Variance	
Simulation	Sample	Theoretical	Sample	Theoretical
I	-0.0021	0.0	0.01890	0.01644
II	0.0080	0.0	0.04467	0.03288
III	0.0433	0.0	0.01823	0.01644
IV	-0.0099	0.0	0.03309	0.03288
V	-0.0441	0.0	0.02917	0.03288

As we can see the simulations have sample means which agree very closely with the theoretical predictions, the maximum error being of the order of 4% this being for the larger values of α The sample variances however are rather inaccurate.

In the case of $\bar{z} = \hat{\Delta}_I + b - \log s_x^2$ we have

TABLES 8 AND 9

	Mean		Variance	
Simulation	Sample	Theoretical	Sample	Theoretical
I	0.0014	0.0	0.0129	0.0064
II	0.9471	8.9962	0.1013	1.0945
III	1.0039	12.7220	0.0462	2.1584
IV	0.0119	0.0	0.0193	0.0127
V	0.1499	1.5321	0.0270	1.5440

It is evident that for II, III, V the sample values diverge from the expected values, the remaining cases which are white noise produce much more accurate results. From this we are inclined to think that the assumptions made about the covariance between $\hat{\Delta}_I$ and $\log s_x^2$ are not entirely correct.

Noting that

$$\text{var}\left[\log s_x^2\right] = \psi'\left(\frac{N-1}{2}\right) \simeq \frac{2}{N}$$

we have

TABLE 10

Simulation	Sample covariances		Relative error ((%age)
I	0.01129	0.01	12
II	0.0295	0.02	45
III	0.0128	0.01	28
IV	0.0194	0.02	6
V	0.0048	0.02	76

From the tabulations above we can see that the rather large discrepancies in the variance cannot be completely explained by errors in approximating the covariances. However it is worth noting that the relative error in the sample covariance estimate is 6% and 12% respectively for white noise sequences and 28,45 and 76% for the remainder. In addition ordering the simulations by the relative error in the sample covariances does give the same ordering as is obtained by rank by the error in variance.

Thus it would appear that when our covariance approximations are dubious, which seems to be for all white noise cases, then the mean and variance of $\bar{z}$ and consequently of z are not as predicted.

Rather more disturbing is the rather poor agreement with the percentage points of the normal distribution shown in table 11.

Our results appear to be at variance with Davis and Jones (1968). While we would agree the fit for white noise is perhaps not unreasonable for small values of α the power seems poor.

TABLE 11

Simulation	I	II	III	IV	V
n	200	100	200	100	100
α	0.0	0.8	-0.8	0.0	0.4

Percentage point of Normal Distn.	Percentage of sample values less than the percentage point				
0.1	1.00	0.00	0.00	1	0.25
0.25	1.25	0.00	0.00	1.5	0.50
0.50	2.00	0.00	0.00	1.75	0.50
1.00	3.00	0.00	0.00	1.75	1.25
2.50	6.00	0.00	0.00	5.25	2.00
5.00	8.25	0.00	0.00	8.0	3.25
10.00	13.5	0.00	0.00	14.75	4.75
25.00	26.5	0.25	0.00	24.5	7.75
50.00	45.0	0.5	0.00	46.25	18.5
75.00	65.25	1.0	0.00	65.75	32.5
90.00	78.50	1.0	0.00	81.25	41.75
95.00	89.50	1.25	0.00	90.00	57.5
97.50	94.00	1.5	0.00	95.00	65.5
99.00	96.75	1.75	0.00	96.75	69.75
99.50	97.75	2.5	0.00	97.25	81.25
99.75	98.75	2.75	0.00	98.25	86.75
99.90	99.75	3.75	0.00	99.75	93.25

We feel that the principle problems with their test as proposed is that asymptotic normality requires much longer realisations, and in addition they advocate discarding the values of $I_N(\omega, x)$ at $\omega = 0$ and $\omega = \pi$ when N is even. It is our opinion that this makes our estimates unrealiable when only short realisations are available. We mention once again that the correlation approximation seems unreliable.

It would seem that we need sharper distributional assumptions than asympotic normality and we now try to obtain some approximations.

TABLE 12

Simulation 1 400 replications

The model generated was of the form

$$X_t = \varepsilon_t \qquad t = 1, \ldots 200$$

the ε_t being independent, normally distributed random variates.

Using the periodogram estimate

	mean	variance	
Sample	-0.0021	0.01890	Estimation of σ^2_ε.
Theoretical	0	0.01644	
Sample	0.0014	0.01286	Estimation of σ^2.
Theoretical	0.0	0.00644	

Sample correlation between components of ℓ

$$= 0.74908$$

TABLE 13

Simulation 2 400 replications

The generated model was

$$X_t = 0.8\, X_{t-1} + \varepsilon_t \qquad t = 1 \ldots 100$$

and the $\{\varepsilon_t\}$ are independent identically distributed Normal variates

	Mean	Variance	
Sample	0.008	0.04467	) Estimation of $\log e_1$
Theoretical	0	0.03288	)
Sample	0.9471	0.1073	} z
Theoretical	1.01	0.352	} z

Sample correlation between components of l

$$= 0.44026$$

TABLE 14

Simulation 3

Generated model

$$X_t = -0.8 X_{t-1} + \varepsilon_t$$

the $\{\varepsilon_t\}$ being independent identically distributed normal variates

Using the periodogram estimate

	Mean	Variance	
Sample	0.0433	0.01823	$\hat{\Delta}_I$
Theoretical	0.00	0.01644	
Sample	1.0039	0.0462	z
Theoretical	1.01	0.352	

Sample correlation 0.4415

TABLE 15

Simulation 4 400 replications

Model Generated

$$X_t = \varepsilon_t \qquad t = 1 \ldots 100$$

the $\{\varepsilon_t\}$ being independent, normally distributed random variates

	Mean	Variance	
Sample	-0.0099	0.03309	$\hat{\Delta}_I$
Theoretical	0.0	0.03288	
Sample	0.0119	0.01925	z
Theoretical	0.0	0.01270	

Correlation = 0.173587

TABLE 16

Simulation 5

Model generated

$$X_t = 0.4 X_{t-1} + \varepsilon_t \qquad t = 1 \ldots 100$$

$\{\varepsilon_t\}$ being independent, normally distributed random deviates

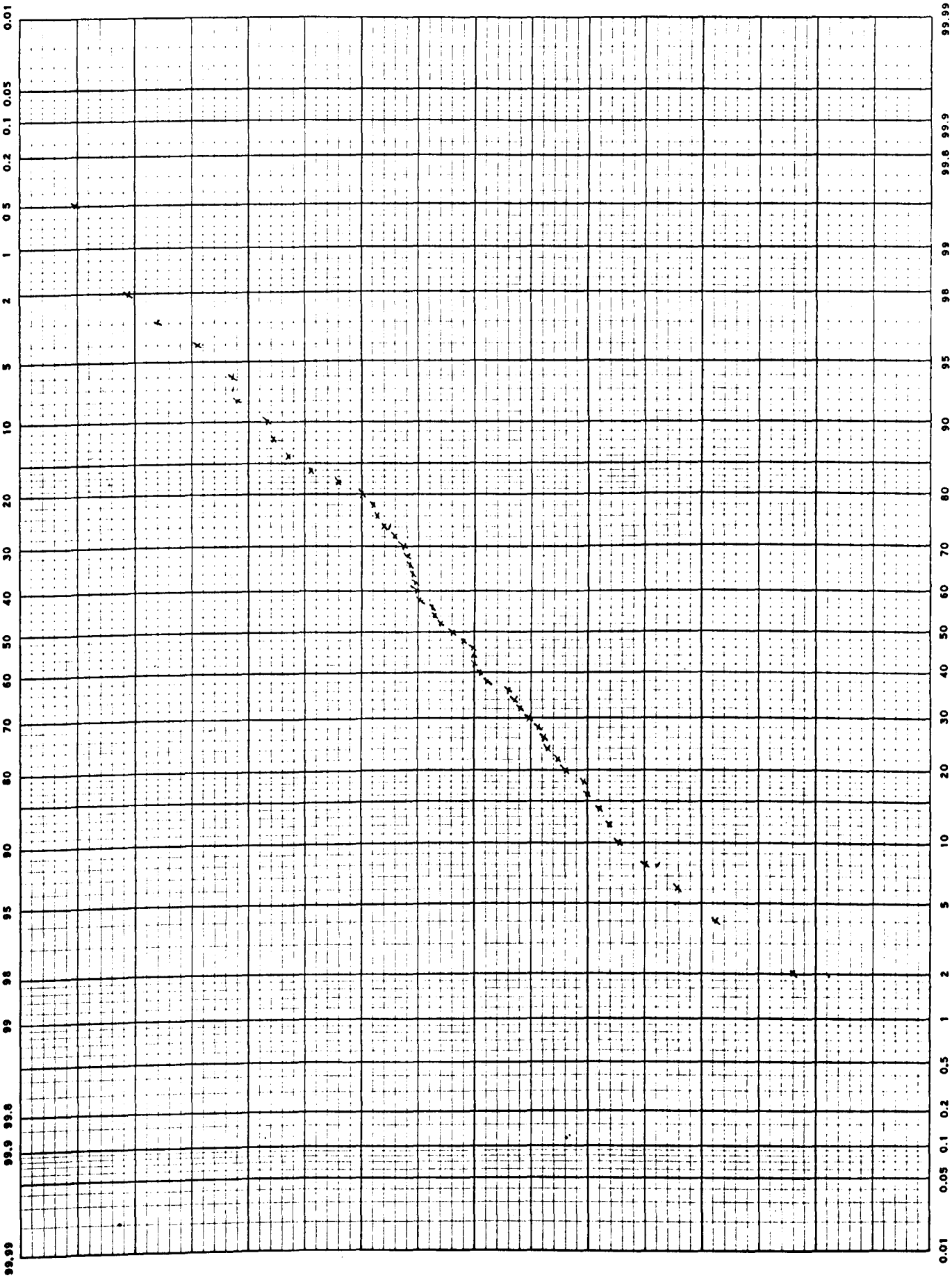
Periodogram estimates

	Mean	Variance	
Sample	-0.0441	0.02917	$\hat{\Delta}_I$
Theoretical	0.0	0.03288	
Sample	0.1499	0.027028	z
Theoretical	1.5321	1.54397	

Correlation between the above variates = 0.239473

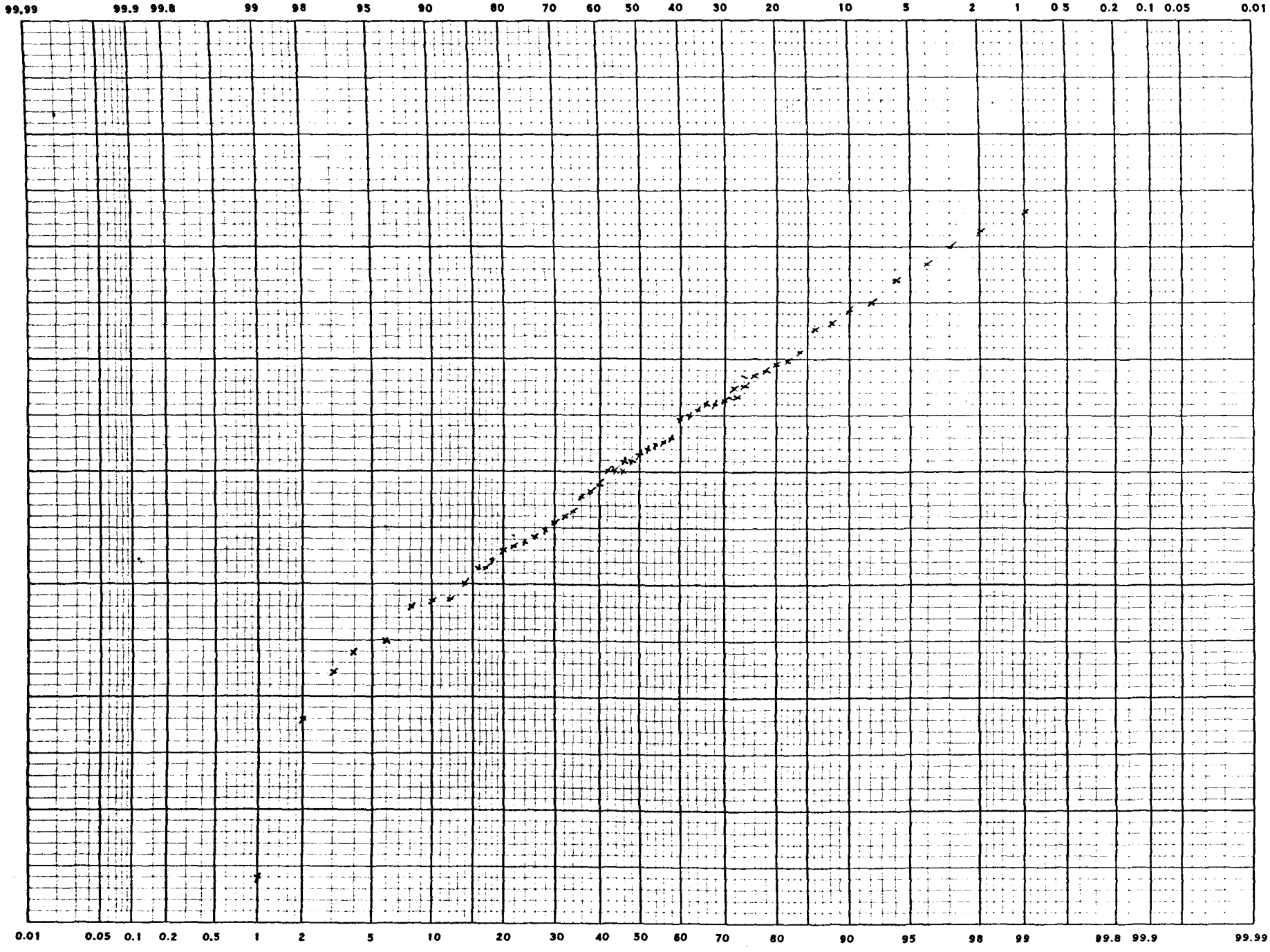
Fig 4

Simulation 1 $\hat{\Delta}_I$



I simulation

WELL Graph Data Ref. 5571 80 Equal Divisions x Probability

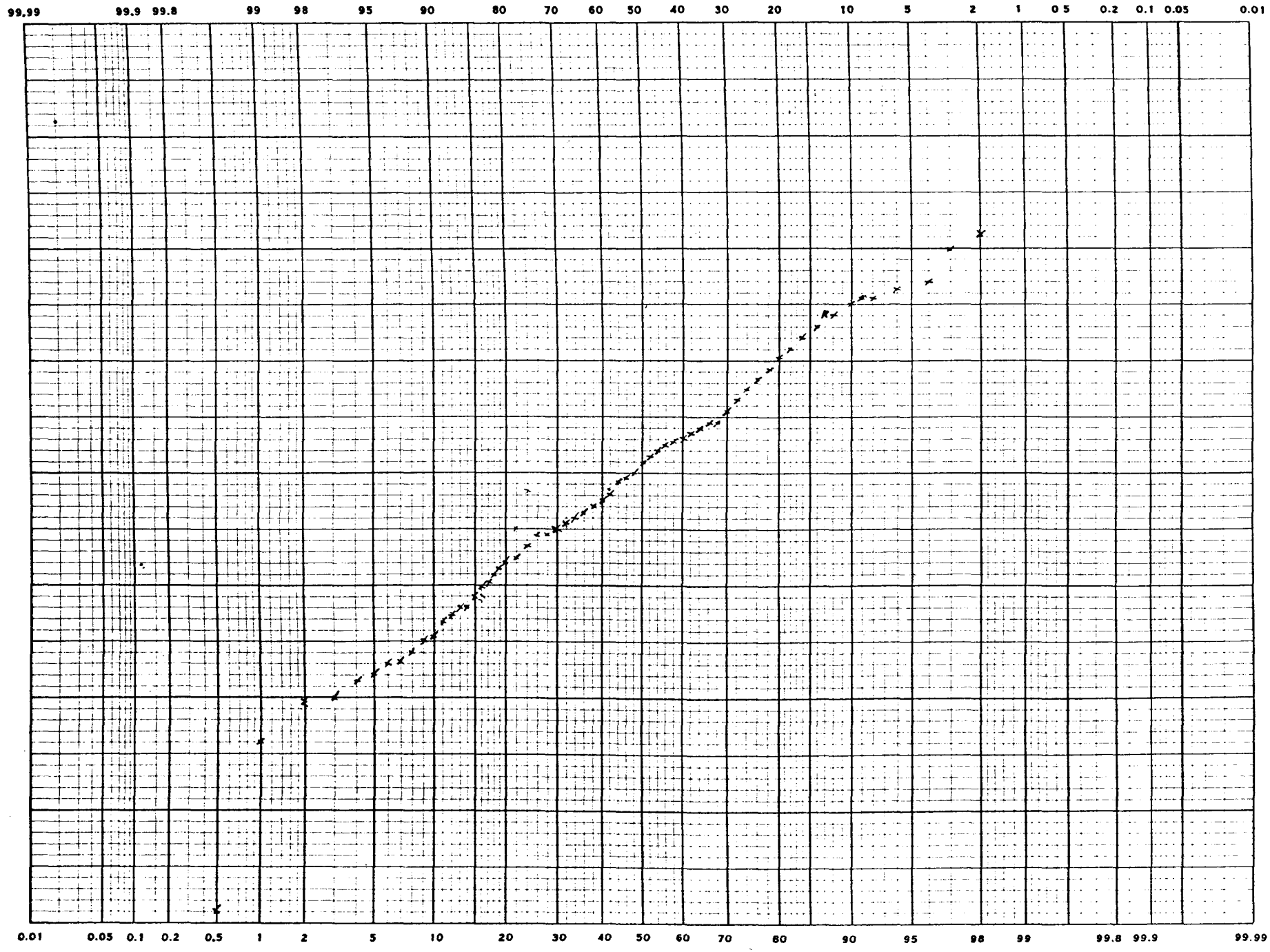


Simulation 1 Nonnormalised z

Fig 4a

WELL Graph Data Ref. 5971

80 Equal Divisions x Probability



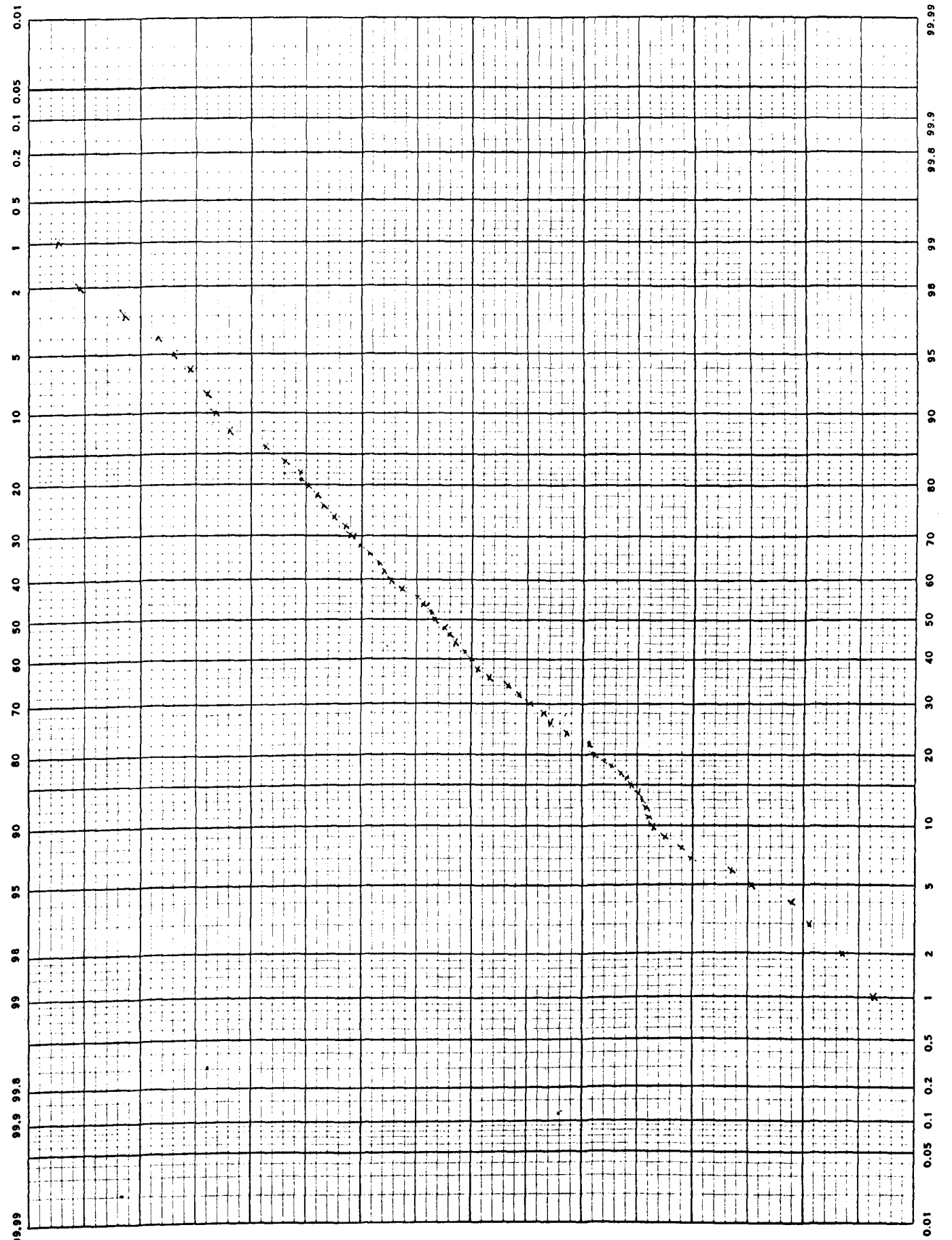
Simulation 2 $\hat{\Delta}_r$

Fig 3

TEXT CUT OFF IN THE ORIGINAL

Simulation 2 Nonnormalised z

Fig 6



CHAPTER 5

As we have seen the distributions of $\hat{\Delta}_I$ and $\hat{\Delta}_C$ converge to normality as the length of realisation N of our process tends to infinity. However we have also seen that convergence to normality is very slow and for many time series problems we have available only short series. In consequence we shall now present some approximations to the exact distributions of $\hat{\Delta}_I$ and $\hat{\Delta}_C$ and as we shall see they do appear to be very good approximations.

Since we have in effect used the central limit theorem to involk normality one approach is to look for a sharper approximation using Edgeworth type expansions. To this end we shall outline some of the ideas behind the derivation of these expansions.

Let $\Phi(x)$ be a distribution function, not necessarily Normal, with characteristic function $\psi(t)$ and let $F(x)$ be the distribution to be approximated. We shall assume $F(x)$ has a characteristic function $\phi(t)$ Denoting the cumulants by $\{\gamma_r\}$ and $\{\kappa_r\}$ respectively we have the formal identity

$$\phi(t) = \exp\left[\sum_{r=1}^{\infty} (\kappa_r - \gamma_r)\frac{(it)^r}{r!}\right]\psi(t) \qquad 5.1$$

since $$\log\phi(t) = \sum_{r=1}^{\infty} \kappa_r \frac{(it)^r}{r!}$$

and $$\log\psi(t) = \sum_{r=1}^{\infty} \gamma_r \frac{(it)^r}{r!}$$

Now if $\Phi(x)$ and all its derivatives vanish as x tends to the limits of its range then $(it)^r\psi(t)$ is the characteristic function of

or if we introduce the differential operator D $(-1)^r \Phi^{(r)}(x)$

then

$$(it)^r \psi(t) = \int_{-\infty}^{\infty} \Psi(x) D^r e^{itx} dx \qquad 5.2$$

In consequence we may write

$$F(x) = \exp\left[\sum_{r=1}^{\infty} (K_r - \gamma_r)\frac{(-D)^r}{r!} \right] \Psi(x) \qquad 5.3$$

Hence one can quite formally construct a distribution with prescribed cumulants by choosing $\Psi(x)$ and formally expanding 5.3. In the development of such asympotic series the most important distribution $\Psi(x)$ has been the Normal distribution, as one might expect, see for example Edgeworth (1905) and Charlier (1905).

Charlier expanded 5.3 and collected terms by the order of derivatives which gives the "Gram-Charlier" A series. It is in fact a least squares expansion in the derivatives of the Normal integral $\Psi(x)$ with respect to a weight function $1/\Psi'(x)$ Cramer (1937) and Szegö (1939) have shown that the A series converges for functions $F(x)$ whose tails tend to zero faster than $[\Psi'(x)]^{\frac{1}{2}}$ and that convergence occurs for all distributions on finite intervals, however Wallace (1958) comments that non-convergence is common in other cases of interest. We note in passing that $\Psi(x)$ is usually chosen to have the same mean and variance as the distribution $F(x)$ this having no effect on the convergence or otherwise.

Edgeworth (1905) followed a slightly different argument by considering improvements to the central limit theorem. To outline his development let the distribution $F(x)$ to be approximated by the distribution $F_n(x)$ of the standardised sum

$$Y_n = \left\{ \sum_{i=1}^{n} \left(X_i - E[X_i] \right) \right\} \Big/ \sqrt{\sum_{i=1}^{n} \operatorname{var}[X_i]} \qquad 5.4$$

and let $\Psi(x)$ be the standard normal distribution

Then in 5.1

$$\kappa_1 - \gamma_1 = \kappa_2 - \gamma_2 = 0 \qquad 5.5$$

and

$$\kappa_r - \gamma_r = \frac{\lambda^r}{n^{r/2-1}} \qquad 5.6$$

where we have assumed that the component variables are independent and identically distributed with mean μ and variance σ^2 while the higher moments are given by

$$\kappa_r = \sigma_r \lambda_r \qquad r \geq 3 \qquad 5.7$$

If we now collect terms in the formal expansion of 5.2 according to powers of n we obtain an asymptotic expansion of the characteristic function of the form

$$\phi_n(t) = \left[1 + \sum_{r=1}^{\infty} \frac{P_r(it)}{n^{r/2}} \right] e^{-t^2/2} \qquad 5.8$$

where $P_r(z)$ is a polynomial of degree $3r$ with coefficients depending on the cumulants of order 3 to $r+2$ The corresponding expression for the distribution function can be shown to be

$$F_n(x) = \Psi(x) + \sum_{r=1}^{\infty} \frac{P_r(-D)\,\Psi(x)}{n^{r/2}} \qquad 5.9$$

Thus we are in fact adding approximating terms to the central limit theorem, for taking terms in r to the power zero we have the familiar result

$$F_n(x) \Rightarrow \Psi(x)$$

Explicity 5.9 becomes

$$F_n(x) = \Psi(x) - \frac{\lambda_3 \Psi^{(3)}(x)}{6\sqrt{n}} + \frac{1}{n}\left[\frac{\lambda_4 \Psi^{(4)}(x)}{4!} + \frac{\lambda_3^2 \Psi(x)}{72} \right] + \ldots \qquad 5.10$$

Cramer (1928) proved that the series is valid but gave no explicit bounds on the errors. Apart from assuming that one more cumulant existed than is used in any partial sum the proof assumed the "Cramer" condition that the characteristic function $\varphi(t)$ of each component random variable satisfied

$$\lim_{|t|\to\infty} \sup |\varphi(t)| < 1 \qquad 5.12$$

In fact this condition is satisfied for all component distributions which have an absolutely continuous part but is not satisfied for discrete distributions. Cramer (1937) also showed that the asymptotic expansions remained valid for sums of non-identically distributed random variables, however the conditions required are very much more restrictive.

It was also shown by Gnedenke and Kolmogerov (1954) that the termwise differentiated Edgeworth series is a valid expansion for the probability density function.

Essen (1945) has studied the problem of developing asymptotic expansions when the Cramer condition is not satisfied. He showed that the error in using the first approximation

$$\sqrt{f}(x) - \frac{\lambda_3 \Phi^{(3)}(x)}{\sigma^3\sqrt{n}} \qquad 5.13$$

is of smaller order than $1/\sqrt{n}$ providing only that the third moment is finite and that the distribution is not a lattice distribution i.e. takes all probability on a set of equally spaced points. However Wallace (1958) has pointed out that the usual Edgeworth expansions can be modified by the addition of discontinuous terms so that the resultant expansion is valid uniformly for all x.

Since the Gram-Charlier series is just a rearrangement of the Edgeworth expansion all the above properties also hold. However for our purposes it is convenient to use the Edgeworth series.

In a great many statistical applications the problem is the reverse one in that the percentage points of the distribution are needed. From the Edgeworth expansions of $F_n(x)$ an asymptotic expansion of the desired quantile x_p can be obtained by formal substitutions and expansions of the form

$$x_p = z_p + \frac{S_1(z_p)}{\sqrt{n}} + \frac{S_2(z_p)}{n} + \cdots \qquad 5.14$$

where the $S_i(z)$ are polynomials and z_p is the corresponding quantile of the Normal distribution. Cornish and Fisher (1937) carried out these inversions treating each cumulant of $F_n(x)$ according to the size of the leading term. Abramowitz and Segun (1968) give explicit form to x_p and we follow their description here.

If x_p is such that $F_n(x_p) = 1 - p$ then x_p is given by

$$x_p = \mu + \sigma w \qquad 5.15$$

where

$$\begin{aligned} w = z_p &+ \gamma_1 h_1(z_p) \\ &+ \gamma_2 h_2(z_p) + \gamma_1^2 h_{11}(z_p) \\ &+ \gamma_3 h_3(z_p) + \gamma_1\gamma_2 h_{12}(z_p) + \gamma_1^3 h_{111}(z_p) \\ &+ \gamma_4 h_4(z_p) + \gamma_2^2 h_{22}(z_p) + \gamma_1^2\gamma_2 h_{112}(z_p) + \gamma_1^4 h_{1111}(z_p) \\ &+ \cdots \end{aligned} \qquad 5.16$$

z_p being the corresponding percentage point of the standard normal distribution and

$$\gamma_{r-2} = \frac{K_r}{(K_2)^{r/2}} \qquad r = 3, 4, \ldots \qquad 5.17$$

while

$$h_1(x) = \frac{1}{6} H_2(x)$$

$$h_2(x) = \frac{1}{24} H_3(x)$$

$$h_{11}(x) = -\frac{1}{36}\left[2H_3(x) + H_1(x)\right]$$

$$h_3(x) = \frac{1}{120}\left[H_4(x)\right]$$

$$h_{111}(x) = \frac{1}{324}\left[12H_4(x) + 19H_2(x)\right]$$

$$h_4(x) = \frac{1}{720} H_5(x)$$

$$h_{21}(x) = -\frac{1}{384}\left[3H_5(x) + 6H_3(x) + 2H_1(x)\right]$$

$$h_{13}(x) = -\frac{1}{180}\left[2H_5(x) + 3H_3(x)\right]$$

$$h_{112}(x) = \frac{1}{288}\left[14H_5(x) + 37H_3(x) + 8H_1(x)\right]$$

$$h_{1111}(x) = -\frac{1}{7776}\left[252H_5(x) + 832H_3(x) + 227H_1(x)\right] \qquad 5.18$$

where the $H_n(x)$ are the Hermite polynomials

$$H_n(x) = n! \sum_{m=0}^{\left[\frac{n}{2}\right]} \frac{(-1)^m x^{n-2m}}{m!\, 2^m (n-2m)!} \qquad 5.19$$

The drawbacks of the Edgeworth and Cornish-Fisher expansions are several. The expansions are not easy to evaluate in view of their complexity and the special functions required. More serious is that the Edgeworth expansion is not a probability distribution and does not give rise to monotonic functions. In fact they can give rise to values outside the unit interval $[0,1]$ this being particularly so in the tails of the distribution.

This does not conflict with the proofs given above since the proofs of convergence etc., refer only to the absolute value of two functions approaching zero.

Barton and Dennis (1953) outline the conditions under which the Edgeworth and Gram Charlier expansions can be expected to give unimodal and positive definate results. Figure 7 summarises their results giving regions in the β_1, β_2 plane for which the expansions are positive definate and unimodel.

For some practical use of these expansions we refer the interested reader to Teichroew (1956) and Greary (1947).

As an alternative to the computation of the appropriate asymptotic expansions another approximation method is to fit the appropriate curve from the Pearson system. That is the system of frequency curves satisfying the differential equation

$$\frac{1}{y}\frac{dy}{dx} = -\frac{(x+c_1)}{c_0 + c_1 x + c_2 x^2} \qquad 5.20$$

which can be obtained from the hypergeometric series, see Kendall and Stuart (1969).

Examining 5.20 we see the mode of the distribution is at $x = -c_1$ and if we shift the origin to the mode we can see that

$$\frac{1}{y}\frac{dy}{dx} = -\frac{x}{c_0 + c_2 x^2} \qquad 5.21$$

and

$$\frac{d^2y}{dx^2} = -\frac{y\left[c_0 - (1+c_2)x^2\right]}{(c_0 + c_2 x^2)^2} \qquad 5.22$$

thus the curves have points of inflection at

$$x = \pm \frac{c_0}{1+c_2} \qquad 5.23$$

Further if we express c_0, c_1, c_2 in terms of the moments of the distribution we have

$$c_0 = \frac{\sigma(4\beta_2 - 3\beta_1)}{2(5\beta_2 - 6\beta_1 - 9)} \qquad 5.24$$

$$c_1 = \frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)} \qquad 5.25$$

$$c_3 = \frac{2\beta_2 - 3\beta_1 - 6}{2(5\beta_2 - 6\beta_1 - 9)} \qquad 5.26$$

where as before

$$\beta_1 = \mu_3^2 / \mu_2^{3/2}$$

$$\beta_2 = \mu_4 / \mu_2^2$$

Thus each curve is uniquely specified by its first four moments. This can be a useful property if one is using sample moments, however in our case since all the moments are known exactly errors in estimation of high order moments are not a real consideration.

Looking at 5.20 we can see that the form of the function is largely determined by the values of the roots of the quadratic

$$c_0 + c_1 x + c_2 x^2 \qquad 5.27$$

or equivalently by the ratio

$$c_1^2 / 4 c_0 c_2$$

This latter quantity has been termed the criterion K by Elderton and Johnson (1969) and can be written in the alternative form

$$K = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} \qquad 5.28$$

The three main "types" or solutions to 5.20 arise as follows

(i) When $K<0$ the roots of 5.25 are real and of opposite sign and the curved obtained is the Pearson type 1. This is a Beta distribution which is written

$$f(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} \qquad 0 \le x \le 1 \qquad 5.29$$

is standardform.

(ii) When $0<K<1$ then the roots of 5.25 are complex and we have a Pearson type 4 which can be written

$$f(x) = k\left(1 + \frac{x^2}{c_0^2}\right)^{-m} \exp\left[-\nu \arctan\left(\frac{x}{c_0}\right)\right] \qquad 5.30$$

the k, m, ν being parameters determined by β_1, β_2

(iii) When $K>1$ both roots of 5.27 are real and of the same sign, giving rise to a Pearson type 6 curve.

$$f(x) = y_0 (x-a)^q x^{-p} \qquad 5.31$$

where again the constants y_0, q, p are determined by the moments.

These three curves are the main "Pearson types", however there are at least 9 further "transitional" types that have been classified, including the Normal and Gamma distributions. Some of these types are shown in Table 17 together with some brief details of their shape. For more detail the reader is advised to consult Pearson and Hartley (1970) or Elderton and Johnson (1969). Figure 7 shows the boundaries of the various curves on the β_1 and β_2 plane.

There are obviously some difficulties in obtaining the appropriate curve from 5.25 given values of β_1 and β_2. However often

all we require are percentage points of a probability distribution and in this case things are made very simple by the extensive tables published by Johnson et al (1963) and the extended version of these given in Pearson and Hartley (1970). These tables give values of the percentage points of the standardised deviate for 15 differing percentages. The use of these tables makes significance testing very straightforward and increases the use of the Pearson curve approximations considerably.

Pearson curves have been used as approximations to unknown distributions with considerable success. Stephens (1963) used such curves to find the distribution of Watsons goodness of fit statistic and the curves have been used by many other authors.

A further elegent method of approximation is that using the method of "steepest descents" in evaluating integrals. This technique has been used in approximating the distribution of the correlation co-efficient, see Kendall and Stuart(1969) but the method appears to have orginated in Jefferies and Jefferies (1950). The application of this technique to statistical problems has been studied by Daniels (1954) whose approach we now follow.

Let us assume that the distribution to be approximated comprising component random variableswith a probability distribution and a moment generating function $M(t)$

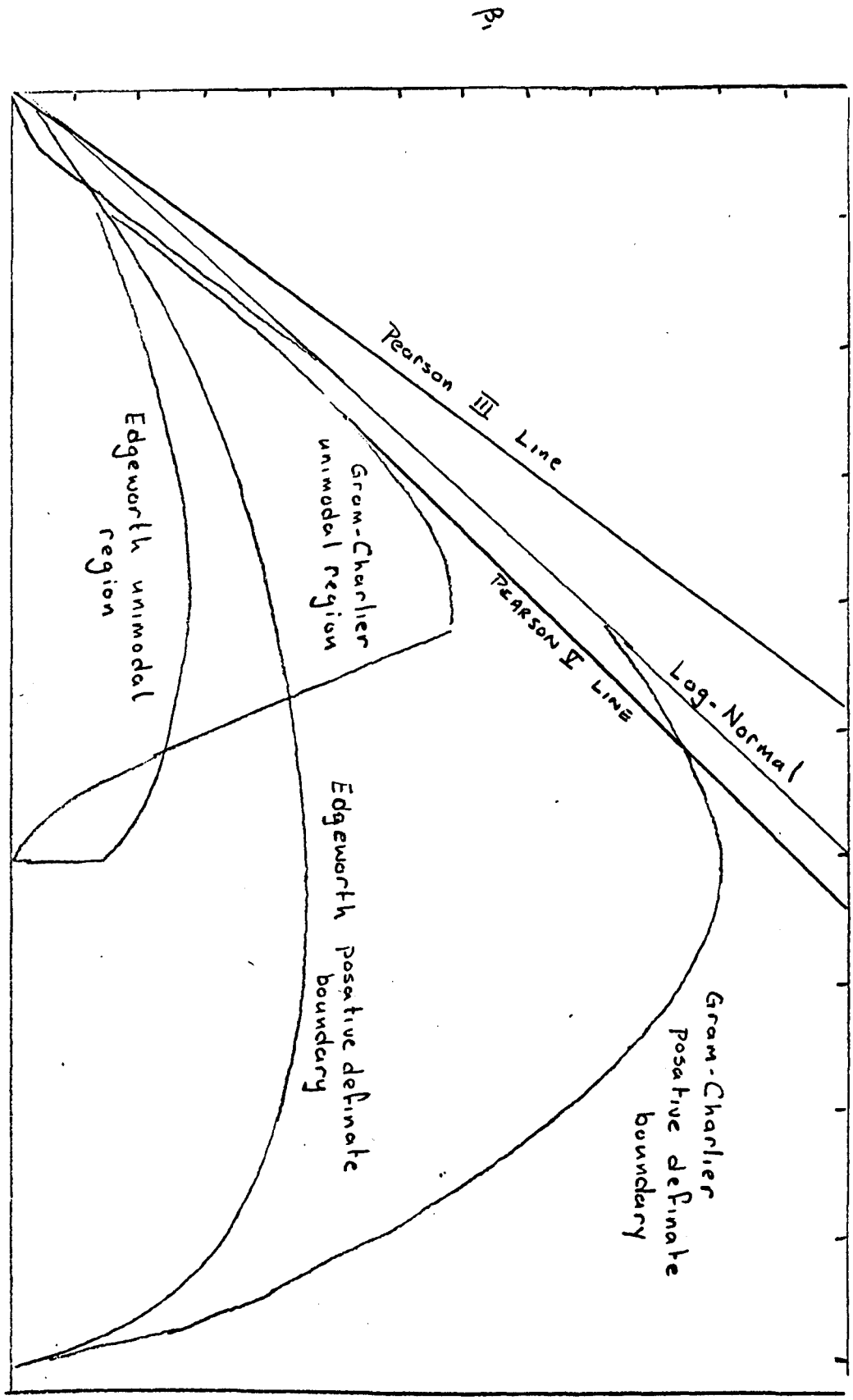
$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \qquad 5.32$$

which exists in some non-vanishing interval including the origin as $t \to \infty$.

TABLE 17

Type	Equation	Origin	K	Limits and Form
1	$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}$	Mode	$\kappa < 0$	$-a_1 \leq x \leq a_2$ Skew may be U or J or bell shaped
2	$y = y_0 \left(1 - \frac{x^2}{a^2}\right)^m$	Mean	$\kappa = 0 \quad \beta_1 = 0$ $\beta_2 < 3$	$-a \leq x \leq a$ Symmetrical can be U shaped for otherwise bell shaped
3	$y = y_0 e^{-\gamma x} \left(1 + \frac{x}{a}\right)^{\gamma a}$	Mode	$2\beta_2 = 6 + 3\beta_1$	$-a \leq x < \infty$ Bell or J shaped
4	$y = y_0 e^{-\nu \tan^{-1} \frac{x}{a}} \left(1 + \frac{x^2}{a^2}\right)^{-m}$ $+ \nu a$	Mean $+ \frac{\nu a}{2m-2}$	$0 < K < 1$	$-\infty < x < \infty$ Skew and bell shaped
5	$y = y_0 e^{-\gamma/x} x^{-p}$	At start of curve	$\kappa = 1$	$0 < x < \infty$ Bell shaped

FIG. 7


β1
β2
Pearson III Line
Log-Normal
PEARSON V LINE
Gram-Charlier unimodal region
Edgeworth unimodal region
Edgeworth posative definate boundary
Gram-Charlier posative definate boundary

Consider the mean of n independent random variables $X_1 \ldots X_n$ then its density function $f_n(x)$ is given by

$$f_n(x) = \frac{n}{2\pi} \int_{-\infty}^{\infty} M^n(it) e^{-it\bar{x}n} \, dt \qquad 5.33$$

or using the Laplace transformation

$$f_n(x) = \frac{n}{2\pi i} \int_{T-i\infty}^{T+i\infty} e^{n[K(t) - t\bar{x}]} \, dt \qquad 5.34$$

$K(t)$ being the cumulant generating function.

For n large we can approximate the integrand by choosing a contour of integration to pass through the saddle point of the integrand so that the contribution to the integral of the part away from the saddle point is negligable.

By looking at possible paths through the saddle point, given by

$$K'(t) = \bar{x}$$

Daniels showed that the path of steepest descent gives an asymptotic expansion of the form

$$f_n(x) = g_n(x) \left\{ 1 + \frac{1}{n} \left[\frac{1}{8} \lambda_4(t_0) - \frac{5}{24} \lambda_3^2(t_0) + \cdots \right] \right\} \qquad 5.35$$

where

$$g_n(x) = \left[\frac{n}{2\pi K''(t_0)} \right]^{1/2} e^{n[K(t_0) - t_0 \bar{x}]} \qquad 5.36$$

$$\lambda_j(t) = \frac{K^{(j)}(t)}{[K''(t)]^{j/2}} \qquad j \geq 3 \qquad 5.37$$

and $\hat{t}_0$ is the saddle point.

In addition he showed that only one real root of

$$K'(t) = \bar{x}$$

exists in the same interval as $M(t)$.

Using in our case the cumulant generating function

$$K(t) = \ln\left(\Gamma\left(\frac{t}{n} + \nu/2\right)\right) - \ln\left[\Gamma(\nu/2)\right] + t \ln 2 \qquad 5.38$$

from 3.27 we have to solve

$$K''(t) = \psi'(t + \nu/2) = \bar{x} \qquad 5.39$$

to find the saddle point. The analytic solution to 5.39 is not easily found, even using the usual Stirling approximation to $\Gamma(z)$ we still must solve a transcendental equation. We feel that some progress might be made using the theory of integral equations but feel that the lack of a-solution necessitates using one of the approximation methods discussed earlier. This is rather unfortunate since 5.35 is an asymptotic expansion which converges much faster than the Edgeworth or Charlier series and has the added attraction of being non-negative.

Having reduced ourselves to Edgeworth series and Pearson curves we now proceed to apply them to the distributions of $\hat{\Delta}_I$ and $\hat{\Delta}_E$

To use the Edgeworth expansion as we know all the moments of the distribution are bounded we need only show that the component distributions satisfy the Cramer condition that

$$\lim_{t \to \infty} |\phi(t)| \leq 1 \qquad 5.40$$

with equality if the distribution is a lattice distribution c.f.Lukas (1960)
In our case

$$\phi(t) = \frac{2^{it}}{\Gamma(\nu/2)} \Gamma(it + \nu/2) \qquad 5.41$$

or to be more precise since we are dealing with the consistent estimator

$$\phi(t) = \frac{\Gamma(it + \nu/2)}{\Gamma(\nu/2)} \qquad 5.42$$

For $\nu = 2$ that is for $\overset{\blacktriangle}{\Delta}_I$ we have simply

$$\phi(t) = \frac{\Gamma(it+1)}{\Gamma(1)} = \frac{it\,\Gamma(it)}{\Gamma(1)}$$

hence

$$|\phi(t)|^2 = \frac{\pi}{t \sinh \pi t}$$

and clearly

$$\lim_{|t| \to \infty} |\phi(t)| < 1$$

In general we have to proceed as follows

$$\phi(t) = \frac{\Gamma(it + \nu/2)}{\Gamma(\nu/2)}$$

and

$$|\phi(t)|^2 = \left| \frac{\Gamma(it + \nu/2)}{\Gamma(\nu/2)} \right|^2$$

$$= \prod_{n=0}^{\infty} \left[1 + \frac{t^2}{(\nu/2 + n)^2} \right]^{-1} . \qquad 5.43$$

Now the convergence of 5.43 is completely dominated by the convergence or otherwise of

$$\sum_{n=0}^{\infty} \log \left[1 + \frac{t^2}{(n+\nu/2)^2} \right] \qquad 5.44$$

Examination of 5.40 and use of the Gauss test for absolute convergence shows that 5.43 is absolutely convergent to zero

and hence

$$\lim_{t \to \infty} |\phi(t)| < 1$$

since

$$\lim_{t \to \infty} |\phi(t)| \neq 0$$

Thus we can construct Edgeworth type expansions.

For the Pearsons approximations we see from Chapter 3 that our values of β_1 and β_2 lie in the range of the tables of Johnson et al (1963) enabling us to interpolate our percentage points as required.

To see how well these approximations worked we tried some simulation experiments and compared the sample results for the distribution of $\hat{\Delta}_1$ with Edgeworth and Pearson predictions. To avoid overwhelming the reader with figures we shall leave discussion of the values for $\hat{\Delta}_2$ until Chapter 7 for as we shall see there are some rather tricky problems involved.

From tables 17 and 18 we can see that the Pearson and Edgeworth approximations provide reasonable fits to the simulation results. The Edgeworth expansion does appear to do rather less well in the upper tail, the reason apparently being that the actual distribution is skewed while the Edgeworth expansion is rather symmetric. This latter fact is well illustrated by Figures 9 and 10.

For this reason together with the fact that the Pearson curve values are much easier to evaluate we shall from now on use only the Pearson curve approximation and for convenience the percentage points of the approximating distribution can be found in Appendix 3.

Another point of interest is the performance of the trapizodal estimate. Davis and Jones (1968) advocate an estimator which has the periodogram ordinates at frequencies 0 and π omitted. We feel that this is unnecessary since the distribution of the end points is known and may possibly introduce errors. To confirm this we attempted some further comparisons using a very small length of realisation

TABLE 18

Simulation results using Edgeworth Series for truncated estimator with no end points.

For various autoregressive models of the form $X_t = \alpha X_{t-1} + \epsilon_t \quad t=1 \ldots 10\alpha$ the proportion of values of $\hat{\Delta}_r$ not exceeding the percentage points are compared with predictions using an Edgeworth expansion to evaluate the percentage points.

	True Percentage	Sample Values		
No. of Replications		400	400	400
α		0.0	0.4	0.8
	0.5	0.5	1.0	0.75
	1.0	1.25	1.5	1.0
	2.5	3.0	3.75	4.25
	5.0	5.5	8.0	8.5
	10.0	11.0	12.5	13.5
	25.0	26.25	27.25	26.0
	50.0	51.75	61.0	46.0
	75.0	76.25	63.25	72.0
	90.0	91.5	96.0	87.5
	95.0	95.0	98.5	94.25
	97.5	98.0	100	96.5
	99.0	99.0	100	98.15
	99.5	100.0	100	99.5
	99.75	100.0	100	99.5

TEXT CUT OFF IN THE ORIGINAL

TABLE 19

Simulation results for the fits of Pearson curves using a truncated estimator δ_I with no end points

For various autoregressive models of the form

$$X_t = \alpha X_{t-1} + \epsilon_t \qquad t = 1 \cdots N$$

the proportion of values not exceeding the percentage points of the appropriate Pearson curve are compared with the true percentages

	True Percentage	Sample Values		
No. of Replications		400	400	400
	α	0.0	0.4	0.8
	0.25	0.25	1.0	0.5
	0.5	0.5	1.0	0.5
	1.0	0.75	1.5	0.75
	2.5	2.25	3.5	3.5
	5.0	5.0	7.45	8.25
	10.0	11.5	12.75	13.5
	25.0	28.5	29.5	28.0
	50.0	52.75	61.5	52.25
	75.0	75.75	85.5	72.25
	90.0	89.75	95.75	86.5
	95.0	94.5	98.0	92.5
	97.5	97.5	99.5	96.0
	99.0	98.5	100	96.5
	99.5	99.0	100	98.75
	99.75	99.75	100	99.25

The sample results are given below for the percentage points. For 400 replications of the model

$$Y_t = \alpha Y_{t-1} + \varepsilon_t \qquad t = 1 \ldots 32 \quad \alpha = 0$$

the sample mean and variance were

Mean		Variance	
Sample	Theoretical	Sample	Theoretical
-0.0200	0.0000	0.1146	0.1415

the agreement being very much closer for a smaller value of N than in any of the cases using the Davis and Jones form given in Chapter 4, see tables 12 - 16. Even using the approximate Pearson curve percentage points for a Davis and Jones estimate the fit as shown below is much better.

Some additional tables 20, 21, 22 give further evidence of the much better results using the trapizodal estimate. Clearly for large samples the superiority is

PERCENTAGE	SAMPLE	PERCENTAGE	SAMPLE
0.25	0.25	75.0	75.75
0.5	0.25	90.0	92.25
1.0	1.00	95.0	94.25
2.5	3.25	97.5	98.00
5.0	5.5	99.0	98.50
10.0	10.0	99.5	99.00
25.0	29.0	99.75	99.5
50	52.5		

much less marked but it would seem that for all sample sizes met in practice the trapizodal form should always be used.

To obtain some idea of the difference in the moments between the two forms we can show, Abramowitz and Segun (1968).

$$K_r(T) = K_1 + \frac{1}{n}\left[\log 2 + \psi(0.5)\right]$$

$$\simeq K_1 - \frac{1.2704}{n}$$

and

$$K_r(T) = K_r\left[1 + \frac{2^r - 1}{n}\right]$$

where $K_r(T)$ denotes the trapizodal form and K_r the estimate with truncated end points.

Having approximated the distribution of $\hat{\Delta}_I$ we have in effect provided ourselves with a method of testing a hypothesis of the form

$$H_0 : \log e_1 = \sigma^2$$

against

$$H_1 : \log e_1 \neq \sigma^2$$

This test has the useful property of not requiring the order of the model to be specified. Since it is usual to have to fit a model to estimate $\log e_1$ it would appear that the use of $\hat{\Delta}_I$ is a considerable advantage especially as it would appear that errors in estimation in the parametric situation are greatly increased when the model is miss-specified.

While the test above is of interest it is a little restricted since in most circumstances one would wish to compare differing estimates of $\log e_1$ However it may be of interest to users of stepwise regression techniques in lagged situations who know a priori the minimum mean squared error. One can always estimate $\log e_1$ for the residual series once one has fitted a time series model, and for a linear model

$$A(B)\,X_t = \theta(B)\,\varepsilon_t$$

with $\operatorname{var}(\varepsilon_t^2) = \sigma^2$ then $E\left[\hat{\Delta}_I\right] = \sigma^2$

when the residuals are obtained from fitting the correct model. If the mean of $\hat{\Delta}_r$ differs significantly from the a priori value then one has a miss-specified model. This does enable one to avoid both under and overfitting but we feel it will provide more information in the overfitted case.

What is of more interest however is the comparison of $\hat{\Delta}_r$ calculated before the model is fitted and S_R^2 the sample variance of the residual series, for if the correct model is fitted

$$E[S_R^2] = \sigma^2$$

This has in essence been treated in the examination of the test for white noise discussed in the previous chapter. However as we have seen the test proposed is less than satisfactory and we now turn our attention to more efficient alternatives.

TABLE 20

of Δ_1 Comparison of theoretical percentages with sample values using Pearson curves

Model $X_t = \alpha X_{t-1} + \varepsilon_t$ $t = 1, \ldots 32$

%age	α 0.0	0.5	0.9
0.25	0.25	0.0	0.25
0.5	0.25	0.25	0.5
1.0	1.00	1.00	0.75
2.5	3.00	3.75	3.00
3.0	5.00	6.00	5.5
10.0	9.00	12.5	9.25
23.0	25.75	24.75	23.00
50.0	51.25	51.00	54.00
75.0	77.5	76.75	77.5
90.0	91.25	91.25	89.5
95.0	95.25	97.00	95.25
97.5	98.5	97.75	98.75
99.0	99.5	99.25	99.5
99.5	99.75	99.25	99.75
99.75	99.75	99.75	100

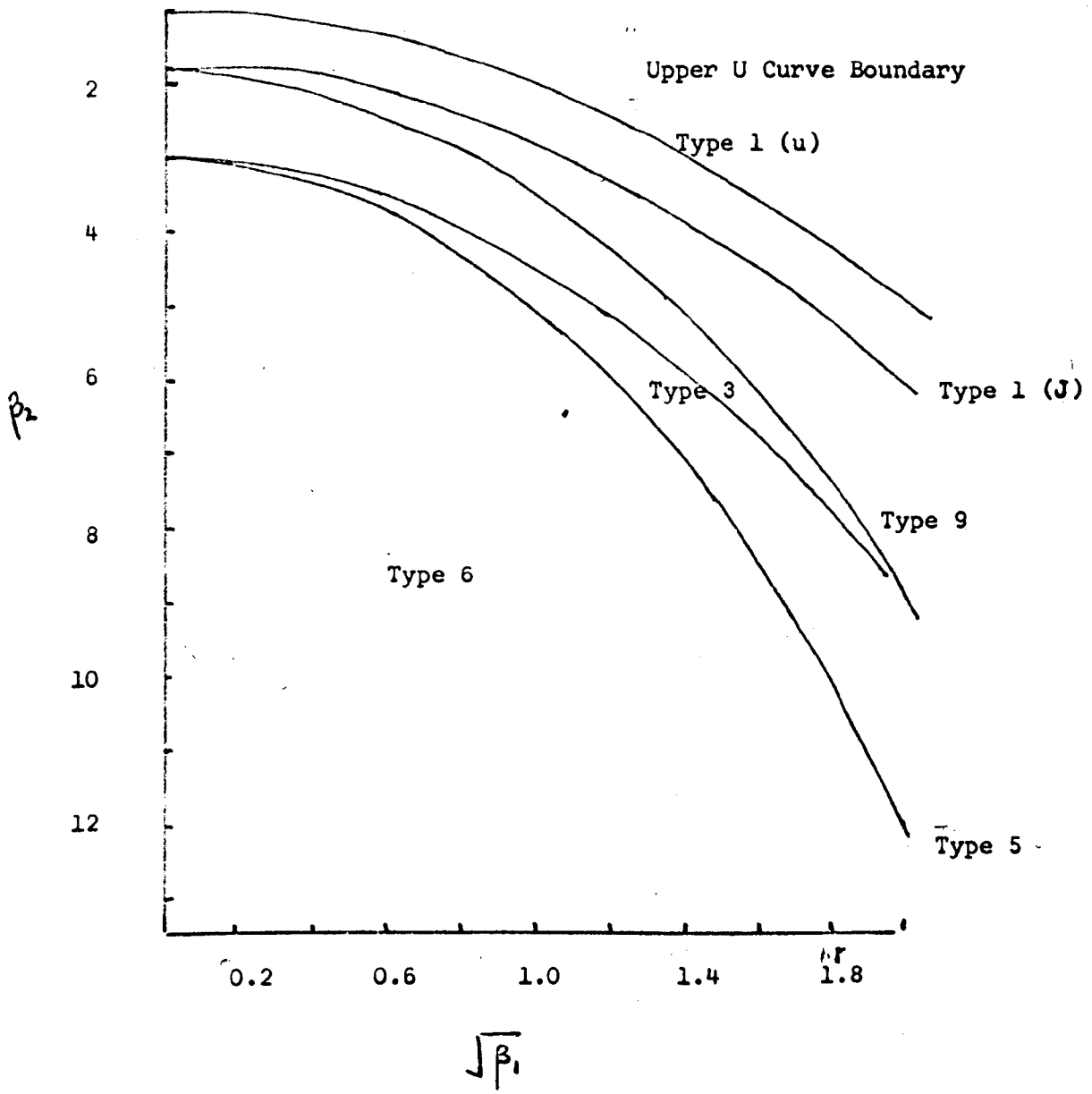
TABLE 21

Simulation results for the fit of two moving average models of the form

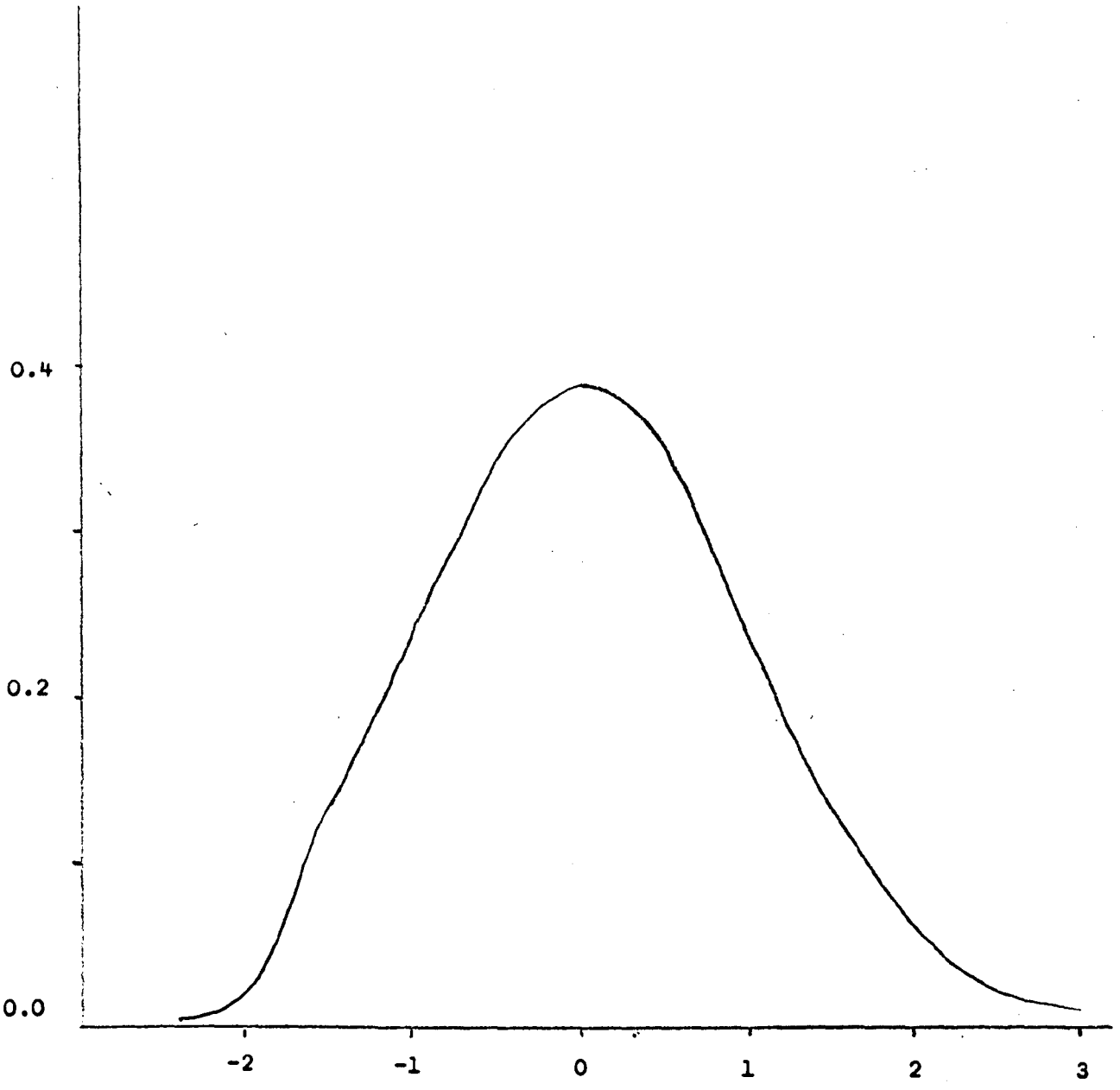
$$X_t = \varepsilon_t - 0.5\varepsilon_{t-1} \qquad t = 1 \ldots N$$

Replication	500	400
N	32	64
Percentage		
0.25	0.2	0.5
0.5	0.4	0.5
1.0	0.75	1.0
2.5	1.25	2.0
5.0	3.4	4.75
10.0	8.0	8.00
25.0	21.4	24.75
50.0	49.8	50.0
75.0	77.2	75.0
90.0	92.0	90.25
95.0	96.0	95.0
97.5	98.75	97.0
99.0	99.0	99.0
99.5	100	99.75
99.73	100	100

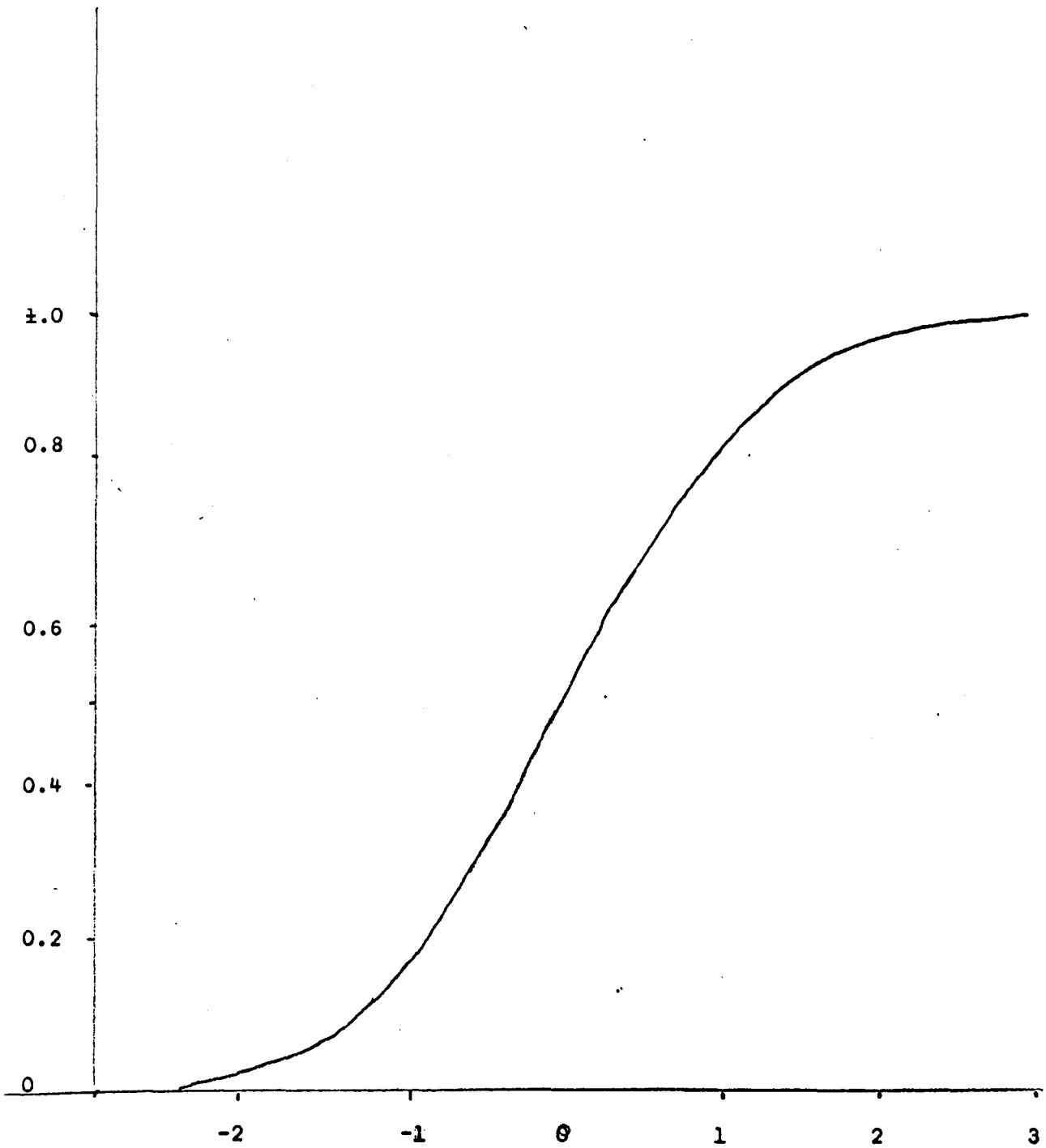
FIG 8


Upper U Curve Boundary
Type 1 (u)
Type 1 (J)
Type 3
Type 9
Type 6
Type 5
2
4
6
8
10
12
β_2
0.2
0.6
1.0
1.4
1.8
$\sqrt{\beta_1}$

P.D.F Edgeworth Series



C.D.F Edgeworth series



CHAPTER 6

As we have seen in Chapter 4 it is possible to devise a test for white noise by comparing the estimate of the minimum mean square error of prediction with the sample variance. Unfortunately as was pointed out this approach is unsatisfactory for several reasons and in consequence we shall now investigate a slightly different approach which appears to be superior in almost every respect. Our test for white noise will use the approximations for the distribution of

$$\hat{\Delta}_I = -\frac{1}{m} \sum_{p=0}^{m} \log I_N(\omega, x) - \log 2 + \gamma$$

which we derived in the previous chapter and this enables us to consider the small sample as well as the asymptotic case.

Consider a stationary non-deterministic process $\{X_t : t = 1 \cdots p+q+I\}$.

Using the first segment of the process $X_1 \ldots X_p$ we estimate the logarithm of the minimum mean square error of prediction $\log \sigma_1$ using $\hat{\Delta}_I$ defined above. Then taking the last segment of the series $X_{p+I+1}, X_{p+I+2}, \cdots X_{p+q+I}$ we estimate the variance of the series using the sample variance.

As we have seen the sample variance can be considered to be a sum of periodogram ordinates and in addition the periodograms computed from differing segments of the sequence are uncorrelated, thus

$$\operatorname{var}(\hat{\Delta}_I \pm \log S_x^2) = \operatorname{var}(\hat{\Delta}_I) + \operatorname{var}(\log S_x^2) \qquad 6.1$$

Now let

$$D = \hat{\Delta}_I - \log S_x^2 - b$$

where

$$b = -\log\left(\frac{N-1}{2}\right) + \psi\left(\frac{N-1}{2}\right) \qquad 6.2$$

as in 4.15 when $N = q$

We know that for a truncated estimater

$$\mathrm{var}(\hat{\Delta}_I) = \frac{\psi'(1)}{m} \qquad m = \left[\frac{p}{2}\right] \tag{6.3}$$

thus as $m \to \infty$ we have as an approximation

$$z = \frac{D}{\sqrt{\frac{\psi'(1)}{m} + \psi'\left(\frac{N-1}{2}\right)}} \tag{6.4}$$

has zero mean and unit variance and has a distribution which is approximately standard normal. If we use the trapizodal form then we replace our approximation $\frac{\psi'(1)}{m}$ by

$$\frac{\psi'(1)}{m} + \frac{\psi'(0.5)}{2m^2} \tag{6.5}$$

This use of segments of the total realisation enables us to avoid estimating the correlation. While the variance of D is inflated by comparison with the form ℓ given by 4.16 we feel that this is not unreasonable in view of the rather poor estimates of covariance available.

Our next problem is to choose the relative magnitudes of p and q. One possibility which is intuitively reasonable is to choose p and q such that

$$\mathrm{var}(\hat{\Delta}_I) \simeq \mathrm{var}(\log s_x^2) \tag{6.6}$$

when $\{X_t\}$ is a white noise sequence. Using the asymptotic expansion

$$\psi'(z) = \frac{1}{z} + \frac{2}{z^2} + O\left(\frac{1}{z^3}\right) \tag{6.7}$$

see Erdelyi (1953), and taking the dominant term in the expansion

$$\psi'\left(\frac{N-1}{2}\right) \simeq \frac{2}{N-1} \simeq \frac{1}{N}$$

while $$\frac{\psi'(1)}{m} \simeq \frac{1.6}{m}$$

Thus $$\frac{3.2}{p} = \frac{2}{q}$$

giving $$p \simeq 1.6q \qquad 6.8$$

In view of 6.6 we can choose a suitable p and q, however for short series the attraction of choosing p and q equal does appear to dominate.

While we can, as we have seen, appeal to the asymptotic results for the distribution of D it would be preferable to obtain some approximation to the small sample distribution since in effect we are using half our sample points for the limiting distribution.

If we make the not unreasonable assumption that $\hat{\Delta}_I$ and $\log s_x^2$ can be assumed to be approximately independent, an assumption that can be made more realistic by choosing $I > 0$ then we can obtain the following expressions for the cumulants of D

$$\begin{aligned} K_1(D) &= K_1(\hat{\Delta}_I) - K_1(\log s_x^2) \\ K_2(D) &= K_2(\hat{\Delta}_I) + K_2(\log s_x^2) \\ K_3(D) &= K_3(\hat{\Delta}_I) - K_3(\log s_x^2) \\ K_4(D) &= K_4(\hat{\Delta}_I) + K_4(\log s_x^2) \end{aligned} \qquad 6.9$$

and in general

$$K_r(D) = K_r(\hat{\Delta}_I) + (-1)^r K_r(\log s_x^2)$$

Here $K_s(\Sigma)$ denotes the s^{th} cumulant of the distribution of the random variable Σ. Since the component cumulants in 6.9 are known we can evaluate the $K_s(D)$ for all s

For $\{X_t\}$ white noise then

$$K_2(D) = \frac{\psi'(1)}{m} + \psi'\left(\frac{N-1}{2}\right) \qquad 6.10$$

and $$K_r(D) = \frac{\psi^{(r-1)}(1)}{m^{r-1}} + (-1)^{(r-1)}\,\psi^{(r-1)}\left(\frac{N-1}{2}\right) \qquad 6.11$$

were for algebraic simplicity we have used the approximate forms, however the modification is minor. Since the $\psi^{(r)}(z)$ are tabulated by Abramowitz and Segun (1968) or may be computed from the following asymptotic expansions, see Erdelyi (1953)

$$\psi''(z) = -\frac{1}{z^2} - \frac{1}{z^3} - \frac{1}{2z^4} + \frac{1}{6z^6} - O\left(\frac{1}{z^8}\right) \qquad 6.12$$

$$\psi'''(z) = \frac{2}{z^3} + \frac{3}{z^4} + \frac{2}{z^5} - \frac{1}{z^7} + O\left(\frac{1}{z^9}\right) \qquad 6.13$$

in general

$$\psi^{(n)}(z) = (-1)^{n+1}\, n! \sum_{k=0}^{\infty} (z+k)^{-n+1} \qquad 6.14$$

in principle the $K_j(D)$ are easily evaluated for all j, N, m.

Using the dominant terms in the expansions and setting $N = 2m$

$$K_2(D) \simeq \frac{0.64}{m}$$

$$K_3(D) \simeq -\frac{3.4}{m^2}$$

$$K_4(D) \simeq \frac{0.41}{m^3} \qquad 6.15$$

giving

$$\beta_1 = -\frac{6.6}{\sqrt{m}} \qquad 6.16$$

while

$$\beta_2 = 3 + \frac{0.98}{m} \qquad 6.17$$

thus in the limit as $N \to \infty$ $\beta_1 = 0$, $\beta_2 = 3$ which gives us the same parameters as the normal distribution and it would seem a reasonable inference that D has a limiting normal distribution.

For convenience table 22 gives the value of the psi-gamma function $\psi\left(\frac{N-1}{2}\right)$ and its derivatives for a range of values of N.

To construct a test of significance based on D we can as before consider the null hypothesis

$$H_0: \operatorname{var}(X_r) = \sigma_X^2 = \sigma^2 \qquad 6.18$$

TABLE 22

The psi-gamma function $\psi(z)$ for arguments of the form $\frac{N-1}{2}$ over a range of values of N.

N	$\psi(\frac{N-1}{2})$	$\psi'(\frac{N-1}{2})$	$\psi''(\frac{N-1}{2})$	$\psi'''(\frac{N-1}{2})$
32		0.06666	-0.004444	0.00059
64		0.03222	-0.00104	0.00007
128	4.1431	0.01584	-0.00025	0.00001
256	4.8442	0.00784	-0.00006	0.0000
512	3.1781	0.00389	-0.00001	0.0000
50	3.1781	0.04163	-0.00173	0.00014
100	3.8918	0.02038	-0.00042	0.00002
200	4.5941	0.01007	-0.00010	0.00000
300		0.00668	-0.00004	0.00000

σ^2 being the "innovation variance" against an alternative of the form

$$H_1 : \sigma_x^2 \neq \sigma^2 \qquad 6.19$$

For large samples, under H_0, D is approximately normal with zero mean and variance given by 6.5 However if our alternative is that $\{X_t\}$ has the form

$$A(B)\, X_t = \varepsilon_t \qquad 6.20$$

that is

$$\operatorname{var}(X_t) = \operatorname{var}\left(\frac{\varepsilon_t}{A(B)}\right) \qquad 6.21$$

then we need to find the distribution of D under H_1 to have any idea of the power of the test. Thus in effect we require the distribution of $\log s_x^2$ under H_1

Proceeding in the same manner as in Chapter 4 we can assume that the distribution of $\log s_x^2$ is capable of being approximated by a normal distribution with mean and variance given by

$$\mu = \log \sigma_x^2 \qquad 6.22$$

$$\mu_2 = \frac{V}{\sigma_x^4 N} \qquad 6.23$$

where

$$V = \lim_{N \to \infty} V_N = \sum_{t=-\infty}^{\infty} R_t^2 \qquad 6.24$$

Thus for

$$X_t = \alpha X_{t-1} + \varepsilon_t \qquad |\alpha| < 1 \qquad 6.25$$

we have

$$\sigma_x^2 = \frac{\sigma^2}{1-\alpha^2} \qquad 6.26$$

and

$$\operatorname{var}(\log s_x^2) = \frac{\sum R_t^2}{\sigma^4 N} \qquad 6.27$$

since $R_t = \alpha R_{t-1}$ we have

$$\operatorname{var}(\log s_x^2) = \left(\frac{1+\alpha^2}{1-\alpha^2}\right)\frac{1}{N} \qquad 6.28$$

and $\log s_x^2$ is asymptotically normal with mean

$$\log\left(\frac{\sigma^2}{1-\alpha^2}\right)$$

and variance

$$\frac{\psi'(1)}{m} + \frac{1+\alpha^2}{(1-\alpha^2)N}$$

Even under the assumption of normality it is clear that the first order autoregressive is a very simple case, and in general for higher order autoregressive or mixed models the calculation of the mean and variance of $\log s_x^2$ become very much more complicated.

However even in these cases, it is possible to make some progress provided we can write down the spectral density. For example consider

$$A(B)\,X_t = \theta(B)\,\varepsilon_t \qquad 6.29$$

with spectrum $$f(\omega) = \left|\frac{\theta(e^{i\omega})}{A(e^{i\omega})}\right|^2 \frac{\sigma^2}{2\pi}$$

Then we can write

$$\sigma_x^2 = \int_{-\pi}^{\pi} f(\omega)\,d\omega \qquad 6.30$$

and since $f(\omega)$ is known we can evaluate 6.30 either analytically or use a suitable numerical integration algorithm such as those given by Clenshaw and Curtis (1960).

For V we can use Parcevals Theorem to give

$$V = \sum_{t=-\infty}^{\infty} R_t^2 = \int_{-\pi}^{\pi} |f(\omega)|^2 d\omega \qquad 6.31$$

Once again the integral in 6.31 can be evaluated analytically since $f(\omega)$ is a rational function, however we would surmise that for many classes of functions $f(\omega)$ some numerical technique would probably prove more convenient, perhaps even more so than in the evaluation of the variance.

As an example we consider the following simple second order case

$$(1-\alpha B)(1-\beta B)X_t = \varepsilon_t \qquad 6.32$$

Then

$$\sigma_x^2 = \int_{-\pi}^{\pi} \frac{\sigma^2 d\omega}{2\pi |1-\alpha e^{-i\omega}|^2 |1-\beta e^{-i\omega}|^2}$$

and using the transformation $z = e^{i\omega}$ and Cauchy's residue theorum we have

$$\sigma_x^2 = \frac{(1+\alpha\beta)\sigma^2}{(1-\alpha\beta)(1-\alpha^2)(1-\beta^2)} \qquad 6.33$$

for the stationary case.

The evaluation of V is rather more complex;

$$V = \frac{\sigma^4}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{d\omega}{|1-\alpha e^{-i\omega}|^4 |1-\beta e^{i\omega}|^4} \qquad 6.34$$

$$= \frac{\sigma^4}{c^2 4\pi^2} \int_0^{2\pi} \left\{ \frac{\alpha}{1+\alpha^2+2\alpha\cos\omega} - \frac{\beta}{1+\beta^2+2\beta\cos\omega} \right\}^2 d\omega$$

and

$$c = (\alpha-\beta)(1-\alpha\beta)$$

Writing

$$\Psi(\alpha) = \frac{1}{1+\alpha^2+2\alpha\cos\omega}$$

$$V = \frac{\sigma^2}{4\pi^2 c^2} \int_0^{2\pi} \left[\alpha^2\Psi^2(\alpha) + \beta^2\Psi^2(\beta) - 2\alpha\beta\Psi(\alpha)\Psi(\beta) \right] d\omega \qquad 6.35$$

for $|\alpha|<1$ and $|\beta|<1$

$$\int_0^{2\pi} \Psi(\alpha)\Psi(\beta)\,d\omega = \frac{2\pi(1+\alpha\beta)}{(1-\alpha\beta)(1-\alpha^2)(1-\beta^2)} \qquad 6.36$$

$$\int_0^{2\pi} \Psi^2(\alpha)\,d\omega = \frac{2\pi}{(1-\alpha^2)^2}\left\{ 1 + 2\left(\frac{\alpha^2}{1-\alpha^2}\right)\right\} \qquad 6.37$$

giving

$$V = \frac{1}{2\pi c^2}\left\{ \frac{\alpha^2(1+\alpha^2)}{(1-\alpha^2)^3} + \frac{\beta^2(1+\beta^2)}{(1-\beta^2)^3} - \frac{2(1+\alpha\beta)\alpha\beta}{(1-\alpha\beta)(1-\alpha^2)(1-\beta^2)} \right\}. \qquad 6.38$$

It should be noted that 6.38 is real even if the α, β are complex. Since the α, β are conjugate pairs one can after some algebra show that the complex part of V vanishes.

Since we have only large sample approximations it is worth attempting to find an approximation to the distribution of $\log s_x^2$ based on the cumulants of $\log s_x^2$ under H_1.

Now
$$s_x^2 \simeq \frac{4\pi}{n} \sum_{k=0}^{n} f(\omega)\, I_N(\omega, \varepsilon) \qquad 6.39$$

and denoting the cumulants of s_x^2 by $\bar{K}_1, \bar{K}_2 \cdots$

we have
$$\begin{aligned}
\bar{K}_1 &= \frac{2}{n} \sum_{k=0}^{n} 2\pi f(\omega) K_1 \simeq \frac{K_1}{\pi} \int_0^{\pi} 2\pi f(\omega)\, d\omega \\
\bar{K}_2 &= \frac{4}{n^2} \sum_{k=0}^{n} [2\pi f(\omega)]^2 K_2 \simeq \frac{2K_2}{\pi} \int_0^{\pi} [2\pi f(\omega)]^2 d\omega \\
\bar{K}_3 &= \frac{8}{n^3} \sum_{k=0}^{n} [2\pi f(\omega)]^3 K_3 \simeq \frac{4K_3}{\pi} \int_0^{\pi} [2\pi f(\omega)]^3 d\omega \qquad 6.40
\end{aligned}$$

where $K_1, K_2 \cdots$ denote the cumulants of $I_N(\omega, \varepsilon)$.

In principle these integrals can be evaluated since $f(\omega)$ is known. In addition we can obtain the cumulants of $\log s_x^2$ by using a Taylor series approximation based on the equations 6.40. However this means our calculations have become very cumbersome and in addition the chance of making errors has become rather too large.

As an alternative we suggest approximating the distribution of $\log s_x^2$ by $a + bz$ where z is a chi-squared variable with $\nu = \left[\frac{N-1}{2}\right]$ degrees of freedom, the a and b being chosen to match the degrees of freedom. Alternatively since $\operatorname{var}(\ln s_x^2) < \operatorname{var}(\hat{\Delta})$ one could

estimate $\hat{\Delta}_I$ and $\log s_x^2$ from segments of equal length and use the normal approximation.

To give some idea of the efficiency of the test outlined above we conducted some simulation exercises. Several white noise series

$$X_t = \varepsilon_t \qquad t = 1 \ldots N \tag{6.41}$$

were generated and the values of $\hat{\Delta}_I$ obtained were compared with those for some first order autoregressive series

$$(1 - \alpha B) X_t = \varepsilon_t \qquad t = 1 \ldots N \tag{6.42}$$

and with two second order models, these being

$$(1 - 0.8B + 0.7B^2) X_t = \varepsilon_t \tag{6.43}$$

and

$$(1 - 0.75B + 0.5B^2) X_t = \varepsilon_t \tag{6.44}$$

For long series, that is values of $N = 256$ the means and variances fit the theoretical predictions in the white noise case very closely as can be seen from tables 23 and 24. More interesting however is the very close fit of the best approximating Pearson curve. Table 25 gives the sample values compared with theoretical values using the Pearson significance points tabulated in Appendix 3. As can be seen the fit is very good indeed and it appears that our small sample approximation for the distribution of D works very well under the assumption that ε_t is white noise. Table 24 also gives the mean and variance of D when only one end point is included in the estimation of $\hat{\Delta}_I$. As we can see the sample values in this case are rather poorer thus indicating as we have said before that estimates of $\log e_1$ using $\hat{\Delta}_I$ should make use of both end points. Thus we can make probability statements and construct confidence intervals for $\log e_1$ If we are to use these results in a test of significance then we need some idea of the power of our test.

TABLE 23

White Noise Models

Means

Series Length	No. of Reps	Theoretical Value	Sample Value
128	300	0.0	-0.0022
		0.0	0.0107
128	300	0.0	-0.0036
		0.0	0.0109
128	300	0.0	-0.00442
		0.0	+0.0117
Pooled values for the above		0.0	0.0034
			0.0111
64	400	0.0	0.0078
16	400	0.0	0.0399
		0.0	0.0410

In each case the second value of the sample mean uses an estimate with the upper end point set to zero.

TABLE 24

White Noise Models

Variances

Series Length	No. of Reps	Theoretical Value	Sample Value
128	300	0.0207	0.0196
		0.0207	0.0202
128	300	0.0207	0.0200
			0.0199
128	300	0.0207	0.0199
			0.0182
64	400	0.0415	0.0423
			0.0426
16	400	0.1769	0.1752
			0.1798
Pooled values for 128	900	0.0207	0.0199

Using the Pearson curve approximation for the distribution of D under H_0 we decided to examine the distribution of some simulated values of D based on a variety of first order autoregressive models and some second order models.

For a model of the form $(1 - 0{\cdot}5B)X_t = \varepsilon_t$ with $m = 128$ we tabulate below the percentage of simulated values of D which do not exceed the white noise percentage points (600 replications)

TABLE 25

White Noise % points	0.25	0.5	1.0	2.5	5	10	25	50
% of values of D	24.7	32.3	39.0	52.7	62.7	76.3	87.3	95.0

As we can see from Table 25 the proposed test of significance works very well for large samples. However $m = 128$ implies a total realisation of 512 when $N = 2m$. For more realistic values of m we have for $m = 64$ an equivalent set of results as shown below for models of the form

$$(1 - \alpha B) X_t = \varepsilon_t$$

TABLE 26

The fit of white noise models to Pearson curve approximations

True Percentage	Sample Percentages not exceeding the Pearson percentage points		
Sample size	400	400	600
m	16	64	128
0.25	0.25	0.25	0.5
0.5	0.25	0.25	0.67
1.0	0.75	1.00	1.17
2.5	2.0	2.75	2.00
5.0	4.0	6.25	4.33
10.0	9.25	10.5	9.67
25.0	23.0	28.75	24.67
50.0	49.0	52.00	50.17
75.0	76.0	75.75	74.5
90.0	88.5	91.0	90.5
95.0	92.75	94.5	95.33
97.5	97.0	97.75	98.33
99.0	99.0	99.25	99.17
99.5	99.5	99.75	99.67
99.75	100	99.75	99.67

TABLE 27

White noise Percentages (400 replications)

α	% of values of D	0.25	0.5	1.0	2.5	5	10	25	50
0.2		1.3	1.5	3.0	6.0	8.8	16.5	38.0	64.3
0.4		3.3	5.5	7.8	16.0	21.8	35.3	53.3	79
0.5		8.0	12.7	19.2	28.5	38.5	52.7	71.5	89.2
0.8		90.3	92.5	93.8	96.0	98.3	99.5	100	100
0.9		99.3	99.3	99.5	99.8	100	100	100	100

We see from Table 27 the power of the test increases rapidly as $\alpha \to 1$ and this is illustrated by Figure 11 which gives the power curves extrapolated from the above table. Three curves are given corresponding to Type 1 errors of 0.01 0.025 and 0.05 respectively.

The total realisation in the above case corresponded to 400 We also consider the case where the total realisation was of length 64 and $N = 40$ This was taken as the smallest practicable case since we must in effect estimate the spectrum from 32 observations. However in view of the good approximations we have obtained using such small sample sizes it might be worth looking at some slightly smaller values of M.

TABLE 28

White Noise Percentage	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
$M = 16$								
α \ Sample Percentage								
0.2	0.5	0.75	1.75	4.0	7.0	12.25	27.25	55
0.5	1.0	2.5	4.75	8.25	14.25	20.25	44.5	68.5
0.8	23.25	29.5	36.5	50.25	57.75	68.75	83.25	96.00
0.9	46.5	53	57.5	73.0	79.5	90.75	96.5	99.0
-0.9	55	60.5	66.0	74.75	86.5	92.75	97.25	99.0

(400 replications)

As we can see from the Table 28 above the test works well even with M = 16 and provides a satisfactory alternative to the method of Davis and Jones (1968) outlined in Chapter 4.

To make our simulation experiments more comprehensive we also generated two second order autoregressive models

Model A $$X_t - 0.8 X_{t-1} + 0.7 X_{t-2} = \varepsilon_t$$

and Model B $$X_t - 0.75 X_{t-1} + 0.5 X_{t-2} + \varepsilon_t$$

and in these cases we find the proportion of rejections is given by Table 29.

TABLE 29

White Noise Percentages	0.25	0.5	1.0	2.5	5	10	25	50
Sample Values								
Model A	88.0	90.0	93.0	94.5	98.0	98.5	98.5	100
Model B	44.5	50.0	61.0	71.0	79.5	86.5	95	99.5

and these results with $M = 64$ and a sample size of 600 would appear to indicate a very powerful test.

We conclude that our method outlined above works well in discriminating between white noise and non-white noise time series models. If we could identify the small sample distribution under any alternative we could also discriminate between alternative time series models. We feel that it is an avenue that is worth pursuing.

Since we had the sample distribution of D from our simulations we also compared the mean and variance of our first and second order models with the values predicted by 6.28 and 6.38 The results given in Tables 30 and 31 show that the theoretical values appear to give very good predictions and we would hope from these results that the power of a general test would be reasonable.

Given such good approximations clearly for large M we can use the Normal approximation, in the small sample case we feel that the chi-squared approximation for S_λ^2 is worth investigating.

TABLE 30

FIRST ORDER MODELS

m	α	No. of Replications	Mean Theoretical	Mean Sample
64	0.2	600	0.0408	0.0387
64	0.4	400	0.1744	0.1657
64	0.5	400	0.2877	0.2621
64	0.8	400	1.0217	0.9363
64	0.9	400	1.6607	1.4798
128	0.5	500	0.2877	0.2853
16	0.2	400	0.0408	0.0206
16	0.5	400	0.2877	0.2172
16	0.9	400	1.6607	1.4876

SECOND ORDER MODELS

64	Model A	200	0.9234	0.9036
64	Model B	200	0.5744	0.5507

TABLE 31

FIRST ORDER MODELS

M	α	No. of Replications	Variance Theoretical	Variance Sample
64	0.2	600	0.04263	0.04206
64	0.4	400	0.04728	0.04549
64	0.5	400	0.05174	0.05139
64	0.8	400	0.09688	0.08764
64	0.9	400	0.17455	0.1358
128	0.5	500	0.02587	0.02560
16	0.2	400	0.15594	0.15478
16	0.5	400	0.19314	0.17417
16	0.9	400	0.41978	0.41114

SECOND ORDER MODELS

64	Model A	200	0.06341	0.06220
64	Model B	200	0.05934	0.05933

As we have seen however our test for white noise works well and we can consider models of higher order than the first unlike the situation described by Davis and Jones (1968). The simpler correlation structure of D would also suggest that this approach could be usefully extended.

It has not escaped our notice that for a first order autoregressive model

$$E[D] = -\log(1-\alpha^2) \qquad 6.45$$

so that $$\alpha^2 = 1 - e^{-E[D]}$$

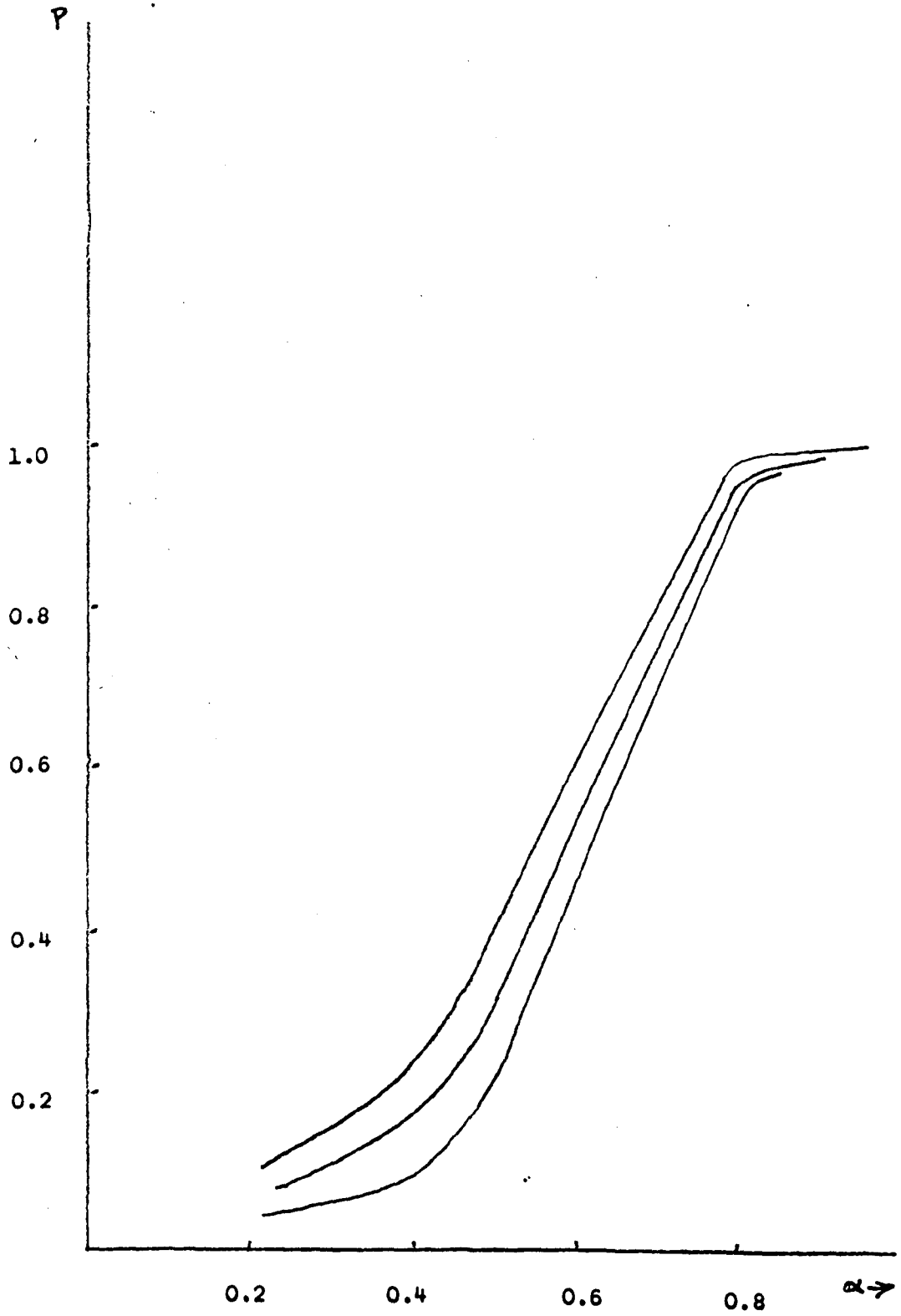
and suggesting that for $\alpha < 1$ a possible estimate of α might be obtained from

$$\hat{\alpha} = \pm\sqrt{1-e^{-D}} \qquad 6.46$$

This is really a very special case and clearly we need additional information so as to determine the sign of 6.46

In fact however we are able to estimate the coefficients of our model directly using a modification of $\hat{\Delta}_1$ and D. Before proceeding further we feel it is an appropriate moment to discuss the use of the spectrum estimates in evaluating $\log e_1$ by $\hat{\Delta}_f$ Thus we shall for a moment defer our discussion of the estimation of parameters and look more closely at $\hat{\Delta}_f$

Power Curves for White Noise Test



CHAPTER 7

As we have seen it is possible to estimate $\log e_1$ using $\hat{\Delta}_I$ and from our various simulation exercises it appears that this estimator behaves as our theoretical development would predict. Moreover we can use $\hat{\Delta}_J$ as the basis of hypothesis tests of some interest. However as we have shown in Chapter 3 one can also construct an estimator

$$\hat{\Delta}_f = \frac{1}{M} \sum_{p=0}^{M}{}' \log \hat{f}(\omega) 2\pi \qquad 7.1$$

of $\log e_1$ based on the traditional "smoothed periodogram" spectral estimate and some of the theory we have developed for $\hat{\Delta}_I$ carries over to $\hat{\Delta}_f$

It must be born in mind that we cannot assume that adjacent spectral estimates are independent since our smoothing kernal has a finite bandwidth and in consequence we are forced to make assmptions when dealing with $\hat{\Delta}_f$ about the degree of correlation between estimates of $f(\omega)$ which are not required when using the periodogram. As a consequence of this the estimate $\hat{\Delta}_f$ is rather more difficult to construct and use.

We now look at $\hat{\Delta}_f$ rather more closely and compare its theoretical behaviour with behaviour in some simulation experiments.

As we have seen we may define $\hat{\Delta}_f$ as

$$\hat{\Delta}_f = \frac{4}{M} \sum_{p=0}^{[\frac{M}{2}]}{}' \log 2\pi \hat{f}(\omega_p) - \log 2 + \psi(\tfrac{1}{2}) \qquad 7.2$$

where M is the number of lags used in the estimation of $f(\omega)$ the power spectrum

We estimate $f(\omega)$ in the usual form

$$\hat{f}(\omega) = \sum_{s=-M}^{M} \hat{R}(s) \lambda\left(\frac{s}{M}\right) \cos s\omega$$

where $\lambda(k)$ is a suitably chosen weighting sequence, see Parzen (1962). We choose the Parzen (1961) sequence defined as

$$\lambda(k) = \begin{cases} 1 - 6k^2 + 6|k|^3 & |k| < 0.5 \\ 2(1-|k|)^3 & 0.5 < |k| < 1.0 \\ 0 & \text{otherwise} \end{cases}$$

and a covariance estimate of the form

$$\hat{R}_s = \frac{1}{N} \sum_{t=1}^{N-|s|} X_t X_{t+|s|}$$

since as we have seen in Chapter 1 the spectral estimates $\hat{f}(\omega)$ are non-negative.

For an estimate of the form 7.1 we have seen that the cumulants are given by 3.102 or for an estimate of the form

$$\hat{\Delta}_r = \frac{2}{M} \sum_{1}^{\lfloor \frac{M}{2} \rfloor} \log \hat{f}(\omega) - \log 2 + \psi(\tfrac{1}{2})$$

with mixing end points by 3.103 that is

$$\kappa_1 = \psi(\tfrac{1}{2})$$

$$\kappa_2 = \frac{1}{m} \psi'(\tfrac{1}{2})$$

$$\kappa_r = \frac{1}{m^{r-1}} \psi^{(r-1)}(\tfrac{1}{2})$$

where

$$m = \left[\frac{M}{2}\right] \qquad 7.3$$

Thus from these cumulants we can for given M estimate the distribution of $\hat{\Delta}_r$ However the form of 7.2 implies that we have estimated $\hat{f}(\omega_p)$ for $\omega_p = \frac{2\pi}{M}$ For some spectral windows this is not unreasonable for one might reasonably assume that adjacent estimates of $f(\omega)$ are uncorrelated. From Table 4 however we can see that the bandwidth of the Parzen kernal is $\frac{8\pi}{3M}$ and from Table 3 we can see

that at $\frac{2\pi}{M}$ there is still a fair degree of correlation. The implication therefore is that our estimate should be modified to the form

$$\hat{\Delta}_F = \frac{3}{M} \sum_{p=0}^{[\frac{M}{3}]}{}' \log 2\pi \hat{f}(\omega_p) - \log 2 + \psi(\nu/2) \qquad 7.4$$

where the separation is now $\frac{3\pi}{M}$

In general we require that the minimum separation between adjacent spectral estimates should be at least as large as the bandwidth of the spectral window used. If this is not so then the correlation introduced can swamp our approximations as will be seen in some of the simulations.

A further complication is introduced by the choice of the parameter M which must be chosen to balance the resolution against variance of the spectral estimate. Clearly we would wish to minimise bandwidth since this reduces the correlation problem and introduces more terms into the summation 7.4 We feel that one should choose M/N to be large.

The estimator $\hat{\Delta}_I$ can be viewed as a smoothed function of the periodogram, rather in the same manner as the spectral estimate $\hat{f}(\omega)$. However $\hat{\Delta}_F$ is the result of a double smoothing operation on the periodogram. This introduction of an extra smoothing operation appears to introduce an extra degree of complexity. However we have to bear in mind the loss of resolution when the ratio N/M is large.

To obtain some idea of the accuracy of our theoretical results we conducted some simulation experiments. Using first an estimator

without end points we generated 400 series of length 100, these series $\{\varepsilon_t\}$ consisting of independent standard normal deviates. For each series $\hat{\Delta}_t$ was computed and the sample mean on variance of these results is compared with our theoretical predictions in Table 32 below

TABLE 32

Mean		Variance	
Sample	Theoretical	Sample	Theoretical
-0.13818	0.0	0.02066	0.01195

In this case the spectrum was evaluated at increments of $\frac{2\pi}{M}$ where M was chosen as being 33.

As is evident from the above the estimator $\hat{\Delta}_t$ does not compare well with our periodogram based estimate $\hat{\Delta}_I$. One may go further and conduct the same experiment but evaluate the spectrum at increments of π/M thus giving M terms in the summation. Again we use M = 33. As we might expect the results for the first two sample moments of $\hat{\Delta}_t$ given in Table 33 below are quite at variance with the theoretical ones.

TABLE 33

Mean		Variance	
Sample	Theoretical	Sample	Theoretical
-0.14585	0.00000	0.02460	0.00579

Thus as one might expect the values of Δ_t have inflated variance because of the correlation between spectral estimates.

We attempted a similar experiment generating 1000 series of the form

$$X_t = 0.5\, X_{t-1} = \varepsilon_t$$

where the ε_t again were independent standard normal and our results are tabulated in Table 34

TABLE 34

Mean		Variance	
Sample	Theoretical	Sample	Theoretical
-0.12910	0.00000	0.04585	0.02380

these being roughly comparable with the results in Table 32.

The results above while showing that the variance of $\hat{\Delta}_t$ is inflated also indicate that for the simulation cases indicate that the estimator is also biased to some extent. This is rather disturbing since this is unexpected.

We also tried simulating the models

$$X_t = \varepsilon_t$$
$$X_t = 0.5\, X_t + \varepsilon_t \qquad t = 1 \ldots 100$$

and computed $\hat{\Delta}_t$ using a separation of $\frac{3\pi}{M}$ In these cases the mean and variance were much improved as may be seen from Table 35.

TABLE 35

Model	Mean		Variance	
	Sample	Theoretical	Sample	Theoretical
White Noise	0.0472	0.0000	0.0212	0.0202
Autoregressive	0.0609	0.0000	0.0226	0.0202

These values are quite good and appear to support our speculation that one requires separations of at least $\frac{3\pi}{M}$ However the sample fit to the approximating Pearson curve is exceedingly poor. For example only 37% of sample values lie to the left of the predicted median.

One would hope that $\hat{\Delta}_t$ would behave in a more suitable fashion if alternative values of M and the separation were chosen and in addition the separation was taken to be at least the bandwidth of the smoothing window used. The problem then for short series, say N = 100 is that one reduces the number of terms in the summation 7.1

While much more investigation of $\hat{\Delta}_t$ is possible especially from the simulation viewpoint we feel that it is probably not worth while since the periodogram based estimate $\hat{\Delta}_I$ works so well.

We also considered the case of an estimator $\hat{\Delta}_t$ with end points included. For 1000 replications of a model

$$X_t = 0.5\ X_{t-1} + \varepsilon_t \qquad t = 1 \ldots 50$$

we have sample results for the mean and variance of $\hat{\Delta}_t$ as given in Table 35 below where M = 16 and the separation is $\frac{2\pi}{M}$

TABLE 35

Mean		Variance	
Sample	Theoretical	Sample	Theoretical
0.09007	0.00000	0.07279	0.03359

As we see the addition of the end points decreases the error in the mean, and since only 16 points are used to compute $\hat{\Delta}_F$ one would expect the contribution from the end points to be large. However the sample variance is still much larger than one would predict from our study of the estimated moments. Thus even the full form of $\hat{\Delta}_F$ compares badly with $\hat{\Delta}_I$

As we indicated earlier this is perhaps to be expected however in view of our simulations we feel that one is better off considering For very long realisations however, one has more room to manoeuvre in choosing M and the separation between adjacent spectral estimates and we feel that the estimate $\hat{\Delta}_F$ might exhibit more reliable behaviour.

In view of the inflation of the variance we have not included tables of fit to the appropriate Pearson curve since as one might expect the fit is poor. However for completeness we have included in the appendix the percentage points of the appropriate Pearson curve for a small set of values of M the number of spectral estimates used in evaluating $\hat{\Delta}_F$

In Chapter 4 we also outlined some theoretical results which suggested that

$$z' = \frac{\hat{\Delta}_F - \log s_x^2 - \beta}{\sqrt{\frac{\psi'(1/2)}{M} - \psi'\left(\frac{N-1}{2}\right)}} \qquad 7.5$$

could be used in the same manner as $\hat{\Delta}_1$ to test for white noise and further indicated that as $N \to \infty$ then z' is approximately standard normal when $\{y_t\}$ is a white noise sequence.

However to use 7.5 we have to choose m very carefully, for if $M = \frac{M}{K}$ then

since $$\psi'(z) = \frac{1}{z}$$

the denominator in 7.5 becomes

$$D_n = \frac{\psi'(\nu/2)}{m} - \psi'\left(\frac{N-1}{2}\right) = \frac{2}{M\nu} - \frac{2}{N-1} \qquad 7.6$$

and for a Parzen window $\nu = 3.7\frac{N}{M}$

thus $$D_n = \frac{2K}{3.7N} - \frac{2}{N-1} \qquad N \geq 133 \qquad 7.7$$

Now it is evident that we require D to be greater than zero and in consequence we must choose $K \geq 4$ as may be seen from 7.7

Thus to make use of the Davis and Jones form of test we must choose a separation of $\frac{4\pi}{M}$ which is reasonable provided N is large but for smaller N it is apparent that the periodogram form described in Chapter 4 is preferable. In addition we have assumed that m is sufficiently large for S_x^2 to be expressed as a sum of spectral estimates. We can avoid this assumption as follows .

Now we know that $\hat{R}(0)$ is equivalent to S_x^2 and in addition from Bartlett (1955)

$$\operatorname{cov}\left[\hat{R}_\tau \hat{R}_\nu\right] \sim \frac{1}{N}\left[\sum_{u=-\infty}^{\infty}\left(R_u R_{u+\tau-\nu} + R_{u+\tau}R_{u-\nu} + K R_\tau R_\nu\right)\right] \qquad 7.8$$

thus for $\{X_t\}$ white noise

$$\text{cov}[\hat{R}_0\ \hat{R}_\tau] \sim \frac{2}{N}R_0^2 \qquad \tau = 0$$

$$\sim 0 \qquad \tau \neq 0$$

Thus

$$\text{cov}[\hat{R}_0\ 2\pi\hat{P}(\omega)] \sim \frac{2R_0^2}{N}\cdot 2\pi$$

Now using a Taylor approximation we can find that

$$\text{cov}(\log\hat{R}_0, \log 2\pi\hat{P}(\omega)) \simeq \frac{2}{N}$$

thus assuming that $\hat{P}(\omega_p)$ have approximately the same distribution for all $\omega_p \in [0,\pi]$

$$\text{cov}[\log s_x^2, \hat{\Delta}_P] \simeq \frac{2}{N}$$

and in consequence

$$\text{var}[\log s_x^2 - \hat{\Delta}_P] \simeq \frac{\psi'(V_2)}{m} + \frac{\psi'(N-1)}{2} - \frac{4}{N}$$

which becomes for large samples

$$\text{var}(\log s_x^2 - \hat{\Delta}_P) \simeq \frac{\psi'(V_2)}{m} - \frac{2}{N}$$

which agrees with our earlier formulation.

The use of $\hat{\Delta}_P$ for tests of white noise is thus not altogether simple and it does seem that one is better using the techniques of Chapter 3. In view of our small sample approximation of the distribution of $\hat{\Delta}_I$ and the tests proposed in Chapter 6 we feel there is little point in pursuing the use of $\hat{\Delta}_P$ and in consequence for the remainder of this work we shall use only periodogram estimates.

We also would conclude that $\hat{\Delta}_t$ is of use only if one has very long realisations. However in these cases one would expect that the periodogram would be far easier and more economical to compute and in this case once again the estimate $\hat{\Delta}_T$ appears to have powerful advantages. Thus our intuitively appealing estimate $\hat{\Delta}_t$ appears to be of limited value if it is of any value at all.

CHAPTER 8

Having derived statistics involving estimates of the minimum mean-squared error of prediction for a one step predictor we now turn our attention to estimating $\log e_k$ the corresponding quantity for a k step predictor. As we have seen previously this latter quantity is intimately related to the first k parameters of the related moving average representation of a stationary time series and it follows we must consider the related problem of evaluating the parameters of a time series model.

The estimation of parameters in the time domain has been extensively studied by Durbin (1960), Whittle (1953), Box and Jenkins (1972) and Akaike (1968) as well as many other authors. We shall however consider the use of the spectrum in estimating parameters and the number of approaches from this viewpoint have been much fewer, the principle reference being Jones (1964).

The principal advantage in using the spectrum is that one need not specify a priori the order of the model and one avoids one considerable source of error. While our approach is necessarily limited we feel that the methods outlined below are worth pursuing.

Following Jones (1964) we consider a stationary non-deterministic process $\{Y_t\}$ with a power spectrum given by $f(\omega)$. In consequence we can assume that $\{Y_t\}$ has a one-sided moving average representation, c.f. Appendix 2 given by

$$Y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \qquad 8.1$$

where the series $\{\varepsilon_t\}$ is white noise and $\mathrm{var}(\varepsilon_t) = \sigma^2$ for all t

We shall also assume there is an autoregressive representation of possibly infinite order which we write as

$$\sum_{j=0}^{\infty} a_j X_{t-j} = \varepsilon_t \qquad 8.2$$

Now the prediction error for a k step predictor e_k is given by

$$e_k = \sigma^2 \sum_{j=0}^{k-1} b_j^2 \qquad 8.3$$

thus to look at the e_1 we need to determine the $b_0, b_1, \ldots$

If we denote

$$B(z) = \sum_{j=0}^{k} b_j z^j \qquad 8.4$$

and

$$A(z) = \sum_{j=0}^{k} a_j z^j \qquad 8.5$$

then from 8.1 and 8.2 we have

$$A(z) = B^{-1}(z) \qquad 8.6$$

or equivalently

$$A(z)\,B(z) = 1 \qquad 8.7$$

and from 8.6 and 8.7 we find that

$$\sum_r \sum_s a_r b_s z^{r+s} = 1 \qquad 8.8$$

Using 8.8 one can by equating coefficients of powers of z obtain a series of equations relating the a_j and b_j these are

$$\begin{aligned} a_0 b_0 &= 1 \\ a_1 b_0 + a_0 b_1 &= 0 \\ a_2 b_0 + a_1 b_1 + a_0 b_2 &= 0 \\ a_3 b_0 + a_2 b_1 + a_1 b_2 + a_0 b_3 &= 0 \\ &\ldots \end{aligned} \qquad 8.9$$

Thus if we can estimate either the $\{ b_j ; j=0 \cdots \}$ or the $\{ a_j ; j = \cdots \}$ then one can fairly easily determine the other set of coefficients. In view of this result we shall concentrate on estimating the coefficients of $B(z)$ as they are more intimately connected with the e_k in 8.3

Now from Kolmogorovs Theorem, (c.f Appendix 2) we have

$$f(\omega) = \frac{|B(e^{-i\omega})|^2 \sigma^2}{2\pi} \qquad 8.10$$

and since $f(\omega)$ is integrable we know it has a Fourier series representation of the form

$$\log f(\omega) = \sum_{r=-\infty}^{\infty} c_r e^{-i\omega r} \qquad 8.11$$

From 8.10 and 8.11 we thus conclude that

$$\frac{\sigma^2}{2\pi} |B(e^{-i\omega})|^2 = \exp \sum_{r=-\infty}^{\infty} c_r e^{-i\omega r} \qquad 8.12$$

which is our starting point for what follows. Writing 8.12 in the form

$$\frac{\sigma^2}{2\pi} B(e^{-i\omega}) B(e^{i\omega}) = \exp \Big\{ c_0 + \sum_{r=1}^{\infty} c_r e^{-i\omega r} + \sum_{r=1}^{\infty} c_r e^{i\omega r} \Big\} \qquad 8.13$$

since $f(\omega)$ and consequencely $\log f(\omega)$ is symmetric about $\omega = 0$ and noting that

$$\exp(c_0) = \frac{\sigma^2}{2\pi}$$

from 2.42 we have

$$B(z) = \exp \Big\{ \sum_{r=1}^{\infty} c_r z^r \Big\} \qquad 8.14$$

thus we have factorised the spectrum. Note for convenience we have assumed that $b_0 = 1$

that is

$$B(z) = 1 + \sum_{j=1}^{\infty} b_j$$

This involves no loss of generality since we can incorporate b_0 into the ε_t and write

$$\sigma^2 = b_0 \sigma'^2$$

where σ'^2 is the variance with $b_0 \neq 1$

Now from 8.11 we have the inverse transformation

$$c_\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) e^{i\omega\nu} d\omega \qquad \nu = 1, 2, \ldots \qquad 8.15$$

which becomes

$$c_\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) \cos\omega\nu \, d\omega \qquad 8.16$$

since $f(\omega)$ is symmetric. In consequence we can write 8.14 as

$$1 + b_1 z + b_2 z^2 + \cdots = \exp\{(c_1 z + c_2 z^2 + \cdots)/2\} \qquad 8.17$$

which is Kolmogorovs (1939) result since

$$\log f(\omega) = \frac{c_0}{2} + c_1 \cos\omega + c_2 \cos 2\omega + \cdots$$

and

$$c_k = \frac{1}{\pi} \int_{-\pi}^{\pi} \log f(\omega) \cos k\omega \, d\omega \qquad 8.18$$

Using 8.17 Jones (1964) evaluated the b_j and the corresponding moving average coefficients using estimates of c_k obtained from 8.18 evaluating the integral as a sum of spectral estimates. His technique, while well suited for computation and providing useful results, proves somewhat intractable when one wishes to investigate the underlying distribution theory. Since our main interest is in the distribution of the b_j's etc., we follow a slightly difference approach.

First we need to obtain the b_j explicitly in terms of the c_j. This can be done by writing 8.17 in the form

$$1 + b_1 z + b_2 z^2 + \cdots = e^{c_1 z/2} e^{c_2 z^2/2} \cdots \qquad 8.19$$

$$= \left(1 + \frac{\gamma_1 z}{1!} + \frac{\gamma_1^2 z^2}{2!}\right)\left(1 + \frac{\gamma_2 z^2}{1!} + \frac{\gamma_2 z^4}{4!} \cdots\right) \cdots \qquad 8.20$$

where $$\gamma_k = \frac{c_k}{2} \qquad k = 1, 2 \ldots \qquad 8.21$$

This may be written

$$1 + \sum_{j=1}^{\infty} b_j z^j = \prod_{j=1}^{\infty} \left\{ 1 + \sum_{k=1}^{\infty} \frac{(\gamma_j z^j)^k}{k!} \right\} \qquad 8.22$$

and equating coefficients we can see that

$$\begin{aligned} b_1 &= \gamma_1 \\ b_2 &= \gamma_2 + \frac{\gamma_1^2}{2!} \\ b_3 &= \gamma_1^3 + \gamma_2 \gamma_3 + \frac{\gamma_1^3}{3!} \end{aligned} \qquad 8.23$$

This is rather tedious and a simpler method can be used which makes use of the relations between raw moments and cumulants tabulated in Kendall and Stuart (1963). As is well known

$$1 + \frac{\mu_1' t}{1!} + \frac{\mu_2' t^2}{2!} + \ldots = \exp\left\{ \frac{\kappa_1 t}{1!} + \frac{\kappa_2 t}{2!} + \ldots \right\} \qquad 8.24$$

and if we write $\ell! \, b_\ell = \mu_\ell'$ and $\ell \kappa_\ell = \gamma_\ell$ we have 8.20

Thus using the relations

$$\begin{aligned} \mu_1' &= \kappa_1 \\ \mu_2' &= \kappa_2 + \kappa_1^2 \\ \mu_3' &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3 \\ \mu_4' &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4 \end{aligned} \qquad 8.25$$

we can obtain the following

$$\begin{aligned} b_1 &= \gamma_1 \\ b_2 &= \gamma_2 + \frac{\gamma_1^2}{2!} \\ b_3 &= \gamma_3 + \gamma_2\gamma_2 + \frac{\gamma_1^3}{3!} \\ &\ldots \end{aligned} \qquad 8.26$$

Note that Kendall and Stuart give the moments up to μ'_{10} in terms of the corresponding cumulants.

As an alternative one can also obtain expressions 8.26 by evaluating the integral

$$b_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left\{\sum_{r=1}^{\infty} c_j e^{-i\omega j}\right\} e^{-i\omega k}\, d\omega \qquad 8.27$$

using the calculus of residues, however the algebra is rather tedious.

In much the same manner one can evaluate the autoregressive coefficients. Using 8.6 and 8.11 we have

$$A(z) = B^{-1}(z) = \exp -\left\{(c_1 z + c_2 z^2 \ldots)/2\right\} \qquad 8.28$$

Thus to evaluate the a_i we use

$$1 + a_1 z + a_2 z^2 + \ldots = \exp -\left\{(c_1 z + c_2 z^2 \ldots)/2\right\} \qquad 8.29$$

and explicit relations between the c_i and the a_i can be obtained in the same manner as for the b_j and c_i. However if the b_j are already known then one can use 8.9

We now consider estimates of

$$\gamma_k = \frac{1}{\pi}\int_0^{\pi} \log f(\omega) \cos k\omega \, d\omega \qquad 8.30$$

Given $\{X_t : t = 1 \ldots N\}$ stationary to second order we can as before estimate $\log f(\omega)$ by the periodogram estimate $\log I_N(\omega_p, x)$ at $\omega = \frac{2\pi p}{N}$ $p = 0, 1 \ldots [\frac{N}{2}]$ and then approximate the integral 8.30 by a finite sum giving

$$\hat{\gamma}_k = \frac{1}{N}\sum_{p=1}^{n} \log I_N(\omega_p, x) \cos k\omega_p \qquad 8.31$$

where $n = \lfloor \frac{N}{2} \rfloor$ and the $I_N(\omega_p, x)$ are computed from $X_1 \ldots X_N$

Now the cumulants of $\log I_N(\omega_p, \varepsilon)$ are given by 3.42 as

$$\begin{aligned} K_1 &= \log 2 + \psi(1) \\ K_2 &= \frac{\psi'(1)}{n} \\ K_3 &= \frac{\psi''(1)}{n} \end{aligned} \qquad 8.32$$

Now making use of 3.10 we have

$$\begin{aligned} \hat{\gamma}_k &= \frac{1}{n} \sum_{p=1}^{n} \{ \log 2\pi f(\omega_p) + \log I_N(\omega_p, \varepsilon) \} \cos k\omega_p \\ &\simeq \frac{1}{n} \int_0^{n} [\log 2\pi f(\omega)] \cos k\omega \, d\omega \\ &\qquad + \frac{1}{n} \sum_{p=1}^{n} \log I_N(\omega_p, \varepsilon) \cos k\omega_p \\ &= \gamma_k + \Delta_k \qquad k \neq 0 \end{aligned} \qquad 8.33$$

in consequence the moments of $\hat{\gamma}_k$ are given by

$$K_1 = \frac{1}{n} \sum_{p=1}^{n} [\log 2 + \psi(1)] \cos k\omega_p + \gamma_k \qquad 8.34$$

$$K_2 = \frac{1}{n^2} \sum_{p=1}^{n} \psi'(1) \cos^2 k\omega_p$$

$$K_3 = \frac{1}{n^r} \sum_{p=1}^{n} \psi^{(r-1)}(1) \cos^r k\omega_p \qquad 8.35$$

To evaluate these values of the cumulants we evidently need to sum series of the form 8.35 Numerically this poses no problem since the $\psi^{(r)}(1)$ are tabulated by Abromowitz and Segun (1968) or can be evaluated using

$$\psi^{(r)}(1) = (-1)^{r+1} r! \, \zeta(r+1) \qquad r = 1, 2, 3 \ldots \qquad 8.36$$

where $\zeta(z)$ is the Zeta function of Reimann and cosines are readily available on most computers to considerable accuracy. For example on an

I.C.L. 1906A $\cos x$ has a maximum error of 5×10^{-11} using single precision arithmetic.

However it is quite possible to sum these series analytically since

$$\sum_{k=1}^{n} \cos^r kx$$

can be transformed into a geometric sum by writing in the form of complex exponentials. For r in the range 1 to 4 we have

$$\sum_{j=0}^{n} \cos jx = \cos\left(\frac{n+1}{2}\right)x \sin\frac{nx}{2} \operatorname{cosec}\frac{x}{2} + 1 \qquad 8.37$$

$$\sum_{j=1}^{n} \cos^2 jx = \frac{n}{2} + \frac{\cos(n+1)x \sin nx}{2\sin x} \qquad 8.38$$

$$\sum_{j=1}^{n} \cos^3 jx = \frac{3}{4}\cos\frac{(n+1)x}{2} \sin\frac{nx}{2} \operatorname{cosec}\frac{x}{2} + \frac{1}{4}\cos\frac{3(n+1)x}{2} \sin\frac{3nx}{2} \operatorname{cosec}\frac{3x}{2} \qquad 8.39$$

$$\sum_{j=1}^{n} \cos^4 jx = \frac{1}{8}\left[3n + 4\cos(n+1)x \sin nx \operatorname{cosec} x + \cos 2(n+1)x \sin 2nx \operatorname{cosec} 2x\right] \qquad 8.40$$

Now writing $x = \frac{\pi v}{n}$ we have two situations

(a) v odd

$$\sum_{r=0}^{n} \cos j\frac{\pi v}{n} = 0 \qquad 8.41$$

$$\sum_{r=1}^{n} \cos j\frac{\pi v}{n} = \frac{n}{2} \qquad 8.42$$

$$\sum_{r=1}^{n} \cos^3 j\frac{\pi v}{n} = -\frac{1}{2} \qquad 8.43$$

$$\sum_{r=1}^{n} \cos 4\, j\frac{\pi v}{n} = \frac{3n}{8} \qquad 8.44$$

while

(b) v even

$$\sum_{r=0}^{n} \cos j\frac{\pi v}{n} = 1 \qquad 8.45$$

$$\sum_{r=1}^{n} \cos^2 j\frac{\pi v}{n} = \frac{n}{2} \qquad 8.46$$

$$\sum_{r=1}^{n} \cos^3 j\frac{\pi v}{n} = 0 \qquad 8.47$$

$$\sum_{r=1}^{n} \cos 4\, j\frac{\pi v}{n} = \frac{3n}{8} \qquad 8.48$$

Thus from (a)

$$\sum_{r=1}^{n-1} \cos \frac{2\pi j r}{n} = 0 \qquad 8.49$$

$$\sum_{r=1}^{n-1} \cos^2 \frac{2\pi j r}{n} = \frac{n}{2} - 1 \qquad 8.50$$

$$\sum_{r=1}^{n-1} \cos^3 \frac{j\pi r}{n} = \frac{1}{2} \qquad 8.51$$

$$\sum_{r=1}^{n-1} \cos 4 \frac{j\pi r}{n} = \frac{3n}{8} - 1 \qquad 8.52$$

while for (b)

$$\sum_{r=1}^{n-1} \cos \frac{j\pi r}{n} = -1 \qquad 8.53$$

$$\sum_{r=1}^{n-1} \cos^2 \frac{j\pi r}{n} = \frac{n}{2} - 1 \qquad 8.54$$

$$\sum_{r=1}^{n-1} \cos^3 \frac{j\pi r}{n} = 1 \qquad 8.55$$

$$\sum_{r=1}^{n-1} \cos 4 \frac{j\pi r}{n} = \frac{3n}{8} - 1 \qquad 8.56$$

and if we redefine

$$\overset{1}{\gamma}_j = \frac{1}{n} \sum_{p=1}^{n-1} \log I_N(\omega_p, x) \cos k\omega_p \qquad 8.57$$

then $\overset{1}{\gamma}_j$ has cumulants given by

$$K_1 = \begin{cases} \gamma_j & j \text{ even} \\ \gamma_j + [\log 2 + \psi(1)]/n & j \text{ odd} \end{cases} \qquad 8.58$$

$$K_2 = \frac{\psi'(1)}{2n} - \frac{\psi'(1)}{n^2} \qquad 8.59$$

$$K_3 = \begin{cases} \frac{\psi''(1)}{n^3} & j \text{ even} \\ \frac{\psi''(1)}{2n^3} & j \text{ odd} \end{cases} \qquad 8.60$$

$$K_4 = \frac{3\psi'''(1)}{8n^3} - \frac{\psi'''(1)}{n^4} \qquad 8.61$$

If we consider a full estimator

$$\hat{\gamma}_j = \frac{1}{n} \sum_{p=0}^{n-1} \log I_N(\omega, x) \cos^j \omega_p \qquad 8.62$$

then the algebra is a little more complex however we find that the cumulants are given by

$$K_1 = \begin{cases} \gamma_j - [2\log 2 + \psi(1) + \psi(0.5)]/n & j \text{ even} \\ \gamma_j & j \text{ odd} \end{cases} \qquad 8.63$$

$$K_2 = \frac{\psi'(1)}{2n} - \frac{\psi'(1)}{n^2} + \frac{\psi''(0.5)}{2n^2} \qquad 8.64$$

$$K_3 = \begin{cases} \frac{\psi''(1)}{n^3} + \frac{\psi''(0.5)}{4n^3} & j \text{ even} \\ \frac{\psi''(1)}{2n^3} \end{cases} \qquad 8.65$$

$$K_4 = \frac{3\psi'''(1)}{8n^3} - \frac{\psi'''(1)}{n^4} + \frac{\psi'''(0.5)}{8n^4} \qquad 8.66$$

While the expressions for the cumulant above are exceedingly messy they are easy to compute and for large n can be somewhat simplified. If we let $n \to \infty$ then we have

$$\beta_1 \to 0 \qquad \beta_2 \to 3$$

which would indicate that for large sample sizes the $\hat{\gamma}_j$ are approximately normally distributed. This is as we would expect since for large n $\log I_N(\omega_p, x)$ is approximately normal and 8.62 is thus a linear combination of independent normal variates.

Our expressions for the cumulants also show that both forms of the estimate of γ_j are asymptotically unbiased and consistent. If we care to redefine the $\hat{\gamma}_j$ for even j as

$$\hat{\gamma}_j = \frac{1}{n} \sum{}' \log I_N(\omega, x) \cos j\omega_p - \left[2 \log 2 + \gamma(1) + \gamma(0.5) \right] \frac{1}{n} \qquad 8.67$$

then all our estimates of the γ_j are unbiased since

$$E[\hat{\gamma}_j] = \gamma_j \qquad 8.68$$

Since we have a sequence of estimates $\hat{\gamma}_j \quad j = 1, 2 \ldots$ another point of interest is the correlation between our estimates.

Consider for simplicity

$$\begin{aligned} \operatorname{cov}[\hat{\gamma}_k \hat{\gamma}_j] &= E[(\hat{\gamma}_k - \gamma_k)(\hat{\gamma}_j - \gamma_j)] \\ &= \frac{1}{n^2} \sum_p \sum_q \log I_N(\omega_p, \epsilon) \cos \omega_p k \log I_N(\omega_q, \epsilon) \cos \omega_q j \\ &= \frac{\sigma_I^2}{n^2} \sum_{p=1}^{n-1} \cos k\omega_p \cos j\omega_p \end{aligned} \qquad 8.69$$

where

$$\sigma_I^2 = E\left[\log I_N(\omega_p,\varepsilon)^2\right] \qquad 8.70$$

Then we can write 8.69 as

$$\operatorname{cov}\left[\hat{\gamma}_k\ \hat{\gamma}_j\right] = \frac{\sigma_I^2}{2n^2}\sum_{p=1}^{n-1}\left\{\cos\omega_p(j+k)+\cos\omega_p(j-k)\right\}$$

$$= -\frac{\sigma_I^2}{2n^2} \qquad j+k \text{ even}$$

$$= 0 \qquad j+k \text{ odd}$$

Hence for large values of n

$$\operatorname{cov}\left[\hat{\gamma}_k\ \hat{\gamma}_j\right] = O\left(\frac{1}{n}\right) \qquad j+k \text{ even}$$

$$0 \qquad j+k \text{ odd}$$

and we see that our estimates are asymptotically uncorrelated. Since they are also asymptotically normal we can consider them to behave as independent normal variates for sufficiently large values of n.

The result also holds for estimates of the form 8.62 but the algebra is rather more complex.

To indicate that these results appear to agree well with practice we devised a small simulation experiment. We generated 400 series of the form

$$X_t = \varepsilon_t - 0.5\,\varepsilon_{t-1} \qquad t = 1 \ldots 64 \qquad 8.71$$

where the ε_t were standard normal deviates.

Then comparing the theoretical results for the mean of γ_j with the sample values

TABLE 36

	Mean	
	Theoretical	Sample
γ_1	-0.5000	-0.4903
γ_2	0.0000	-0.0178
γ_3	0.0000	-0.0340

with the sample covariance matrix is

γ_0	0.0527			
γ_1	0.0026	0.0230		
γ_2	-0.0002	0.0018	0.0245	
γ_3	0.0039	-0.0030	0.0030	0.0264

The diagonal terms agree well with the predicted values of the variance which are

$$\mathrm{var}(\hat{\gamma}_0) = 0.0526$$

$$\mathrm{var}(\hat{\gamma}_k) = 0.0257$$

while the covariances are predicted to be

$$\mathrm{cov}[\hat{\gamma}_k \hat{\gamma}_h] = 0 \qquad k+h \text{ odd}$$

and

$$\mathrm{cov}[\hat{\gamma}_k \hat{\gamma}_h] = -0.0008 \qquad h+k \text{ even}$$

Thus it would seem that we can assume our theoretical values for the mean and variance of $\hat{\gamma}_k$ $k = 0, 1 \ldots$ agree well with practice. The correlation matrix is, giving only the lower triangle

1.0000			
0.0747	1.0000		
-0.0056	0.0050	1.0000	
0.1046	-0.0804	0.0804	1.000

For the same model except with a changed parameter

$$X_t = \varepsilon_t - 0.2\varepsilon_t$$

we have the following results

TABLE 37

	Mean	
	Theoretical	Sample
γ_1	-0.2000	-0.2319
γ_2	-0.1250	-0.10969
γ_3	0.0000	-0.02627

while the covariance matrix was given by

0.05278			
0.00651	0.02161		
-0.00083	0.00247	0.25413	
-0.00146	-0.00181	0.00254	0.20639

giving a correlation matrix

1.0000			
0.0600	1.0000		
-0.0226	0.0676	1.0000	
-0.0442	-0.0548	0.0770	1.0000

once again the structure appears to agree well with theoretical ideas.

In addition we have compared the sample values of $\hat{\gamma}_1$ with the theoretical frequences from the best approximating Pearson curve. The results, as can be seen from table 39 are rather good both for the first

model described above and for two further models based on longer realisations.

As can be seen from the tables in Appendix 3 the Pearson curve rapidly approaches the Normal curve and it would appear that a normal approximation is adequate in almost all circumstances. The tables of $\sqrt{\beta_1}$ and β_2 below reveal the swiftness of the approach to the asymptotic case

TABLE 38

Sample length	$\sqrt{\beta_1}$	β_2
32	0.00211	3.00010
64	0.00046	3.00000
128	0.00010	3.00000
256	0.00002	3.00000

TABLE 39

Pearson curve approximations to the model

$$X_t = \varepsilon_t - 0.5\,\varepsilon_t \qquad t = 1 \ldots N$$

These tables give the percentage of values of $\hat{\Delta}_t$ not exceeding the percentage point of the best fitting Pearson curve

N	64	128	256
NO REPLICATIONS	400	300	500
Theoretical Percentage			
0.25	0.0	0	0.2
0.5	0.25	0	0.4
1.0	0.5	0.3	0.8
2.5	1.0	1.7	1.6
5.0	3.75	2.7	3.4
10.0	7.0	9.7	10.6
25.0	23.0	23.7	24.0
50.0	52.5	50.0	50.8
75.0	78.0	78.0	75.0
90.0	90.75	90.7	90.0
95.0	95.5	96.0	95.8
97.5	98.0	97.1	98.0
99.0	99.0	99.3	99.2
99.5	99.75	100	100
99.75	100	100	100

We also considered two further simulations with larger series lengths, the model being as before

$$X_t = \varepsilon_t - 0.5\varepsilon_{t-1} \qquad t = 1 \cdots N$$

The results are as follows

TABLE 40

	Mean Theoretical	Sample	
		N = 128	N = 256
γ_0	0.5000	-0.4982	-0.4998
γ_1	0.1250	-0.0930	-0.1044
γ_2	0.0000	-0.0592	-0.0642

while the covariance matrices

N = 128

0.02342
0.00138 0.01451
-0.00005 0.00050 0.01217
0.00110 -0.00127 0.00242 0.01253

N = 256

0.01135
0.00046 0.006042
0.00022 0.00041 0.00564
0.00029 -0.00002 0.00033 0.00411

While the estimation of the γ_j is of interest our main purpose is the evaluation of the b_j's in the moving average model. Since as we have seen the $\hat{\gamma}_j$ are asymptotically normal the distributional theory is eased however there are problems in estimating high order b_j's because of non-linear relationships connecting them with the γ_k The large sample case where the $\hat{\gamma}_k$ are approximately independent may well be a reasonable one to handle but it would appear that the estimation of b_j for $j \geq 2$ is very complex.

We shall restrict ourselves to the estimation of b_1 or if we prefer to work with the autoregressive model a_1 since the distribution can be found for our estimate. In addition knowledge of e_1 enables us to evaluate

$$e_2 = e_1(1 + b_1^2)$$

or correspondingly

$$\log e_2$$

Box and Jenkins (1971) have shown that first order moving average models are of some interest in constructing prediction models and in addition one can estimate the ratio

$$\frac{e_1}{e_2}$$

to determine the increase in variance in predicting two steps into the future

Since $b_1 = \gamma_1$ we see that the estimation of b_1 is straightforward. However for b_2 we have

$$b_2 = \gamma_2 + \gamma_{1/2}^2$$

and similarly

$$b_3 = \gamma_3 + \gamma_1\gamma_2 + \gamma_1^3/3!$$

from 8.26. Thus to consider the moments of even b_2 we need to evaluate expressions like

$$P = E\left[(\hat{\gamma}_r - \gamma_r)(\hat{\gamma}_s - \gamma_s)(\hat{\gamma}_t - \gamma_t)\right]$$

$$= \frac{1}{n^3}\sum_{p,q,m} \log I_N(\omega_p,\epsilon)\log I_N(\omega_q,\epsilon)\log I_N(\omega_m,\epsilon) \cdot \cos r\omega_p \cos s\omega_q \cos t\omega_m$$

$$= M_3 \sum_p \cos r\omega_p \cos s\omega_q \cos t\omega_p$$

where $$M = E\left[\log I_N(\omega_p,\epsilon)^3\right]$$

and after some algebra we can show that

$$P = -\frac{M}{4n^3}\,\Sigma\!\!\!\Sigma(r,s,t)$$

where $\Sigma\!\!\!\Sigma(r,st)$ =

- 3 r, s, t all odd
- 3 if one of r, s, t even
- 0 otherwise

In consequence for n large $P \Rightarrow 0$ This would appear to indicate that one could analyse the correlation structure of $\hat{b}_2$ but the extension to $\hat{b}_j > 2$ appears to be quite formidable.

The approximation to the small sample distribution of $\hat{b}_2$ is equally complex but if we assume that $\hat{b}_1$ and $\hat{b}_2$ are independent then one can at least approximate the cumulants of $\hat{b}_2$

What might be slightly more feasible is a procedure to determine whether the $\hat{b}_j$ are zero by successively testing the corresponding $\hat{\gamma}_j$.

For a moving average model of finite length one could repeatedly take differences until $b_1 = 0$. A test for white noise as suggested in Chapter 6 could be made to confirm if the residuals after differencing were indeed white noise.

As we have seen one can at least estimate the leading parameter in a time series model from the spectrum. While the estimation of other parameters appears to be rather complex it would appear that their estimation from the coefficients of the factorised spectrum is, in principle at least, feasible and this approach does seem worth pursuing.

For prediction two steps into the future we can use our estimate of b_1 to evaluate

$$1 + \hat{b}_1^2$$

since the distribution of $\hat{b}_1$ is known and for larger sample sizes we can show that the estimates $\hat{\Delta}_I$ and $\hat{b}_1$ are approximately independent thus enabling one to compare e_1 and e_2 This is probably more easily done by examining

$$\frac{e_2}{e_1} = 1 + b_1^2$$

CHAPTER 9

As we have seen we can estimate the minimum mean square error of prediction e_1 or rather log e_1 and we can approximate the distribution of our estimate very closely. While for our purposes the logarithmic form has been very convenient one can visualise cases in which the distribution of $\hat{e}_1$ itself would be required.

A possible method of approximating the distribution of $\hat{e}_1$ is to obtain its moments and use the Pearson approximation. This may be done as follows:

Consider $z = \log x$ and let the characteristic function of z be $\phi_z(t)$

Then

$$\phi_z(t) = E[e^{it\log x}] = E[x^{it}] \qquad 9.1$$

and the moments of z are related to those of x by the following relation

$$\mu_r'(x) = \phi_z(-ir) \qquad 9.2$$

provided the characteristic function exists at this point. Since the characteristic function in question is essentially a Gamma function with argument

$$\frac{r}{m} + 1$$

it clearly exists and the moments of $\hat{e}_1$ can be found using 9.2. However as we have stated the logarithmic form has proved convenient.

Since we can estimate $\log e_1$ then we can make decisions as to whether it is in fact worth trying to predict a series for a given

cost function. However we see the real use of our estimate of prediction error as a diagnostic procedure when model fitting in addition to the test of white noise outlined in Chapter 6.

As can be seen from Box and Pierce (1970) and Durbin (1970 b) making tests on the residuals of a fitted model can be a troublesome procedure and in addition the distribution theory derived to date is asymptotic. We feel that a very good criteria for a prediction model is its mean square error of prediction compared with the optimium and we suggest the following procedure.

Given $X_1, \ldots X_{2N}$ we estimate $\log e_1$ using $\hat{\Delta}_I$ given by 3.54 using $X_1 \ldots X_N$ Having obtained an estimate of $\log e_1$ we now proceed to fit a prediction model on the same length of series $X_1, \ldots X_N$. Once we have constructed our model we can make forecasts X^*_{N+1} X^*_{N+2} .. X^*_{N+3} .. X^*_{2N} and obtain the forecast errors

$$d_r = X^*_{N+r} - X_{N-r} \qquad 9.3$$

Using the residual series $\{d_t\}$ we can evaluate the minimum mean square error of prediction for the residuals, say $\log e_{1,d}$ using $\hat{\Delta}_{I,d}$ say.

Then assuming the model used for prediction is the correct model

$$E\left[\hat{\Delta}_I - \hat{\Delta}_{I,d}\right] = 0$$

while

$$U = \hat{\Delta}_I - \hat{\Delta}_{I,d} > 0$$

implies that our model overfits while

$$U = \hat{\Delta}_I - \hat{\Delta}_{I,d} < 0$$

implies the model underfits.

Under the null hypothesis we expect $\hat{\Delta}_{I,1}$ $\hat{\Delta}_{I,d}$ to be uncorrelated and to be indeed approximately independent. We can easily approximate the distribution of U under this assumption using our Pearson approximation.

This procedure would appear to avoid the difficulties described by Box and Pierce (1970) but needs some use in practice to see how well it performs.

A slightly different procedure can be obtained by estimating the sample variance of the $\{d_t\}$ series and comparing this with the value $\hat{\Delta}_I$ obtained from the first segment of the series $X_1 \ldots X_N$. For moderatly large sample sizes we know that the sample variance is approximately Normally distributed, Anderson (1971) and we can then base a test on

$$V = \Delta_I - \log s_d^2$$

Ideally one would wish for a small sample test but in this case one need the exact distribution of s_d^2. It seems fairly reasonable to assume that the distribution of s_d^2 is approximated by a χ^2_{N-k} distribution where k is the number of parameters fitted to the model. We have however been unable to provide an adequate justification of this assertion. This does seem worth of further study since our test for white noise is immediately transferable to this situation. For values of k which are small compared to N the moments of $\log s_d^2$ given by the polygamma functions $\psi'\left(\frac{N+k}{2}\right)$ appear to be insensative to small variations in k since

$$\psi\left(\frac{N-k}{2}\right) \simeq \log\left(\frac{N}{2}\right) - \frac{k}{N} - \frac{1}{N}$$

$$\psi'\left(\frac{N-k}{2}\right) \simeq \frac{2}{N-k} \simeq \frac{2}{N} + O\left(\frac{1}{N^2}\right)$$

and it would seem that we could use our test of white noise without much chance of error.

One can also use the minimum mean square error as a measure of the predictability of different series. This might be of some use in considering transformations of series before deciding to construct a model for prediction. This in a sense is the comparison of non-linear prediction methods when one uses non-linear transformations.

In practice one often has to remove deterministic components from time series by regression or the successive differencing methods of Box and Jenkins (1972). In particular one often removes a mean from a given series. While Grenander and Rosenblatt (1957) and Granger (1964) indicate that regression methods can be applied and the residuals from a regression model can be regarded as a stationary time series there is one rather interesting effect on the periodogram when one removes a mean. In this case

$$I_N(\omega,x) = \frac{2}{N}\left|\sum_{t=1}^{N}(X_t-\bar{X})e^{i\omega t}\right|^2 \qquad 9.4$$

and at $\omega = 0$

$$I_N(0,x) = \frac{2}{N}\left|\sum_{t=1}^{N}(X_t-\bar{X})\right|^2 \qquad 9.5$$
$$= 0$$

This clearly has implications in the construction of $\hat{\Delta}_I$ and in this situation we advise the use of a form of estimate with end points deleted viz

$$\hat{\Delta}_I = \frac{1}{n}\sum_{t=1}^{n-1}\log I_N(\omega,x) + \gamma - \log 2 \qquad 9.6$$

where γ is Eulers constant given by

$$\gamma = \lim_{m \to \infty} \left[1 + \frac{1}{2} + \frac{1}{3} \cdots + \frac{1}{m} - \log m \right] \qquad 9.7$$

$$= 0 \cdot 57721\ 56649 \ldots$$

This property of the periodogram is presumably the cause of some dips in the estimated power spectrum around zero frequency.

In addition to using just the estimates of $\log e_1$ we have also derived methods of estimating the coefficients of the moving average representation of a stationary time series. For a model of the form

$$X_t = \varepsilon_t + \sum_{k=1}^{\infty} b_k \varepsilon_{t-k}$$

we have shown that the spectral factorisation enables us to obtain expressions for the b_k as the sums of independent normal variates. While the estimation of b_1 is fairly well defined in Chapter 8 we think that more work could be put into the estimation of the b_k for $k > 1$ Indeed we would like to obtain some results for the simultaneous estimation of the b_k this would appear to be the crux of the estimation problem.

A natural extension of our investigations would be to the multivariate time series situation. As we shall see the extensions are not immediate and there are as one might expect some additional complexities.

Let us define a q variate process $\{ X_{1t}\ X_{2t} \cdots X_{qt} \}\ t = 0, 1 \ldots$ which we shall write as a $q \times 1$ vector $\underline{X}_t$ $t = \ldots 0, 1, \ldots$ and we shall assume

$$E[\underline{X}_t] = 0$$

Let

$$R_{jk}(s) = E[X_{j,t+s} X_{k,t}] \qquad 9.8$$

$$f_{jk}(\omega) = \sum_{s=-\infty}^{\infty} R_{jk}(s) e^{i\omega s} \qquad 9.9$$

and $$\underline{F}(\omega) = (f_{jk}(\omega)) \qquad 9.10$$

Zasuhin (1941) has generalised Wold's representation to the multivariate case giving

$$X_{j,t} = \sum_{m=0}^{\infty} b_{j,km} \varepsilon_{t,t-m} \qquad j = 1,2 \ldots q$$

where the ε_{lm} are uncorrelated and have zero mean.

Whittle (1953) has shown that corresponding to Kolmogorov's result for the mean-square error of prediction

$$V = \exp\left[\frac{1}{q\pi} \int_0^{\pi} \log \det[\underline{f}(\omega)]\, d\omega\right]$$

where V is the "total prediction variance" corresponding to the total variance defined by Wilks (1932). For a matrix moving average representation

$$\underline{X}_t = \underline{B}_0 \underline{\varepsilon}_t + \underline{B}_1 \underline{\varepsilon}_{t-1} \cdots$$

it can be shown that

$$V = \det |\underline{B}_0 \underline{B}'|$$

In addition one can also perform a spectral factorisation in the manner of Chapter 8, this has been described in considerable detail by Wiener and Masani (1957, 58).

One would expect the estimation of V to be much more complex than the corresponding estimates for the univariate case but it would

appear that estimates of V could be of considerable use, see Whittle (1953 b) and Whittle (1953). In addition we feel that estimates of $\log e_i$ for each of the q series could prove of use in constructing multivariate models.

The Wiener and Masani approach also appears very fruitful from a mathematical viewpoint and while these are considerable practical problems these are presumably not insurmountable.

In short we feel that much remains to be done using the estimated spectrum for model identification especially in small sample situations.

APPENDIX I

Consider a stochastic process $\{X_t(\omega)\}$ then in the space of functions on Ω which are $\mathcal{F}$ measurable and quadratically integrable with respect to P we introduce the scalar product

$$(f,g) = E f\bar{g} = \int_{\Omega} \bar{g}(\omega) f(\omega)\, dP(\omega)$$

The norm being $\|f\| = (f,f)^{1/2}$. A sequence of functions converges in the mean to $f(\omega) \in L_2(\Omega)$ if $\|f_n - f\| \to 0$ as $n \to \infty$

Defining the set of all finite linear combinations

$$L(x) = \left\{ \sum_{\nu=1}^{n} c_\nu X_{t_\nu};\ t_\nu \in T,\ c_\nu \in \mathbb{C} \right\}$$

we find that closing this linear manifold with respect to convergence in the mean we obtain a space $L_2(x)$ which has all the properties of a Hilbert space.

We shall now consider only discrete processes which are stationary. Let us assume we have observed $\{X_t, t \le 0\}$ then the observed variables span a subspace $A \subset L_2(x)$ and as our predictor X_h^* of X_h we choose an element in A. In consequence we use only linear predictors and suppose the entire part of the process is known.

To construct our forecast we shall seek to minimise

$$E\left[|X_h^* - X_h|^2 \right] = \|X_h^* - X_h\|^2$$

Such a predictor clearly exists and is unique.

Consider the set of complex valued functions $g(\omega)$ belonging to

$$L_2[dF(\omega); -\pi, \pi]$$

This becomes a Hilbert space $L_2(F)$ if we define the scalar product

$$(g,h) = \int_{-\pi}^{\pi} g(\omega)\,\overline{h(\omega)}\,dF(\omega)$$

It we now define the transformation S

$$S\,e^{it\omega} = X_t$$

and extending it so that

$$S\sum_{\nu=1}^{n} c_\nu e^{it_\nu\omega} = \sum_{\nu=1}^{n} c_\nu X_{t_\nu}$$

we can show that

$$\left\| \sum_{\nu=1}^{n} c_\nu X_{t_\nu} \right\|^2 = \left\| \sum_{\nu=1}^{n} c_\nu e^{it_\nu\omega} \right\|^2$$

Extending S using convergence in the mean we can easily show that S is a (1-1) isometric mapping of $L_2(F)$ onto $L_2(X)$

Hence the prediction problem becomes:
In the subspace of $L_2(F)$ spanned by $\{1, e^{-i\omega}, e^{-2i\omega} \ldots\}$ we seek $\Phi(\omega)$ which minimises

$$\|\Phi(\omega) - e^{i\omega}\|^2 = \int_{-\pi}^{\pi} |\Phi(\omega) - e^{i\omega}|^2\,dF(\omega)$$

and if this function exists the best predictor is given by

$$X_1^* = \int_{-\pi}^{\pi} \Phi(\omega)\,dZ(\omega) \qquad \text{(c.f. 1.11)}$$

Consider

$$e_1 = \lim_{n\to\infty} \|X_1 - \ell(X_0, X_{-1} \ldots X_{-n+1})\|^2$$

where $f(X_i X_{i+1} \ldots X_{i+j})$ denotes an arbitary linear combination of the $X_i, X_{i+1} \ldots X_{i+j}$

and

$$\| X_1 - \sum_{v=0}^{h-1} c_v X_{-v} \| = \int_{-n}^{n} |p_n(e^{i\omega})|^2 dF(\omega)$$

where $p_n(z)$ denotes the polynomial

$$p_n(z) = z - \sum_{v=0}^{n-1} c_v z^{-v}$$

Then we need to minimise the integral

$$\int_{-n}^{n} |1 - \bar{c}_0 e^{i\omega} - \bar{c}_1 e^{2i\omega} \ldots \bar{c}_{n-1} e^{-in\omega}|^2 dF(\omega)$$

The minimum of this integral may be found using the theory of Toeplitz forms c.f. Granander and Szego (1958 pp 44) and may be shown to be given by

$$\exp\left\{ \frac{1}{2\pi} \int_{-n}^{n} \log 2\pi f(\omega)\, d\omega \right\}$$

if it exists. The minimum can also be shown to be zero if and only if

$$\int_{-n}^{n} \log 2\pi f(\omega)\, d\omega = -\infty$$

To construct the best predictor explicitly we require the following theorem:

A necessary and sufficient condition for a stationary process with an absolutly continuous spectrum to be representable as a one-sided moving average is that it be non-deterministic.

Consider a sequence $\{ \xi_t : -\infty < t < \infty \}$ such that

$$E[\xi_s \bar{\xi}_t] = \delta_{st}$$

and

$$E[\xi_s] = 0 \qquad \forall s$$

and we construct

$$X_t = \sum_{v=0}^{\infty} a_j \xi_{t-j}$$

Assuming convergence in the mean then the covariance function is

$$R_{s-t} = \sum_{\nu=0}^{\infty} \bar{a}_\nu a_{\nu+s-t} \qquad 1.$$

or

$$\underline{R} = \bar{\underline{A}}\,\underline{A}$$

where $\underline{A}$ is the matrix given by

$$\underline{A} = \{ a_{\nu-\mu} \; ; \; \nu-\mu \geq 0, 0, \nu-\mu < 0 \}$$

From 1. it follows that

$$f(\omega) = |a(e^{-i\omega})|^2$$

almost everywhere, the polynomial $a(z)$ being defined as

$$a(z) = \sum_{p=0}^{\infty} a_p z^p$$

Since $a(z)$ belongs to the class H_2 we have using the Jacobi-Jensen inequality

$$\int_{-\infty}^{\infty} \log f(\omega)\, d\omega > -\infty$$

We now show that it is sufficient that the process be non-deterministic.

Since $\log f(\omega)$ integrable we have $f(\omega) > 0$ almost everywhere and we define a new process by

$$Z_\xi(\omega) = \int_{-\pi}^{\omega} [\phi(\omega)]^{-1} \, dZ(\omega)$$

where $|\phi(\omega)|^2 = f(\omega)$. Now this integral exists and defines an orthogonal process with

$$\| Z_\xi(\omega) \|^2 = \frac{\lambda + \pi}{2\pi}$$

Now defining the stationary process

$$\xi_t = \int_{-\pi}^{\pi} e^{it\omega} \, dZ_\xi(\omega) \qquad 2.$$

we can show the $\{\xi_t\}$ are orthogonal and normalised.

Under the assumptions above there exists a function $G(\omega) \in L_2$ such that

$$G(\omega) = \frac{1}{2\pi} \sum_{p=0}^{\infty} G_p e^{-ip\omega} \qquad 3.$$

$$f(\omega) = |G(\omega)|^2$$

Choosing $\phi(\omega)$ as $G(\omega)$ we have

$$X_t = \int_{-\pi}^{\pi} e^{it\omega} G(\omega)\, d Z_\xi(\omega)$$

and since

$$G(\omega) = \frac{1}{2\pi} \sum_{p=0}^{n} G_p e^{i\omega p} + \rho_n(\omega)$$

with $\|\rho_n\| \to 0$ we have from 2

$$X_t = \frac{1}{2\pi} \sum_{p=0}^{n} G_p \xi_{t-p} + Z_t$$

To construct the optimum predictor we must minimise

$$\| X_1 - \sum_{r=0}^{n-1} c_r X_{-r} \|^2 = \log e_1$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum u_j e^{i\omega(j-1)} \right|^2 f(\omega)\, d\omega$$

where $u_0 = 1 \qquad u_j = -\bar{c}_{j-1} \qquad 1 \le j < n$

and we predict X_1 using $X_0, X_{-1} \ldots X_{-n+1}$

It can be shown that the minimum μ_n is given by

$$\mu_n^{-1} = S_n(0,0) = \sum_{r=0}^{n} |\phi_r(0)|^2$$

where $\{\phi_p(z)\}$ are the polynomials orthogonal on the unit circle $|z| = 1$ with respect to the weight function $f(\omega)$. The polynomials $S_n(x,y) = \sum_{p=0}^{n} \overline{\phi_p(x)}\, \phi_p(y)$ are the kernal polynomials of $f(\omega)$ c.f. Grenandzer & Szego (1958).

The minimum itself occurs when

$$u_n(z) = \sum_{p=0}^{n} u_p z^p = \mu_n \sum_{p=0}^{n} \overline{\phi_p(0)}\, \bar{\phi}_p(z)$$

and in this case the predictor is

$$X_1^* = \sum_{r=0}^{n-1} c_r X_{-r} = -\int_{-\pi}^{\pi} e^{i\omega} \left[\overline{u_n(e^{i\omega})} - 1 \right] d Z(\omega)\,.$$

Now for $n < m$ we can show that $u_n(z)$ converges, for

$$\int_{-n}^{n} | S_n(0z) - S_m(0z) | f(\omega) d\omega = \sum_{\nu=n+1}^{m} | \phi_\nu(0) |^2 = \mu_m^{-1} - \mu_n^{-1} \to 0$$

where $z = e^{i\omega}$ and $n, m \to \infty$

In addition

$$f(\omega) = \lim_{r \to 1} | g(r e^{i\omega}) |^2 \qquad 4.$$

almost everywhere when $g(z)$ has no zeros inside the unit circle and $g(0)$ is real and positive where $g(z)$ belongs to the class H_2

If we write

$$t_n(z) = g(z) S_n(0,z)$$

equation 4 implies that there is an element $H(z)$ in H_2 such that

$$\| t_n - H \|^2 = \int_{-n}^{n} | t_n(z) - H(z) |^2 d\omega \to 0$$

as $n \to \infty$

Thence for any $\rho \in]0,1[$ the function $t_n(z)$ converges uniformly in $|z| \leq \rho$ to $H(z)$ so that

$$\lim_{n \to \infty} S_n(0z) = \frac{H(z)}{g(z)}$$

We can, using the asymptotic formula for $\phi_p(z)$ show that

$$H(z) = [g(0)]^{-1}$$

see Grenander and Szego (1958) and in consequence

$$\lim_{n \to \infty} \int_{-n}^{n} | S_n(0,z) - [g(0) g(z)]^{-1} |^2 f(\omega) d\omega = 0$$

and since $\mu_n \to |g(0)|^2$

$$\lim_{n \to \infty} \int_{-n}^{n} | u_n(z) - \frac{g(0)}{g(z)} |^2 f(\omega) d\omega = 0$$

Making use of the isometric correspondence between the Hilbert spaces spanned by the process and by the exponential functions respectively we have for the best predictor

$$X_1^* = \underset{n\to\infty}{l.i.m}\,(X_1^*)_n = \int_{-n}^{n} e^{i\omega}\left[1 - \frac{g(0)}{g(e^{i\omega})}\right] dZ(\omega)$$

Notice that the function $g(z)$ introduced in 4 corresponds to the function $G(\omega)$ introduced in 3 in that

$$g(e^{i\omega}) = \overline{G(\omega)}$$

and making use of the orthogonal process $dZ_\xi(\omega)$ we can show that

$$X_1 = \frac{1}{2\pi} \sum_{\nu=0}^{\infty} G_\nu \xi_{1-\nu}$$

giving

$$|X_1^* - X_1|^2 = \frac{1}{4\pi^2}|G_0|^2 = \frac{1}{4\pi^2}\left|\int_{-n}^{n} G(\omega)\,d\omega\right|^2$$

and using the Jacobi-Jensen inequality

$$\|X_1^* - X_1\|^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\omega)\,d\omega\right\}$$

where

$$f(\omega) = 2\pi \frac{dF(\omega)}{d\omega} = 2\pi \frac{d}{d\omega}\|dZ(\omega)\|^2$$

One can prove similar results for the h step predictor. We state the following without proof.

The predictor X_{t+h}^* for h units of time ahead based on $X_t, X_{t+1}\ldots$ can be written as

$$X_{t+h} = \int_{-n}^{n} e^{i(t+h)\omega}\left\{1 - \frac{G_h(\omega)}{G(\omega)}\right\} dZ(\omega)$$

where $$G_h(i\omega) = \frac{1}{2\pi} \sum_{\nu=0}^{h-1} G_\nu e^{-i\omega\nu}$$

or in the equivalent form

$$X^*_{t+h} = \frac{1}{2\pi} \sum_{\nu=h}^{\infty} G_\nu \xi_{t+h-\nu}$$

with the minimum prediction error

$$\| X^*_{t+h} - X_{t+h} \|^2 = \frac{1}{4\pi^2} \sum_{\nu=0}^{h-1} |G_\nu|^2$$

The above outline which has been included for completness follows the derivation of Grenander and Szeger who provide many more results both for the continuous case and for finite approximations to the optimum linear predictor.

APPENDIX 2

Pseudo random numbers $\{X_r\}$ were generated on the I.C.L. 1906A using a variant of the standard multiplicative congruential generator

$$X_{r+1} = b\,X_r \pmod{M}$$

b_0, X_0 and M being suitably chosen constants. For the theoretical background to these generators see Knuth (1969) and Newman and Odell (1971).

To minimise correlation in our sequence of pseudo-random numbers the actual algorithm used combined two generators

$$X_{1,r+1} = b_1\,X_{1,r} \pmod{M}$$

$$X_{2,r+1} = b_2\,X_{2,r} \pmod{M}$$

in the form

$$X_{r+1} = X_{1,r+1} + X_{2,r+1} \bmod (M) \qquad \text{A2.1}$$

with $M = 2^{26}$, $b_1 = 3^{15}$, $b_2 = 5^9$ and $X_{10} = X_{20} = 1234567$

This generator was written by the Nottingham Algorithms Group as part of their effort in building a software library for the I.C.L.1900 series of computers, however the idea of using two generators originated in Neave (1972). This generator has received a great deal of testing and the numbers produced have satisfied all the tests devised for random number generators.

Since A2.1 generates variates from a uniform distribution on $[0,1]$ it is necessary to make a suitable transformation if Normal deviates are required. Unfortunately, in the case of the normal distribution one cannot use the transformation

$$y = F^{-1}(x)$$

to produce deviates Y having a distribution $F(y)$ using X from a uniform $[0,1]$ distribution and one requires more specialised methods.

The transformation used is due to Box & Muller (1958)

$$z_{2t-1} = (-2\log X_{2t-1})^{1/2} \sin(2\pi X_{2t})$$

$$z_{2t} = (-2\log X_{2t-1})^{1/2} \sin(2\pi X_{2t})$$

where the $\{X_t\}$ are from a uniform $[0\,1]$ distribution and the $\{z_t\}$ are random numbers from a standard normal distribution.

This method works well with the particular uniform generator A2.1, however, an alternative due to Marsaglia (1964) may be worth implementing in future.

APPENDIX 3

TABLES OF PERCENTAGE POINTS

SIGNIFICANCE POINTS OF $\hat{\Delta}_I$ NEGLECTING END POINTS FOR $n = 2^k$

LOWER TAIL

Percentage n	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
8	-28.29365	-1.50842	-1.35988	-1.20381	-0.98290	-0.80071	-0.59945	-0.28440	-0.03483
16	-21.1152	-1.01834	-0.92239	-0.82100	-0.67614	-0.55540	-0.42068	-0.20600	0.01743
32	-10.89988	-0.69572	-0.63244	-0.56517	-0.46844	-0.38719	-0.29563	-0.14794	0.00871
64	- 5.50451	-0.47973	-0.43772	-0.39190	-0.32634	-0.27092	-0.2080	-0.10566	0.00435
128	- 3.99099	-0.33309	-0.30415	-0.27322	-0.22828	-0.19010	-0.14661	-0.07523	0.00218
	- 3.45626	-0.23246	-0.21256	-0.19124	-0.16017	-0.13368	-0.10341	-0.05346	0.00109
512	- 3.05482	-0.16283	-0.14904	-0.13424	-0.11262	-0.09415	-0.07298	-0.03793	0.00055
1024	- 2.57817	-0.11437	-0.10476	-0.09443	-0.07932	-0.06639	-0.05154	-0.0266	0.00027
2048	- 2.07456	-0.08048	-0.07376	-0.06653	-0.05593	-0.4685	-0.03641	-0.01904	0.00014
4096	- 1.60825	-0.05676	-0.05200	-0.04692	-0.03947	-0.03308	-0.02473	-0.01348	0.00007
8192	- 1.21347	-0.04001	-0.03669	-0.03311	-0.02787	-0.02337	-0.01818	-0.00954	0.00003

UPPER TAIL

Percentage	75	90	95	97.5	99.0	99.5	99.75	100
8	0.32237	0.55463	0.68197	0.78548	0.89803	0.96990	1.03322	7.46162
16	0.22482	0.39832	0.49611	0.57742	0.66807	0.72706	0.78011	18.28855
32	0.15739	0.28446	0.35753	0.41910	0.48867	0.53471	0.57650	9.76021
64	0.11042	0.20249	0.25610	0.30168	0.35364	0.38836	0.42008	3.45221
128	0.07761	0.14383	0.18270	0.21595	0.25410	0.27973	0.30326	1.96991
256	0.05465	0.10201	0.12998	0.15400	0.18167	0.20034	0.21754	1.93946
512	0.03852	0.07229	0.09230	0.10954	0.12946	0.14291	0.15537	2.09314
1024	0.02718	0.05119	0.06546	0.07778	0.09204	0.10171	0.11064	2.01882
2048	0.01919	0.03623	0.04638	0.05516	0.06533	0.07223	0.07862	1.76562
4096	0.01355	0.02564	0.03285	0.03908	0.04632	0.05124	0.05579	1.44308
8192	0.00958	0.01814	0.02325	0.02768	0.03281	0.03631	0.03954	1.12702

SIGNIFICANCE POINTS FOR $\hat{\Delta}_I$ NO ENDS

LOWER TAIL

Percentage n	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
50	-6.7383	-0.54714	-0.49825	-0.44617	-0.37098	-0.30755	-0.23576	-0.11917	0.00557
60	-5.77212	-0.49647	-0.45238	-0.40539	-0.33745	-0.28004	-0.21497	-0.10904	0.00464
70	-5.18468	-0.45747	-0.41705	-0.37395	-0.31154	-0.25875	-0.19885	-0.10114	0.00398
80	-4.80013	-0.42629	-0.38878	-0.34876	-0.29076	-0.24165	-0.18587	-0.09474	0.00348
90	-4.53277	-0.44064	-0.36530	-0.32800	-0.27362	-0.22572	-0.17513	-0.09043	+0.00309
100	-4.33727	-0.37905	-0.34591	-0.31052	-0.25916	-0.21560	-0.16606	-0.08493	0.00278
110	-4.18801	-0.36057	-0.32913	-0.29553	-0.24674	-0.20537	-0.15826	-0.08105	0.00253
120	-4.06976	-0.34452	-0.31454	-0.28250	-0.23597	-0.19646	-0.15146	-0.07766	0.00232
130	-3.97304	-0.33041	-0.30171	-0.27104	-0.22647	-0.18861	-0.14547	-0.07466	0.00214
140	-3.89173	-0.31788	-0.29032	-0.26086	-0.21802	-0.18162	-0.14013	-0.07199	0.00199
150	-3.82178	-0.30665	-0.28011	-0.25173	-0.21045	-0.17536	-0.13535	-0.06959	0.00186

UPPER TAIL

Percentage	75	90	95	97.5	99.0	99.5	99.75	100
50	0.12525	0.22861	0.22857	0.33941	0.39719	0.43570	0.47081	4.93906
60	0.11411	0.20902	0.26423	0.31114	0.36458	0.40026	0.43284	3.76693
70	0.10549	0.19375	0.24521	0.28900	0.33893	0.37238	0.40293	3.08605
80	0.09855	0.18141	0.22980	0.27104	0.31815	0.34969	0.37857	2.66797
90	0.09282	0.17117	0.21699	0.25609	0.30080	0.33076	0.35821	2.40021
100	0.08798	0.16249	0.20613	0.24338	0.28603	0.31464	0.34087	2.22297
110	0.08382	0.15502	0.19675	0.23242	0.27327	0.30070	0.32586	2.10250
120	0.08020	0.14849	0.18856	0.22282	0.26210	0.28848	0.31270	2.01882
130	0.07701	0.14273	0.18132	0.21033	0.25221	0.27766	0.30103	1.95960
140	0.07417	0.13759	0.17486	0.20676	0.24337	0.26799	0.29060	1.91700
150	0.07162	0.13297	0.16904	0.19993	0.23540	0.25927	0.28119	1.88589

SIGNIFICANCE POINTS FOR $\hat{\Delta}_I$ NO ENDS

LOWER TAIL

Percentage n	0.0	0.25	0.5	1.0	2.5	5.0	100	25.0	50.0
200	-3.59882	-0.26410	-0.24139	-0.21707	-0.18166	-0.15151	-0.11709	-0.06039	0.00139
250	-3.46941	-0.23532	-0.21517	-0.19358	-0.16211	-0.13530	-0.10465	-0.05409	0.00112
300	-3.36891	-0.21421	-0.19593	-0.17633	-0.14774	-0.12336	-0.09548	-0.04943	0.00112
400	-3.20569	-0.18478	-0.16908	-0.15223	-0.12765	-0.10665	-0.08262	-0.04286	0.00070
500	-3.06981	-0.16482	-0.15086	-0.13587	-0.11399	-0.09528	-0.07386	-0.03838	0.00056

UPPER TAIL

Percentage	75	90	95	97.5	99.0	99.5	99.75	100
200	0.05900	0.10998	0.14005	0.16586	0.19557	0.21559	0.23403	1.89461
250	0.05531	0.10322	0.13151	0.15580	0.18378	0.20266	0.22004	1.93214
300	0.05044	0.09429	0.12026	0.14248	0.16816	0.18550	0.20148	1.98802
400	0.04363	0.08173	0.10429	0.12370	0.14611	0.16126	0.17523	2.06044
500	0.03899	0.07314	0.09339	0.11083	0.13098	0.14460	0.15718	2.09137

PERCENTAGE POINTS FOR THE DISTRIBUTION OF Δ_I'
TAPIZODAL FORM

LOWER TAIL

Percentage n	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
16	-16.80002	-1.00588	-0.92066	-0.82915	-0.69546	-0.58121	-0.45033	-0.23368	0.00383
32	-18.07466	-0.67220	-0.61619	-0.55587	-0.46750	-0.39180	-0.30453	-0.15940	0.00099
64	-14.75544	-0.46227	-0.42401	-0.38277	-0.32226	-0.27027	-0.21040	-0.11049	0.00027
128	-10.98789	-0.32244	-0.29583	-0.26712	-0.22499	-0.18877	-0.14703	-0.07731	0.00008
256	- 7.92181	-0.22647	-0.20780	-0.18766	-0.15809	-0.13266	-0.10335	-0.05437	0.00002
512	- 5.64312	-0.15961	-0.14646	-0.13227	-0.11144	-0.09352	-0.07286	-0.03834	0.00001

UPPER TAIL

Percentage	75	90	95	97.5	99	99.5	99.75
16	0.23785	0.44542	0.56820	0.67378	0.79551	0.87773	0.95347
32	0.16048	0.30326	0.38835	0.46193	0.54723	0.60515	0.65873
64	0.11079	0.21005	0.26936	0.32074	0.38040	0.42100	0.45859
128	0.07740	0.14693	0.18851	0.22456	0.26645	0.29497	0.32139
256	0.05440	0.10332	0.13259	0.15797	0.18747	0.20756	0.22617
512	0.03835	0.07285	0.09350	0.11140	0.13221	0.14639	0.15952

PERCENTAGE POINTS FOR THE DISTRIBUTION OF $\hat{\Delta}_L$ TAPIZODAL FORM

LOWER TAIL

Percentage n	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
20	-18.10759	-0.87953	-0.80553	-0.72597	-0.60961	-0.51003	-0.39576	-0.20614	0.00246
30	-18.24687	-0.69685	-0.63872	-0.57614	-0.48447	-0.40586	-0.31547	-0.16505	0.00112
40	-17.22413	-0.59455	-0.54515	-0.49193	-0.41391	-0.34694	-0.26988	-0.14147	0.00065
50	-16.11673	-0.52707	-0.48337	-0.43627	-0.36721	-0.30789	-0.23960	-0.12573	0.00043
60	-15.11793	-0.47831	-0.43871	-0.39602	-0.33340	-0.27960	-0.21764	-0.11427	0.00030
70	-14.24964	-0.44097	-0.40450	-0.36517	-0.30747	-0.25789	-0.20077	-0.10546	0.00023
80	-13.49817	-0.41120	-0.37721	-0.34050	-0.28678	-0.24055	-0.18730	-0.09842	0.00018
90	-12.84458	-0.38674	-0.35479	-0.32033	-0.26976	-0.22630	-0.17622	-0.09262	0.00015
100	-12.27173	-0.36619	-0.33594	-0.030332	-0.25546	-0.21431	-0.16690	-0.08773	0.00012
110	-11.76555	-0.34859	-0.31981	-0.28877	-0.24321	-0.20404	-0.15891	-0.0354	0.00010
120	-11.31475	-0.33331	-0.30580	-0.27612	-0.23257	-0.19512	-0.15197	-0.07990	0.00009

UPPER TAIL

Percentage n	75.0	90.0	95.0	97.5	99	99.5	99.75
20	0.20882	0.39261	0.50170	0.59574	0.70444	0.77804	0.84596
30	0.16627	0.31404	0.40207	0.47817	0.56636	0.62624	0.68161
40	0.14217	0.26905	0.34475	0.41026	0.48626	0.53792	0.58572
50	0.12619	0.23906	0.30645	0.36480	0.43254	0.57861	0.52125
60	0.11460	0.21725	0.27857	0.33168	0.39335	0.43531	0.47416
70	0.10571	0.20048	0.25711	0.30618	0.36318	0.40194	0.43785
80	0.09861	0.18707	0.34995	0.28576	0.33898	0.37520	0.40875
90	0.09277	0.17603	0.22581	0.26895	0.31906	0.35317	0.38477
100	0.08365	0.15878	0.20370	0.24264	0.28788	0.31869	0.3655
120	0.0800	0.15186	0.19483	0.23208	0.27537	0.30484	0.33214

PERCENTAGE POINTS FOR $\hat{\Delta}_f$ (NO ENDS)

m denotes number of spectral points used

Truncation point $M = N/3$

LOWER TAIL

Percentage m	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
8	-14.93241	-0.51839	-0.47531	-0.42889	-0.36086	-0.30247	-0.23527	-0.12331	0.00058
10	-13.80302	-0.44907	-0.41184	-0.37172	-0.31289	-0.26236	-0.20417	-0.10715	0.00035
12	-12.79455	-0.40095	-0.36777	-0.33199	-0.27951	-0.23442	-0.18249	-0.90584	0.00024
14	-11.93505	-0.36521	-0.33502	-0.30246	-0.25469	-0.21363	-0.16633	-0.08739	0.00017
16	-11.20610	-0.33739	-0.30951	-0.27945	-0.23534	-0.19742	-0.15373	-0.08080	0.00013
20	-10.04725	-0.29642	-0.27195	-0.24556	-0.20682	-0.17352	-0.13514	-0.07105	0.00008
40	-7.06493	-0.20192	-0.18527	-0.16732	-0.14095	-0.11828	-0.09214	-0.04848	0.00002
60	-5.73877	-0.16274	-0.14933	-0.13486	-0.11362	-0.09535	-0.07428	-0.03909	0.00001

UPPER TAIL

Percentage m	75.0	90.0	95.0	87.5	99.0	99.5	99.75
8	0.12394	0.23453	0.30051	0.35759	0.42382	0.46883	0.51048
10	0.10753	0.20372	0.26117	0.31091	0.3685	0.40792	0.44428
12	0.09609	0.18219	0.23363	0.27819	0.32994	0.36315	0.39775
14	0.08757	0.16612	0.21306	0.25374	0.30098	0.33314	0.36292
16	0.08093	0.15357	0.19699	0.23462	0.27834	0.30810	0.33566
20	0.07114	0.13504	0.17325	0.20638	0.24486	0.27107	0.29334
40	0.04850	0.09212	0.11821	0.14084	0.16714	0.18506	0.20165
60	0.03910	0.07427	0.09532	0.11357	0.13478	0.14923	0.16262

NULL DISTRIBUTION FOR WHITE NOISE TEST, FULL TRAPIZODAL FORM
EQUAL LENGTH SEGMENTS $P = Q = 2n$

LOWER TAIL

Percentage n	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
16	-32.89607	-1.15642	-1.06117	-0.95749	-0.80553	-0.67512	-0.52509	-0.27514	0.00138
32	-26.26475	-0.81226	-0.74507	-0.67263	-0.56635	-0.47502	-0.36983	-0.19427	0.00042
64	-19.56851	-0.57234	-0.52511	-0.47417	-0.39939	-0.33510	-0.26100	-0.13725	0.00013
128	-14.14675	-0.40606	-0.37075	-0.33481	-0.28206	-0.23669	-0.18439	-0.09701	0.00004

UPPER TAIL

Percentage n	75	90	95	97.5	99	99.5	99.75
16	0.27664	0.52334	0.67048	0.79779	0.94546	0.104581	1.13867
32	0.19472	0.36928	0.47360	0.56398	0.66895	0.74038	0.80653
64	0.13739	0.26085	0.33468	0.39869	0.47307	0.52371	0.57062
128	0.09705	0.18434	0.23656	0.28185	0.33448	0.37033	0.40354

NULL DISTRIBUTION FOR WHITE NOISE TEST (NO END POINTS)
EQUAL LENGTH SEGMENTS WITH $P = Q = 2n$

LOWER TAIL

Percentage n	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50
16	-32.92380	-1.19168	-1.09245	-0.98558	-0.82900	-0.69460	-0.54015	-0.28285	+0.0062
32	-26.46099	-0.82415	-0.75596	-0.68344	-0.57459	-0.48191	-0.37517	-0.2005	0.0045
64	-19.68193	-0.57751	-0.52894	-0.47762	-0.40230	-0.33753	-0.2689	-0.14424	0.0014
128	-14.20529	-0.40553	-0.37210	-0.33603	-0.28309	-0.23755	-0.18506	-0.09736	0.00004

UPPER TAIL

Percentage n	75.0	90.0	95.0	97.5	99.0	99.5	99.75
16	0.28461	0.53806	0.68916	0.81985	0.97138	1.07431	1.16954
32	0.19754	0.37458	0.48036	0.57201	0.67744	0.75086	0.82792
64	0.13849	0.26273	0.33709	0.40156	0.47647	0.52747	0.57472
128	0.09740	0.18501	0.23742	0.28287	0.33570	0.37167	0.40500

PERCENTAGE POINTS FOR THE DISTRIBUTION OF $\hat{\delta}_k$ $k = 2m + 1$
Trapizodal Estimate

LOWER TAIL

Percentage N	0.0	0.25	0.5	1.0	2.5	5.0	10.0	25.0	50.0
32	-23.5137	-0.67562	-0.61991	-0.55980	-0.47158	-0.39571	-0.30825	-0.16215	0.00009
64	-16.44366	-0.46395	-0.42573	-0.38448	-0.32393	-0.27185	-0.21180	-0.11146	0.00001
128	-11.49973	-0.32315	-0.29654	-0.26781	-0.22564	-0.18936	-0.14754	-0.07765	0.00000
256	- 8.07647	-0.22676	-0.20808	-0.18793	-0.15834	-0.13288	-0.10353	-0.05449	0.0000

UPPER TAIL

Percentage Sample Size	0.75	0.90	0.95	0.975	0.99	0.995	0.9975
32	0.16225	0.30814	0.39542	0.47110	0.55906	0.01896	0.67446
64	0.11147	0.21178	0.27180	0.32386	0.38437	0.42559	0.46378
128	0.7765	0.14754	0.18936	0.22563	0.26780	0.29652	0.32313
256	0.05449	0.1033	0.13288	0.15833	0.18792	0.20808	0.22676

REFERENCES

Abramowitz, M and Segun, I (1968)

Handbook of Mathematical Functions
Dover

Akaike, H (1969)

A Method of statistical identification of discrete time parameter linear systems

Ann. Institute of Stat. Math
21 No.2 pp

- (1969)

Power spectrum estimation through autoregressive model fitting

Ann. Inst. Stat. Math
21 No.2 pp 407 - 419

- (1970)

Fitting autoregressive models for prediction

Ann. Inst. Stat. Math.
21 No.2

Anderson, R. (1942)

Distribution of the serial correlation coefficient

Ann. Math. statist
13 pp 1 - 13

Anderson, T (1971)

The Statistical Analysis of Time Series
Wiley

Bartlett, M (1946)

On the theoretical specification and sampling properties of autocorrelated time series

J. Roy. Statist Socc. Supp.
8 pp 27 - 41, 85 - 97

(Corrigenda (1948) 10 p 200)

M (1949)

Periodogram analysis and continuous spectra

Biometrika 37 pp 1 - 16

(1955)

An introduction to stochastic processes with special reference to methods and applications
Cam. Univ. Press
(2nd ed. 1966)

Barton, D and Dennis, K (1953)

The conditions under which Gram-Charlier and Edgeworth curves are positive definate and unimodal

Biometrika 37 425 - 427

Bates J and Granger C (1969)

The combination of forecasts

O.R. Quarterly 20, 4 pp 451 - 468

Beveridge, W (1922)

Wheat prices and rainfall in Europe
J. Roy Statist Soc. 85 p 417 - 459

Billingsley, P (1965)

Ergodic theory and information
Wiley.

Blackman, R and Tukey, J (1959)

The measurement of power spectra from the point of view of communications engineering
Bell Systems Tech.Journal
37 pp 158 - 282 and 485 - 569

Box, G and Jenkins, G (1972)

Time Series analysis forecasting and control
Holden Day

Box G, Jenkins, G and Bacon, D (1967)

Models for forecasting seasonal and non-seasonal time series

Advanced seminar on the spectral analysis of time series
ed. Harris B
Wiley

Box, G and Muller, M (1958)

A note on the generation of random normal deviates
Ann. Math Statist. 29 pp 610 - 611

Box, G and Pierce, D (1970)

Distribution of residual autocorrelations in autoregressive integrated moving average time series models

J. American Statist Assn. 65, 332 pp 1509 - 1525

Brown, R (1959)

Statistical Forecasting and inventory control

MacGraw Hill

- (1963)

Smoothing, Forecasting and prediction of discrete time series
Prentice-Hall

Charlier, C (1905)

Uber das Fehlergesetz
Ark. Mat. Astr och Fys 2, 8

Clenshaw, C and Curtis, A (1960)

A method for numerical integration on an automatic computer
Numerische. Math. 2 pp 197 - 205

Cooley, J and Tukey, J (1965)

A algorithm for the machine calculation of Fourier series
Math. Comp. 19 pp 297 - 301

Copas, J (1966)

Marte Carlo results for estimation in a stable Markov time series
J. Roy, Statist Soc. B pp 110 - 116

Cornish, E and Fisher, R (1937)

Moments and cumulants in the specification of distributions
Rev. Ins. Int. Statist. 5 p 307

Cramer,H (1925)

On some classes of series used in mathematical statistics
Proc. of 6th Scandinavian congress of Mathematicians
Copenhagen pp 399 - 425

- (1928)

On the composition of elementary errors
Skand. Akt 11, pp 13 and 141

- (1937)

Random variables and Probability distributions
Camb. Tracts in Math. No.36

(1946)

Mathematical Methods in Statistics
Princeton University Press

Daniel, P (1946)

Discussion on "Symposium or autocorrelation in time series"
J. Roy Statist Soc. 8 pp 88 - 90

Daniels, H (1954)

Saddlepoint approximations in statistics
Ann. Math. Statist. 25 pp 631

Davis, H (1941)

The analysis of economic time series
Princip. Press.

Davis, H and Jones, R (1968)

Estimation of the innovation variance of a stationary time series
J. Am. Statist. Assn. 63 pp 141 - 149

Davison, L (1965)

Prediction of a time series from a finite part
J. Soc. Ind. Math 13 No.3

Doob, J (1953)

Stochastic Processes
Wiley

Durbin, J (1959)

Efficient estimation of parameters in moving average models
Biometrika 46 pp 301 - 316

(1960a)

The estimation of parameters in time-series regression models
J. Roy Statist Soc. B 22 pp 139 - 153

(1960b)

The fitting of time series models
Rev. Inst. Statist. 28 233 - 244

Durbin, J (1968)

Some recent work on testing for serial correlation in regression analysis.
Summary of lectures given at the Conference on Statistical methods in Econometrics at the Centro Internazional Matematico Esti Rome. June

Durbin, J (1970)

An alternative to the bounds test for testing serial correlation in least squares regression
Econometrica 38 pp 422 - 429

Edgeworth, F. (1905)

The law of error
Trans. Camb. Phil Soc. 20 36 and 113

Eicker, F(1963)
Central limit theorems for families of Sequences of Random Variables
Ann. Math Statist. 34 pp 437 - 447

Elderton, W and Johnson, N (1969)

Systems of frequency curves
Camb. Univ. Press

Erdelyi, A (1953)

Higher trancendental functions Vol.1.
McGraw Hill

Essen, C (1945)

Fourier analysis of distribution functions
Acta Math 77 pp 1 - 125

Freiberger, W and Grenander

The approximate distribution of noise power measurements
Jour. Ind App. Math 17 No. 3 271 - 283

Fisher, R.A. (1929)

Tests of significance in hammeric analysis
Proc Roy Soc. A 125 pp 54 - 59

Fishman, G (1969)

Spectral methods in Econometrics
Harvard Univ. Press

Flecher, R and Powell, M (1903)

A rapidly convergent descent method for minimisation
Camp. Journal. 6 pp 163 - 168

Frieberger, W and Grenander, H (1959)

Approximate distributions of noise power measurements
Jour. Ind app. matt 17 3 pp 271 - 283

Gilchrist, W (1968)

Two Stage forecasting methods
Northern Polytechnic
Statistical Forecasting Research Group report

Gnedenko, B and Kolmogorov, A (1954)

Limit distributions for sums of independent random variables
Addison-Wesley

Granger, C and Hatanaka, M (1964)

Spectral analysis of economic time series
Princeton Univ. Press

Granger, C (1967)

Simple trend fitting for long range forecasts
Management Decision 1964 pp 29 - 34

Granger, C (1969)

Prediction with a generalised cost of error function
O.R. research quarterly V20 (1969) pp 199 - 207

Grander, C and Newbold, P

Some comments on forecast evaluation
Note 8
Nottingham Forecasting Project

Granger, C and Newbold, P (1973)

Economic forecasting from the atheists view point
J. Roy Statisc soc. to appear

Greary, R (1947)

Testing for normality
Biometrika 34 pp 209 - 242

Grenander, U (1954)

On the estimation of regression coefficients in the case of autocorrelated disturbances
Ann. Math Statist 25 pp 252 - 272

Grenander, U, Pollack, H and Slepian, D (1959)

The distribution of quadratic forms in normal variates
J. Soc Ind App. Matt 7 pp 374 - 401

Grenander, U and Rosenblatt, M (1957)

Statistical analysis of stationary time series
Wiley

Grenander, U and Szego, G (1958)

Toeplitz forms and their application
University ofCaliformia Press

Hannan, T (1957)

Testing for serial correlation in least squares regression
Biametrika 44 pp 57 - 66

(1960)
Time Series Analysis
Methuen

(1972)

The analysis of multiple time series
Wiley

Harrison, P (1965)

Short term sales forecasting
Applied Statistics 14 No. 2

Hart, B (1942)

Significance levels of the ratio of mean square successive differences for the variance
Ann. Math Statist 13 pp 445 - 7

Hilbebrand, F (1956)

Introduction to numerical analysis
McGraw Hill

HMSO (1961)

Modern computing methods
Notes on Applied Science 16 (2nd ed) DSIR/HMSO

Holt, C (1957)

Forecasting seasonal and trends by exponentially weighted moving averages
Camegie inst. of Tech report

Hurwicz, L (1954)

Least squares bias in Time series
Statistical inference in Economic models
Cowels Commission ed. Coopmans pp 205 -

Ibragimov, I (1964)

On the asymptotic behaviour of prediction error
Theory Prob. App. 9 pp 627 - 634

Ibraginov and Solev (1967)

The asymptotic behaviour of the prediction error of a stationary sequence with a spectral density of a special type
Theory Prob. App. 12 pp 703 - 707

Isserlis, L (1918)

On a formula for the product moment coefficient of any order of a normal frequency distribution in any number of variables
Biometrika 12 p 134

Jefferies, H and Jefferies, B (1950)

Methods of mathematical physics
Oxford Univ. Press

Jenkins, G (1961)

General consideration in the estimation of spectra
Technometrics 3 pp 133 - 166

Jenkins, G and Priestley, M (1957)

The spectral analysis of time series
J. Roy Statist. Soc 19 pp 1 - 12

Johnson, N, Nixon, E, Amos, D andPearson, E (1963)

Tables of percentage points of Pearson curves, for given β_1 and β_2 expressed in standard measure
Biametrika 50 pp 459 - 498

Jones, R (1964)

Spectral analysis and linear prediction of Meteorological time series
J. Appl. Meteorology 3 pp 42 - 52

Kac, M (1954)

Toeplitz matrices, kernals and a related problem in probability theory
Duke Math. Jour. 21 pp 501 - 9

Kalman, R (1960)

A new approach to linear filtering and prediction problems
ASME trans ptD 82

Kendall, M (1954)

A note on the bias in the estimation of autocorrelation
Biometrika 41 p 397

Kendall, M and Stuart, A (1963)

The Advanced theory of statistics Vo. I
Griffin

(1961) Vol. 2

(1969) Vol. 3

Khintchine, A (1934)

Korrelations theorie der stationaren stochaotichen prozesse
Math. Ann. 109 p 604

Kruth, D (1969)

The art of computer programming Vol. 2
Addison Wesley

Kolmogorov, A (1939)

Sur l'interpretation et extrapolation des suites stationanaires
Comptes Rendus Ac. Sci. Paris 208 pp 2043 - 2045

(1941a)

Stationary sequences in Hilbert space
Bull Math. Univ. Moscoe 2 No.6 pp 1 - 40

(1941b)

Interpolation and Extrapolation von stationaren zufalligen folgen
Bull. Acad. Sci. U.S.S.R. Ser. Math. 5 pp 3 - 14

Kozulgaev, P (1941)

On problems of interpolation and extrapolation of stationary sequences
Dokl. Akad. Nourk S.S.S.R. 30 13

Kunichinka, H (1971)

A note on short term business survey
Paper presented to the 10th CIRET conference
Brussels

Lounguist, A (1971)

The predictive power of quantative investment forecasts as given in Swedish investment surveys
10 CIRET conference
Brussels

Lomnicki, Z and Zaremba, S (1959)

On some moments and distribution occuring in the theory of linear stochastic processes
Monatshefte fur Math. 2 No.63 pp 134

Magness, T.A. (1954)

Spectral response of a quadratic device to non-Gaussian noise
J. App. Phys. 35 pp 1357

Mann, H and Wald, A (1943)

On the statistical treatment of linear stochastic models
Econometrika 11 pp 173 - 220

Marriot, F and Pope, J (1954)

Bias in the estimation of autocorrelation
Biometrika 41 pp 390 - 6

Muth, J (1960)

Optimal properties of exponentially weighted forecasts of time series with permanent and transitory components
J. Am. Statist Assn. 55 pp 299 - 306

Neave, H (1970)

An improved formula for the asympotic variance of spectrum estimates
Ann. Math. Statist 41 pp 70 - 77

(1972)

A random number package
Computer applications in the natural and social sciences
Vol.14.

Newbold, P and Granger, C (1973)

Some comments on forecast evaluation
To appear

Newman, T and Odell, P (1971)

The generation of Random variates
No.21 Griffins Statistical Manographs

Olshen, R (1967)

Asymptotic properties of the periodogram of a discrete stationary process
J. App. Prob. 4 pp 508 - 528

Orcutt, G and Winokur, H (1969)

First order autoregression : Inference estimation and prediction
Econometrika 1969 37 No.1

Parzen, E (1957)

Optimum consistent estimates of the spectrum of a stationary time series
Ann. Math. Statist. 28 2

(1957b)

On choosing an estimate of the spectral density function of a stationary time series
Ann. Math. Statist 28 4

Parzen, E (1958)

On asympotically efficient consistent estimates of the spectral density function of a stationary time series
J. Roy Statist Soc. B 20 2 p 303

(1961a)

Mathematical considerations in the estimation of spectra
Technometrics 1 No.1 p 167

(1961b)

An approach to time series analysis
Ann. Math Statist. 32 pp 951 - 89

(1963)

On statistical spectral analysis
Proc. of a symposium on Stoc Proc. published by Am.Math Soc

Pearson, E and Hartley, H (1970)

Biometrika Tables for Statisticians Vol.1
Cambridge Univ. Press

Pitt, H (1963)

Integration, measure and probability
Oliver and Boyd

Pitman, E (1937)

The "closest estimates of statistical parameters
Proc. Cam. Phil, Soc. 33 pp 212 - 222

Priestley, M (1962)

Basic considerations in the estimation of spectra
Technometrics 4 pp 551 - 564

(1965)

The role of bandwidth in spectral analysis
Applied Stats. 14 1 pp 33 - 47

Reid, D (1969)

A comparative study of time series prediction techniques on economic data
PhD thesis University of Nottingham

Robbins, H (1948)

Application of the method of mixtures to quadratic forms in normal variates
Ann. Math. Stat. 20 pp 552 - 560

Rosenblatt, M (1957)

Some purely deterministic processes
J. Math. Mech 6 pp 801 - 810

Rosenblatt, M (1962)

Random Processes
Oxford University Press

Rozanov Yu (1967)

Stationary random processes
Holden-Day

Schuster, A (1898)

On the investigation of hidden periodicities
Terr. Mag.

Singleton, R (1968)

Algorithm 338, Algol procedures for the fast fourier transform
Comm. ACM 11 pp - 773 - 776

Slepian (1958)

Fluctuations of random noise power
Bell syt. tech. Journal 37 pp 183 - 184

Stephens, H (1963)

The distribution of the goodness of fit stochistic
Biametrika 50 pp 303 - 313

Szego, G (1939)

Orthogonal Polynomials
Am. Math. Soc Collog. pub. No.23.

Teichroew (1956)

Tables of expected values of order statistics and products of order statistics
Am. Math. Statist. 27 No.41

Theil, H (1958)

Economic forecasts and policy
North Holland pub. co.

Thornber, H. (1967)

Finite sample Monte-Carlo Studies an autoregressive illustration
J. Am Statist. Association pp 801 - 818

Tukey, J. (1949)

The sampling theory of power spectrum estimates
"Symp. on applications of autocorrelation analysis to Physical problems
Woods Hole, Mass.

Walker, A (1961)

Large sample estimation of parameters for moving avarerage models
Biometrika 48 pp 343 - 357

(1962)

Large sample estimation of parameters for autoregressive processes with moving average residuals
Biametrika 49 pp 117 - 131

(1965)

Some asympotic results for the periodogram of a short time series
J. Aust. Math. Soc. 5 pp 107 - 128

Wallace, D (1958)

Asymptotic approximations to distributions
Am. Math. Statist 29 pp 635 - 654

Welch, B (1947)

The generalisation of "Students" problem when several different populations are involved
Biometrika 34 pp 28 - 35

White, J (1959)

The limiting distribution of the serial correlation coefficient in the explosive case I and II
Ann. Math. Statist 29 pp 1188 - 97
and 30 pp 831 - 4

(1961)

Asymptotic expansions for the mean and variance of the serial correln coeff
Biometrika 48 pp 85 - 94

Whittaker, E (1935)

Interpolatory function theory
Camb. Univ. Press

Whittaker, E and Watson, G (1965)

A Course of Modern Analysis
Camb. Univ. Press (4th ed.)

Whittle (1953)

Tests of fit in time-series analysis
Biometrika 39 pp 309 - 318

(1953)

The analysis of Multiple Stationary time series
J. Roy Statis Soc. B 13 pp 123 - 139

(1963)

Prediction and Regulation

English University Press

Wienier, N (1930)

Generalised harmonic analysis
Acta math. 55 pp 117 - 258

(1949)

Extrapolation, interpolation and smoothing of Stationary time series
Wiley

Wienier, N and Masani (1957, 58)

The prediction theory of multivariate stochastic processes, I and II
Acta Math. 98 pp 111 - 150
and 99 pp 83 - 137

Wilks, S (1932)

Certain generalisations in the analysis of variance
Biometrika 24 pp 471 - 494

Winters, P (1959)

Forecasting sales by exponentially weighted moving averages
Man. Sci 6

(1962)

Two exponentially weighted forecasting models
Carnegie Inst. of Tech. report

Wold, H (1938)

A study in the analysis of stationary time series
Upsalon

Yaglom, A (1962)

An introduction to the theory of stationary random processes
Trans. RA. Silverman
Prentice Hall

Yule, G (1927)

On a method for investigating periodicities in disturbed series
Philos Trans. Roy Soc. London
Ser. A 226 pp 267 - 298

Zasukin, V (1941)

Comptes Rendus de l'Academic des Sciences de l'URSS
23 pp 435 - 437

Mr. P. Fisk has pointed out that in the development of the finite sample approximations presented on pages 44-50 there is an implicit assumption that $E\left[X_t X_{t-k}\right] = E\left[X_t X_{t+k}\right]$. This is not so and it can be shown that

$$E\left[X_t X_{t-k}\right] = \frac{\sigma^2 \alpha^k (1-\alpha^{2(t-k)})}{1-\alpha^2} + \alpha^{2t-k}\,\sigma^2(X)$$

while $E\left[X_t X_{t+k}\right] = E\left[X_{t+k} X_{(t+k)-k}\right]$

$$= \frac{\sigma^2 \alpha^k (1-\alpha^{2t})}{1-\alpha^2} + \alpha^{2t+k}\,\sigma^2(X)$$

In consequence the argument must be modified as follows, thus proof is that of Mr. Fisk. We seek coefficients a_j $(j=0, \ldots, r)$ which minimize

$$e_k = E\left[\sum_{j=0}^{r} a_j X_{n-j} - X_{n+k}\right]^2 .$$

The estimating equations $\partial e_k/\partial a_j = 0$ $(j=0, \ldots, r)$ may be written as

$$\begin{bmatrix} E(X_n^2) & E(X_n X_{n-1}) & \ldots & E(X_n X_{n-r}) \\ E(X_{n-1} X_n) & E(X_{n-1}^2) & \ldots & E(X_{n-1} X_{n-r}) \\ E(X_{n-2} X_n) & E(X_{n-2} X_{n-1}) & \ldots & E(X_{n-2} X_{n-r}) \\ \cdot & \cdot & \cdot & \cdot \\ E(X_{n-r} X_n) & E(X_{n-r} X_{n-1}) & \ldots & E(X_{n-r}^2) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_r \end{bmatrix} = \begin{bmatrix} E(X_n X_{n+k}) \\ E(X_{n-1} X_{n+k}) \\ E(X_{n-2} X_{n+k}) \\ \cdot \\ E(X_{n-r} X_{n+k}) \end{bmatrix}.$$

or equivalently as

$$\begin{bmatrix} Z+Y_n & \alpha Z+\alpha^{-1} Y_n & \ldots & \alpha^r Z+\alpha^{-r} Y_n \\ \alpha Z+\alpha^{-1} Y_n & Z+Y_{n-1} & \ldots & \alpha^{r-1} Z+\alpha^{-(r-1)} Y_{n-1} \\ \alpha^2 Z+\alpha^{-2} Y_n & \alpha Z+\alpha^{-1} Y_{n-1} & \ldots & \alpha^{r-2} Z+\alpha^{-(r-2)} Y_{n-2} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha^2 Z+\alpha^{-r} Y_n & \alpha^{r-1} Z+\alpha^{-(r-1)} Y_{n-1} & \ldots & Z+Y_{n-r} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_r \end{bmatrix}$$

$$= \begin{bmatrix} \alpha^k Z+\alpha^k Y_n \\ \alpha^{k+1} Z+\alpha^{k+1} Y_n \\ \cdot \\ \cdot \\ \alpha^{k+r} Z+\alpha^{k+r} Y_{n-r} \end{bmatrix}$$

Here $Z = \sigma^2(1-\alpha^2)^{-1}$ and $Y_s = \alpha^{2s}\left[\sigma^2(X) - \sigma^2(1-\alpha^2)^{-1}\right]$.

If we note that

$$Y_{n-s} = \alpha^{-2s} Y_n$$

we may write these estimating equations as

$$(\underset{\sim}{A} Z + \underset{\sim}{B} Y_n)\ \underset{\sim}{a} = \underset{\sim}{R}$$

with $\underset{\sim}{a}' = (a_0, a_1, \ldots, a_r)$,

$$\underset{\sim}{A} = \begin{bmatrix} 1 & \alpha^2 & \ldots & \alpha^r \\ \alpha & \alpha & \ldots & \alpha^{r-1} \\ \alpha^2 & 1 & \ldots & \alpha^{r-2} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha^r & \alpha^{r-2} & \ldots & 1 \end{bmatrix},$$

$$\underset{\sim}{B} = \begin{bmatrix} 1 & \alpha^{-1} & \ldots & \alpha^{-r} \\ \alpha^{-1} & \alpha^{-2} & \ldots & \alpha^{-(r+1)} \\ \alpha^{-2} & \alpha^{-3} & \ldots & \alpha^{-(r+2)} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha^{-r} & \alpha^{-(r+1)} & & \alpha^{-2r} \end{bmatrix}.$$

$$= \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & \alpha^{-1} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \alpha^{-r} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \cdot \\ 1 \end{bmatrix} [1,1,1,\ldots,1] \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & \alpha^{-1} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \cdot & & & \\ 0 & 0 & \ldots & \alpha^{-r} \end{bmatrix}$$

$$= C\iota\iota' \underset{\sim}{C}$$

and

$$\underset{\sim}{R} = \alpha^k Z \begin{bmatrix} 1 \\ \alpha \\ \alpha^2 \\ \vdots \\ \alpha^r \end{bmatrix} + \alpha^k Y_n \underset{\sim}{C}\,\underset{\sim}{\iota}.$$

The estimating equations may be written as

$$(\underset{\sim}{A}\,\underset{\sim}{a}-\alpha^k \begin{bmatrix} 1 \\ \alpha \\ \vdots \\ \alpha^r \end{bmatrix}) Z + (\underset{\sim}{B}\,\underset{\sim}{a}-\alpha^k \underset{\sim}{C}\,\underset{\sim}{\iota}) Y_n = \underset{\sim}{0}$$

which will be satisfied for arbitrary values of σ^2 and $\sigma^2(X)$ if the coefficient vectors for Z and Y_n are separately null. But

$$\underset{\sim}{A}\,\underset{\sim}{a}-\alpha^k \begin{bmatrix} 1 \\ \alpha \\ \vdots \\ \alpha^r \end{bmatrix} = \underset{\sim}{0}$$

is satisfied by $\underset{\sim}{a} = \alpha^k \underset{\sim}{j}_1$ where $\underset{\sim}{j}_1$ has a unit in the first location and zeros elsewhere. This choice of $\underset{\sim}{a}$ gives

$$\underset{\sim}{B}\,\underset{\sim}{a}-\alpha^k \underset{\sim}{C\iota} = \alpha^k \underset{\sim}{C\iota}\,\underset{\sim}{\iota}'\,\underset{\sim}{C}\,\underset{\sim}{j}_1 - 1 = 0$$

so $\underset{\sim}{a}=\alpha^k \underset{\sim}{j}_1$ is a solution of the estimating equation giving as the optimal predictor

$$X^*_{t+k} = \alpha^k X_t.$$

This same result could be obtained more directly by noting that

$$\underset{\sim}{a} = (\underset{\sim}{A}\, Z + \underset{\sim}{B}\, Y_n)^{-1}\, \underset{\sim}{R}$$

$$= Z^{-1}(\underset{\sim}{A} + (Y_n/Z)\underset{\sim}{B})^{-1}\, \underset{\sim}{R}.$$

But

$$(\underset{\sim}{A} + (Y_n/Z)\, \underset{\sim\sim}{C\iota\iota'C})^{-1}$$

$$= \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\, \underset{\sim\sim}{C\iota\iota'C}\underset{\sim\sim}{A}^{-1} \left[(Z/Y_n) + \underset{\sim}{\iota'C}\underset{\sim\sim}{A}^{-1}\, \underset{\sim\sim}{C\iota}\right]^{-1} .$$

Noting that

$$\underset{\sim}{A}^{-1} \begin{bmatrix} 1 \\ \alpha \\ \alpha^2 \\ \vdots \\ \alpha^r \end{bmatrix} = \underset{\sim}{j}_1$$

and so

$$\underset{\sim}{\iota'C}\underset{\sim\sim}{A}^{-1} \begin{bmatrix} 1 \\ \alpha \\ \alpha^2 \\ \vdots \\ \alpha^r \end{bmatrix} = 1$$

we have

$$\underset{\sim}{a} = \alpha^k Z^{-1}\left[\underset{\sim}{A}^{-1} - A^{-1}\, \underset{\sim\sim}{C\iota\iota'C}\underset{\sim\sim}{A}^{-1} \{(Z/Y_n) + \underset{\sim}{\iota'C}\underset{\sim\sim}{A}^{-1}\underset{\sim\sim}{C\iota}\}^{-1}\right]$$

$$\times \left[Z \begin{bmatrix} 1 \\ \alpha \\ \vdots \\ \alpha^r \end{bmatrix} + Y_n\, \underset{\sim\sim}{C\iota}\right]$$

$$= \alpha^k \underset{\sim}{j}_1 + \alpha^k (Y_n/Z)\, \underset{\sim}{A}^{-1}\, \underset{\sim\sim}{C}\underset{\sim}{\iota} - \alpha^k Z^{-1}\{(Z/Y_n) + \underset{\sim}{\iota}'\underset{\sim\sim}{CA}^{-1}\underset{\sim\sim}{C}\underset{\sim}{\iota}\}^{-1}$$

$$\times\left[\underset{\sim}{A}^{-1}\, \underset{\sim\sim}{C}\underset{\sim}{\iota}\, Z + Y_n \underset{\sim}{A}^{-1}\, \underset{\sim\sim}{C}\underset{\sim}{\iota}\, (\underset{\sim}{\iota}'\underset{\sim\sim}{CA}^{-1}\, \underset{\sim\sim}{C}\underset{\sim}{\iota})\right]$$

$$= \alpha^k \underset{\sim}{j}_1 \quad .$$

REFERENCES

Bartlett, M. (1950)

Periodogram analysis and continuous spectra

Biometrika 37 pp 1-16.

Gurland, J. (1955)

Distributions of definate and indefinate quadratic forms

Ann. Math. Statist. 26 pp 122-127.

Lukacs, E. (1960)

Characteristic Functions

Griffin.

Hartree, D. (1958)

Numerical Analysis

Oxford (2nd ed).

Marsaglia, G., Maclaren, M. and Bray, T. (1964)

A fast procedure for generating normal random variables

C.A.C.M. 7 (1964) pp 4-10.

Neave, H. R. (1966)

A missingdata problem in time-series analysis

Unpublished Ph.D. thesis

University of Nottingham.

Wilkinson, J. (1965)

The Algebraic Eigenvalue Problem

Oxford.