

REVIEW AND APPLICATION OF STATISTICAL METHODS
UNIVARIATE TIME-SERIES ANALYSIS

Gary James Kelly

A Dissertation Submitted to the Faculty of Science
University of the Witwatersrand, Johannesburg
for the degree of Master of Science

Johannesburg 1984

DECLARATION

I declare that this dissertation is my own, unaided work.

It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

GARY JAMES KELLY

25th day of February, 1984

ABSTRACT

This dissertation reviews the theory and methods of three areas in time-series analysis - periodogram analysis, spectral analysis and Box-Jenkins modelling. Once the basic mathematical necessities and definitions have been set down, the relationship between the periodogram and spectrum is illustrated.

The statistical properties of both the periodogram and spectrum are provided, as well as the practical problems involved when analysing data using these techniques. As a result of the complexities of the spectrum, more emphasis is placed on periodogram analysis.

The time-domain approach to time-series analysis consists of an introduction to, and explanation of the Box-Jenkins modelling theory. Techniques (for example, the autocorrelation function and inverse autocorrelation function) and the theory behind them - for identifying, estimating and checking models, are given.

Forecasting and Box-Jenkins seasonal models are only touched upon, as are nonlinear time-series models. An application of all the major techniques is provided.

For my family

TABLE OF CONTENTS

	Page
PREFACE	x
List of abbreviations	xiii
List of mathematical symbols	xv
List of figures	xvi
List of tables	xx
 Chapter	
1. INTRODUCTION AND NECESSARY THEORY	1
1.1 Introduction	1
1.2 The meaning of the Fourier transform	4
1.3 Stochastic and stationary processes	5
1.4 The autocovariance and autocorrelation functions	8
1.5 An outline proof of Fourier's theorem	9
1.6 The nonperiodic case (outline proof of theorem 2)	13
1.7 The spectral density function	16
1.8 Summary	18
 2. THE WIENER-KHINTCHINE THEOREM AND THE SPECTRAL DENSITY FUNCTION	19
2.1 Relating the SDF to the autocovariance	19
2.2 The Wiener-Khinchine theorem	24
2.3 Discrete processes	30
2.4 Estimation of the autocovariance and hence SDF and PSD	33
2.5 Estimation of the SDF	37
2.6 Summary	38

Chapter	Page
3. PERIODOGRAM ANALYSIS	39
3.1 Introduction	39
3.2 Harmonic analysis	40
3.3 Aliasing	44
3.4 The discrete Fourier transform	45
3.4.1 Notes	50
3.5 Properties of a_1 and b_1	54
3.6 Periodogram analysis	56
3.7 Sampling properties of the periodogram	59
3.8 Summary and discussion	66
4. THE SPECTRUM	70
4.1 The periodogram as an estimate of the spectrum	71
4.2 Consistent estimates of the sif	72
4.3 Spectral window	75
4.3.1 Selected windows	76
5. THE PROBLEMS INVOLVED IN SPECTRAL ANALYSIS	84
5.1 INTRODUCTION	84
5.2 SCALE PARAMETER BIAS	85
5.3 A measure of the precision of spectral estimates	86
5.4 WINDOW SELECTION	88
5.5 CYCLE OF WORK	88
5.6 IDENTIFYING AND REMOVING	88
5.6.1 IDENTIFYING	88
5.6.2 REMOVING	88
5.7 SUMMARY	89

Chapter	Page
6. ESTIMATION IN THE TIME DOMAIN - BOX-JENKINS MODELS	101
6.1 Introduction	101
6.2 The MA process	105
6.3 The AR process	107
6.3.1 The Yule-Walker equations	109
6.3.2 The duality between an MA(1) and an AR(1)	109
6.3.3 The AR(2) process	110
6.4 The partial autocorrelation function (pacf)	113
6.5 ARMA processes and the Box-Jenkins models	115
6.5.1 ARMA processes	115
6.5.2 The spectra of AR, MA and ARMA models	116
6.6 A general note on stationarity and invertibility	118
6.7 Notes on appendices A9-A11	122
7. THE BOX-JENKINS STRATEGY	123
7.1 Phase one: identifying stationarity	123
7.2 Phase two: identification of the model	128
7.2.1 The inverse acf and inverse pacf	133
7.3 Phase three: estimation of model parameters	135
7.3.1 Calculating the unconditional sum of squares (USS)	134
7.3.2 Variance and covariance of ML estimates	137
7.3.3 Maximum likelihood estimation of θ , ϕ and σ^2	138
7.4 Phase four: checking the fit of the model and improving the model	140
7.4.1 Model inadequacy	140
7.4.2 Selecting the best model	144

Chapter	Page
7.5 Phase five - forecasting	147
7.6 Seasonal Box-Jenkins models	149
7.6.1 The acrf	149
7.6.2 The pacrf	150
7.6.3 The iacrf	150
7.6.4 The ipacrf	150
8. NONLINEAR MODELLING AND MISCELLANEOUS NOTES	152
8.1 Bilinear time-series models	152
8.2 Threshold autoregressive models	156
8.3 Missing observations	158
8.4 R and S arrays	158
9. AN APPLICATION	159
9.1 The data	159
9.2 Transforming the data	160
9.3 Detecting any deterministic models	160
9.4 The Box-Jenkins analysis	160
9.4.1 Stationarity	160
9.4.2 Model identification and specification	164
9.4.3 Estimation	165
9.4.4 AR modelling and AR-MA model	165
9.5 Comment	165
9.5.1 Model adequacy	166
9.5.2 Logarithmic transformation	166
9.6 Conclusions	166

<u>APPENDICES</u>	page
A1 Stochastic limiting operations	A1.1-A1.3
A2 The variance and covariance of acvf estimates	A2.1-A2.3
A3 The variance and covariance of acrf estimates	A3.1-A3.2
A4 The variance and covariance of the periodogram . .	A4.1-A4.2
A5 Testing the significance of periodogram ordinates	A5.1-A5.3
A6 A guide for detecting periodicity	A6.1
A7 Sampling properties of spectral estimates	A7.1-A7.12
A8 Approximate variance in spectrum estimates	A8.1-A8.4
A9 Obtaining initial estimates in an ARMA process ...	A9.1-A9.2
A10 A recursive method for calculating estimates of AR parameters	A10.1-A10.2
A11 Parameter estimates and properties of specific BJ models	A11.1-A11.4
A12 Estimation of the iacrf and the asymptotic behaviour of the iacr's	A12.1-A12.5
A13 The distribution of residual autocorrelations ..	A13.1-A13.2
A14 ARMA processes with non-normal residuals	A14.1-A14.2
A15 Estimating the order of AR and MA models - the $\max \chi^2$ method	A15.1-A15.4
A16 Transformed series	A16.1-A16.3
A17 Forecasting	A17.1-A17.4
A18 Description of computer programs	A18.1-A18.22
A19 Results relevant to the chapter 9 airline data	A19.1-A19.16
<u>BIBLIOGRAPHY</u>	B1 - B9

PREFACE

A discrete univariate time-series is a sequence of observations ordered and equally spaced in time. This dissertation reviews the theory and methods of three areas in time-series analysis and presents an application thereof. The object of this study is to present a unified treatment of the techniques involved and to gain a working knowledge of these procedures. The application of these techniques has necessitated the development of a suite of computer programs for performing time-series analysis.

There are two major approaches to the study of time-series processes - time-domain and frequency-domain analysis. In the context of this work, I shall consider the former as consisting of an approach generally called Box-Jenkins modelling and the latter of the two techniques - periodogram analysis and spectrum analysis.

Once the basic mathematical necessities and definitions have been set down, the periodogram is introduced as a tool useful for identifying (and removing) any periodic tendencies in the data. Verification of its statistical properties and notes relating to its use are given. Thereafter, the link between the periodogram and (theoretical) spectrum is forged.

In considering the spectrum and spectral analysis, the mathematical theory behind these concepts is discussed and an introduction to, and the properties of, 'windowed' spectral estimates are given. The approach used here is from a periodogram viewpoint. The problems in setting up a spectral analysis are also considered. This completes the analysis in the frequency-domain.

The time-domain approach consists of an introduction to the Box-Jenkins modelling theory and its statistical properties. Tools (and the theory behind them) for identifying, estimating and 'checking' models are given. 'Checking' will involve ascertaining whether the residuals from the fitted model are approximately white noise variables.

Box-Jenkins seasonal models will only be mentioned in passing as the general approach adopted here is to remove any seasonal components by using the periodogram (i.e. harmonic analysis) and to apply the Box-Jenkins theory to the residuals. The reason for this is that if the time-series is seasonal the usual Box-Jenkins approach of differencing implies that several terms may be lost in order to induce the necessary state of stationarity for Box-Jenkins stationary models. This is not necessary when using periodogram analysis. However, one price you may have to pay for this is a lack of parsimony, i.e. there may not be a minimum number of parameters in the model.

Spectral analysis is considered as subsidiary to periodogram analysis due to the problems (both practical and theoretical) involved in calculating the spectrum. However one may 're-generate' the data based on the fitted model and hence compute a spectral estimate and compare this with the approximate spectrum of the original data.

As in the case of Box-Jenkins seasonal models, forecasting (future) values will only be touched upon. As the basis of this study is rather on modelling. However, it follows that if a reasonable model has been formulated, forecasts should not prove too difficult to obtain. (The only difficulty arising may be in

The structure of the book is as follows. Chapter 1 of a preliminary chapter on the theory of time-series is given in the appendix.

The techniques presented are given. Examples for some of the particular approaches and methods are given in the text. This application has necessitated the creation of a battery of computer programs for time-series analysis. These programs have been written in a conversational mode and are as automatic as possible at the moment. An idea for the future is to make this package fully automatic thus enabling the layman to perform a time-series analysis. A description of the programs is given in an appendix.

It is worthwhile stressing that only discrete univariate time-series are considered in this work. Any continuous time-series may be converted into a discrete one by sampling at equal time intervals. Thus the theory presented here, is, in a way, indirectly applicable to continuous time-series. No multivariate time-series theory is considered as univariate series are more common, and more work has been done in this field.

The main sources for reference material were Box and Jenkins (1974), Priestley (1981) (both volumes 1 and 2), Bloomfield (1976) and to a lesser degree, Chatfield (1980). Gottman (1981) proved to be an invaluable asset as far as a non-theoretical (practical) approach was concerned and also served as a readable introductory text.

List of Abbreviations

acr	autocorrelation
acrf	autocorrelation function
acrgf	autocorrelation generating function
acv	autocovariance
acvf	autocovariance function
acvgf	autocovariance generating function
AR	autoregressive
ARIMA	autoregressive integrated moving average
ARMA	autoregressive - moving average
BJ (B-J)	Box-Jenkins
cdf	cumulative density function
edf	empirical distribution function
iacr	inverse acf
iacrf	inverse acrf
iacrgf	inverse acrgf
iacv	inverse acv
iacvf	inverse acvf
iacvgf	inverse acvgf
ipacrf	inverse partial autocorrelation function
i.t.o.	in terms of
LHS	left hand side
LSE	least-squares estimate
MA	moving average
MLE	maximum likelihood estimate
MSE	mean-square error
nsdf	normalized spectral density function

pacf	partial autocorrelation function
pdf	probability density function
resp	respectively
RHS	right hand side
RSS	residual sum of squares
rv (r.v.)	random variable
sdf	spectral density function
SS	sum of squares
USS	unconditional sum of squares
w.l.o.g.	without loss of generality
W-K	Wiener-Khintchine
w.r.t.	with respect to
Y-W	Yule-Walker

List of mathematical symbols

$\forall$	for all
$\lim_{k \rightarrow \infty}$	the limit as k tends to ∞
$\rightarrow$	tends to
$\propto$	proportional to
$\approx$	approximately equal
$\sim$	distributed as
$\asymp$	asymptotically distributed as
$\bar{x}$	the sample mean or the complex conjugate of x
$O(k)$	of order k
$\therefore$	therefore
$\nrightarrow$	does not tend to
$\Re$	real part of
$\Im$	imaginary part of
$\frac{1}{2}$	approximately distributed as
$\log$	$\log_e$
$\log_{10}$	$\log_{10}$
$\iff$	if and only if
$\exists$	there exists
$\in$	is an element of
$\equiv$	equivalent to
$[k]$	the next integer $\leq k$ if $k \geq 0$
$\approx$	approximately
$\ a\ $	the Euclidean norm of vector a
a'	the transpose of vector a
$\chi^2_{(k)}$	$P(\chi^2_{(k)} \leq x) = 1 - e^{-x/2}$ (and $k = \text{degrees of freedom}$)

	<u>List of figures</u>	<u>page</u>
Fig 4.1	The periodogram lag window	76
4.2	The truncated periodogram lag window	77
4.3	The Bartlett lag window	77
4.4	Danielli's lag window	78
4.5	The Blackman-Tukey lag window	79
4.6	The Parzen (b) lag window, and the Tukey-Hanning lag window	81
5.1	The truncated periodogram spectral window	82
6.1	A plot of the airline passenger data	160
6.2	Quantile-box plot of the airline passenger data	161
6.3	Quantile-box plot of $\log_{10}$ of the airline passenger data	164
6.4	Quantile-box plot of the square-root of the airline passenger data	165
6.5	Plot of $\log_{10}$ of the airline passenger data	167
6.6	Autocorrelation of $\log_{10}$ of the airline passenger data	169
6.7	Spectrum estimate of the transformed data (Parzen (b) window, truncation point 75)	171
6.8	Spectrum estimate of the transformed data (Parzen (b) window, truncation point 142)	172
6.9	$\log_{10}$ of the periodogram of the transformed data	173
6.10	$\log_{10}$ of the periodogram of the transformed data (50% tapering)	175
6.11	$\log_{10}$ of the periodogram of the transformed data (50% tapering)	176

9.12	$\log_{10}$ of the periodogram of the (once) de-sined transformed data (Parzen (b) window, truncation point 60)	178
9.13	Spectrum estimate of the (once) de-sined transformed data (Parzen (b) window, truncation point 60)	179
9.14	$\log_{10}$ of the periodogram of the (twice) de-sined transformed data (20% tapering)	181
9.15	Spectrum estimate of the (twice) de-sined transformed data (Tukey-Hanning window, truncation point 75)	182
9.16	$\log_{10}$ of the periodogram of the (thrice) de-sined transformed data	183
9.17	Spectrum estimate of the (thrice) de-sined transformed data (Tukey-Hanning window, truncation point 80)	187
9.18	Spectrum estimate of the (thrice) de-sined transformed data (Tukey-Hanning window, truncation point 80)	188
9.19	Spectrum estimate of the (thrice) de-sined transformed data (Tukey-Hanning window, truncation point 75)	189
9.20	Autocorrelation of the (thrice) de-sined transformed data	193
9.21	Partial autocorrelation of the (thrice) de-sined transformed data	195
9.22	Inverse autocorrelation of the (thrice) de-sined transformed data	196
9.23	Inverse partial autocorrelation of the (thrice) de-sined transformed data	198
9.24	Residual autocorrelation with asymptotic 99% confidence limits	199

	Page
Fig. 9.25 Residual autocorrelogram of (0,0,3)(1,0,1)19 BJ model with 99% confidence limits	201
9.26 Residual autocorrelogram of (3,0,0)(1,0,1)19 EJ model with 99% confidence limits	202
9.27 Residual autocorrelogram of (0,0,1)(1,0,1)19 BJ model with 99% confidence limits	203
9.28 Residual autocorrelogram of (0,0,0)(1,0,1)19 BJ model with 99% confidence limits	204
9.29 Spectrum estimate of the residuals (Parzen (b) window, truncation point 60)	206
9.30 Spectrum estimate of the residuals (Parzen (b) window, truncation point 67)	208
9.31 Spectrum estimate of transformed data minus 3 sinusoids (Parzen (b) window, truncation point 67)	209
9.32 Spectrum estimate of transformed data minus 3 sinusoids minus BJ one-step-ahead forecasts (Parzen (b) window, truncation point 67)	210
9.33 Spectrum estimate of transformed data minus 3 sinusoids minus the simulated BJ series (Parzen (b) window, truncation point 67)	211
9.34 $P(\cdot)$ with lower truncation point 22 and upper truncation point 53	219
9.35 $P(\lambda)$ with lower truncation point 22 and upper truncation point 59	220
9.36 Autocorrelogram (up to lag 148) of the transformed data minus the harmonic component detected by the $P(\cdot)$ test	222

Fig. A19.1 Autocorrelogram of the first 150 observations (in the original form)	A19.8
A19.2 Spectrum estimate for the original data (Parzen (b) window, truncation point 75)	A19.9
A19.3 Spectrum estimate for the original data (Parzen (b) window, truncation point 112)	A19.10
A19.4 Results from the program BUYBAL for the original data	A19.11
A19.5-A19.7 Spectrum estimates for the transformed series (Parzen (b) window, truncation points 7, 15 and 30 resp.)	A19.12
A19.8-A19.10 Spectrum estimates for the transformed series (Parzen (b) window, truncation points 60, 75 and 112 resp.)	A19.13
A19.11 Autocorrelogram (up to lag 148) of the transformed data	A19.14

List of tables

	<u>Page</u>
9.1 A comparison of quantile-box plots	166
9.2 Finding the degree of differencing	191
9.3 Overfitting in AR models	199
9.4 SARIMA models	199
9.5 One step ahead forecasts	205
9.6 AR modelling (using backforecasting)	217
 A19.1 Listing of the original data	 A19.1
A19.2 Listing of the transformed ($\log_{10}$) data	A19.2
A19.3 Stationarity 'test' on original data (subseries length 13)	A19.3 A19.4
A19.4 Stationarity 'test' on original data (subseries length 26)	A19.4 A19.5
A19.5 Stationarity 'test' on original data (subseries length 52)	A19.5 A19.6
A19.6 Stationarity 'test' on transformed data (subseries length 13)	A19.6 A19.7
A19.7 Stationarity 'test' on transformed data (subseries length 26)	A19.7 A19.8
A19.8 Stationarity 'test' on transformed data (subseries length 52)	A19.8 A19.9

CHAPTER 1: INTRODUCTION AND NECESSARY THEORY

1.1 INTRODUCTION

When we look at any object, one of the first things we notice is its colour. We generally see a mixture of basic colours, but cannot say how much red or blue light the object contains. Instead we observe the overall effect.

To distinguish the various colours, the light from the object must be passed through a prism and we may then determine the 'strength' of different colours. This information is called the spectrum of light.

In a similar fashion any wave-like phenomenon can also be analysed by breaking the wave into a composition of sine and cosine waves of various frequencies and amplitudes. Now wave-like phenomena are generally associated with ocean or radio waves, and we usually see some pattern in these waves - i.e. they repeat themselves after certain time intervals. This type of phenomenon is called a periodic function.

Definition 1

A function $f(x)$ (where x is a measure of 'time') is periodic if $f(x) = f(x+h)$ (for some number h) and h is said to be the period of $f(x)$.

$a \cos(wx)$ is one of the most familiar periodic functions. If it has period h , then a equals $h/2\pi$ and w is called the angular frequency (expressed in radians per unit time). a is called the amplitude.

This function, together with $b \sin(wx)$, can be considered as building blocks for all other periodic functions, since any 'well-behaved' periodic function of finite period h , can be expressed as a (possibly infinite) sum of weighted sine and cosine waves. This remarkable result will be discussed more fully in section 1.5 but can be written more formally as

Theorem 1 Fourier's theorem

If $f(x)$ is a well-behaved periodic function of period h , then $f(x)$ may be written as a Fourier series in the form

$$f(x) = \sum_{k=0}^{\infty} \left(a_k \cos\left(\frac{2\pi k x}{h}\right) + b_k \sin\left(\frac{2\pi k x}{h}\right) \right) \quad (1.1)$$

where $a_0, a_1, \dots$ and $b_0, b_1, \dots$ are constants to be determined.

(For proof see section 1.5)

In a similar fashion if $f(x)$ is nonperiodic, we may still regard it as periodic but with infinite period and hence

Theorem 2

If $f(x)$ is a nonperiodic function, then it can be rewritten as

$$f(x) = \int_0^{\infty} \left(a(w) \cos(wx) + b(w) \sin(wx) \right) dw \quad (1.2)$$

where $a(w)$ and $b(w)$ are functions whose form is still to be determined.

(For proof see section 1.6)

absolutely integrable over $(-\infty, \infty)$ with a finite number of discontinuities

For computational purposes we may define

$$J_r = \begin{cases} \frac{1}{2}(a_r - ib_r) & r > 0 \\ a_0 & r = 0 \\ \frac{1}{2}(a_{|r|} + ib_{|r|}) & r < 0 \end{cases} \quad (1.3)$$

or analogously

$$J(w) = \begin{cases} \frac{1}{2}(a(w) - ib(w)) & w > 0 \\ a(0) & w = 0 \\ \frac{1}{2}(a(|w|) + ib(|w|)) & w < 0 \end{cases} \quad (1.4)$$

and then we may rewrite (1.1) and (1.2), respectively, as

$$f(x) = \sum_{k=-\infty}^{\infty} J_k e^{i w_k x} \quad (1.5)$$

(where $w_k = 2\pi k/h$ and $w_{-k} = -w_k$)

$$\text{and} \quad f(x) = \int_{-\infty}^{\infty} J(w) e^{i w x} dw \quad (1.6)$$

Definition 2

$J(w)$ (or analogously J_k) is called the Fourier transform of $f(x)$

Thus if we are prepared to accept theorems 1 and 2, it turns out that any well-behaved function may be written as some function of its Fourier transform. So if we have a set of data which follows some functional form, we may estimate this functional form merely by estimating the Fourier transform of the data.

So given any set of data ordered in time, it is possible to 'model' the process which generated it simply by approximating the generating process by the Fourier transform of the data.

THE MEANING OF THE FOURIER TRANSFORM

Theorems one and two are valid (and this validity will be shown later). We assume that any periodic and nonperiodic functions can be written as a sum of sine and cosine functions, then it may be shown (see section 1.3) that the 'energy' given off by the process at any time interval is equal to the sum of the amounts of energy given off by each of the sine and cosine terms.

The energy given off by any sine or cosine term is proportional to the square of its amplitude and so for periodic processes, the contribution of the r th term in the case where $f(x)$ is periodic, is given by

$$E_r = \frac{1}{2} \pi a_r^2 \quad (\text{by theorem 1})$$

which is proportional to $a_r^2 + b_r^2$

and is equivalent to a_r^2

Plotting a_r^2 against r (or ω_r) will give the relative contribution of various sine and cosine terms to the total energy.

This plot is called an energy spectrum.

The relative contributions can only be considered for discrete frequencies (we say a periodic process has a discrete spectrum). Nonperiodic processes must be considered for continuous frequencies (see representation (1.2)), and we say these processes have a continuous spectrum.

Since $|J(\omega)|^2 = |J(-\omega)|^2$ and $|J(\omega)|^2 = |J(-\omega)|^2$ so the spectrum is symmetric and only need be plotted for positive frequencies.

For a more complete description of the Fourier transform see, for example, 'The Fourier Transform' by R. V. Churchill (1958).

1.3 STOCHASTIC AND STATIONARY PROCESSES

So far no mention has been made of the whole time-series, so we are entitled to ask

"How does the theory of sections 1.1 and 1.2 help us?"

We shall see from the following definitions (some of which we shall only see the relevance of at a later stage) exactly what bearing this theory has.

Definition 3

A random (or stochastic) process is such that for each t , the value $X(t)$ is a random variable. $X(t)$ is called a random process if the situation is such that for each t , we cannot determine theoretically a precise value of $X(t)$. Instead we have a range of values with an associated pdf (probability density function).

Definition 4

An observed record of a random process is merely one realization out of a whole collection of possible realizations.

Definition 5

For each t , $X(t)$ is a r.v. (random variable) and so accordingly $X(t)$ will have the same pdf as $X(s)$ (for all s) $p_t(x)$, say, (if $X(t)$ is continuous) or $p_t(x_i)$, say, (if $X(t)$ is a discrete r.v.)

At this point, I have changed notation from $f(x)$ to $X(t)$ since a time-series is dependent on time, say, and at each time point $X(t)$ is a random variable i.e. $f(t) = X(t) \forall t$.

Definition 6

A discrete time-series is a realization of a random process $X(t)$ only considered at the discrete time points $\dots, -1, 0, 1, 2, \dots$ and denoted by X_t .

The theory of sections 1.1 and 1.2 concerning a function $f(x)$ is really for a deterministic function $f(x)$ (i.e. a function which for a particular value of x , we can determine $f(x)$). Does this then mean we may not apply this theory to a discrete time series which is a realization of a random process?

No. We are saved by the implications of Wiener's theory of generalized harmonic analysis (which is formulated in section 2.2) which states: each realization of a 'stationary' process can be expressed as a Fourier-Stieltjes (Wiener(1930)) transform of the form

$$X(t) = \int e^{itw} dZ(w)$$

where $Z(w)$ is not necessarily differentiable but has the same form as $J(w)$ given by (1.4).

Saved? No - we have not yet defined what we mean by 'stationary'. If this stationarity condition does not impose too many restraints on our single realization (i.e. our time-series) we may then apply the theory of sections 1.1 and 1.2 and hence estimate some form of model.

Definition 7

A stationary process has the characteristic that its statistical properties do not change with time i.e. if X_t is stationary, then $X(1), X(2), \dots$ must each have the same pdf.

Definition 8

$X(t)$ is called strictly stationary if the joint distribution of $X(t_1), \dots, X(t_n)$ is identical to that of $X(t_1+k), \dots, X(t_n+k)$ (for any set $t_1, \dots, t_n$ and any k).

Definition 9

$X(t)$ is said to be stationary up to order m if for any set $t_1, \dots, t_n$ and any k , all the joint moments up to m of $X(t_1), \dots, X(t_n)$ exist and equal the corresponding moments of $X(t_1+k), \dots, X(t_n+k)$.

Definition 10

A process is called Gaussian if all joint distributions of $X(t_1), \dots, X(t_n)$ (for any and sets $t_1, \dots, t_n$) are multivariate normal.

Thus we may in principle ascertain whether our time-series is stationary (if it is not we need some technique to induce stationarity) and hence calculate Fourier transforms of the data.

1.4 THE AUTOCOVARANCE AND AUTOCORRELATION FUNCTIONS

A special case of the stationarity concept is for stationarity of order $m = 2$. By definition this implies that

- (i) $\{X(t)\}$ has the same mean, μ , $\forall t$,
- (ii) $\{X(t)\}$ has the same variance, σ^2 , $\forall t$, and
- (iii) the covariance between the values at any two time points, s and t say, depends only on $(s-t)$.

So if $\{X(t)\}$ is stationary up to order two, two functions, namely the autocovariance function $\{\gamma(\tau)\}$ and the autocorrelation function $\{\rho(\tau)\}$, may be derived.

Definition 11

$$\begin{aligned}\gamma(\tau) &= \text{cov}\{X(t), X(t+\tau)\} \\ &= E\{(X(t) - \mu)(X(t+\tau) - \mu)\} \quad (\forall t, \text{ say } t)\end{aligned}$$

is defined to be the covf and is independent of t by the definition of 2nd order stationarity.

Similarly

Definition 12

$$\rho(\tau) = \gamma(\tau)/\gamma(0) \quad (\forall \tau) \text{ is defined to be the acrf.}$$

When $\{X(t)\}$ is a discrete time process, i.e. $X(t) = X_n$, $(t=0, \pm 1, \pm 2, \dots)$

the covf has the following properties

- (i) $\gamma(0) = \sigma^2$ the variance of X_n ,
- (ii) $|\gamma(\tau)| \leq \gamma(0)$ ($\forall \tau$) since $|\rho(\tau)| \leq 1$ for all τ ,
- (iii) if X_n is real-valued then $\gamma(-\tau) = \gamma(\tau)$ $\forall \tau \geq 0$ and
- (iv) $\gamma(\tau)$ is positive semidefinite.

Similar properties may be derived when $\{X(t)\}$ is a continuous-time process.

The value and uses of these functions is not immediately clear, but they will be utilized at a later stage.

1.5 AN OUTLINE PROOF OF FOURIER'S THEOREM (THEOREM 1)

Theorem 1, Fourier's theorem, states that if $f(x)$ is a well-behaved periodic function of period h , then $f(x)$ may be written as a Fourier series in the form

$$f(x) = \sum_{k=0}^{\infty} \left(a_k \cos\left(\frac{2\pi k x}{h}\right) + b_k \sin\left(\frac{2\pi k x}{h}\right) \right) \quad (1.7)$$

(where a_k, b_k are some constants).

Without loss of generality, suppose $h = 2\pi$. Now the following functions, namely $\cos(t), \sin(t), \cos(2t), \sin(2t), \dots$, are all periodic with period 2π , and any linear combination of them will also be periodic with period 2π . Then we must show that if

$X(t)$ is a function of t with period of 2π , then we can represent $X(t)$ by

$$X(t) = \frac{1}{2} a_0 + \sum_{k=1}^{\infty} \left(a_k \cos(kt) + b_k \sin(kt) \right) \quad \forall t. \quad (1.8)$$

This will prove Fourier's theorem.

The reason for using $\frac{1}{2} a_0$ instead of a_0 is simply for notational purposes.

Note that a_k and b_k are called Fourier coefficients and the term corresponding to $k = 1$ is called the fundamental, that

again, I use $X(t)$ instead of $f(x)$ for notational continuity

corresponding to $k = 2$ the first harmonic, that corresponding to $k = 3$, the second harmonic, etc.

If we multiply both sides of (1.8) by $\cos(jt)$ (j some integer ≥ 0) and integrate over t from $-\pi$ to π , then if the RHS of (1.8) can be integrated term by term (which is possible if the series is uniformly convergent over the interval $(-\pi, \pi)$) then we have

$$\int_{-\pi}^{\pi} X(t) \cos(jt) dt = \frac{1}{2} a_0 \int_{-\pi}^{\pi} \cos(jt) dt + \sum_{k=1}^{\infty} \left[a_k \int_{-\pi}^{\pi} \cos(kt) \cos(jt) dt + b_k \int_{-\pi}^{\pi} \sin(kt) \cos(jt) dt \right] \quad (1.9)$$

$$j = 0, 1, 2, \dots$$

(if we assume the Fourier representation (1.8) is permissible)

Now since

$$\begin{aligned} \int_{-\pi}^{\pi} \cos(kt) \cos(jt) dt &= 0 & j \neq k \\ \int_{-\pi}^{\pi} \sin(kt) \cos(jt) dt &= 0 & \forall j, k \\ \int_{-\pi}^{\pi} \sin(kt) \sin(jt) dt &= 0 & j \neq k \end{aligned}$$

then we may rewrite (1.9) as

$$\begin{aligned} \int_{-\pi}^{\pi} X(t) \cos(jt) dt &= a_j \int_{-\pi}^{\pi} \cos^2(jt) dt \\ &= \pi a_j \end{aligned}$$

$$\text{i.e. } a_j = \frac{1}{\pi} \int_{-\pi}^{\pi} X(t) \cos(jt) dt \quad j = 0, 1, 2, \dots \quad (1.10)$$

[the use of $\frac{1}{2}a_0$ instead of a_0 now becomes apparent - since we now note that (1.10) includes the case $j = 0$]

In a similar manner, it may be shown that

$$b_j = \frac{1}{\pi} \int_{-\pi}^{\pi} X(t) \sin(jt) dt \quad j = 1, 2, \dots \quad (1.11)$$

[(1.10) and (1.11) are called the Euler-Fourier formulae]

By assuming the series $X(t)$ is uniformly convergent on $(-\pi, \pi)$ we eliminate the problem of integrating term by term. However we are still not justified in writing $X(t)$ i.t.o. a Fourier representation (1.7). Or are we?

If we define a_j and b_j by (1.10) and (1.11) respectively then it may be shown that with these coefficients the RHS of (1.8) represents the function $X(t)$. The only questions to be asked are

(a) do (1.10) and (1.11) always yield finite values?

and

(b) if so, does (1.8) (i.e. the RHS of (1.8)) always converge to $X(t)$?

(a) is answered by insisting that $X(t)$ be absolutely integrable over $(-\pi, \pi)$. (b) is solved by restricting $X(t)$ to a finite number of discontinuities and also a finite number of maxima and minima (these conditions are termed the Dirichlet conditions).

Under the above assumptions, we can claim that any (periodic) function has a Fourier series representation and hence any (deterministic) time-series also has one. If we estimate the a_k and b_k , we may determine at what frequencies the time-series has (significant) power.

In summary then, if $X(t)$ is a well-behaved periodic function, of period 2π , we may represent $X(t)$ by

$$X(t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} [a_k \cos(kt) + b_k \sin(kt)] \quad \forall t$$

if

- (i) $X(t)$ is uniformly convergent on $(-\infty, \infty)$
- (ii) $X(t)$ is absolutely integrable over $(-\infty, \infty)$
- (iii) $X(t)$ has a finite number of discontinuities in $(-\infty, \infty)$
- (iv) $X(t)$ has a finite number of maxima and minima in $(-\infty, \infty)$

Notes

(a) Jordan's test (Theorem 3)

If $X(t)$ is of bounded variation in the neighbourhood of $t = t_0$, then its Fourier series converges to the sum

$$\frac{1}{2} [X(t_0 + 0) + X(t_0 - 0)]$$

$X(t)$ is of bounded variation in (a, b) if for any subdivision of this interval, $a = t_0 < t_1 < \dots < t_n = b$

we have

$$\sum_{i=0}^{n-1} |X(t_{i+1}) - X(t_i)| \leq M$$

is bounded by a constant whose value is independent of the form of the subdivision

For proof, see Titchmarsh (1939)

(b) Theorem 4

If $X(t)$ has a finite number of discontinuities as well as a finite number of maxima and minima in $(-\infty, \infty)$, then its Fourier series converges (for all t) to

$$\frac{1}{2} (X(t-0) + X(t+0))$$

(for proof, use theorem 3)

hence

(c) Theorem 5

If $X(t)$ is continuous and of bounded variation throughout $(-\infty, \infty)$, then its Fourier series converges for all t to $X(t)$.

1.6 THE NONPERIODIC CASE (OUTLINE PROOF OF THEOREM 2)

The results of the last section only hold for the case where $f(t)$ (or $X(t)$) is periodic (with finite period h). If $X(t)$ is deterministic but not periodic, we would also like some Fourier representation if possible.

Simply define (or choose) some interval $(-\tau, \tau)$ and then define

$$Y(t) = \begin{cases} X(t) & -\tau \leq t \leq \tau \\ Y(t+2p\tau) = Y(t) & |t| > \tau, p = 1, 2, 3, \dots \end{cases}$$

Then $Y(t)$ is periodic with period 2τ , and so can be expressed in terms of a Fourier series:

So

$$Y(t) = \sum_{k=-\infty}^{\infty} (a_k \cos(2\pi w_k t) + b_k \sin(2\pi w_k t))$$

where $w_k = \frac{k}{2\tau}$

Defining J_k as in (1.3) i.e.

$$J_k = \begin{cases} \frac{1}{2}(a_k - ib_k) & k > 0 \\ \frac{1}{2}a_0 & k = 0 \\ \frac{1}{2}(a_{|k|} + ib_{|k|}) & k < 0 \end{cases}$$

then

$$Y(t) = \sum_{k=-\infty}^{\infty} J_k e^{2\pi i w_k t}$$

and so (see (i) and (ii) below)

$$Y(t) = \sum_{k=-\infty}^{\infty} \left(\int_{-\tau}^{\tau} X(t) e^{-2\pi i w_k t} dt \right) e^{2\pi i w_k t} \quad (-\tau \leq t \leq \tau)$$

where $w_k = w_k - w_{k-1} = \frac{1}{2\tau}$

(i) Since if $Z(t)$ has period 2τ (i.e. not necessarily 2π) $Z(t)$ may be written as

$$Z(t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{k\pi t}{\tau}\right) + b_k \sin\left(\frac{k\pi t}{\tau}\right) \right)$$

(by simply extending the proof in section 1.3 to deal with the case of period 2τ)

where

$$a_i = \frac{1}{\tau} \int_{-\tau}^{\tau} Z(t) \cos\left(\frac{k\pi t}{\tau}\right) dt$$

$$b_i = \frac{1}{\tau} \int_{-\tau}^{\tau} Z(t) \sin\left(\frac{k\pi t}{\tau}\right) dt \quad i = 1, 2, \dots$$

Note that this is not the same as (1.3) since (1.3) employs a_0 instead of $\frac{1}{2}a_0$. $\frac{1}{2}a_0$ is used for computational purposes.

(ii) Using (1), it may be shown that

$$\begin{aligned} J_k &= \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(t) e^{-2\pi i \omega_k t} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\tau) e^{-2\pi i \omega_k \tau} d\tau \quad (1.11) \end{aligned}$$

Now taking the limit as $r \rightarrow \infty$ just gives

$$X(t) = \sum_{j=-\infty}^{\infty} J_j e^{2\pi i \omega_j t} / \omega_j$$

(since $\omega_k^* \rightarrow 0$ and as $(\omega_1, \dots, \omega_{n-1}, \omega_n, \omega_1, \dots)$ becomes a continuous set of points and hence summation becomes integration)

$$\text{where } J(\omega) = \int_{-\infty}^{\infty} X(t) e^{-2\pi i \omega t} dt \quad (1.12)$$

(These integrals exist only if $X(t)$ is absolutely integrable over $(-\infty, \infty)$)

Let $x^* = \frac{t}{2\pi}$ then

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(x^*) e^{i x^* t} dx^*$$

$$\text{where } g(x^*) = \int_{-\infty}^{\infty} X(t) e^{-i x^* t} dt$$

which then proves theorem 2.

(Note that $g(x^*) = \frac{1}{2\pi} J(\frac{x^*}{2\pi})$)

THE SPECTRAL DENSITY FUNCTION

Let us consider a stationary stochastic process, or rather a single realization $x(t)$ of $X(t)$, thereof (and w.l.o.g. assume $E(X(t)) = 0$). If we cannot assume $X(t)$ is periodic and nor can we assume $X(t)$ is absolutely integrable. Thus it would seem we cannot represent $x(t)$ as a Fourier series (since $X(t)$ must either be periodic, or aperiodic and absolutely integrable over $(-\infty, \infty)$ - see section 1.1 for the latter restriction).

To clear this problem, simply define another process, $X^*(t)$ and let it be

$$X^*(t) = \begin{cases} X(t) & -\infty < t < \infty \text{ some } T \\ 0 & \text{otherwise} \end{cases} \quad (1.12)$$

Then we can write

$$X^*(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\omega) e^{i\omega t} d\omega$$

$$\text{where } g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X^*(t) e^{-i\omega t} dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t) e^{-i\omega t} dt \quad (1.14)$$

Since $X(t)$ may or may not possess a Fourier integral, the limit $\lim_{T \rightarrow \infty} \int_{-T}^T |g(\omega)|^2 d\omega$ does not exist. But the total energy given out by $X(t)$ in the interval $(-T, T)$ is equal to

$$\int_{-T}^T X^2(t) dt$$

1.7 THE SPECTRAL DENSITY FUNCTION

If we consider a stationary stochastic process, or rather a single realization (i.e. $X(t)$) thereof (and w.l.o.g. assume $E(X(t)) = 0$ $\forall t$), we cannot assume $X(t)$ is periodic and nor can we assume $X(t)$ is absolutely integrable. Thus it would seem we cannot represent $X(t)$ i.t.o a Fourier series (since $X(t)$ must either be periodic, or nonperiodic but absolutely integrable over $(-\infty, \infty)$ - see section 1.5 for the latter restriction).

To skirt this problem, simply define another process, $X^*(t)$ say, such that

$$X^*(t) = \begin{cases} X(t) & -\tau \leq t \leq \tau \text{ some } \tau \\ 0 & \text{otherwise} \end{cases} \quad (1.13)$$

Then we may write

$$X^*(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(w) e^{iwt} dw$$

$$\text{where } g(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X^*(t) e^{-iwt} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\tau}^{\tau} X(t) e^{-iwt} dt \quad (1.14)$$

Since $X(t)$ may or may not possess a Fourier integral, the limit (as $\tau \rightarrow \infty$) of $|g(w)|^2$ does not exist. But the total energy off by $X(t)$ in the interval $(-\tau, \tau)$ is equal to

$$\int_{-\tau}^{\tau} X^2(t) dt$$

(by the basic laws of electrical circuit theory - Priestley(1981))

$$\text{and } \int_{-\infty}^{\infty} X^2(t) dt = \int_{-\infty}^{\infty} |g(\omega)|^2 d\omega.$$

However under suitable conditions, the limit (as $T \rightarrow \infty$) of $|g(\omega)|^2/2T$ may be finite $(\forall \omega)$.

If this limit exists, it will only be for the specific realization of $X(t)$ considered. In order to construct some quantity representative of the whole stochastic process, consider rather the average of $|g(\omega)|^2/2T$ over the different realizations. Thus we define

$$h(\omega) = \lim_{T \rightarrow \infty} \left\{ E \left[\frac{|g(\omega)|^2}{2T} \right] \right\} \quad \text{say} \quad (1.15)$$

When $h(\omega)$ exists, the average contribution to the total power from components in $X(t)$, with frequencies between ω and $\omega + d\omega$, is $h(\omega)d\omega$.

Definition 1.3

$h(\omega) (= \lim_{T \rightarrow \infty} \left\{ E \left[\frac{|g(\omega)|^2}{2T} \right] \right\})$ is called the (power) spectral density function (power sd) of $X(t)$.

$$\text{If } H(\omega) = \int_{-\infty}^{\infty} h(\omega) d\omega$$

then $H(\omega)$ is the average contribution to the total power from all components with frequencies $\leq \omega$. When $h(\omega)$ exists, $\frac{d}{d\omega} H(\omega) = h(\omega)$.

[Note: if $X(t)$ is deterministic with period 2π , then total energy over $[-\pi, \pi]$ is $\int_{-\pi}^{\pi} X^2(t) dt$.

If we assume $X(t)$ has the Fourier representation given by (1.8), then

$$\int_{-\pi}^{\pi} X^2(t) dt = \left\{ \frac{1}{2} a_0^2 + \sum_{k=1}^{\infty} (a_k^2 + b_k^2) \right\}$$

Define
$$e_k = \begin{cases} a_0^2 & k=0 \\ \frac{1}{2}(a_k^2 + b_k^2) & k=1,2,\dots \end{cases}$$

then total energy over $(-\pi, \pi) = 2\pi \sum_{k=0}^{\infty} e_k$
and the total energy given off per unit time interval

$$= \sum_{k=0}^{\infty} e_k$$

(A plot of e_k against k gives the discrete power spectrum). So we have a breakdown of the energy into contributions made by particular terms.

1.8 SUMMARY

In this chapter we have introduced the concepts of stationarity and two important functions - the spectral density function (sdf) and the autocovariance function (acvf). The importance of these two functions will be discussed in more detail in the next chapter.

We have also seen how any process (whether random or deterministic, periodic or nonperiodic) may be expressed in terms of a Fourier series representation i.e. a sum (or integral) of sine and cosine functions, evaluated over a range of frequencies. We can compare the power of different frequencies by plotting the Fourier transform of the series.

This plot (or power spectrum) gives an idea of the contribution of certain sine and cosine functions (terms) to the overall process

CHAPTER 2: THE WIENER-KHINTCHINE THEOREM AND THE SPECTRAL DENSITY FUNCTION

In the previous chapter we saw two seemingly unrelated functions, the autocovariance function (acvf) and the spectral density function (sdf). In this chapter, we shall prove the existence of most important relationship between the sdf and the acvf. This result is one of the most important (if not the most important) results used when analysing time-series in the frequency domain.

Further results concerning the sdf and acvf will be given, as well as properties and estimation procedures. It must be borne in mind that the objective of all this theory is firstly to help us understand the concepts involved and secondly to take this understanding and these concepts and apply them to our observed time-series and hence formulate a model which "describes" the time-series.

2.1 RELATING THE SDF TO THE ACVF

The relationship between the acvf and sdf can be expressed in the following theorem

Theorem 6

If $X(t)$ is a (zero-mean) stationary process (where t is a continuous time variable) with sdf $h(w)$ and acvf $\gamma(\tau)$ then

$$h(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i w \tau} \gamma(\tau) d\tau \quad (2.1)$$

Definitions 11 and 13 respectively

Proof To prove this theorem we need to make use of a lemma given below

Lemma 1 If $f_1(t)$ and $f_2(t)$ are real-valued functions with Fourier transforms and

$$F_1(w) = \int_{-\infty}^{\infty} e^{-iwt} f_1(t) dt$$

$$\text{and } F_2(w) = \int_{-\infty}^{\infty} e^{-iwt} f_2(t) dt$$

$$\text{then } F_1(w) \overline{F_2(w)} = \int_{-\infty}^{\infty} e^{-iwt} k(t) dt$$

where $\overline{F(w)}$ denotes the complex conjugate of $F(w)$ and

$$k(t) = \int_{-\infty}^{\infty} f_1(x) f_2(x-t) dx$$

(For proof of this lemma, see Priestley(1981), pg. 211).

As a result of this lemma, if $f_1(t) = f_2(t)$, then

$$|F_1(w)|^2 = \int_{-\infty}^{\infty} e^{-iwt} k(t) dt \quad (2.2)$$

$$\text{where } k(t) = \int_{-\infty}^{\infty} f_1(x) f_1(x-t) dx$$

Applying (2.2) to (1.14), gives us

$$|g(w)|^2 = \int_{-\infty}^{\infty} e^{-i w v} \left\{ \int_{-\infty}^{\infty} \frac{X^*(u)}{2} \frac{X(u-v)}{2} du \right\} dv$$

where $X^*(t)$ is as defined in (1.13) and $g(w)$ as in (1.14)

$$\text{So } \frac{|g(w)|^2}{2} = \frac{1}{2} \int_{-\infty}^{\infty} e^{-i w v} (v) dv$$

where
$$\gamma(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X^*(u) X^*(u-v) du$$

Thus

$$B(v) = \lim_{\tau \rightarrow \infty} \left\{ E \left[\frac{g(v)}{2\pi} \right] \right\} = \lim_{\tau \rightarrow \infty} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda v} L(\lambda) d\lambda \right\} \quad (2.3)$$

(by (1.13)).

But

$$\gamma(v) = \begin{cases} \frac{1}{2\pi} \int_{-\infty}^{\infty} X(u) X(u-v) du & |v| \leq 2\pi \\ -(\gamma(-v)) & |v| > 2\pi \\ 0 & |v| \geq 2\pi \end{cases}$$

(by the definition of (1.13)).

Considering $|v| \leq 2\pi$ and $|v| > 2\pi$ separately,

$$B(v) = \begin{cases} \gamma(v) \left(1 - \frac{|v|}{2\pi}\right) & |v| \leq 2\pi \\ 0 & |v| > 2\pi \end{cases}$$

(since $X(t)$ is zero-mean stationary and by the definition of $\gamma(v)$ in section 1.4)

hence

$$B(v) = \lim_{\tau \rightarrow \infty} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(1 - \frac{|v|}{2\pi}\right) e^{-i\lambda v} \gamma(v) d\lambda \right\} \quad (\text{by (2.3)})$$

To evaluate this limit, we consider the limiting form of $\left(1 - \frac{|v|}{2\pi}\right)$ as $\tau \rightarrow \infty$ - it tends to 1 as $\tau \rightarrow \infty$ for all fixed v .

So if $\gamma(v) \rightarrow 0$ fast enough for its Fourier transform to exist, we get the result.

This theorem gives us a very important tool. It provides the condition necessary for $B(v)$ to exist - $X(t)$ should be such

where
$$\bar{\gamma}(v) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(u) X^*(u-v) du$$

Thus

$$h(w) = \lim_{T \rightarrow \infty} \left\{ E \left[\frac{g(w)^2}{2\pi} \right] \right\} = \lim_{T \rightarrow \infty} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-1wv} E(\bar{\gamma}(v)) dv \right\} \quad (2.3)$$

(by (1.15))

But

$$\bar{\gamma}(v) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} X(u) X(u-v) du & |v| \leq \pi \\ 0 & |v| > \pi \end{cases}$$

(by the definition of (1.13))

Considering $|v| \leq \pi$ and $|v| > \pi$ separately,

$$E(\bar{\gamma}(v)) = \begin{cases} \gamma(v) \left(1 - \frac{|v|}{\pi}\right) & |v| \leq \pi \\ 0 & |v| > \pi \end{cases}$$

(since $X(t)$ is zero-mean stationary and by the definition of $\gamma(v)$ in section 1.4)

hence

$$h(w) = \lim_{T \rightarrow \infty} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1 - \frac{|v|}{\pi}\right) e^{-1wv} \gamma(v) dv \right\} \quad (\text{by (2.3)})$$

To evaluate this limit, we consider the limiting form of $\left(1 - \frac{|v|}{\pi}\right)$ as $\pi \rightarrow \infty$ - it tends to 1 as $\pi \rightarrow \infty$ for all fixed v .
So if $\gamma(v) \rightarrow 0$ fast enough for its Fourier transform to exist, we get the result:

This theorem gives us a very important tool. It provides the condition necessary for $h(w)$ to exist - $X(t)$ should be such

that its acvf has a Fourier transform. This Fourier transform exists if $\gamma(\tau)$ is absolutely integrable (see section 1.6), i.e. (more formally)

Theorem 7

$h(w)$ exists $\forall w$ if

$$\int_{-\infty}^{\infty} \gamma(\tau) d\tau < \infty$$

Note that if $X(t)$ is a real-valued process then (2.1) reduces to

$$h(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(w\tau) \gamma(\tau) d\tau \quad (2.4)$$

(since $h(w)$ is an even function of w) and similarly

$$\gamma(\tau) = \int_{-\infty}^{\infty} \cos(w\tau) h(w) dw$$

(since if $\int_{-\infty}^{\infty} \gamma(\tau) d\tau < \infty$

and if $\gamma(\tau)$ is continuous at 0 (and hence, everywhere), we may write

$$\gamma(\tau) = \int_{-\infty}^{\infty} e^{i w \tau} h(w) dw$$

by (1.14) - since $X(t)$ is absolutely integrable).

Thus knowledge of either $h(w)$ or $\gamma(t)$ implies knowledge of the other!

If we define

$$H(w) = \int_{-\infty}^w h(\theta) d\theta \quad (2.5)$$

then

$$H(w) = \int_{-\infty}^w \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(\theta\tau) \gamma(\tau) d\tau \right\} d\theta$$

if $X(t)$ is real-valued (which is the case we shall consider in this study). But $h(w)$ is an even function (since $h(-w) = h(w)$ and $\gamma(\tau)$ is also an even function), so

$$H(w) = \frac{1}{2} \int_{-\infty}^{\infty} \frac{\sin(w\tau)}{\tau} \gamma(\tau) d\tau \quad \forall w > 0$$

$$\text{Since } H(w_2) - H(w_1) = \frac{1}{2} \int_{w_1}^{w_2} h(\theta) d\theta$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} \frac{e^{-iw_2\tau} - e^{-iw_1\tau}}{-i\tau} \gamma(\tau) d\tau$$

(assuming we can interchange the order of integration)

Now $H(0) = \frac{\sigma^2}{2}$ (since $H(0) = \int_{-\infty}^{\infty} h(w)dw = \frac{\sigma^2}{2}$)

but $h(w)$ is an even function so $\int_{-\infty}^{\infty} h(w)dw = 2 \int_0^{\infty} h(w)dw = \frac{\sigma^2}{2}$

and $H(w) = \frac{\sigma^2}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\omega\tau} - 1}{-i\tau} \gamma(\tau) d\tau$

Replace $\sin(\cdot)$ by $\frac{e^{i\theta} - e^{-i\theta}}{2i}$ to get result.

It also follows that

$$(i) \quad H(\infty) = \frac{\sigma^2}{2} \quad (H(0) = \int_{-\infty}^{\infty} h(w)dw)$$

$$\text{and} \quad (ii) \quad H(-\infty) = 0$$

Definition 14

The normalized sdf (or nsdf), $f(w)$, is defined by

$$f(w) = h(w) \frac{\sigma^2}{2} \quad (2.6)$$

As a result of the form of $h(w)$, it has the following properties

$$(i) \quad \int_{-\infty}^{\infty} f(w)dw = 1 \quad \text{i.e. } f(w) \text{ has the form of a pdf (also using the next property)} \quad (2.7)$$

$$(ii) \quad f(w) \geq 0 \quad \forall w \quad (2.8)$$

(since $|g(w)|^2 \geq 0$ ($g(w)$ as in (1.14))

$$\Rightarrow h(w) \geq 0$$

$$\Rightarrow f(w) \geq 0)$$

(iii) If $X(t)$ is real-valued,

$$f(w) = f(-w) \quad (2.9)$$

$$(iv) \quad f(w) = \frac{1}{\pi} \int_0^{\infty} e^{-i\omega\tau} \gamma(\tau) d\tau \quad (2.10)$$

where $\gamma(\cdot)$ is the acrf (autocorrelation function)

$$(\text{hence } \gamma(\cdot) = \int_{-\infty}^{\infty} e^{i\omega\tau} f(w)dw) \quad (2.11)$$

$$(v) \quad F(w) = \int_{-\infty}^w f(\theta) d\theta \quad \text{so}$$

$$f(w) = \frac{d}{dw} F(w) \quad (\text{if it exists}) \quad (2.12)$$

$$\text{and (a) } 0 \leq F(w) \leq 1 \quad \forall w$$

$$(b) \quad F(w) \text{ is a nondecreasing function of } w$$

(vi) If $X(t)$ is real-valued

$$(A) \quad f(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(w\tau) \gamma(\tau) d\tau \quad (2.13)$$

$$(B) \quad \gamma(\tau) = \int_{-\infty}^{\infty} \cos(w\tau) f(w) dw \quad (2.14)$$

$$(c) \quad F(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin(\tau w)}{\tau} \gamma(\tau) d\tau = \frac{1}{2} \quad (2.15)$$

(from the result on the previous page)

In this section we have seen the very important relationship that exists between the acvf and the sdf and we have also provided a condition sufficient for $h(w)$ (and hence $f(w)$) to exist (namely that $\gamma(\cdot)$ is absolutely integrable). The only remaining thing to be proved is that $h(w)$ (or $f(w)$) always exists for a stationary process. This is proved in the next section in the Wiener-Khintchine theorem.

2.2 THE WIENER-KHINTCHINE THEOREM

is a very important theorem and is used to establish the fact that if a process is stationary, $f(w)$, the nsdf, exists provided $F(w)$ is absolutely continuous or equivalently that $\gamma(\cdot)$ (the acrf) satisfies the condition that

$$\int_{-\infty}^{\infty} |\gamma(\tau)| d\tau < \infty$$

($\phi(\tau)$ is absolutely integrable).

If $\phi(\tau)$ does not exist, we are still able to express $\phi(\tau)$ i.t.o.

the definite

$$\phi(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} dF(\omega) \quad (2.16)$$

which is termed a Fourier-Stieltjes (Wiener(1930)) transform

and is defined even when $F(\omega)$ is not differentiable everywhere.)

More formally

Theorem 3 (Wiener-Khintchine (W-K) theorem)

A necessary and sufficient condition for $\phi(\tau)$ to be the a.c.f. of some stochastically continuous process $X(t)$ (see appendix A1), is that there exists a function $F(\omega)$ having the properties of a distribution function on $(-\infty, \infty)$ such that for all τ , $\phi(\tau)$ may be expressed in the form

$$\phi(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} dF(\omega)$$

Proof

(a) Necessity

A theorem by Bochner (1936) states that any positive semi-definite function, which is continuous everywhere, must have a Fourier-Stieltjes transform of the form (2.16), with $F(\omega)$ having the properties of a distribution function. $\phi(\tau)$ is continuous by the assumption that $X(t)$ is stochastically continuous ($X(t)$ is stochastically continuous at $t = t_0$ if

$\phi(r)$ is positive semi-definite in the sense that for any time points $r_1, r_2, \dots, r_n$ and real $Z_1, \dots, Z_n$

$$\sum_{i=1}^n \sum_{j=1}^n \phi(r_i - r_j) Z_i Z_j \geq 0$$

$\phi(r)$ any function

$$\lim_{t \rightarrow t_0} \text{l.i.m.} E \{ (X(t) - X(t_0))^2 \} = 0$$

- see appendix A1 for a definition of l.i.m. as well as for other stochastic limiting operations) and $\Gamma(\cdot)$ is positive semi-definite by defn. 12, hence the W-K theorem is a special case of Bochner's theorem.

A restricted form of this result is shown in theorem 5 (section 2.1) and we can consider the more general case. Using the proof of theorem 6, write

$$E \left\{ \frac{|g(w)|^2}{2\pi \sigma^2} \right\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(v) e^{-i w v} dv \quad (2.17)$$

where

$$\chi(v) = \begin{cases} \Delta(v) \left[1 - \frac{|v|}{2\pi} \right] & |v| \leq 2\pi \\ 0 & |v| > 2\pi \end{cases}$$

Let $f(w) = E \left\{ |g(w)|^2 \right\}$ and multiply both sides of (2.17) by $e^{i w t} \left(1 - \frac{|w|}{2\pi} \right)$. Integrating from $w = -\infty$ to ∞ gives

$$\int_{-\infty}^{\infty} \left(1 - \frac{|w|}{2\pi} \right) f(w) e^{i w t} dw = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(v) \left\{ \frac{\sin(\frac{1}{2}(v-t))}{\frac{1}{2}(v-t)} \right\}^2 dv \quad (2.18)$$

(since $\int_{-\infty}^{\infty} \left(1 - \frac{|w|}{2\pi} \right) e^{-i w (v-t)} dw = \delta \left\{ \frac{\sin(\frac{1}{2}(v-t))}{\frac{1}{2}(v-t)} \right\}$)

Now $f(w)$ is nonnegative for all w , so if

$$\left(1 - \frac{|w|}{2\pi} \right) f(w)$$

is suitably scaled, it may be regarded as a pdf over the range $(-\infty, \infty)$. As a result, the LHS of (2.18) has the properties of

an arbitrary constant ($\neq 0$)

a (scaled) characteristic function

As $\epsilon \rightarrow 0$,

$$\frac{\epsilon}{2\pi} \int_{-\infty}^{\infty} \frac{\sin(\frac{\epsilon}{2}(v-t))}{\frac{\epsilon}{2}(v-t)} dt$$

behaves like a Dirac delta function, $\text{Dir}(v-t)$.

If $\phi(t)$ is any function continuous at $t = 0$, the Dirac delta function, $\text{Dir}(t)$, is such that

$$\int_{-\infty}^{\infty} \text{Dir}(t)\phi(t)dt = \phi(0)$$

[Outline proof: Since the Dirac delta function $\text{Dir}(v)$ is usually defined by

$$\text{Dir}(v) = \begin{cases} 0 & v \neq 0 \\ \infty & v = 0 \end{cases}$$

such that $\int_{-\infty}^{\infty} \text{Dir}(v)dv = 1$, we then need to show

$$\lim_{\epsilon \rightarrow 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin(\frac{\epsilon}{2}(v-t))}{\frac{\epsilon}{2}(v-t)} dt = \text{Dir}(v)$$

$$\text{i.e.} \quad \lim_{\epsilon \rightarrow 0} \frac{2\sin^2(\frac{\epsilon}{2}(v-t))}{\epsilon^2} = \text{Dir}(v)$$

For $v \neq 0$, this is certainly true (since the sine function is bounded). For $v = 0$, it may be shown that

$$\lim_{\epsilon \rightarrow 0} \frac{2\sin^2(\frac{\epsilon}{2}(v-t))}{\epsilon^2} = 0$$

So as $t \rightarrow \infty$, the RHS of (2.18) converges (uniformly) to $\chi(t)$ (provided $\chi(t)$ is continuous).

Hence $\chi(t)$ is proportional to the characteristic function and since $\chi(0) = \phi(0) = 1$, $\chi(t)$ is (exactly) a characteristic function.

Also, as $t \rightarrow \infty$, $\chi(t)$ converges (uniformly) to $\phi(t)$ and so $\phi(t)$ is also a characteristic function. Hence there must exist some distribution function $F(w)$ which is such that $\phi(t)$ is its Fourier-Stieltjes transform.

(b) Sufficiency

Let Y be a random variable having $F(y)$ as its distribution function satisfying (2.16) and let Z be a random variable having a Uniform $(-T, T)$ distribution (independent of Y).

Define

$$X(t) = e^{i(Z + Yt)} \quad (2.19)$$

$$E(X(t)) = E(e^{iZ})E(e^{iYt})$$

$$\text{but } E(e^{iZ}) = 0$$

$$\therefore E(X(t)) = 0 \quad \forall t$$

$$\begin{aligned} \text{and } V(X(t)) &= E[|X(t)|^2] \\ &= E[1] \quad (\text{by defn.}) \\ &= 1 \end{aligned}$$

$$(\text{since } |X(t)|^2 = \cos^2(Z + Yt) + \sin^2(Z + Yt) = 1)$$

Also,

$$\gamma_X(v) = E[\overline{X(t)}X(t+v)] = E[e^{iYv}]$$

and

$$E\{e^{iYv}\} = \int_{-\infty}^{\infty} e^{iwy} dF(w) \\ = \phi(v) \quad wv$$

So $X(t)$ defined by (2.19) has the given function $\phi(\tau)$ as its acrf.

2.3 DISCRETE PROCESSES

Without actually having stated it, the processes considered so far in this chapter have all been continuous ones i.e. $X(t)$ has been defined for a continuous-time variable t . However our main interest lies in discrete time-series and hence processes which are defined only for integer t values i.e. $X(t)$, $t = 0, \pm 1, \pm 2, \dots$

Denoting these processes by $\{X_n\}$, an analogue of the Wiener-Khintchine theorem is required (since we can no longer talk i.t.o. integrals and continuous acrf's, but rather we are forced to consider summations and discrete acrf's). This analogue is known as Wold's theorem.

Theorem 2 (Wold's theorem)

A necessary and sufficient condition for the sequence

$\phi(r)$; $r = 0, \pm 1, \dots$ to be the acrf for some discrete time process

$\{X_n$; $t = 0, \pm 1, \dots\}$, is that there exist a function $F(w)$ having the properties of a distribution function on $(-\pi, \pi)$ such that

$$\phi(r) = \int_{-\pi}^{\pi} e^{irw} dF(w) \quad r = 0, \pm 1, \dots \quad (2.20)$$

Proof

(a) Necessity

Plot $\phi(r)$ against r and join consecutive points by straight lines; this will give $\phi^*(\tau)$, say, $\forall$, and in particular

$$\phi^*(\tau) = (s+1-\tau)\phi(s) + (\tau-s)\phi(s+1) \quad \text{where } s \leq \tau < s+1$$

$$\text{and } \phi^*(-\tau) = \phi^*(\tau) \quad \forall.$$

Then $\phi^*(\tau)$ is continuous everywhere and Priestley (1963) has shown that there exists a stationary continuous-time process, which has $\phi^*(\tau)$ as its a.c.f. Hence by the W-K theorem, there exists a nondecreasing function, $F^*(w)$ say, such that

$$\phi^*(\tau) = \int_{-\infty}^{\infty} e^{i w \tau} dF^*(w).$$

This is true $\forall$ and so is certainly true when τ is an integer, r say.

So

$$\phi(r) = \int_{-\infty}^{\infty} e^{i w r} dF^*(w) \quad \text{for } r = 0, \pm 1, \dots$$

hence

$$\begin{aligned} \phi(r) &= \sum_{n=-\infty}^{\infty} \frac{(2\pi+1)^n}{(2\pi+1)^n} e^{i w r} dF^*(w) \\ &= \int_{-\infty}^{\infty} e^{i w r} \left\{ \sum_{n=-\infty}^{\infty} dF^*(w+2\pi n) \right\} \end{aligned}$$

$$\text{since } e^{i(w+2\pi n)r} = e^{i w r}$$

Let $dF(w) = \int_s dF^*(w + 2s)$ (with $F(-) = 0$) and then (2.20) follows.

The (b) sufficiency condition is similar to that of the W-K theorem.

Comparing (2.20) with (1.5) and (1.6) and assuming $F(w)$ is differentiable everywhere (thus $f(w) = \frac{d}{dw} F(w)$)

then

$$c(r) = \int_{-\pi}^{\pi} e^{iwr} f(w) dw. \quad (2.21)$$

$c(r) 2\pi$ is the coefficient of e^{-iwr} in a Fourier series expansion of $f(w)$,

hence

$$f(w) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c(r) e^{-iwr} \quad -\pi < w < \pi \quad (2.22)$$

Note that if X_t is real-valued,

then

$$f(w) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{r=1}^{\infty} c(r) \cos(rw) \quad (2.23)$$

and

$$F(w) = \int_{-\pi}^w f(u) du = \frac{u+w}{2\pi} + \frac{1}{\pi} \sum_{r=1}^{\infty} c(r) \frac{\sin(rw)}{r} \quad -\pi < w < \pi \quad (2.24)$$

Thus if we have a (stationary) discrete time-series we can compute the theoretical acrf or the theoretical sdf of the process, provided we have knowledge of one or the other. The usefulness of this technique will be shown at a later stage.

2.4 ESTIMATION OF THE ACVF (AND HENCE ACRF AND SDF)

In general, we only have a finite realization of the time series i.e. a sample from a stationary process $\{X_t\}$ of size N , say. Suppose this sample can be represented by $x_1, x_2, \dots, x_N$. Now since we only have a sample, we cannot compute the theoretical acvf or sdf, but rather only estimates of parts thereof.

Forming the observations $x_1, x_2, \dots, x_N$ into (non-independent) pairs

$$(x_1, x_{r+1}), (x_2, x_{r+2}), \dots, (x_{N-r}, x_N) \quad r > 0$$

then in general, these pairs will come from different bivariate distributions. However since $\{X_t\}$ is (second-order) stationary, all these bivariate distributions will have the same covariance, $\gamma(r)$ say.

Defining $\{z_1\}$ to correspond to $\{x_1, x_2, \dots, x_{N-r}\}$ and $\{y_1\}$ to $\{x_{r+1}, x_{r+2}, \dots, x_N\}$, then it is possible to estimate $\gamma(r)$ by

$$\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y}) \quad (n = N - r)$$

But it is senseless using two separate estimates $\bar{z}$ and $\bar{y}$ since $\{z_1\}$ and $\{y_1\}$ both came from the same distribution (assuming second order stationarity) - so estimate $\bar{z}$ and $\bar{y}$ by $\bar{X}$ (the sample mean of $x_1, x_2, \dots, x_N$)

Thus

$$\hat{\gamma}(r) = \frac{1}{N-r} \sum_{t=1}^{N-r} (x_t - \bar{X})(x_{t+r} - \bar{X}) \quad (2.25)$$

2.4 ESTIMATION OF THE ACVF (AND HENCE ACRF AND SDF)

In general, we only have a finite realization of the time series
i.e. a sample from a stationary process $\{X_t\}$ of size N , say.

Suppose this sample can be represented by $x_1, x_2, \dots, x_N$. Now
since we only have a sample, we cannot compute the theoretical acvf
or sdf, but rather only estimates of parameters thereof.

Forming the observations $x_1, x_2, \dots, x_N$ into (non-independent) pairs

$$(x_1, x_{r+1}), (x_2, x_{r+2}), \dots, (x_{N-r}, x_N) \quad r \geq 0$$

then in general, these pairs will come from different bivariate
distributions. However since $\{X_t\}$ is (second-order) stationary,
all these bivariate distributions will have the same covariance,
 $\gamma(r)$ say.

Defining $\{z_i\}$ to correspond to $\{x_1, x_2, \dots, x_{N-r}\}$ and $\{y_i\}$ to
 $\{x_{r+1}, x_{r+2}, \dots, x_N\}$, then it is possible to estimate $\gamma(r)$ by

$$\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y}) \quad (n = N - r)$$

But it is senseless using two separate estimates $\bar{z}$ and $\bar{y}$ since
 $\{z_i\}$ and $\{y_i\}$ both came from the same distribution (assuming
second order stationarity) - so estimate $\bar{z}$ and $\bar{y}$ by $\bar{x}$ (the sample
mean of $x_1, x_2, \dots, x_N$).

Thus

$$\hat{\gamma}(r) = \frac{1}{N-r} \sum_{i=1}^{N-r} (x_i - \bar{x})(x_{i+r} - \bar{x}) \quad (2.25)$$

$$= c_r \text{ say } (r = 0, \pm 1, \dots, \pm(N-1))$$

Since $v(r)$ is an even function, it is reasonable to make $\gamma(r)$ and even function also

$$\text{i.e. } \gamma(-r) = \gamma(r) \quad \forall \text{ integer } r$$

Properties of c_r :

$$(i) \quad \text{If } \mu \text{ is replaced by } \bar{X} \text{ in (2.25) (to give a new estimate, } c_r^* \text{ say)} \\ \text{then } E(c_r^*) = \gamma(r), \quad \forall r \quad (2.26)$$

$$(ii) \quad \text{If } \bar{X} \text{ is not replaced by } \dots \text{ then}$$

$$E(c_r) = \gamma(r) - V(\bar{X}) \\ = \gamma(r) - \frac{2r^2}{N} f(0) \quad (2.27)$$

Outline proof: if X_t has sdf $f(w)$, then

$$f(w) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma(r) e^{-iwr} \quad (\text{by definition})$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{r=-(N-1)}^{N-1} \gamma(r) = \int_{-\infty}^{\infty} \gamma(r) dr = 2\pi f(0)$$

$$\text{Also, } V\left\{ \sum_{i=1}^p a_i x_i \right\} = \sum_{i=1}^p \sum_{j=1}^p a_i a_j r_{ij} \quad \text{where } r_{ij} = \text{cov}(x_i, x_j) \\ p \text{ any integer } > 0$$

$$\text{so } V(\bar{X}) = \frac{1}{N^2} \sum_{r=-N}^N (N - |r|) \gamma(r)$$

$$= 2 f(0) \frac{2}{N} \quad (\text{as } N \rightarrow \infty)$$

since

$$= c_r \text{ say } (r = 0, \pm 1, \dots, \pm(N-1))$$

Since $\gamma(r)$ is an even function, it is reasonable to make $\bar{\gamma}(r)$ and even function also

$$\bar{\gamma}(-r) = \bar{\gamma}(r) \quad \forall \text{ integer } r.$$

Properties of c_r^*

(i) If μ is replaced by $\bar{X}$ in (2.25) (to give a new estimate, c_r^* say) then $E(c_r^*) = \gamma(r)$, $\forall r$ (2.26)

(ii) If $\bar{X}$ is not replaced by μ , then

$$\begin{aligned} E(c_r^*) &= \gamma(r) - V(\bar{X}) \\ &= \gamma(r) - \frac{2^{-1}}{N} f(0) \end{aligned} \quad (2.27)$$

Outline proof: if $\{X_t\}$ has sdf $f(w)$, then

$$f(w) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma(r) e^{-iwr} \quad (\text{by definition})$$

and

$$\frac{1}{N} \sum_{r=-(N-1)}^{(N-1)} \left(1 - \frac{|r|}{N}\right) \gamma(r) = \sum_{r=0}^{N-1} \gamma(r) = 2^{-1} f(0)$$

Also, $V \left\{ \sum_{i=1}^p a_i x_i \right\} = \sum_{i=1}^p \sum_{j=1}^p a_i a_j \gamma_{ij}$ where $\gamma_{ij} = \text{cov}(x_i, x_j)$
 p any integer > 0

$$\begin{aligned} V(\bar{X}) &= \frac{1}{N} \sum_{r=-(N-1)}^{(N-1)} \left(1 - \frac{|r|}{N}\right) \gamma(r) \\ &= 2^{-1} f(0) \frac{2}{N} \quad (\text{as } N \rightarrow \infty) \end{aligned}$$

since

(a) if $\lim_{N \rightarrow \infty} \left\{ \sum_{|r| < N} \left(1 - \frac{|r|}{N}\right) u(r) \right\}$ is finite

then $V(\tilde{X}) = 0$ as $N \rightarrow \infty$

and

(b) if X_t has a purely continuous spectrum with nsdf $f(\omega)$ then

$$f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} u(r) e^{-i\omega r}$$

and it may be shown that

$$\lim_{N \rightarrow \infty} \left\{ \sum_{|r| < N} \left(1 - \frac{|r|}{N}\right) u(r) \right\} = \sum_{-\infty}^{\infty} u(r) \\ = 2\pi f(0)$$

Now

$$\begin{aligned} & \frac{N - |r|}{N} (X_t - \bar{X})(X_{t+r} - \bar{X}) \\ &= \frac{N - |r|}{N} \left\{ (X_t - \bar{X}) + (\bar{X} - \bar{X}) \right\} \left\{ (X_{t+r} - \bar{X}) + (\bar{X} - \bar{X}) \right\} \\ &= \frac{N - |r|}{N} \left\{ (X_t - \bar{X})(X_{t+r} - \bar{X}) + (N - |r|)(\bar{X} - \bar{X})^2 \right\} \\ &= (N - |r|) u_1 + (N - |r|)(\bar{X} - \bar{X})^2 \end{aligned}$$

hence result (a)

Another estimate for (a). In fact the more widely used one,

It may also be shown (see appendix A2) that

(a) $V(c'_r)$ is of order $\frac{1}{N}$, and

(b) $V(c_r)$ is of order $\frac{1}{N - |r|}$

Thus using (a) and (b), as $|r| \rightarrow N-1$, $V(c'_r)$ remains constant

So if $|r|$ is small relative to N , there is little difference between the estimates c_r and c'_r . However if $|r|$ is large relative to N , the larger bias of c_r is compensated for by its smaller variance. So considering the merit of the positive semidefinite property, it seems preferable to use c'_r instead of c_r as an estimate of $\gamma(r)$.

(Jenkins and Watts (1968) conjecture that c_r generally has a higher MSE).

2.5 ESTIMATION OF THE ACRF

Since $\gamma(r) = \gamma(r)(0)$, obvious estimates for $\gamma(r)$ will be

$$\begin{aligned} (i) \quad \hat{\gamma}_1(k) &= r_k \\ &= c_k c_0 \end{aligned} \tag{2.32}$$

and

$$\begin{aligned} (ii) \quad \hat{\gamma}_2(k) &= r'_k \\ &= c'_k / c'_0 \end{aligned} \tag{2.33}$$

where $c'_0 = c_0$ $k = 0, \pm 1, \dots, \pm(N-1)$

The preferable estimate (for the same reasons as in the previous section) is r'_k

The variance and covariance of acf estimators are given in appendix A3.

2.6 SUMMARY

In this chapter we have seen the relationship that exists between the acvf and the sdf. Thus knowledge of one implies knowledge of the other (given certain conditions). We have also seen the necessary conditions for the existence of the sdf (or nsdf). These conditions were given in the Wiener-Khintchine theorem which handles the concept of the spectrum in the continuous-time domain.

For discrete-time processes, a different theorem was needed, namely Wold's theorem, and there again sufficient and necessary conditions were given for a sequence $\{r(r)\}$ to be the acrf of a discrete-time process $\{X_t\}$.

Finally, estimates of the acvf and acrf, based on a sample $x_1, x_2, \dots, x_N$ (from a discrete-time process), were formulated. The variance and covariances of these estimates are given in appendices A2 and A3.

In the next chapter we shall see some of the practical uses of these tools.

CHAPTER 3: PERIODOGRAM ANALYSIS

3.1 INTRODUCTION

Some of the more important aspects of the theory seen so far can be summed up as follows:

- (a) Any periodic function $g(x)$ can be written i.t.o. a Fourier series i.e.

$$g(x) = \sum_{r=0}^{\infty} \left\{ a_r \cos\left(\frac{2\pi r x}{h}\right) + b_r \sin\left(\frac{2\pi r x}{h}\right) \right\} \quad (3.1)$$

(h = period of $g(x)$) provided $g(x)$ satisfies certain conditions (see section 1.5, page 9).

Rescaling (3.1), we may write

$$g(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos(w_k x) + b_k \sin(w_k x)) \quad (3.2)$$

where $w_k = 2\pi k / h$

- (b) For any stationary stochastic process with acvf $\gamma(k)$, there exists a monotonically increasing function $F(w)$ such that

$$\gamma(k) = \int_{-\pi}^{\pi} e^{i w k} dF(w)$$

If

$$f(w) = \frac{d}{dw} F(w) \quad \text{exists } \forall w,$$

then

$$\gamma(k) = \int_{-\pi}^{\pi} e^{i w k} f(w) dw. \quad (3.3)$$

Periodogram analysis is involved with estimating a_k , b_k and ω_k in (3.2) i.e. if we have a set of data $x_0, x_1, \dots, x_{N-1}$, we would like to write

$$g(t) = x_t \quad \forall t$$

and then estimate the parameters in representation (3.2).

On the other hand, spectral analysis is concerned with estimating $f(\omega)$ in (3.3), based on sample estimates of the $\gamma(k)$.

This chapter is devoted to obtaining estimates of the form (3.2) i.e.

$$x_t = \sum_{k=1}^{\infty} (a_k \cos(\omega_k t) + b_k \sin(\omega_k t)) \quad (3.4)$$

$$t = 0, 1, \dots, N-1$$

In chapter four we shall see the relationship between periodogram analysis and spectral analysis.

3.2 HARMONIC ANALYSIS

The simplest periodic data are of the form

$$x_t = R \cos(\omega t - \phi) \quad (\text{all } t)$$

A sinusoid of frequency ω (or period $\frac{2\pi}{\omega}$) may be written as

$$g(t) = R \cos(\omega t + \phi)$$

where

R = amplitude

ϕ = phase

Now if we consider a daily time-series of 600 observations, say, which when plotted contains 20 equally spaced peaks, we might expect any periodicity to have a period of around

$$\frac{600}{20} = 30 \text{ days.}$$

Thus the t^{th} data point should contain a component of the form

$$R \cos(\omega t + \phi) \text{ where } \omega = \frac{2\pi}{30}$$

A possible model for the data could be

$$x_t = \mu + R \cos(\omega t + \phi) + \epsilon_t \quad t = 0, 1, \dots, 599$$

(where ϵ_t = experimental error and $E(\epsilon_t) = 0 \quad \forall t$)

Initially, if ω is kept fixed, the unknown parameters are μ, R and ϕ . For least squares estimation we need to find μ, R and ϕ to minimize

$$S(R, \mu, \phi) = \sum_{t=0}^{599} (x_t - \mu - R \cos(\omega t + \phi))^2$$

i.e. the least squares estimators (LSE's)

Reparametrizing $\mu + R \cos(\omega t + \phi)$ to the standard linear form gives

$A \cos(\omega t) + B \sin(\omega t)$ (i.e. a transformation to polar co-ordinates)

where $A = R \cos \phi$

and $B = -R \sin \phi$

and keeping $\omega = 0$ (temporarily) and ω fixed, it is a simple

matter to calculate the LSE's of A and B (and hence R and ϕ)

i.e.

$$\frac{\partial S}{\partial A} = -2 \sum_{t=0}^{599} \cos(\omega t) (x_t - A \cos(\omega t) - B \sin(\omega t))$$

$$\frac{\partial S}{\partial B} = -2 \sum_{t=0}^{599} \sin(\omega t) (x_t - A \cos(\omega t) - B \sin(\omega t))$$

Equate these two equations to zero and solve simultaneously. Thus

$$\hat{A} = \frac{1}{N} \left\{ \sum_{t=1}^N \cos(wt) \sum_{t=1}^N (\sin(wt))^2 - \sum_{t=1}^N \sin(wt) \sum_{t=1}^N \cos(wt) \sin(wt) \right\} \quad (3.5)$$

$$\hat{B} = \frac{1}{N} \left\{ \sum_{t=1}^N \sin(wt) \sum_{t=1}^N (\cos(wt))^2 - \sum_{t=1}^N \cos(wt) \sum_{t=1}^N \cos(wt) \sin(wt) \right\} \quad (3.6)$$

$$\text{where } \Delta = \left(\sum_{t=1}^N (\cos(wt))^2 - \sum_{t=1}^N (\sin(wt))^2 - (\sum_{t=1}^N \cos(wt) \sin(wt))^2 \right)$$

Finding R and $\hat{\phi}$ is easy:

$$\hat{\phi} = \begin{cases} \arctan(-B/A) & A \neq 0 \\ \arctan(-B/A) + \pi & A < 0, B \leq 0 \\ \arctan(-B/A) - \pi & A < 0, B > 0 \\ -\pi/2 & A = 0, B > 0 \\ \pi/2 & A = 0, B < 0 \\ \text{arbitrary} & A = 0, B = 0 \end{cases} \quad (3.7)$$

(since R is nonnegative and the basic equation for $\hat{\phi}$ is $\tan \phi = -B/A$)

$$\text{and } R = (\hat{A}^2 + \hat{B}^2)^{1/2} \quad (3.8)$$

Useful approximations to these solutions may be found by noting that

$$\sum_{t=1}^N (\cos(wt))^2 = \frac{N}{2} \left\{ 1 + D_N(2w) \cos((N-1)w) \right\}$$

$$\sum_{t=1}^N \cos(wt) \sin(wt) = \frac{N}{2} D_N(2w) \sin((N-1)w)$$

$$\sum_{t=1}^N (\sin(wt))^2 = \frac{N}{2} \left\{ 1 - D_N(2w) \cos((N-1)w) \right\}$$

where
$$D_N(w) = \frac{\sin(\frac{Nw}{2})}{N \sin(\frac{w}{2})} \quad (3.9)$$

which is a version of the Dirichlet kernel (see Titchmarsh(1939)).
See, for example, Bloomfield (1976).

Note that $D_N(\frac{2k\pi}{N}) = 0 \quad \forall$ integer $k \neq 0$ or a multiple of N

and
$$ND_N(w) \leq \frac{1}{\sin(\frac{w}{2})}$$

any terms in the LSE's involving D_N are, for large N and w not too close to zero, always small when compared with $\frac{N}{1}$

By omitting such terms, and allowing μ to be nonzero, we obtain

$$N\bar{\mu} = \sum x_t = N\bar{X}$$

$$\frac{NA}{2} = \sum x_t \cos(wt)$$

$$\frac{NB}{2} = \sum x_t \sin(wt)$$

Thus

$$\bar{\mu} = \bar{X} \quad (3.10)$$

$$A = \frac{2}{N} \sum (x_t - \bar{X}) \cos(wt) \quad (3.11)$$

$$B = \frac{2}{N} \sum (x_t - \bar{X}) \sin(wt) \quad (3.12)$$

Similarly

$$R = (A^2 + B^2)^{\frac{1}{2}} \quad (3.13)$$

and $\hat{\mu}$ is in terms of A and B .

Extending this procedure to take w into account, we now need to minimize

$$S(\mu(w), A(w), B(w), w) \text{ w.r.t. } \mu, A, B \text{ and } w$$

or equivalently minimize

$$S(0(\omega), A(\omega), B(\omega), \omega) \quad \text{w.r.t. } \omega$$

The residual sum of squares (RSS) is

$$S(\tilde{x}(\omega), \tilde{A}(\omega), \tilde{B}(\omega), \omega) \quad \text{and}$$

$$\begin{aligned} \text{RSS} &= S(\tilde{x}, \tilde{A}^*, \tilde{B}^*, \omega) \\ &= S(\tilde{x}, 0, 0, \omega) - \frac{N}{2} \left\{ \tilde{A}^{*2} + \tilde{B}^{*2} \right\} \end{aligned}$$

(since the sum of squares of residuals, RSS, may be written

$$\text{RSS} = \text{sum of squares of (original data - mean)}$$

$$= \text{sum of squares associated with frequency } \omega$$

$$\begin{aligned} &= \sum (x_t - \bar{x})^2 - \frac{1}{2} \left\{ \left(\sum (x_t - \bar{x}) \cos(\omega t) \right)^2 \right. \\ &\quad \left. + \left(\sum (x_t - \bar{x}) \sin(\omega t) \right)^2 \right\} \\ &= \sum (x_t - \bar{x})^2 - \frac{1}{2} \left\{ \left(\sum (x_t - \bar{x}) \cos(\omega t) \right)^2 + \left(\sum (x_t - \bar{x}) \sin(\omega t) \right)^2 \right\} \quad \triangleq \text{as before} \\ &= S(\tilde{x}, 0, 0, \omega) - \frac{N}{2} \tilde{R}^{*2} \end{aligned}$$

The 'best' value of ω , is that value which minimizes the RSS, or equivalently, maximizes $\tilde{R}^{*2}$ (i.e. $(\tilde{R}^*(\omega))^2$)

If the residuals from the model are plotted,

further periodicity may be apparent. If this is so, remove any periodicity from the residuals in exactly the same manner.

So the final form of the model is

$$x_t = \mu + \sum_{j=1}^M (A_j \cos(\omega_j t) + B_j \sin(\omega_j t)) + \tilde{x}_t \quad (3.14)$$

(where ω_j is an integer ≥ 1) $t = 0, \dots, N-1$

4.3 ALIASING

What is the range of ω ? Units of frequency are radians per unit time and hence nonnegative, thus $\omega \geq 0$.

Can we restrict ω further?

Since $\cos(w) = \cos(-w)$

and $-\sin(w) = \sin(-w)$,

w and $-w$ are indistinguishable and are said to be aliases of each other. If the sampling interval of our data is 1 (usually 1), then the t^{th} observation is made at time t . Suppose that the data consist of a pure cosine wave i.e.

$$\cos(wt) = x_t.$$

As w increases from 0, the wave oscillates more and more rapidly until at $w = \pi$, we have $x_t = (-1)^t$ (the most rapid oscillation).

Increasing w further - say

$$\frac{\pi}{2} < w < \frac{3\pi}{2}$$

and letting $w' = \frac{2\pi}{1} - w$

$$\begin{aligned} \text{then } x_t &= \cos(wt) \\ &= \cos(w't) \end{aligned}$$

Similarly, $\sin(wt) = \sin(w't)$

So w and w' are aliases of each other, and so every frequency outside the range $0, \pi$ has an alias in this range (termed its principal alias).

- is known as the Nyquist frequency (and equals half the sampling rate).

3.4 THE DISCRETE FOURIER TRANSFORM (DFT)

As already noted in section 3.1, by the theory of chapter one it is possible to express any function (periodic or nonperiodic) i.t.o. a Fourier series (provided the function satisfies the Dirichlet conditions - see section 1.5). So if $x(t)$ is our function, we may

Since $\cos(w) = \cos(-w)$

and $-\sin(w) = \sin(-w)$,

w and $-w$ are indistinguishable and are said to be aliases of each other. If the sampling interval of our data is δ (usually 1), then the t^{th} observation is made at time $t\delta$. Suppose that the data consist of a pure cosine wave i.e.

$$\cos(wt\delta) = x_t.$$

As w increases from 0, the wave oscillates more and more rapidly until at $w = \frac{\pi}{\delta}$, we have $x_t = (-1)^t$ (the most rapid oscillation).

Increasing w further - say

$$\frac{\pi}{\delta} + w = \frac{2\pi}{\delta}$$

and letting $w' = \frac{2\pi}{\delta} - w$

$$\begin{aligned} \text{then } x_t &= \cos(wt\delta) \\ &= \cos(w't\delta) \end{aligned}$$

Similarly, $\sin(wt\delta) = \sin(w't\delta)$

So w and w' are aliases of each other, and so every frequency outside the range $[0, \frac{\pi}{\delta}]$ has an alias in this range (termed its principal alias).

$\frac{\pi}{\delta}$ is known as the Nyquist frequency (and equals half the sampling rate).

3.4 THE DISCRETE FOURIER TRANSFORM (dFt)

As already noted in section 3.1, by the theory of chapter one it is possible to express any function (periodic or nonperiodic) i.t.o. a Fourier series (provided the function satisfies the Dirichlet conditions - see section 1.5). So if $g(t)$ is our function, we may

write

$$g(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(w_k t) + b_k \sin(w_k t)) \quad \text{say.}$$

Now this can be written in the computationally more convenient form

$$\sum_{k=-\infty}^{\infty} J_k e^{i w_k t} \quad (\text{see section 1.1})$$

where $w_k = 2\pi k/h$ (if our function has period h) and

$$J_k = \frac{1}{2}(a_k - i b_k) \quad (3.15)$$

$$J_{-k} = \bar{J}_k \quad \text{the complex conjugate.}$$

How do we estimate a_k and b_k ?

Well, in sections 1.5 (proof of Fourier's theorem) and 1.6 (the nonperiodic case) it was shown that irrespective of whether $g(t)$ is periodic or nonperiodic,

$$J_k = \frac{1}{h} \int_{-h/2}^{h/2} g(t) e^{-i w_k t} dt \quad (3.16)$$

(if $g(t)$ has period h)

$$\text{and } J_k = \int_{-\infty}^{\infty} g(t) e^{-2\pi i w_k t} dt = J(w_k) \quad (3.17)$$

(if $g(t)$ is nonperiodic)

suppose we assume $g(t)$ is periodic. This makes more sense, since we shall hopefully be able to devise some test to determine whether the contribution of terms associated with w_k is significantly different from zero; if all these contributions are not significant, then we may conclude $g(t)$ is nonperiodic. Furthermore suppose the period $h = 2\pi$ then by the definition of (3.15) and also by section 1.5

$$\left. \begin{aligned} a_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} g(t) \cos(kt) dt & k = 0, 1, 2, \dots \\ b_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} g(t) \sin(kt) dt & k = 1, 2, \dots \end{aligned} \right\} \quad (3.18)$$

(and $g(t) = x_k$ for $t = t_k$)

Now suppose our observed time-series is only defined at $2N$ (for computational convenience) equally spaced points. Define $t_k = \frac{k\pi}{N}$ for $k = 0, 1, \dots, 2N-1$ (say instead of $k = 1, 2, \dots, 2N$ for convenience)

$$g(t_k) = \bar{w}\left(\frac{k\pi}{N}\right)$$

$$= \sum_{j=-(N-1)}^{N-1} w_j e^{i j t_k} = \frac{J_N e^{i N t_k} - J_{-N} e^{-i N t_k}}{2} \quad (3.19)$$

where $w_j = j$

Rewriting J_N and J_{-N} i.t.o. (3.15) (expressing the exponential terms i.t.o. cosine and sine functions) and multiplying both sides of (3.19) by $e^{-i w_l t_k}$ and summing over k , we obtain

$$\sum_{k=0}^{2N-1} e^{-i w_j t_k} g(t_k) = \sum_{j=-(N-1)}^{N-1} \sum_{k=0}^{2N-1} J_j e^{i(w_j - w_j) t_k} \\ + \frac{a_N}{2} \sum_{k=0}^{2N-1} (-1)^k e^{-i w_j t_k} \quad j = 0, \pm 1, \dots, \pm N \quad (3.20)$$

However the last term in (3.20) vanishes ($\forall w_j$) since it is a geometric sequence and equals

$$\frac{a_N}{2} \left(\frac{1 - e^{-i(2\pi/N)}}{1 - e^{-i(2\pi/N)}} \right) \quad (3.20a)$$

and $e^{-i2\pi} = 1$ when 2π is an integral multiple of 2π .

Also, if $w_j - w_k = w^*$ say

$$\text{then } \sum_k e^{i w^* t_k} = 0 \quad \text{if } w^* \neq 0 \quad (\text{using (3.20a)})$$

$$\text{so } \sum_k e^{-i w_j t_k} g(t_k) = \sum_k J_j = 2N(J_j)$$

$$\text{i.e. } J_j = \frac{1}{2N} \sum_k e^{-i j t_k} g(t_k) \quad \text{since } w_j = j \quad (3.21) \\ j = 0, \pm 1, \dots, \pm(N-1)$$

and hence

$$a_j = \frac{1}{N} \sum_k \cos\left(\frac{k j}{N}\right) g(t_k) \quad j = 0, \dots, N-1 \quad (3.22)$$

$$b_j = \frac{1}{N} \sum_k \sin\left(\frac{k j}{N}\right) g(t_k) \quad j = 1, \dots, N-1 \quad (3.23)$$

where $b_0 = 0$

$$\text{and } (-1)^N a_N = a_0 + 2(a_1 - a_2 + \dots + (-1)^{N-1} a_{N-1})$$

$$\dots + (-1)^N a_{N-1} = 0 \quad (\text{by definition of (3.15)})$$

$$\text{i.e. } J_j = \frac{1}{2N} \sum_k g(t_k) e^{-i w_j t_k}$$

$$a_j = \frac{2}{2N} \sum_{k=0}^{N-1} g(t_k) \cos w_j t_k \quad j = 0, \dots, N-1$$

$$b_j = \frac{2}{2N} \sum_{k=0}^{N-1} g(t_k) \sin w_j t_k \quad j = 1, \dots, N-1$$

$$a_N = \frac{1}{2N} \sum_{k=0}^{N-1} g(t_k) (-1)^k$$

$$b_N = 0$$

Thus if our observed time series is $x_0, x_1, \dots, x_{N-1}$ and $g(t) = x_t$ $\forall t$ then defining

$$a_0 = \bar{x} \quad (3.24)$$

$$a_j = \frac{2}{N} \sum_{t=0}^{N-1} x_t \cos(w_j t) = a_{-j} \quad (0 < j < \frac{N}{2}) \quad (3.25)$$

$$b_j = \frac{2}{N} \sum_{t=0}^{N-1} x_t \sin(w_j t) = -b_{-j} \quad (0 < j < \frac{N}{2}) \quad (3.26)$$

and if N even then

$$a_{N/2} = \frac{1}{N} \sum_{t=0}^{N-1} (-1)^t x_t \quad (3.27)$$

where $w_j = 2\pi j/N = j^{th}$ Fourier frequency (3.28)

then

$$x_t = a_0 + \sum_{0 < j < \frac{N}{2}} (a_j \cos(w_j t) + b_j \sin(w_j t)) \quad (3.29)$$

$$+ (-1)^t a_{N/2}$$

(the last term appears only if N even)

3.4.1 NOTES

- (i) (3.29) can be rewritten as

$$x_t = \frac{1}{N} \sum_{j=0}^{N-1} J_j e^{i \omega_j t} \quad (3.30)$$

- (ii)
- $J_{N-j} = \bar{J}_j$
- the complex conjugate of
- J_j
- .

- (iii)
- $J_0, \dots, J_{N-1}$
- is called the discrete Fourier transform of
- $x_0, \dots, x_{N-1}$
- .

- (iv) The discrete Fourier transform at any frequency
- ω
- is defined by

$$J(\omega) = \frac{1}{N} \sum_{t=0}^{N-1} x_t e^{-i \omega t} \quad (3.31)$$

thus

(v) Writing $J_1 = \frac{1}{N} \sum_{t=0}^{N-1} x_t e^{-i \omega_1 t}$

then $J_1 = \frac{1}{4} R_1$ where R_1 is as defined in (3.13)

and $R_1 = \frac{1}{4} (a_1^2 + b_1^2)$

- (vi) The periodogram is often defined by

$$I(x_1) = \frac{N}{2} R_1^2$$

The use of the periodogram will be discussed more fully at a later stage.

- (vii) If $x_0, \dots, x_{N-1}$ is a periodic series with period h and $x_t = x_{t+kh}$, some integer k , then, by definition

$$x_t = \sum_{k=0}^{h-1} x_k e^{i \frac{2\pi k}{h} t} \quad t = 0, 1, \dots, h-1$$

$$\text{and } 2\pi/h = 2\pi k/N = \omega_k$$

So

$$J_j = \begin{cases} Q & j = k\lambda \quad (k \text{ an integer}) \\ 0 & \text{otherwise} \end{cases}$$

i.e. the transform vanishes except at the frequency $\omega_k = \frac{2\pi}{h}$ and its multiples!

(viii) Leakage

Consider the following situation:

suppose $x_t = \cos(\omega t + \phi)$ $t = 0, \dots, N-1$

$$\therefore x_t = \frac{1}{2} \left(e^{i(\omega t + \phi)} + e^{-i(\omega t + \phi)} \right)$$

Since $e^{i\phi}$ is just a scaling factor (i.e. constant), it is only necessary to find the Fourier transform of $e^{i\omega t}$

Let $x_t = z_t + y_t$ $\forall t$

$$\text{where } z_t = \left(\frac{1}{2} e^{i\phi} \right) e^{i\omega t}$$

$$\text{and } y_t = \left(\frac{1}{2} e^{-i\phi} \right) e^{-i\omega t}$$

$$\text{Now } J_j = \frac{1}{N} \sum_{t=0}^{N-1} x_t e^{-i\omega_j t} \quad (\text{by definition})$$

$$= \frac{1}{N} \sum_{t=0}^{N-1} z_t e^{-i\omega_j t} + \frac{1}{N} \sum_{t=0}^{N-1} y_t e^{-i\omega_j t}$$

$$\therefore J_j = \frac{c_1}{N} e^{-i(\omega_1 - \omega_j)t} + \frac{c_2}{N} e^{-i(\omega_j + \omega)t}$$

$$(\text{where } c_1 = \frac{1}{2} e^{i\phi} \text{ and } c_2 = \frac{1}{2} e^{-i\phi})$$

$$= J_{z;j} + J_{y;j} \quad \text{say.}$$

Now it may be shown (see, for example, Bloomfield, (1976))

that

$$\sum_{i=0}^{N-1} \cos(i\alpha) = \cos \left(\frac{(N-1)\alpha}{2} \right) \frac{\sin(\frac{N\alpha}{2})}{\sin(\frac{\alpha}{2})}$$

$$\text{(since } \sum_{i=0}^{N-1} e^{i\alpha i} = e^{i(N-1)\alpha/2} \left[\frac{e^{iN\alpha/2} - e^{-iN\alpha/2}}{e^{i\alpha/2} - e^{-i\alpha/2}} \right] \text{)}$$

and similarly that

$$\sum_{i=0}^{N-1} \sin(i\alpha) = \sin \left(\frac{(N-1)\alpha}{2} \right) \frac{\sin(\frac{N\alpha}{2})}{\sin(\frac{\alpha}{2})}$$

Thus, after manipulation, it may be shown that

$$\begin{aligned} J_{x(1)} &= \frac{C_2}{N} \left(\cos \left(\frac{(N-1)x^{(1)}}{2} \right) \frac{\sin(\frac{Nx^{(1)}}{2})}{\sin(\frac{x^{(1)}}{2})} \right. \\ &\quad \left. - i \sin \left(\frac{(N-1)x^{(1)}}{2} \right) \frac{\sin(\frac{Nx^{(1)}}{2})}{\sin(\frac{x^{(1)}}{2})} \right) \end{aligned}$$

and

$$\begin{aligned} J_{y(1)} &= \frac{C_2}{N} \left(\cos \left(\frac{(N-1)x^{(2)}}{2} \right) \frac{\sin(\frac{Nx^{(2)}}{2})}{\sin(\frac{x^{(2)}}{2})} \right. \\ &\quad \left. - i \sin \left(\frac{(N-1)x^{(2)}}{2} \right) \frac{\sin(\frac{Nx^{(2)}}{2})}{\sin(\frac{x^{(2)}}{2})} \right) \end{aligned}$$

where $x^{(1)} = \omega_1 - \omega$

$x^{(2)} = \omega_2 - \omega$

$$\text{So } J_j = \frac{c_1}{N} e^{\frac{(N-1)w(1)}{2}} \frac{\sin(\frac{Nw(1)}{2})}{\sin(\frac{w(1)}{2})} \\ + \frac{c_2}{N} e^{\frac{(N-1)w(2)}{2}} \frac{\sin(\frac{Nw(2)}{2})}{\sin(\frac{w(2)}{2})}$$

Now if w is itself a Fourier frequency, w_k say, then J_j is zero when $j \neq k$ and is nonzero for $j = k$ (in fact it equals $c_1 + c_2 = \cos \phi$).

If w is not a Fourier frequency then all the J_j are nonzero.

The appearance of nonzero terms in the transform is due to a sinusoid of a different frequency - this is defined to be 'leakage'. Fourier analysis (or periodogram analysis) is carried out in order to separate the effects of different frequencies and so leakage is a confusing phenomenon.

If we define

$$D_N(w) = \sum_{t=0}^{N-1} e^{jw(N-t)} = (N \sin(\frac{wN}{2})) e^{jw(N-1)/2}$$

then for N large and w small,

$$D_N(w) \approx \frac{\sin(\frac{wN}{2})}{\frac{wN}{2}} \quad (3.32)$$

So the amplitude of the Fourier transform of $x_0, \dots, x_{N-1}$ (where $x_t = \cos(wt + \phi)$ $\forall t$) is a function of (3.32), centred at w and rescaled so that the zeroes are separated by $\frac{2}{N}$ (the distance between adjacent Fourier frequencies).

$$\text{So } J_j = \frac{c_1}{N} e^{\frac{(N-1)w^{(1)}}{2}} \frac{\sin(\frac{Nw^{(1)}}{2})}{\sin(\frac{w^{(1)}}{2})} \\ + \frac{c_2}{N} e^{\frac{(N-1)w^{(2)}}{2}} \frac{\sin(\frac{Nw^{(2)}}{2})}{\sin(\frac{w^{(2)}}{2})}$$

Now if w is itself a Fourier frequency, w_k say, then J_j is zero when $j \neq k$ and is nonzero for $j = k$ (in fact it equals $c_1 + c_2 = \cos \phi$).

If w is not a Fourier frequency then all the J_j are nonzero.

The appearance of nonzero terms in the transform is due to a sinusoid of a different frequency - this is defined to be 'leakage'. Fourier analysis (or periodogram analysis) is carried out in order to separate the effects of different frequencies and so leakage is a confusing phenomenon.

If we define

$$D_N(w) = \sin(\frac{wN}{2}) \quad (N \sin(\frac{w}{2}))$$

then for N large and w small,

$$D_N(w) \approx \frac{\sin(\frac{wN}{2})}{\frac{wN}{2}} \quad (3.32)$$

So the amplitude of the Fourier transform of $x_0, \dots, x_{N-1}$ (where $x_t = \cos(wt + \phi)$ $\forall t$) is a function of (3.32), centred at w and rescaled so that the zeroes are separated by $\frac{2\pi}{N}$ (the distance between adjacent Fourier frequencies).

So, when w is not a Fourier frequency, only the values of the transform that are close to w are large.

Even if x_t consists of some sum of functions (i.e. sine and cosine functions) at several frequencies, the above will still hold true. So if we plot the periodogram as defined in note (vi), or even a function thereof, large values of $I(w)$ will indicate nonzero frequencies in the data. This will be discussed more fully in section 3.6.

3.5 PROPERTIES OF a_k and b_k

In the previous section, estimates of a_k and b_k in the model

$$x_t = \sum_{k=0}^K (a_k \cos(w_k t) + b_k \sin(w_k t)) + \varepsilon_t \quad (3.33)$$

ε_t = purely random process $t = 0, \dots, N-1$

$$E(\varepsilon_t) = 0 \quad w_k = 2\pi k/N \quad k = 0, \dots, N/2$$

were found.

In this section, it will be verified that the estimates given by (3.24) - (3.27) are unbiased, by using a different procedure. This verification is not necessary, but the technique involved facilitates computation of the properties of the estimates.

(3.33) above may be regarded as a standard multiple regression model. So define

$$X = Z\beta + \varepsilon$$

where

$$\mathbf{X}' = (x_0, x_1, \dots, x_{N-1}) \quad (\text{dimension } N)$$

$$\mathbf{Z}' = (z_0, z_1, \dots, z_{N-1}) \quad (")$$

$$\mathbf{u}' = (a_0, a_1, b_1, \dots, a_k, b_k) \quad (\text{dimension } 2k+1)$$

and

$$\mathbf{Z} = \begin{bmatrix} 1 & \cos(0 \cdot \omega_1) & \sin(0 \cdot \omega_1) & \dots & \sin(0 \cdot \omega_k) \\ 1 & \cos(1 \cdot \omega_1) & \sin(1 \cdot \omega_1) & \dots & \sin(1 \cdot \omega_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos((N-1) \cdot \omega_1) & \sin((N-1) \cdot \omega_1) & \dots & \sin((N-1) \cdot \omega_k) \end{bmatrix} \quad (N \times (2k+1))$$

where, without loss of generality, we assume N is odd.

Then, by definition of the linear regression model,

$$\hat{\mathbf{u}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} \quad (\text{see for example, Draper and Smith (1966)})$$

since $\mathbf{Z}$ is of full rank and $\{z_t\}$ is a white noise process,

$$\begin{aligned} \text{hence} \quad a_0 &= \bar{X} \\ a_1 &= \frac{2}{N} \sum_{t=0}^{N-1} x_t \cos(\omega_1 t) \\ b_1 &= \frac{2}{N} \sum_{t=0}^{N-1} x_t \sin(\omega_1 t) \end{aligned}$$

and if N is even

$$a_{N/2} = \frac{2}{N} \sum_{t=0}^{N-1} x_t (-1)^t \quad \text{as before.}$$

If we assume the ε_t come from a purely random process

$$\begin{aligned} \text{i.e.} \quad E(\varepsilon_t) &= 0 \quad \forall t \\ V(\varepsilon_t) &= \sigma^2 \quad \forall t \\ \text{Cov}(\varepsilon_t, \varepsilon_s) &= 0 \quad \forall t \neq s \end{aligned} \quad (3.34)$$

then it follows that

$$\begin{aligned} E(a_1) &= a_1 \quad \forall 1 \\ E(b_1) &= b_1 \quad \forall 1 \end{aligned} \quad (3.35)$$

$$\begin{aligned}
 \text{and } V(a_i) &= \left(\frac{2}{N}\right)^2 \sigma^2 \int_0^T \cos^2(\omega_i t) dt \\
 &= \frac{2\sigma^2}{N} = V(b_i) \quad i \neq 0 \text{ and } i \neq \left[\frac{N}{2}\right] \\
 &= \frac{\sigma^2}{N} \quad i = 0 \text{ or } i = \left[\frac{N}{2}\right]
 \end{aligned} \tag{3.36}$$

$$\text{Also } \text{Cov}(a_i, b_j) = 0 \quad \forall i, j$$

[The variances of estimates are obtained simply by noting that the covariance matrix of $\hat{\underline{a}}$ is given by $(Z'Z)^{-1} \underline{J}^2$ (by definition). But $(Z'Z)$ is a diagonal matrix due to the orthogonality relations given below

$$\begin{aligned}
 \text{(a) } \int_0^T \cos(\omega_j t) \cos(\omega_k t) dt &= \begin{cases} 0 & k \neq j \\ \frac{T}{2} & k = j \neq 0 \text{ or } \left[\frac{N}{2}\right] \\ T & k = j = 0 \text{ or } \left[\frac{N}{2}\right] \end{cases} \\
 \text{(b) } \int_0^T \sin(\omega_j t) \sin(\omega_k t) dt &= \begin{cases} 0 & k \neq j \\ \frac{T}{2} & k = j \neq 0 \text{ or } \left[\frac{N}{2}\right] \\ 0 & k = j = 0 \text{ or } \left[\frac{N}{2}\right] \end{cases} \\
 \text{(c) } \int_0^T \cos(\omega_j t) \sin(\omega_k t) dt &= 0 \quad \forall j, k \\
 \text{(d) } \int_0^T \cos(\omega_j t) dt &= 0 \quad \forall j \\
 \text{(e) } \int_0^T \sin(\omega_j t) dt &= 0 \quad \forall j
 \end{aligned} \tag{3.36a}$$

3.6 PERIODOGRAM ANALYSIS

If we decide to model our observed time-series $x_0, x_1, \dots, x_{N-1}$ by

$$x_t = \sum_{i=0}^k (a_i^* \cos(\omega_i^* t) + b_i^* \sin(\omega_i^* t)) + \epsilon_t \tag{3.37}$$

(x_t a purely random process) $\forall t$, some k , then in general, the $(k, w_1^*, w_2^*, \dots, w_N^*)$ are unknown.

We define

$$w_j^* = \omega_j = 2^{-j/N}; \quad j = 0, 1, \dots, \left[\frac{N}{2}\right]$$

but this may lead to the inclusion in the model of 'insignificant' terms i.e. terms where the contribution of 'functions' associated with frequency w_j^* is not significantly different from zero (or to the exclusion of significant terms).

In other words

$$a_j^* \cos(w_j^* t) + b_j^* \sin(w_j^* t) \approx 0.$$

Thus we wish to model

$$x_t = \sum_{j=0}^{\left[\frac{N}{2}\right]} (a_j \cos(w_j t) + b_j \sin(w_j t)) \quad \forall t$$

$$\omega_j = 2^{-j/N} \tag{3.38}$$

and then only actually include those terms in the model which are different (significantly) from zero.

It turns out that the way around this problem is reasonably easy. Model by equation (3.38) and find estimates of a_j and b_j by using the a_j^* and b_j^* as defined by (3.24)-(3.29).

If w_j is a true model frequency, then a_j and b_j should be close to a_j^* and b_j^* , and so the squared amplitude (i.e. $(\hat{a}_j^2 + \hat{b}_j^2)$), or any linear function thereof, will be nonzero. If w_j is not a model frequency, then this effectively estimates the coefficients of a term that should not be in the model, so

$$\hat{a}_j^2 + \hat{b}_j^2 = 0.$$

Thus, for $w_1, w_2, \dots$, a plot of $(\hat{a}_j^2 + \hat{b}_j^2)$ (or any linear function thereof) against j , will be such that if w_j is near a true w_j^* of the model, then its squared amplitude will be nonzero. See also notes (vi) and (viii) in section 3.4.1.

Define the (sample) periodogram of our time-series $x_0, x_1, \dots, x_{N-1}$ by

$$I(w_j) = (a(w_j))^2 + (b(w_j))^2 \quad (3.39)$$

where

$$\left. \begin{aligned} a(w) &= \sqrt{\frac{2}{N}} \sum_{t=0}^{N-1} x_t \cos(wt) = \sqrt{\frac{N}{2}} a_w \\ b(w) &= \sqrt{\frac{2}{N}} \sum_{t=0}^{N-1} x_t \sin(wt) = \sqrt{\frac{N}{2}} b_w \end{aligned} \right\} \quad (3.40)$$

(where a_w and b_w are as in (3.25) and (3.28)). (Note that this differs from the definition given in note (vi), section 3.4.1.) Alternatively, we may write

$$I(w) = \frac{2}{N} \left| \sum_{t=0}^{N-1} x_t e^{-iwt} \right|^2 \quad (3.41)$$

$I(w)$ is defined for all w in $(-\pi, \pi)$, but it is only evaluated at the frequencies w_j (where $w_j = 2\pi j/N$ and $j = 0, 1, \dots, [N/2]$) since $I(w)$ is an even function and due to the definition of $\hat{a}_j$ and $\hat{b}_j$ in the model (3.38).

If w_j coincides with one of the true w_j^* in the model, w_j^* say, then

$$\sqrt{\frac{N}{2}} \hat{a}_j = a(w_j^*) \quad \text{and}$$

$$\sqrt{\frac{N}{2}} \hat{b}_j = b(w_j^*)$$

Thus, if $w_1 = w_0$

$$E(I(w_1)) = \frac{N}{2} E(a_1^2) = E(b_1^2)$$

$$\text{But } E(a_1^2) = (E(a_1))^2 + \sigma(a_1)$$

$$= a_1^2 + \frac{2\sigma^2}{N} \quad (\text{by (3.36)}) \quad 1 \neq 0 \text{ and } 1 \neq \left(\frac{N}{2}\right)$$

$$\text{so } E(I(w_1)) = \frac{N}{2} (a_1^2 + b_1^2) + 2\sigma^2$$

$$= \frac{N}{2} R_1^2 + 2\sigma^2 \quad \text{say} \quad (3.42)$$

If $w_1 \neq w_0$ (V) then

$$R_1^2 \approx 0$$

so

$$E(I(w_1)) = 2\sigma^2$$

So as w_1 approaches each of the w_0 , the ordinate is of order N and the periodogram has a large peak. Thus, the location of the true frequencies are noted (approximately), whenever a peak occurs in the periodogram.

3.7 SAMPLING PROPERTIES OF THE PERIODOGRAM

The model we have used for our series of observations $x_1, \dots, x_{N-1}$ is

$$x_t = \sum_{i=0}^k (a_i \cos(w_i t) + b_i \sin(w_i t)) + \epsilon_t$$

where ϵ_t = purely random process

and this can be rewritten as

$$x_t = \sum_{i=1}^k R_i \cos(w_i t + \phi_i) + \epsilon_t$$

where the w_j here are the true w_j (and not necessarily $2\pi/N$) and the $\{\phi_k\}$ are independently and rectangularly distributed on $(-\pi, \pi)$.

Let $R_1 = 0 \quad \forall t$, and suppose x_t is a Gaussian random process.

Thus

$$x_t \sim N(0, \sigma_x^2) \quad \forall t \text{ independently and so}$$

$$x_t \sim N(0, \sigma_x^2) \quad \forall t \text{ independently (since the } R_1 = 0)$$

Now $a(w_j)$ and $b(w_j)$, as defined by (3.39) and (3.40), are linear combinations of the x_t 's thus $a(w_j)$ and $b(w_j)$ are normally distributed with zero mean and variance given by

$$V(a(w_j)) = V(b(w_j)) = \begin{cases} \frac{2\sigma_x^2}{N} \cos^2(w_j t) & j \neq 0 \text{ and } j \neq \left[\frac{N}{2}\right] \\ \frac{2\sigma_x^2}{N} & \text{otherwise} \end{cases} \quad (3.43)$$

So

$$a(w_j) \sim \begin{cases} N(0, \frac{2\sigma_x^2}{N}) & j \neq 0 \text{ and } j \neq \left[\frac{N}{2}\right] \\ N(0, \frac{2\sigma_x^2}{N}) & \text{otherwise} \end{cases} \quad (3.44)$$

and similarly for $b(w_j)$

Now by using the orthogonality relations for the functions $\cos(w_j t)$ and $\sin(w_j t)$ (see (3.36a)), we can show that

$$\begin{aligned} \text{cov}(a(w_j), b(w_l)) &= \frac{2\sigma_x^2}{N} \int_0^T \cos(w_j t) \sin(w_l t) dt \\ &= 0 \quad \forall j, l \end{aligned} \quad (3.45)$$

Even though $a(w_j)$ and $b(w_l)$ are defined as random variables, they retain constant values for each realization

and

$$\begin{aligned} \text{cov}\{a(w_j), a(w_j)\} &= \text{cov}\{b(w_j), b(w_j)\} \\ &= 0 \quad \forall j \neq \frac{N}{2} \end{aligned} \quad (3.46)$$

But, $a(w_j)$ and $b(w_j)$ are linear combinations of the x_i and hence have a multivariate normal distribution. They are uncorrelated by (3.45) and hence independent. Thus

$$\begin{aligned} I(w_j) &= a^2(w_j) + b^2(w_j) \quad (\text{by (3.39)}) \\ &\sim \sigma^2 \chi^2_2 \\ &= \frac{2}{x} \chi^2_2 \quad j \neq 0 \text{ or } j \neq \frac{N}{2} \end{aligned} \quad (3.47)$$

since $\sigma^2 = \frac{2}{x}$

$$\begin{aligned} \text{and} \quad I(w_j) &= a^2(w_j) \quad j = 0 \text{ or } j = \frac{N}{2} \\ &= 2\sigma^2 \chi^2_1 \\ &= 2\frac{2}{x} \chi^2_1 \end{aligned} \quad (3.48)$$

$$(\text{since } b(w_j) = 0 \quad \text{if } j = 0 \text{ or } j = \frac{N}{2})$$

So

$$E(I(w_j)) = 2\frac{2}{x} \quad \forall j \quad (3.49)$$

and

$$V(I(w_j)) = \begin{cases} 4\sigma^4 & j \neq 0, \text{ or } j \neq \frac{N}{2} \\ 8\sigma^4 & \text{otherwise} \end{cases} \quad (3.50)$$

By (3.41), we may write $I(w_j)$ i.t.o. $e^{-i w_j t}$ i.e.

$$I(w_j) = \frac{2}{N} \sum_{k=1}^N x_k e^{-i w_j t} \cdot 2$$

or more generally

$$I(w) = \frac{2}{N} \left| \sum_t x_t e^{-iwt} \right|^2$$

But then

$$\begin{aligned} I(w) &= \frac{2}{N} \left(\sum_t x_t e^{-iwt} \right) \left(\sum_s x_s e^{iws} \right) \\ &= \frac{2}{N} \left[\sum_t x_t \cos(wt) - i \sum_t x_t \sin(wt) \right] \\ &\quad \times \left[\sum_s x_s \cos(ws) + i \sum_s x_s \sin(ws) \right] \\ &= \frac{2}{N} \sum_t \sum_s x_t x_s [\cos(wt)\cos(ws) + \sin(wt)\sin(ws)] \\ &= \frac{2}{N} \sum_t \sum_s x_t x_s \cos((t-s)w) \\ &= 2 \sum_{h=-(N-1)}^{(N-1)} \left\{ \frac{1}{N} \sum_{t=0}^{N-|h|-1} x_t x_{t+h} \right\} \cos(hw) \\ &= 2 \sum_h c'_h \cos(hw) \end{aligned} \quad (3.51)$$

(where c'_h is the sample acvf as in section 2.4.)

So we have succeeded in writing the periodogram i.t.o the sample acvf. Thus since we can express the sdf in terms of the acvf, it seems likely that some relationship exists between the sdf and the periodogram. We shall see this relationship in the next chapter.

Now all the above results hold if all the R_k in

$$x_t = \sum_k R_k \cos(2\pi f_k t + \phi_k) + \epsilon_t \quad \forall t$$

are zero. Suppose not all of the R_k are zero.

Define $x_t = z_t + \xi_t \quad \forall t$

where $z_t = R_1 \cos(\omega_1 t + \phi_1)$

The ϕ_i are just some constants, and so ξ_t is independent of ϕ_i for all i and t . Thus $\{z_t\}$ and $\{\xi_t\}$ are independent processes.

At this stage it becomes necessary to quote a result not previously given. It will be given in the form of a lemma, the proof of which follows by definition of the acvf.

Lemma 2
.....

If $x_t = z_t + v_t$ where z_t and v_t are independent processes, then

$$\gamma_x(h) = \gamma_z(h) + \gamma_v(h)$$

where $\gamma_z(h)$ is the acvf of the process $\{z_t\}$.

Using this lemma, we may write

$$\gamma_x(h) = \gamma_z(h) + \gamma_v(h)$$

But

$$\gamma_z(h) = \frac{1}{2} R_1^2 \cos(\omega_1 h)$$

and

$$\gamma_v(h) = \frac{1}{2} R_2^2 \cos(\omega_2 h)$$

$$\text{where } R_2^2 = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

($\gamma_v(h)$ is as above since $\{v_t\}$ is a purely random process)

Now

$$I(\omega) = 2 \sum_{h=-(N-1)}^{(N-1)} \cos(h\omega) \quad \text{by (3.51)}$$

$z_t = R_1 \cos(\omega_1 t + \phi_1)$; the $\{\phi_i\}$ are independent rectangular r.v's.

and so

$$E(I(w)) = \frac{2}{h} E(c'_h) \cos(hw) \\ = 2 \sum_{h=1}^N (1 - \frac{|h|}{N}) \gamma_x(h) \cos(hw) \quad (\text{by (2.29)})$$

$$\text{but } \gamma_x(h) = \gamma_+(h) + \gamma_-(h)$$

and substituting gives

$$E(I(w)) = \sum_{h=1}^N R_h^2 \left\{ \left(1 - \frac{|h|}{N}\right) \cos(hw_1) \cos(hw) \right\} + 2c \quad (3.52)$$

We now only need simplify the term inside the braces () in (3.52)

It turns out (3.52) can be written as

$$E(I(w)) = 2c + \frac{1}{2N} \sum_{h=1}^N R_h^2 (g_h + h_h) \quad (3.53)$$

$$\text{where } g_h = \frac{\sin^2 \left(\frac{N(w+w_1)}{2} \right)}{\sin^2 \left(\frac{w+w_1}{2} \right)}$$

$$\text{and } h_h = \frac{\sin^2 \left(\frac{N(w-w_1)}{2} \right)}{\sin^2 \left(\frac{w-w_1}{2} \right)}$$

Outline proof:

We need to simplify

$$\sum_{h=-(N-1)}^{(N-1)} \left(1 - \frac{|h|}{N}\right) \cos(hw_1) \cos(hw)$$

$$\text{Now } \sum_{h=-(N-1)}^{(N-1)} (N - |h|) a(h) = \sum_{h=0}^{N-1} (N-h) a(h) + \sum_{h=1}^{N-1} (N-h) a(-h) \\ \text{where } a(h) = \cos(hw_1) \cos(hw) \quad \text{and } a(-h) = \cos(-hw_1) \cos(-hw) = \cos(hw_1) \cos(hw)$$

$$\text{and } \sum_{s=-u}^u \cos(s) = \frac{\sin((u+\frac{1}{2}))}{\sin(\frac{1}{2})}$$

Also
$$\sum_{u=0}^{N-1} \frac{\sin((u+\frac{1}{2})\theta)}{\sin(\frac{\theta}{2})} = \frac{\sin^2(\frac{N\theta}{2})}{\sin^2(\frac{\theta}{2})}$$

and
$$\cos \theta_1 \cos \theta_2 = \frac{1}{2} (\cos(\theta_1 + \theta_2) + \cos(\theta_1 - \theta_2))$$

Combining the above facts will provide the result (3.53).

So, as $w \rightarrow w_j$, the term $(g_j + h_j)$ in (3.53) gives a peak of size N . Thus, if x_j has a sine cosine function of frequency w_j^* , and if w_j^* is of the form $2^{-j}N$ (some j), then $I(w_j)$ will have a peak of height $NR_j^2/2$ at j .

Having derived $E(I(w))$, all that is now required to complete this section is (a) $\text{Cov}(I(w_j), I(w_l))$ and (b) $V(I(w_j))$.

These results will be stated here and the proofs given in appendix A4. (Assume $R_k = 0 \quad \forall k$)

(a) $\text{Cov}(I(w_j), I(w_l))$

$$= \frac{4K_4}{N} + \frac{4X^4}{N} \left(1 - \frac{h}{N}\right) (e_h + f_h) \quad j \neq 0, l \neq 0$$

$$(\text{where } e_h = \cos(h(w_j + w_l)))$$

$$f_h = \cos(h(w_j - w_l))$$

$$K_4 = E(x_k^4) = \mu_4 - 3\sigma_x^4$$

and we have assumed the x_k are independent $\forall t$.)

$$\text{Cov}(I(w_j), I(w_l)) = \frac{4K_4}{N} + \frac{4X^4}{N} (a + b) \quad (3.54)$$

$$a = \frac{\sin^2(\frac{N(w_j + w_l)}{2})}{\sin^2(\frac{w_j + w_l}{2})} \quad \text{and} \quad b = \frac{\sin^2(\frac{N(w_j - w_l)}{2})}{\sin^2(\frac{w_j - w_l}{2})}$$

Setting $w_1 = w_2 = w$ and assuming w of the form $2\pi r/N$ (some integer r) then

$$(b) \quad V(I(w)) = \begin{cases} 80 \frac{4}{N} + \frac{4K}{N} & w = 0 \text{ or } \pi \\ 40 \frac{4}{N} + \frac{4K}{N} + O\left(\frac{1}{N^2}\right) & \text{otherwise} \end{cases} \quad (3.55)$$

3.8 SUMMARY AND DISCUSSION

Since all functions (periodic or nonperiodic) can be written in the form

$$g(t) = \sum_{k=0}^{\infty} (a_k \cos(w_k t) + b_k \sin(w_k t))$$

(by the theory of chapters one and two), in this chapter we have seen that, having decided on the following model

$$x_t = \sum_{k=0}^K (a_k \cos(w_k t) + b_k \sin(w_k t)) + \epsilon_t, \quad \forall t$$

it is possible to estimate a_k , b_k and w_k . This can be done by modelling our observed time series $x_0, \dots, x_{N-1}$ by

$$x_t = \sum_{k=0}^{\frac{N}{2}-1} (a_k \cos(w_k t) + b_k \sin(w_k t)) + \epsilon_t \quad \forall t$$

where $w_k = 2\pi k/N$.

Estimates a_i and b_i were obtained and these turned out to be

$$\begin{aligned} a_i &= \frac{2}{N} \sum_{t=0}^{N-1} x_t \cos(w_i t) & i &= 0, \dots, \left[\frac{N}{2}\right] \\ b_i &= \frac{2}{N} \sum_{t=0}^{N-1} x_t \sin(w_i t) & i &= 1, 2, \dots, \left[\frac{N}{2}\right] \end{aligned}$$

A function called the periodogram, $I(w)$, is formed where

$$I(w_j) = a^2(w_j) + b^2(w_j)$$

and

$$\begin{aligned} a(w_j) &= \sqrt{\frac{2}{N}} \sum_{t=0}^{N-1} x_t \cos(w_j t) = \sqrt{\frac{N}{2}} a_j \\ b(w_j) &= \sqrt{\frac{2}{N}} \sum_{t=0}^{N-1} x_t \sin(w_j t) = \sqrt{\frac{N}{2}} b_j \end{aligned}$$

and then $I(w_j)$ is plotted against j .

If one of the 'true' w_i takes the form

$$w_j = 2^{-j} N = w_i \quad (\text{say})$$

then $I(w)$ exhibits a (distinct) peak at j . Thus we can estimate the w_i , a_i and b_i .

However, we may incorporate irrelevant frequencies into the model, so what we really need is a test for periodogram ordinates to determine which values are significantly different from zero (see appendix A5) and hence our final model is

$$x_t = \sum_{i=0}^{\left[\frac{N}{2}\right]} \{a_i \cos(w_i t) + b_i \sin(w_i t)\} + \varepsilon_t$$

$$t = 0, \dots, N-1$$

where some of the a_i 's and b_i 's may be (approximately) zero.

The sampling properties of $I(w)$ were derived and a link was also established showing a relationship between $I(w)$ and the sample acvf (and hence possibly with the sdf). It turns out that $I(w)$ is

distributed as $h(w)X_n^2$ (this will be shown in the next chapter) where $h(w)$ is the sdf of X_1 . Also, $V(I(w)) = h^2(w)$ (also shown in the next chapter) thus the variance of the periodogram estimate does not decrease as our sample size increases. Since estimates of frequencies near one another have independent distributions, the plot of $I(w_j)$ against j (or w_j) will be unstable (this fact is also seen from the $V(I(w))$ given by (3.55)).

For example, if the periodogram of a white noise process (i.e. a purely random process) were to be computed, since the sdf of this process is a straight line (since $h(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} \gamma(\tau) d\tau$ but $\gamma(\tau)$ is constant (zero) $\forall \tau > 0$) we would expect the periodogram peaks to 'die out' as the sample size increased - this is not the case. The periodogram remains as wild for large samples as for smaller ones (and in fact has coefficient of variation = 100%).

Since the peaks are randomly placed, by taking an average of all possible white noise periodograms (for varying sample sizes) at each frequency a straight line (approximately) should be obtained. This gives some indication of the relationship between the sdf and the periodogram.

Thus a way of reducing (or eliminating) the variance of the periodogram ordinates, is to average neighbouring values (since if independent estimates are averaged, the variance of the average is less than the variance of each of the values in the average). This 'averaging' is called spectral analysis, which is discussed in the next chapter.

[We have seen that the periodogram effectively determines whether a frequency w_j (or rather components associated with w_j) is 'necessary' in our model

$$x_t = \sum_{j=0}^N (a_j \cos(w_j t) + b_j \sin(w_j t)) + \epsilon_t \quad \forall t.$$

A simple technique devised by Buys-Ballot (1847) is given in appendix A6. This technique allows us to determine whether the observed time-series has period T , for any given integer T .

CHAPTER 4: THE SPECTRUM

In chapters one and two it was shown that if $\{X_t\}$ is a stationary process with spectral density function (sdf) $h(\omega)$ and autocovariance function (acvf) $\gamma(k)$ (k an integer), then

$$h(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma(t) \cos(\omega t) dt \quad (4.1)$$

if X_t is real-valued (see section 2.3 and specifically (2.22) which relates the nsdf (normalized sdf) to the acrf (autocorrelation function)). An obvious estimate of $h(\omega)$ would be

$$\begin{aligned} \hat{h}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\gamma}(t) \cos(\omega t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} c_t' \cos(\omega t) dt \end{aligned} \quad (4.2)$$

(where c_t' is the (sample) acvf based on our realization of $\{X_t\}$ and is defined in section 2.4).

In chapter three, where we dealt with the (sample) periodogram $I(\omega)$, it was shown that

$$I(\omega) = 2 \sum_{h=-(N-1)}^{(N-1)} c_h' \cos(h\omega) \quad (4.3)$$

where $I(\omega)$ is based on a realization $x_0, x_1, \dots, x_{N-1}$ (see section 7 and specifically (3.51)).

Thus, when we compare (4.2) with (4.3), it is obvious that some relationship exists between $I(\omega)$ and $h(\omega)$ (an estimate of the sdf). This relationship was fleetingly touched upon in the

summary of the last chapter (section 3.8), and will be dealt with more formally here. Once we have established this relationship, the uses and interpretations of the sdf will follow from those of the periodogram, thus making clear the reason for defining this function. More 'efficient' estimates of the sdf will also be considered in this chapter.

4.1 THE PERIODOGRAM AS AN ESTIMATE OF THE SPECTRUM

Suppose, w.l.o.g., that for our real-valued (discrete) process

$$\{X_t\}, E(X_t) = \mu \quad \forall t \text{ and } h(\omega) \text{ is continuous } \forall \omega.$$

Now $\mu(t)$ can be estimated by $\bar{x}_t$ ($t=0, \pm 1, \dots, \pm(N-1)$) if we have a realization $x_0, x_1, \dots, x_{N-1}$, thus the obvious estimate of (4.2) is

$$\hat{I}(\omega) = \frac{1}{2\pi} \sum_{t=-(N-1)}^{(N-1)} c_t' \cos(t\omega) \quad (4.4)$$

Consider $I(\omega)$ as in (3.39) and define

$$I^*(\omega) = \frac{1}{2\pi} I(\omega) \quad (4.5)$$

$$= \frac{1}{2\pi} \sum_{t=-(N-1)}^{N-1} c_t' \cos(t\omega) \quad (4.6)$$

$$= h(\omega) \quad (\text{defined in (4.4)})$$

Hence

$$E(I^*(\omega)) = \frac{1}{2\pi} \sum_{t=-(N-1)}^{N-1} E(c_t') \cos(t\omega) = h(\omega)$$

(since $E(I^*(\omega)) = \frac{1}{2\pi} \sum_{t=-(N-1)}^{N-1} E(c_t') \cos(t\omega)$ and see equation (2.29)).

$$\text{and } \lim_{N \rightarrow \infty} E(I^*(\omega)) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} c_t' \cos(t\omega) \quad (4.7)$$

Thus $I^*(w)$ is an asymptotically unbiased estimate of $h(w)$. Now in section 3.7, we saw that

$$V(I(w)) \rightarrow 0 \text{ as } N \rightarrow \infty \quad (\text{by (3.55)})$$

hence

$$V(I^*(w)) \rightarrow 0 \text{ as } N \rightarrow \infty$$

thus $I^*(w)$ is an inconsistent estimate of $h(w)$ (also see appendix A7, result 9). What we really desire is a consistent estimate of $h(w)$ (and hence $f(w)$).

Now $V(I^*(w))$ does not tend to zero since it contains 'too many' acv^s (autocovariances) - this is since $I^*(w)$ involves $N \text{acv}^s$ and although the variance of each acv is of order $\frac{1}{N}$ (see section 2.4), the cumulative effect of N of these acv^s produces a variance of order 1.

4.2 CONSISTENT ESTIMATES OF THE SDF

By the above argument, it seems reasonable to omit some of the acv^s in (4.6). This will reduce the variance, but will influence the expected value also.

If $h(w)$ is continuous (i.e. $\{X_t\}$ has a continuous spectrum) then

$$r(t) \rightarrow 0 \text{ as } |t| \rightarrow \infty$$

(by the implications of the Wiener-Khintchine theorem)

and so if only the acv^s in the tail of the (sample) acvf , are omitted, then (hopefully), the bias will not be too serious.

Hence a 'reasonable' estimate of $h(w)$ would seem to be

$$\hat{h}(w) = \frac{1}{2\pi} \sum_{t=-M}^M c_t^* \cos(wt) \quad (4.8)$$

where $M = N-1$ is some fixed truncation point yet to be specified.

Equation (4.8) can be considered as a truncated periodogram.

Note that

$$E(\hat{h}(w)) \rightarrow h(w) \text{ as } M \rightarrow \infty$$

thus if M is some function of N such that as $N \rightarrow \infty$, $M \rightarrow \infty$, then

$\hat{h}(w)$ is an asymptotically unbiased estimate of $h(w)$.

Heuristically, since $h(w)$ contains only $2M + 1$ acv^s and $I^*(w)$ contains $2N$ acv^s (or more correctly, since $c'_t = c'_{-t}$, $M + 1$ acv^s as compared with N acv^s), it might be reasonable to suppose

$$V(h(w)) = \frac{M}{N} V(I^*(w))$$

which is of order $\frac{M}{N}$.

So if $M \rightarrow \infty$ as $N \rightarrow \infty$ and $\frac{M}{N} \rightarrow 0$ as $N \rightarrow \infty$, then the bias and variance of $\hat{h}(w)$ should tend to zero and thus $\hat{h}(w)$ will be a consistent estimate of $h(w)$.

The obvious extension of (4.8) would seem to be

$$h(w) = \frac{1}{2} \sum_{t=-(N-1)}^{(N-1)} c'_t \cos(tw) \quad -\pi \leq w \leq \pi \quad (4.9)$$

To obtain (4.8) from (4.9), define

$$c'_t = \begin{cases} 1 & |t| \leq M \\ 0 & \text{otherwise} \end{cases}$$

We can say (4.9) is just a weighted average of acv^s or periodogram ordinates. When we recall the summary of the last chapter (i.e. section 3.3), this agrees with the ideas proposed there.

Since c'_t (and hence c'_t) and $\cos(tw)$ are even functions, one of the restrictions we may wish to impose is that c'_t also be an

even function. If this is the case, then it is possible to rewrite (4.9) as

$$h(w) = \frac{1}{2\pi} \sum_{|t| \leq N} \lambda_t c'_t e^{-itw} \quad (4.10)$$

So

$$I^*(w) = \frac{1}{2\pi} \sum_{|t| \leq N} \lambda_t c'_t e^{-itw} \quad -\pi \leq w \leq \pi \quad (4.11)$$

where

$$\lambda_t = 1 \quad \forall t$$

Inverting equation (4.11) gives

$$c'_t = \int_{-\pi}^{\pi} I^*(w) e^{itw} dw \quad |t| \leq N$$

(i.e. multiply both sides of (4.11) by e^{itw} , sum over t , and integrate over w)

so

$$\begin{aligned} h(w) &= \frac{1}{2\pi} \sum_{|t| \leq N} \lambda_t \left\{ \int_{-\pi}^{\pi} I^*(\theta) e^{it\theta} d\theta \right\} e^{-itw} \\ \hat{h}(w) &= \int_{-\pi}^{\pi} I^*(\theta) \left\{ \frac{1}{2\pi} \sum_{|t| \leq N} \lambda_t e^{-it(w-\theta)} \right\} d\theta \\ &= \int_{-\pi}^{\pi} I^*(\theta) W_N(w-\theta) d\theta \quad \text{say} \end{aligned} \quad (4.12)$$

(where $W_N(w-\theta) = \frac{1}{2\pi} \sum_{|t| \leq N} \lambda_t e^{-it(w-\theta)}$)

i.e. $\hat{h}(w)$ can not only be expressed as a weighted sum of acv^s , but also as a weighted integral of the (modified) periodogram, $I^*(\theta)$. The weighting function, $W_N(w-\theta)$, turns out to be the discrete Fourier transform of λ_t . It is possible to consider (4.12) as a 'smoothing of the periodogram' process (again supporting the ideas presented in section 3.8).

[Note that since

$$h(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c_t' e^{-iwt} dt$$

$$\text{and } w_N(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c_t e^{-iwt} dt \quad (4.13)$$

then inverting (4.13) gives

$$c_t = \int_{-\infty}^{\infty} w^{it\theta} w_N(\theta) d\theta \quad |t| \leq N \quad (4.14)$$

If c_t is a (real) even function of t , then so is $w_N(\theta)$ and hence

$$w_N(\theta) = \frac{1}{2\pi} \int_{-N}^N c_t \cos(t\theta) dt \quad (4.15)$$

c_t is called the lag window and $w_N(\theta)$ the spectral window]

4.3 SPECTRAL WINDOWS

The reason for the name 'spectral window' makes intuitive sense - we think of $w_N(\theta)$ as being zero outside some (small) interval, $(-\delta, \delta)$ say, and so (4.12) can be regarded as enabling us to look at $I^*(\cdot)$ through a small 'window' from $(w-\delta)$ to $(w+\delta)$.

Obviously, the form of $w_N(\theta)$ (or equivalently c_t) may vary, since we are considering $h(w)$ as either a weighted sum or weighted integral of the (modified) periodogram ordinates of $I^*(w)$. Thus the 'weight' we attach to each acv is of some importance. Certain weighting functions will have better properties than others.

Several authors have evolved their own windows and presented claims as to why their spectral windows are 'suitable'. The pros and cons of the windows will not be discussed in any detail in

this work, but rather just their general forms will be given.

As λ_t is just a weighting function, the 'form' of some of the lag-windows will make intuitive sense.

4.1.1 Selected Windows

1. The general periodogram

$$\lambda_t = 1 \quad \forall t \quad (4.16)$$

hence

$$W_N(\theta) = \frac{1}{2\pi} \sum_{|t| < N} \cos(t\theta) \quad (\text{by definition}) \quad (4.16a)$$

$$= \frac{1}{2\pi} \left(\frac{\sin(N - \frac{1}{2})\theta}{\sin(\frac{\theta}{2})} \right)$$

$$[\text{since } \sum_{t=0}^{N-1} e^{it\theta} = (e^{iN\theta} - 1)/(e^{i\theta} - 1)]$$

$$\text{hence } \sum_{t=0}^{N-1} \cos(t\theta) = \cos\left(\frac{(N-1)\theta}{2}\right) \frac{\sin(\frac{N\theta}{2})}{\sin(\frac{\theta}{2})}$$

$$\text{and } \sum_{|t| < N} \cos(t\theta) = 2 \sum_{t=0}^{N-1} \cos(t\theta) - 1$$

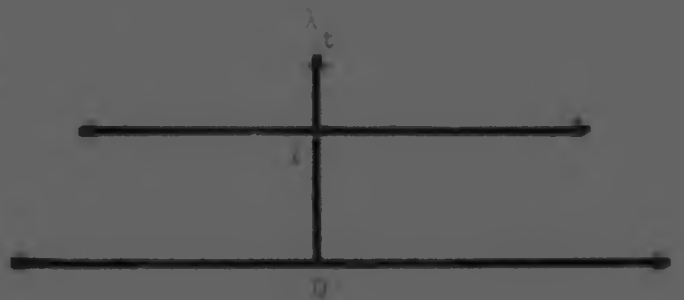


Fig 4.1

2. The truncated periodogram

$$\lambda_t = \begin{cases} 1 & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.17)$$

λ_t is only defined for integer t , however in the graphs, λ_t is 'defined' (mapped) for all t .

hence

$$W_N(\theta) = \frac{1}{2\pi} \left\{ \frac{\sin((M + \frac{1}{2})\theta)}{\sin(\frac{\theta}{2})} \right\} \quad (4.17a)$$

(proven by using the same procedure as for the general periodogram)

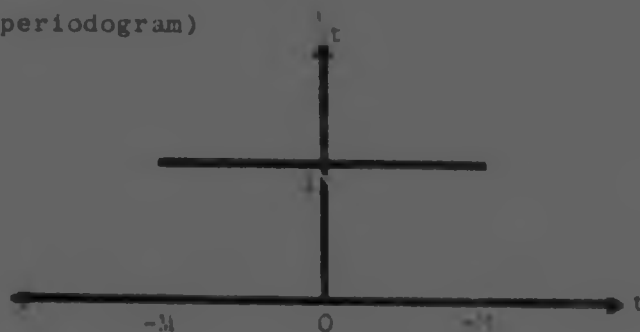


Fig. 4.2

3. Bartlett's window (a) (Bartlett (1950))

$$w_t = \begin{cases} 1 - \frac{|t|}{M} & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.18)$$

hence

$$W_N(\theta) = \frac{1}{2\pi} \left\{ \frac{\sin(\frac{\theta}{2})}{\sin(\frac{\theta}{2})} \right\}^2 \quad (4.18a)$$

$$\text{since } W_N(\theta) = \frac{1}{2\pi} \int_{-M}^M \left(1 - \frac{|t|}{M}\right) \cos(t\theta) dt$$

and this has been found already - see 'outline proof' after

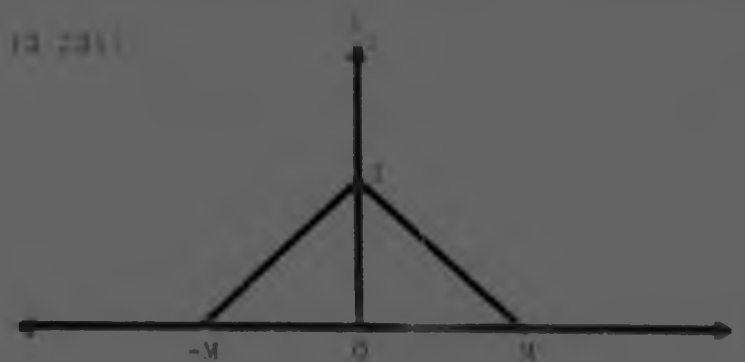


Fig. 4.3

4. Bartlett's window (b) (Bartlett (1950))

$$w_t = \begin{cases} (1 - \frac{|t|}{M})(1 - \frac{|t|}{N}) & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.19)$$

(for $W_N(\cdot)$, see Bartlett (1950); it is not given here since Bartlett's (a) window is the more renowned)

[This window enables one to work with c_t as an estimate of $\gamma(t)$ (instead of c_t') when estimating $h(w)$, the estimate of the sdf.

Bartlett originally divided the N observations into N/M subsamples, with each subsection containing M observations. He then computed, for each w , the average of the corresponding periodogram ordinates for each section.]

5. Daniell's window (Daniell (1946))

$$W_N(\cdot) = \frac{\sin(\frac{\pi}{M})}{\frac{\pi}{M}} \quad t = N \quad (4.20)$$

he e

$$W_N(\cdot) = \begin{cases} \frac{1}{2} & -\frac{\pi}{M} \leq \theta \leq \frac{\pi}{M} \\ 0 & \text{otherwise} \end{cases} \quad (4.20a)$$

(Daniell (1946), defined (4.20a) first and then worked backwards to obtain (4.20))

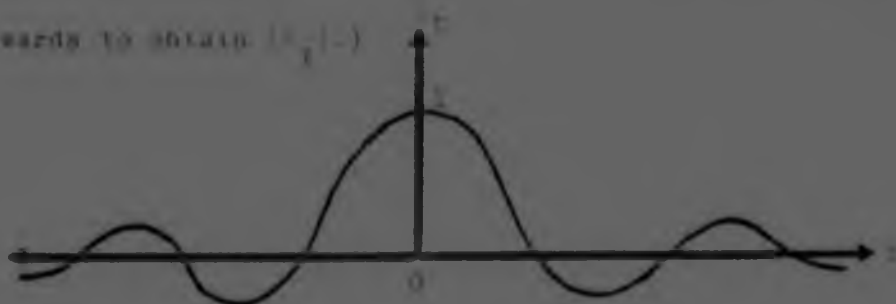


Fig 4.4

$h(w)$ is estimated by averaging the periodogram over a small interval centred at w (from $w - \frac{\pi}{M}$ to $w + \frac{\pi}{M}$ say) i.e.

$$h(w) = \frac{1}{2} \int_{w - \frac{\pi}{M}}^{w + \frac{\pi}{M}} I^*(\theta) d\theta$$

hence the form of W_N . For this window, M is not really a truncation point but instead determines the degree of smoothing

applied to $I^*(\theta)$. Small M implies greater smoothing.

6. The Blackman-Tukey window (Blackman and Tukey (1959))

$$\lambda_t = \begin{cases} 1 - 2a + 2a \cos\left(\frac{\pi t}{M}\right) & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.21)$$

$$0 < a \leq \frac{1}{4}$$

hence

$$w_N(\theta) = \frac{1}{2\pi} \int_{-M}^M \left\{ (1 - 2a) + 2a \cos\left(\frac{\pi t}{M}\right) \right\} e^{-it\theta} dt \quad (4.21a)$$

$$= aD\left(\theta - \frac{\pi}{M}\right) + (1 - 2a)D(\theta) + aD\left(\theta + \frac{\pi}{M}\right)$$

(i.e. a weighted average of the truncated windows - (4.17) - over $\theta - \frac{\pi}{M}$ and $\theta + \frac{\pi}{M}$)

$$\text{where } D(\pi) = \frac{1}{2\pi} \frac{\sin\left((M + \frac{1}{2})\pi\right)}{\sin\left(\frac{\pi}{2}\right)} \quad (= w_N(\theta) \text{ for (4.17a)})$$

(= Dirichlet's kernel)

(for proof use a similar technique to that of (4.16a))

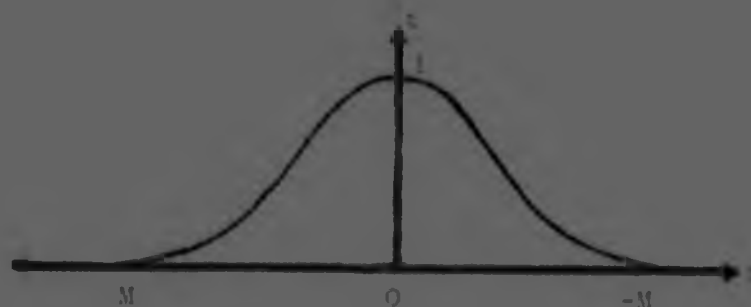


Fig. 4.5

The next two windows are just special cases of (4.21) (and hence (4.21a)) and so no graph of λ_t will be given.

(6a) The Tukey-Hamming window (Tukey (1949))

In (4.21), set $a = 0.23$

thus

$$\lambda_t = \begin{cases} 0.54 + 0.46 \cos\left(\frac{\pi t}{M}\right) & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.22)$$

(6b) The Tukey-Hanning window (Gottman(1981))

In (4.21), set $a = 0.25$

thus

$$w_t = \begin{cases} \frac{1}{2} \left(1 + \cos\left(\frac{\pi t}{M}\right) \right) & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.23)$$

7. Parzen's window (a) (Parzen (1957a))

$$w_t = \begin{cases} 1 - \left(\frac{t}{M}\right)^2 & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.24)$$

hence

$$w_N(\theta) = \frac{1}{2\pi} \left\{ \frac{\sin(M\theta) \cos\left(\frac{\theta}{2}\right)}{2M^2 \sin^2\left(\frac{\theta}{2}\right)} + \frac{\cos(M\theta)}{M \sin^2\left(\frac{\theta}{2}\right)} \right\} \quad (4.24a)$$

[Also see Jenkins (1961)]

w_t plotted against t has much the same form as figure 4.5.

8. Parzen's window (b) (Parzen (1961))

$$w_t = \begin{cases} 1 - 6\left(\frac{t}{M}\right)^2 + 6\left(\frac{t}{M}\right)^3 & |t| \leq \frac{M}{2} \\ 2\left(1 - \frac{t}{M}\right)^3 & \frac{M}{2} \leq |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.25)$$

hence

$$w_N(\theta) = \frac{3}{8M^3} \left(\frac{\sin\left(\frac{\theta}{2}\right)}{\frac{1}{2} \sin\left(\frac{\theta}{2}\right)} \left(1 - \frac{2}{3} \sin^2\left(\frac{\theta}{2}\right) \right) \right) \quad (4.25a)$$

(see Parzen (1961))

Figure 4.6 shows two functions - the Tukey-Hanning and the Parzen (b) windows. This enables us to make comparisons between the two windows

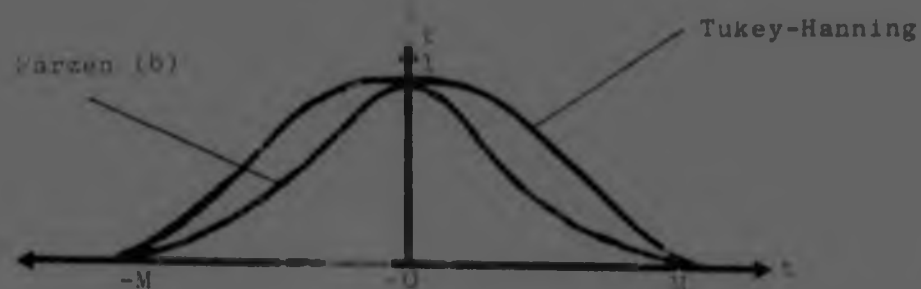


Fig. 4.6

9. The Bartlett-Friestley window (Priestley (1962))

$$\lambda_t = \frac{3M^2}{(\pi t)^2} \left\{ \frac{\sin(\frac{\pi t}{M})}{\frac{\pi t}{M}} - \cos(\frac{\pi t}{M}) \right\} \quad \text{for } |t| \leq M \quad (4.26)$$

hence

$$w_N(\theta) = \begin{cases} \frac{3M}{4\pi} \left(1 - \left(\frac{M^2}{\pi^2} \right)^2 \right) & \text{if } |\theta| \leq \frac{\pi}{M} \\ 0 & \text{otherwise} \end{cases} \quad (26a)$$

(see Priestley (1962))

As in the Daniell window, Priestley 'created' the quadratic spectral window $w_N(\theta)$ (which is an optimal window w.r.t. the relative MSE - see Epanechnikov (1969)) and then worked backwards to derive λ_t . As in the graph of the Daniell window (figure 4.4), λ_t does not truncate at M and $-M$. For both these cases, it is easier to plot the spectral window $w_N(\theta)$ (Daniell's is just linear, and Bartlett Priestley's is quadratic).

10. The Bohman window (Bohman (1961))

$$\lambda_t = \begin{cases} \left(1 - \frac{|t|}{M} \right)^2 \cos\left(\frac{\pi t}{M}\right) + \frac{\sin\left(\frac{\pi t}{M}\right)}{\pi} & \text{if } |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.27)$$

hence

$$\begin{aligned} \lambda_M(\omega) &= \frac{1}{2\pi} \sum_{|t| \leq M} \left(1 - \frac{|t|}{M}\right) \cos\left(\frac{\pi t}{M}\right) \cos(\omega t) \\ &= \frac{1}{2\pi} \sum_{|t| \leq M} \frac{\sin\left(\frac{\pi t}{M}\right)}{M} \cos(\omega t) \end{aligned} \quad (4.27a)$$

which is complicated (see Neave (1972)).

11. The Tukey-Parzen window (Parzen (1958))

$$\lambda_t = \begin{cases} 1 - 2a + 2a \cos\left(\frac{\pi t}{M}\right) & |t| \leq M \\ 0 & \text{otherwise} \end{cases} \quad (4.28)$$

and $a = 0.282$

This is of the same form as (4.21)

Neave (1972) has compared the performances of some of these windows - namely (3), (5), (6a), (6b), (7), (8), (10) and (11).

He concluded that the Bohman and Parzen (6) windows were superior to the others by considering the "negated derivative" of the lag windows, but that even these could be improved upon. To a lesser extent, the Tukey-Hamming window was also 'satisfactory'.

More discussion on the choice of spectral window (or lag window) will follow in the next chapter. The sampling properties of spectral estimates are given in appendix A7, but with no details or discussion. These sampling properties are given in the form of a series of results and mainly concern the variance and covariance of spectral estimates under certain conditions.

See appendix A7 for a brief explanation

In the next chapter, we shall see some of problems involved in spectral estimation. Since the object of this study is to model an observed time-series 'successfully', and since it turns out that modelling is 'easier' using a periodogram approach rather than a spectral analysis approach, the properties of spectral estimates are only covered in the barest detail in appendix A7.

CHAPTER 5. THE PROBLEMS INVOLVED IN SPECTRAL ANALYSIS

5.1 INTRODUCTION

Strictly speaking, the spectrum does not really enable us to model an observed time-series. Rather, if one calculates some spectrum estimate, we may be able to gain an idea of the model by comparing our spectrum estimate with known spectra, and also gain an idea about dominant frequencies. Alternatively, and this is the approach used in this study, we may remove all seasonal components (if any) from the model by using periodogram analysis (see chapter three).

We are then left with a set of residuals $a_0, a_1, \dots, a_{N-1}$ say. Using techniques to be described a little later (i.e. in chapter six onwards), it may be possible to model these residuals even further by Box-Jenkins modelling and hence reduce the residual variance.

Thus we shall hopefully be in the position where we have a model consisting of seasonal components (i.e. cosine and sine functions) and some nonseasonal components. Using this model, we can now 'regenerate' the data (using the model parameters) and compare the resulting spectrum with that of the original data (or even calculate the spectrum directly from the model).

However, certain problems, which are to be discussed in this chapter, still exist. These problems emphasize the difficulty of trying to compute the spectrum of the process, from which the time-series has come, 'accurately'.

See chapter 9.4, Comment, for a discussion of this technique

5.2 SCALE PARAMETER WINDOWS

It will be noted that for most of the spectral windows given in section 4.3.1., the truncation point, M , merely acted as a scale parameter i.e. contracted the function λ_t .

For such windows, it is possible to write λ_t in the form

$$\lambda_t = k\left(\frac{t}{M}\right)$$

where $k(u)$ is a fixed continuous even function in u , with $k(0) = 1$. [$k(u)$ is often called a lag window generator.]

For example, window (2) was the truncated periodogram. The form of $k(u)$ is

$$k(u) = \begin{cases} 1 & |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(since the original form of λ_t was

$$\lambda_t = \begin{cases} 1 & |u| \leq M \\ 0 & \text{otherwise} \end{cases}$$

So let $u = \frac{t}{M}$ and $k(u)$ follows.)

$w_N(\theta)$, the spectral window was originally expressed as

$$w_N(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda_t e^{-it\theta} dt \quad (\text{see (4.12)})$$

and we can now rewrite this as

$$w_N(\theta) = M \left\{ \frac{1}{2\pi M} \int_{-\infty}^{\infty} k\left(\frac{t}{M}\right) e^{-i\left(\frac{t}{M}\right)M\theta} dt \right\} \quad (5.1)$$

(assuming $k\left(\frac{t}{M}\right) = 0 \quad \forall t > M$)

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} k\left(\frac{t}{M}\right) e^{-it\theta} dt$$

$k(w)$ only applies when M really acts like a truncation point.

Also

$$K_N(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k(u) e^{-i u \theta} du \quad (5.2)$$

= Fourier transform of $k(u)$

hence $K_N(\theta)$ may be expressed i.t.o, $K_N(\theta)$

For M large, (5.1) may be approximated (i.e. the sum may be approximated) by (5.2) thus

$$W_N(\theta) \approx M K_N(M\theta) \quad (5.3)$$

When lag windows can be written in scale parameter form, more precise treatment of the sampling properties of spectral estimates is available. These results are given in 'result 10', appendix A7

The two main questions to be answered are

$$(i) \text{ which window do I use?} \quad (5.4)$$

and

$$(ii) \text{ what value should } M \text{ take?} \quad (5.5)$$

5.3 A MEASURE OF THE PRECISION OF SPECTRAL ESTIMATES

Obviously, the single question which (5.4) and (5.5) produce is

$$(iii) \text{ What } h(w) \text{ do I use?} \quad (5.6)$$

A way of solving this problem is to require that $h(w)$ has a certain precision. Parzen (1957) has suggested two techniques for assessing precision namely

$$(a) \quad P_1(w) = \frac{h(w)}{h(w)} = \frac{h(w)}{h(w)} \quad (5.7)$$

$k(w)$ only applies when M really acts like a truncation point.

Also

$$K_N(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k(u) e^{-i u \theta} du \quad (5.2)$$

= Fourier transform of $k(u)$

hence $W_N(\theta)$ may be expressed i.t.o, $K_N(\theta)$

For M large, (5.1) may be approximated (i.e. the sum may be approximated) by (5.2) thus

$$W_N(\theta) \approx M K_N(M\theta) \quad (5.3)$$

When lag windows can be written in scale parameter form, more precise treatment of the sampling properties of spectral estimates is available. These results are given in 'result 10', appendix A7.

The two main questions to be answered are

$$(i) \text{ which window do I use?} \quad (5.4)$$

and

$$(ii) \text{ what value should } M \text{ take?} \quad (5.5)$$

3 A MEASURE OF THE PRECISION OF SPECTRAL ESTIMATES

Obviously, the single question which (5.4) and (5.5) produce is

$$(iii) \text{ What } h(w) \text{ do I use?} \quad (5.6)$$

A way of solving this problem is to require that $h(w)$ has a certain precision. Parzen (1957) has suggested two techniques for assessing precision namely

$$(a) \quad P_1(w) = \frac{1}{2\pi} \frac{v(w)}{h(w)} + \frac{[h(w)]}{h(w)} \quad (5.7)$$

where $v^2(w) = V(h(w))$ = the variance of $h(w)$

$b(w) = E(h(w)) - h(w)$ = bias

and

$z_{\alpha} = 1 - \alpha/2$ fractile of the $N(0,1)$ distribution

Motivation:

Asymptotically

$$P\left\{\frac{h(w) - h(w)}{h(w)} \leq P_1(w)\right\} > \quad (5.8)$$

since

$$\begin{aligned} P\left\{\frac{h(w) - h(w)}{h(w)} \leq z\right\} \\ > P\left\{\frac{h(w) - h(w)}{h(w)} \leq z - \frac{b(w)}{h(w)}\right\} \end{aligned} \quad (5.9)$$

But $h(w)$ is asymptotically normal, and so the RHS of (5.9) is

set to 1 by defining

$$z = z \frac{v(w)}{h(w)} - \frac{b(w)}{h(w)} \quad \text{hence result}$$

Parzen's (1957) other measure of precision is

$$\begin{aligned} (b) \quad P_2(w) &= E(h(w) - h(w))^2 / h^2(w) \\ &= [v^2(w) + b^2(w)] / h^2(w) \end{aligned} \quad (5.10)$$

$$\text{Since } \lim_{N \rightarrow \infty} \left\{ \frac{N}{M} v^2(w) \right\} = (1 + \epsilon) h^2(w) \int_0^1 k^2(u) du$$

$$\text{and } b(w) = M^{-r} k^{(r)} h^{(r)}(w)$$

(by 'result 10', appendix A7)

where

$$\delta = \begin{cases} 1 & w = 0, \pm\pi \\ 0 & \text{otherwise} \end{cases}$$

$k(u)$ = scale parameter window

$$h^{(r)}(w) = \frac{1}{2\pi} \sum_{|t| < \infty} |t|^r \gamma(t) e^{-itw}$$

$$k^{(r)} = \lim_{x \rightarrow 0} \left\{ \frac{1 - k(x)}{|x|^r} \right\}$$

and r is the largest integer (>0) such that $k^{(r)}$ exists (and is finite) - see 'result 10', appendix A7

then

$$P_1(w) = z \left\{ \frac{M}{N} \int_{-\infty}^{\infty} k^2(x) dx \right\}^2 = \frac{k^{(r)}}{M^r} \frac{1}{(h_r)^r} \quad w \neq 0, -$$

where $h_r = |h(w) h^{(r)}(w)|^r$ (with r as before)

Also

$$P_2(w) = \frac{M}{N} \left\{ \int_{-\infty}^{\infty} k^2(x) dx \right\} + \left(\frac{k^{(r)}}{M^r h_r} \right)^2 \quad w \neq 0, -$$

Thus $P_1(w)$ and $P_2(w)$ are dependent on $h(w)$ (i.t.o. h_r), on $k^{(r)}$ and $\int_{-\infty}^{\infty} k^2(x) dx$.

Obviously, because of this dependence, neither $P_1(w)$ nor $P_2(w)$ is easily calculated. To evaluate $P_1(w)$ or $P_2(w)$, knowledge of $h(w)$ is required ($\forall w$), due to the dependence on h_r - this is an unrealistic requirement.

5.4 WINDOW BANDWIDTH

In appendix A7 (specifically, result 10) it is shown that

$$V(h(w)) = v^2(w) \cdot O\left(\frac{M}{N}\right) \quad (5.11)$$

and

$$b(w) = \text{bias of } h(w) = O(M^{-r}) \quad (5.12)$$

Hence, as M increases, the variance increases and the bias decreases, as M decreases, the variance decreases and the bias increases. Thus no matter how we choose M , $h(w)$ is either biased or its variance is large. We require some 'happy medium'. (This problem is analagous to power and significance levels in hypothesis testing).

Suppose $h(w)$ has two peaks, w_1 and w_2 , say. We want $h(w)$ to 'reproduce' $h(w)$ accurately, also with peaks at w_1 and w_2 . Now, by definition

$$E(h(w)) = \int_{-\pi}^{\pi} h(\theta) \eta_M(w - \theta) d\theta \quad (\text{see results 1 and 2, appendix A7}).$$

Suppose $\eta_M(\cdot)$ is the Daniell window (given by (4.20a)), thus

$$\eta_M(\theta) = \begin{cases} \frac{M}{2} & \text{if } |\theta| \leq \frac{\pi}{M} \\ 0 & \text{otherwise} \end{cases}$$

so

$$E(h(w)) = \frac{M}{2} \int_{w - \frac{\pi}{M}}^{w + \frac{\pi}{M}} h(\theta) d\theta$$

= average value of $h(\cdot)$ over an interval of width $\frac{2}{M}$ centred at w .

If M is so small that

$$\frac{2\pi}{M} > |w_2 - w_1|$$

then the two peaks will be indistinguishable i.e. they merge into one peak. So we need M large enough to be able to 'resolve' (distinguish) the two peaks i.e. we need

$$\frac{2\pi}{M} < |w_2 - w_1|$$

$$\text{i.e. } M > 2\pi / |w_2 - w_1|$$

Also, $\hat{h}(w)$ must, as accurately as possible, reproduce the shape of $h(w)$. This implies that $\hat{h}(w)$ must reproduce each peak and each trough! Thus M needs to be chosen such that the width of $\hat{h}(w)$ is not greater than the 'bandwidth' of the narrowest peak or trough.

Define the 'bandwidth', B , of a peak (or trough) at w_0 by the distance between w_1 and w_2 (where w_1 and w_2 are such that

$$w_1 = w_0 - \frac{B}{2}, \quad w_2 = w_0 + \frac{B}{2}$$

$$h(w_1) = h(w_2) = \frac{1}{2} h(w_0))$$

Priestley (1981, p 314), has defined B by

$$B = 2 h(w_0) / h'(w_0) \quad (5.13)$$

Thus if $h(w)$ consists of several peaks and troughs, define

$$B = 2 \left\{ \inf_w h(w) / h''(w) \right\}^{1/2}$$

and

$$h''(w) = - \frac{1}{2\pi} \int_{-\infty}^{\infty} t^2 \gamma(t) e^{-itw} dt \quad (5.14)$$

(by the definition of $h(w)$)

However we do not know $h(w)$. If we make B^* too small (and hence M large) any 'minor' fluctuation in $h(w)$ will be shown out of all proportion in $h(w)$.

But

$$|h''(w)| \leq \frac{1}{2} \left\{ \sum t^2 |\gamma(t)| \right\} \quad (\text{since } |\sum t^2 \gamma(t)| \leq \sum |t^2 \gamma(t)|)$$

thus if $\gamma(t)$ decays quickly, $|h''(w)|$ is small and B^* is large i.e. the rate at which $\gamma(t) \rightarrow 0$ is inversely proportional to B^* . Thus if we can estimate this rate of decay, it is possible to estimate B^* .

For windows like the truncated periodogram ((4.17a)), M represents the point beyond which $\gamma(t) \approx 0$. Thus if we can find this point, it may be possible to calculate B^* . Alternatively, choose B^* large (and so M small) and progressively decrease B^* , each time calculating $h(w)$. It may be possible to detect some point at which this 'unsmoothing' has been taken too far.

Finally, it is sensible to choose M to give a compromise between the bias and variance of $h(w)$. So we may choose M to minimize either $P_1(w)$ or $P_2(w)$ (defined by (5.7) and (5.10) resp.) If we have chosen our window, $W_w(t)$, evaluate $P_1(w)$ or $P_2(w)$ by using asymptotic values for the bias and variance, and then minimize $P_1(w)$ (or $P_2(w)$) w.r.t. M (i.e. find the derivative of either criterion w.r.t M and set this derivative equal to zero).

5.5 CHOICE OF WINDOW

So far we have only considered the estimation of M , the truncation point. However the other problem of which window to choose still exists. In fact, before we can even consider estimating M , we need to have decided upon a window.

If we consider the truncated periodogram window (given by (4.17a)), a graph of

$$w_N(\theta) = \frac{1}{2\pi} \left\{ \frac{\sin((M + \frac{1}{2})\theta)}{\sin(\frac{\theta}{2})} \right\}$$

will indicate a main 'lobe' at $\theta = 0$, but negative sidelobes at (approximately) $\frac{\pm 2\pi}{2M+1}$ and positive sidelobes at $\frac{\pm 4\pi}{2M+1}$ etc.
i.e.

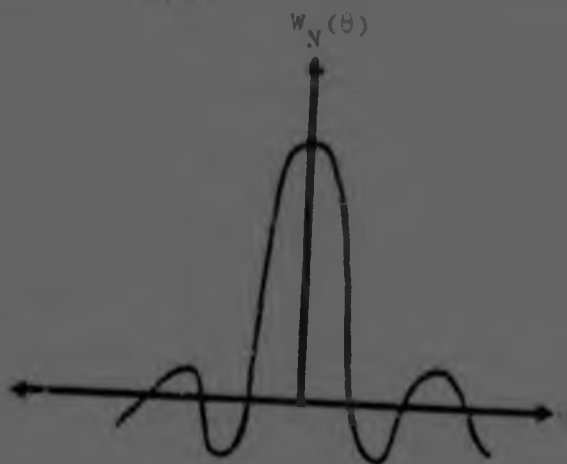


Fig. 5.1

Thus there is considerable contribution to $E(h(w))$ from the values of $h(\cdot)$ at $w = \frac{1}{2M+1}$, $w = \frac{3}{2M+1}$

Now if $\frac{d\theta}{d\omega} = 0$, say, happens to coincide with a peak in $h(\theta)$, then this could seriously affect the value of $E(h(\omega))$ i.e. leakage will occur.

Often, leakage reduction is used as a criterion for window selection - e.g. if sidelobe reduction is required (i.e. sidelobe leakage reduction), then the Daniell window (4.20a) which has no sidelobes is best. (However this window has large bias i.e. it suffers from periodogram leakage due to the bias in the periodogram).

The distortion due to sidelobes can be controlled by M . As M increases, the window contracts and so reduces the distance between subsidiary lobes and the main lobe.

Any objective approach to selection should take into account both the bias and variance of the window estimate. This approach should then give a criterion which determines the form of window which will optimize this criterion.

In appendix A8, the approximate variance of spectral estimates is given for the set of windows described in section 4.3.1. If one is prepared to ignore the bias in spectral estimates, these variances may be used as a criterion for window choice.

5.6 PREWHITENING AND TAPERING

These are techniques which try to improve the spectral estimates by transforming the data before any analysis is attempted.

8.6.1 PREWHITENING

This technique takes the original process $\{x_t\}$ and passes it through a filter to give a new process $\{y_t\}$ i.e. a transformation of the form

$$y_t = \sum_u a_u x_{t-u} \quad \forall t \quad \text{is made}$$

Then

$$\begin{aligned} h_y(\omega) &= \text{sdf of } \{y_t\} \\ &= |A(\omega)|^2 h_x(\omega) \end{aligned}$$

$$\text{where } A(\omega) = \sum_u a_u e^{-i\omega u}$$

Proof

$$h_y(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{r_y(t)} r_y(t) \cos(\omega t) dt$$

$$\text{but } r_y(t) = E((y_s - \bar{y})(y_{s+t} - \bar{y}))$$

$$\text{where } \bar{y} = E(y_1)$$

Without loss of generality, assume

$$E(x_t) = 0 \quad \forall t$$

then

$$\begin{aligned} r_y(t) &= E \left(\sum_p \sum_q a_p a_q x_{s-p} x_{s+t-q} \right) \\ &= E \left(\sum_p \sum_q a_p a_q x_{(s-p)-(s-t+q)} \right) \\ &= E \left(\sum_p \sum_q a_p a_q x_{(q-p-t)} \right) \\ &= \sum_p \sum_q a_p a_q \int_{-\pi}^{\pi} \cos(\omega(q-p-t)) h_x(\omega) d\omega \\ &= \int_{-\pi}^{\pi} \cos(\omega(q-p-t)) \left(\sum_p \sum_q a_p a_q e^{i\omega(q-p-t)} \right) h_x(\omega) d\omega \quad (\text{real part of}) \end{aligned}$$

$$= \int_{-\pi}^{\pi} \frac{1}{2\pi} e^{-i\omega t} \left[\frac{1}{p} a_p e^{-i\omega p} \right] \left[\frac{1}{q} a_q e^{i\omega q} \right] h_x(\omega) d\omega$$

$$= \int_{-\pi}^{\pi} \cos(\omega t) \left[\frac{1}{2} a_p a_q e^{i\omega(p-q)} \right] h_x(\omega) d\omega$$

$$h_y(t) = \left[\frac{1}{2} a_p a_q \right] h_x(t)$$

[For more on the choice of the transformation i.e. the a_u , see Bloomfield (1976), Chapter 6 . Complex demodulation]

5.6.2 TAPERING

Here the purpose is to reduce the bias in the spectral estimates, as well as to reduce leakage.

Consider the following situation:

suppose $x_t = e^{i\lambda t}$ $t = 0, 1, \dots, N-1$

$J(\omega)$ = Fourier transform of x_t

$$= \frac{1}{N} \sum_{t=0}^{N-1} x_t e^{-i\omega t}$$

$$= \frac{1}{N} \sum_{t=0}^{N-1} e^{it(\lambda - \omega)}$$

$$= \frac{1}{N} \frac{1 - e^{iN(\lambda - \omega)}}{1 - e^{i(\lambda - \omega)}} \quad \text{say, where } \lambda - \omega$$

$$= e^{i(N-1)(\lambda - \omega)/2} \frac{\sin(N(\lambda - \omega)/2)}{N \sin((\lambda - \omega)/2)} \quad \text{if } \lambda \neq \omega \quad (5.15)$$

(see note (viii) in section 3.4.1)

Let $J^*(w) = J(w)e^{i(N-1)w/2}$ (5.1)

then if $\phi \neq w$

$$J^*(w) = e^{i(N-1)\phi/2} \frac{\sin(\frac{N}{2})}{N\sin(\frac{1}{2})}$$

and

$$J^*(w - \frac{2\pi}{N}) = e^{i(N-1)\phi/2} \frac{\sin(\frac{N}{2}(\phi - w + \frac{2\pi}{N}))}{N\sin(\frac{1}{2}(\phi - w + \frac{2\pi}{N}))}$$

$$= -e^{i(N-1)\phi/2} \frac{\sin(\frac{N}{2}(\phi - w))}{N\sin(\frac{1}{2}(\phi - w) + \frac{\pi}{N})}$$

$$= -e^{i(N-1)\phi/2} \frac{\sin(\frac{N}{2}(\phi - w))}{N\sin(\frac{1}{2}(\phi - w))} \quad (\text{if } N \text{ is large})$$

$$\therefore J^*(w - \frac{2\pi}{N}) = -J^*(w)$$

$$\text{Similarly } J^*(w + \frac{2\pi}{N}) = -J^*(w)$$

$$\text{Define } JH(w) = \frac{1}{4} J^*(w + \frac{2\pi}{N}) + \frac{1}{2} J^*(w) + \frac{1}{4} J^*(w - \frac{2\pi}{N}) \quad (5.17)$$

$$\text{then } JH(0) = 0$$

$$JH(w) = \frac{1}{4N} e^{i(N-1)\phi/2} \frac{\sin(\frac{N}{2}(\phi - w))}{\sin(\frac{1}{2}(\phi - w))} + \left\{ \frac{2}{\sin(\frac{1}{2}(\phi - w))} - \frac{1}{\sin(\frac{1}{2}(\phi - w) - \frac{\pi}{N})} - \frac{1}{\sin(\frac{1}{2}(\phi - w) + \frac{\pi}{N})} \right\}$$

If N is large, then

$$\sin(\frac{1}{2}(\phi - w) - \frac{\pi}{N}) = \sin(\frac{1}{2}(\phi - w)) - \frac{\pi}{N} \cos(\frac{1}{2}(\phi - w))$$

Operation (5.17) is known as 'hanning' (Bloomfield (1976), pg 83)

hence the notation JH

and

$$\sin\left(\frac{\phi - w}{2} + \frac{\pi}{N}\right) = \sin\left(\frac{\phi - w}{2}\right) + \frac{\pi}{N} \cos\left(\frac{\phi - w}{2}\right)$$

hence, after simplification

$$JH(w) = \frac{-\pi^2}{2N^3} e^{i(N-1)\phi/2} \sin\left(\frac{N}{2}(\phi - w)\right) \frac{(\cos(\frac{\phi - w}{2}))^2}{(\sin(\frac{\phi - w}{2}))^3} \quad (5.18)$$

The leakage in $JH(w)$, given approximately by (5.18) is better behaved than that in $J^*(w)$ (given by (5.15)) (it contains a factor $1/N^3$ rather than $1/N$ and so it is smaller $\forall N$; also, $(\sin(\frac{\phi - w}{2}))^3$ decays as $1/|\phi - w|^3$ instead of $1/|\phi - w|$ as in $J^*(w)$ - thus the leakage is contained closely around $w = \phi$).

So, if we use (5.17) and (5.16) for any series $\{X_t\}$, then

$$JH(w) = \frac{1}{4} J^*(w_1) + \frac{1}{4} J^*(w_2) + \frac{1}{2} J^*(w) \quad (5.19)$$

$$w_1 = w - \frac{2\pi}{N}$$

$$w_2 = w + \frac{2\pi}{N}$$

$$\text{Now } J^*(\theta) = J(\theta) e^{i(N-1)\theta/2}$$

where

$$J(\cdot) = \frac{1}{N} \sum_{t=0}^{N-1} X_t e^{-i\theta t}$$

thus (5.19) becomes

$$JH(w) = \frac{1}{N} e^{iw(N-1)/2} \sum_t X_t \left[\frac{1}{4} e^{-i(N-1)(w - \frac{2\pi}{N})/2} e^{-it(w - \frac{2\pi}{N})} + \frac{1}{4} e^{-i(N-1)(w + \frac{2\pi}{N})/2} e^{-it(w + \frac{2\pi}{N})} + \frac{1}{2} e^{-iwt} \right]$$

$$\begin{aligned}
&= \frac{1}{N} e^{i\omega(N-1)/2} \left[x_1 e^{-i\omega t} \left(\frac{1}{2} + \frac{1}{4} e^{-i\left(\frac{N-1}{N} - \frac{2\pi t}{N}\right)} \right) \right. \\
&\quad \left. + \frac{1}{4} e^{-i\left(\frac{2\pi t}{N} - \frac{N-1}{N}\right)} \right]
\end{aligned}$$

But

$$\begin{aligned}
e^{-i(a-b)} &= e^{-i(b-a)} = \cos(b-a) + i\sin(b-a) \\
&= \cos(b-a) - i\sin(b-a) \\
&= 2(\cos(a)\cos(b) - \sin(a)\sin(b)) \\
&= 2(\cos(a-b))
\end{aligned}$$

$$\text{Let } a = \frac{N-1}{N} \text{ and } b = \frac{2\pi t}{N}$$

$$\therefore a - b = \frac{N-1}{N} - \frac{2\pi t}{N}$$

$$= \frac{2\pi}{N} \left(\frac{N-1}{2} - t \right) = \pi \text{ say}$$

$$JH(\pi) = \frac{1}{N} e^{i\omega(N-1)/2} \left[x_1 e^{-i\omega t} \left(\frac{1}{2} + \frac{1}{2} \cos(\pi) \right) \right]$$

$$= \frac{1}{N} e^{i\omega(N-1)/2} x_1 e^{-i\omega t} \left(\frac{1}{2} + \frac{1}{2} \cos(d) \right) \quad (5.20)$$

$$\text{where } d = \frac{2\pi}{N} \left(t - \frac{1}{2} \right)$$

Then

$$JH^*(\omega) = e^{-i\omega(N-1)/2} JH(\omega) \quad (5.21)$$

is the Fourier transform of the series

$$y_t = x_1 \left[\frac{1 - \cos\left(\frac{2\pi}{N} \left(t + \frac{1}{2}\right)\right)}{2} \right] \quad t = 0, \dots, N-1 \quad (5.22)$$

$$= x_1 x_1 \text{ say}$$

(5.21) is called the "Hanned transform" of x_t , and (5.22) shows that the alternative way to calculate $JH^*(w)$ is to multiply x_t by z_t $\forall t$ and then to take the Fourier transform of the series. $\{z_t\}$ is called a data window, and $JH^*(w)$ the Fourier transform of the tapered data.

Usually, we would like the transform z_t to leave the majority of the data unchanged - since the Fourier transform of a tapered sinusoid is the Fourier transform of the data window $\{z_t\}$ centred at the frequency of the sinusoid. So we only want to 'taper' the extreme values in $\{x_t\}$ i.e. the end values.

A convenient window that does this can be obtained from $\{z_t\}$ by inserting 1's in the middle stretch of data - thus we are only tapering the ends. The new data window is then

$$z_t = \begin{cases} \frac{1}{2} \left[1 - \cos \left(\frac{\pi}{k} \left(t - \frac{1}{2} \right) \right) \right] & t = 0, \dots, k-1 \\ 1 & t = k, \dots, N-k-1 \\ \frac{1}{2} \left[1 - \cos \left(\frac{\pi}{k} \left(N - t + \frac{1}{2} \right) \right) \right] & t \geq N-k \end{cases} \quad (5.23)$$

and k is such that $\frac{2k}{N}$ of the data is 'tapered'.

Takey (1967) suggests that 10% or 20% tapering may be satisfactory.

5.7 SUMMARY

In this chapter, we have seen some of the problems associated with computing an estimate of the spectrum. ~~Like the~~ ~~sdf~~ really does not assist in indicating model parameters (except that it does indicate dominant frequencies in the data, and thus can only be used like the periodogram). However, even considering the inherent problems in the periodogram (like leakage, failure of

the variance of periodogram estimates to decrease etc), in my opinion it is far easier to use the periodogram than the sdf (even though the periodogram is an unbiased, inconsistent estimate of the sdf).

The periodogram does not have the serious problems associated with the sdf mentioned above, and if tapering (section 5.4.2) is used some of the problems associated with leakage will be (in fact will be) eliminated to a large extent.

However if one decides that a spectral estimate is necessary then it is probably best to use the Parzen window (b) (see (4.25)) and a range of truncation points, M_i where $i = 1, 2, \dots, 5$ say (see Priestley (1981), Neave (1972), Gottman (1981)).

CHAPTER 6: ESTIMATION IN THE TIME-DOMAIN - BOX-JENKINS MODELS

6.1 INTRODUCTION

So far we have only been dealing with time-series in the frequency-domain i.e. we have been fitting models of the form

$$x_t = \sum_{j=0}^m \left[a_j \cos(\omega_j t) + b_j \sin(\omega_j t) \right] \quad (6.1)$$

some $m, \forall t$

In the frequency-domain, the time-series is considered i.t.o. waves (i.e cosine and sine functions) and the frequencies related to these waves. If the data are studied in terms of the auto-correlation function, then inferences based on this function are generally called an analysis in the time-domain.

If we assume that any oscillations of major frequencies (that explain variability in the time-series) have been removed (or 'modelled out') - that is if any such oscillations exist - we are left with a series of residuals $\dots, b_{-1}, b_0, b_1, \dots$ (or the original series $\dots, x_{-1}, x_0, x_1, \dots$) where

$$b_t = x_t - \sum_{j=0}^m \left[a_j \cos(\omega_j t) + b_j \sin(\omega_j t) \right] \quad (6.2)$$

$\forall t$

Now it is certainly feasible that these residuals follow some model (assuming we are dealing with a stationary process)

e.g.

$$b_t = a + b_1 + b_2$$

$$b_t = a + b_1 + b_2 + b_3 + \dots$$

$$b_t = a + b_1 + b_2 + b_3 + \dots$$

etc

(b_t = error terms)

In time-series analysis in the time-domain, the object is to try and find a model, based on inferences from the acrf, and then calculate it's parameters,

The strategies which will be used here are the strategies formalized by Box and Jenkins (1970,1976) and henceforth to be known as Box-Jenkins models (BJ models).

These models can be categorized into three classes

- (i) autoregressive (AR) processes
- (ii) moving average (MA) processes
- (iii) mixed AR-MA processes (ARMA).

The form of these processes is given in the following definitions.

Definition 15

- (i) The backward shift operator, B , is defined by

$$B^k x_t = x_{t-k}$$

k a non-negative integer

- (ii) The inverse operation is defined by the forward shift operator, F , and is such that

$$F^k x_t = x_{t+k}$$

k a non-negative integer

Definition 16

The backward difference operator, ∇ , is defined by

$$\nabla x_t = (1-B)x_t = x_t - x_{t-1}$$

$$\text{and } \nabla^k x_t = (1-B)^k x_t = \sum_{i=0}^k (-1)^i \binom{k}{i} x_{t-i} \quad (k \text{ some integer})$$

I have used the 2nd edition (1976) as my main source of reference for Box-Jenkins models

Definition 17

An autoregressive process of order p (or AR(p) process), $\{X_t\}$, (where $\{X_t\}$ is (second-order) stationary - see defn. 9, pg. 7 ($m=2$)) is such that

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t \quad (6.1)$$

where $E(X_t) = 0$ (w.l.o.g) $\forall t$ and $\{a_t\}$ is a purely random process. (i.e. white noise).

∴ if $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ (B = backward shift operator - see defn. 15, pg. 102) then

$$\phi(B)X_t = a_t \quad (6.3)$$

Definition 18

A moving average process of order q (or MA(q) process), $\{X_t\}$, is such that

$$X_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

where $\{a_t\}$ and $\{X_t\}$ are as in definition 17

If $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ (B = backward shift operator)

then

$$X_t = \theta(B)a_t \quad (6.4)$$

Definition 19

A mixed autoregressive - moving average process of orders p and q (or ARMA(p, q) process), $\{X_t\}$, is such that

$$\phi(B)X_t = \theta(B)a_t \quad (6.5)$$

with $\{X_t\}$, $\{a_t\}$, $\phi(B)$ and $\theta(B)$ as in definitions 17 and 18.

From this point, X_t will be considered as the time-series irrespective of whether the original series has been analysed using spectrum analysis i.e. X_t is the end result of the spectrum analysis.

Definition 20

The covariance matrix associated with a stationary process $\{X_t\}$, which has realization $x_0, x_1, \dots, x_{N-1}$, is given by

$$C_N = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(N-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \dots & \gamma(0) \end{bmatrix}$$

where $\gamma(1)$ is the acv at lag 1.

Note that

$$C_N = \frac{1}{N} P_N$$

where

$$P_N = \begin{bmatrix} \phi(0) & \phi(1) & \dots & \phi(N-1) \\ \phi(1) & \phi(0) & \dots & \phi(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(N-1) & \phi(N-2) & \dots & \phi(0) \end{bmatrix}$$

where $\phi(1)$ = acf at lag 1

and $\phi(0) = 1$ (and $\phi_X^2 = \phi(0)$)

P_N is called the autocorrelation matrix. Any acv or acf matrix for a stationary process $\{X_t\}$ is positive semi-definite and thus the determinant is always positive (unless the $a_t = 0 \forall t$).

[As in earlier chapters, c_k and r_k will be used as estimates of $\gamma(k)$ and $\phi(k)$]

All three of these processes (i.e. AR, MA and ARMA) turn out to be special cases of the general linear process

Second-order stationary

$$x_t = \sum_{j=0}^{\infty} a_j a_{t-j} \quad (\text{MA}(\infty)) \quad (6.6)$$

where $a_0 = 1$
and $\{a_t\}$ as before

In fact, Yule (1927), has shown that any (second-order) stationary process $\{X_t\}$ may be expressed in the form (6.6.). [Thus any AR(p) process can be expressed as an MA(∞) process (and vice versa) - see, for example, T.W. Anderson (1971)].

Due to these facts, the MA process will be considered now.

6.2 THE MA PROCESS

If $\{X_t\}$ is an MA(q) process (where from now on, unless stated otherwise, $\{X_t\}$ is assumed to be (second-order) zero-mean stationary, and q is finite), then by (6.4)

$$x_t = \theta(B) a_t \quad \forall t \quad (6.7)$$

where $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$
and $\{a_t\}$ = white noise.

Note that if we 'know' $\{a_t\}$, then for any $t_1, t_2, \dots, t_q$, $\{X_t\}$ in (6.7) is stationary. How do we estimate the θ_j ?

What we really need are estimates of $\theta_1, \dots, \theta_q$ as well as some idea about σ_a^2 (since $E(a_t) = 0 \quad \forall t$).

Firstly, suppose $q = 1$ i.e. we have an MA(1) process. Thus

$$x_t = a_t - \theta_1 a_{t-1} \quad \forall t.$$

$$\gamma(k) = E(x_t x_{t+k})$$

has zero-mean.

$$\begin{aligned}
 \gamma(k) &= E[(a_t - \theta_1 a_{t-1})(a_{t+k} - \theta_1 a_{t+k-1})] \\
 &= E(a_t a_{t+k}) - \theta_1 E(a_{t-1} a_{t+k}) \\
 &\quad - \theta_1 E(a_t a_{t+k-1}) \\
 &\quad + \theta_1^2 E(a_{t-1} a_{t+k-1})
 \end{aligned}$$

Setting $k = 0$,

$$\begin{aligned}
 \gamma(0) &= \sigma_X^2 \\
 &= (1 + \theta_1^2) \sigma_a^2
 \end{aligned}$$

$$\text{since } \gamma_a(k) = 0 \quad \forall k \neq 0$$

(by definition of (a_t))

For $k = 1$, it may be shown that $\gamma(1) = \theta_1 \sigma_a^2$ (due to the form of

$\gamma(k)$ and since $\gamma_a(k) = 0 \quad \forall k \neq 0$)

and if $k > 1$, $\gamma(k) = 0$

It turns out that if (X_t) is an MA(q) process that

$$\gamma(k) = 0 \quad \forall k > q \quad (\text{Box and Jenkins (1976)})$$

i.e. an MA(q) process has an acvf that truncates after lag q .

Also,

$$\gamma(k) = E\left[(a_t - \sum_{i=1}^q \theta_i a_{t-i})(a_{t+k} - \sum_{j=1}^q \theta_j a_{t+k-j})\right]$$

thus if $k = 0$

$$\begin{aligned}
 \gamma(0) &= \sigma_X^2 \\
 &= (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2
 \end{aligned} \tag{6.8}$$

and in general (see Box and Jenkins (1976) pg.68)

$$\begin{aligned}
 \gamma(k) &= (-\theta_k + \theta_1 \theta_{k-1} - \dots + \theta_q \theta_{q-k+1}) \sigma_a^2 \\
 &= \sigma_a^2 \sum_{i=1}^{q-k} (\theta_{k+i} - \theta_i \theta_{k-i+1}) \quad \forall k \leq q
 \end{aligned} \tag{6.9}$$

If we replace $\gamma(k)$ by σ_a^2 in (6.9) we have $(q+1)$ nonlinear equations for $(q+1)$ unknowns, θ_i . (Strictly speaking, $\theta_0 = 1$ and so is not an unknown.)

6.3 THE AR PROCESS

If $\{x_t\}$ is an AR(p) process then by (6.3)

$$\phi(B)x_t = a_t \quad \psi_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\text{and } \{a_t\} = \text{white noise series}$$

Again, as in section 6.2, we need to identify something about the AR(p) process in order to distinguish it from other processes (the MA(q) has acvf cut-off at lag q).

Consider the AR(1) process

$$1 - \phi_1 x_t - \phi_1 x_{t-1} = a_t$$

Multiplying this by x_{t-k} and taking expectations gives

$$\gamma(k) = \phi_1 E(x_{t-k} x_{t-1}) + \text{cov}(x_{t-k}, a_t)$$

But $\text{cov}(x_{t-k}, a_t) = 0$ (since x_{t-k} depends only on $a_{t-k}, a_{t-k-1}, \dots$ which are uncorrelated with a_t provided $k > 0$)

$$\gamma(k) = \phi_1 \gamma(k-1)$$

$$\text{or } \gamma(k) = \phi_1^k \gamma(0)$$

hence

$$\rho(k) = \phi_1^k \quad (\text{if } p = 1) \quad (6.9a)$$

and thus the correlogram of an AR(1) process does not truncate like an MA(1).

$$\text{Now } a_t = (1 - \phi_1 B)x_t = (1 - \phi_1 B)x_t$$

$$x_t = (1 - \phi_1 B)^{-1} a_t \quad (\text{MA}(\infty) \text{ see, for example, Anderson(1971) or Yule (1927)})$$

$$\text{if } |\phi_1| < 1$$

$$V(x_t) = \gamma(0)$$

$$= \sigma_a^2 \sum_{j=0}^{\infty} (\phi_1^j)^2$$

which is finite only if $|\phi_1| < 1$

So an AR(1) process is stationary only if $|\phi_1| < 1$ (unlike MA(q) processes which are stationary for all q).

6.3.1 THE YULE-WALKER EQUATIONS

Does some set of equations exist (as in the MA case) whereby for a general AR(p) process we can derive estimates of ϕ_i ($i=1,2,\dots,p$)?

Suppose $\{x_t\}$ is an AR(p) process, then

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t$$

and $\gamma(k) = E(x_t x_{t-k})$

But $\gamma(k) = E(x_t x_{t-k})$

$$\gamma(k) = \phi_1 \gamma(k-1) + \dots + \phi_p \gamma(k-p) + E(a_t x_{t-k}) = 0$$

($E(a_t x_{t-k}) = 0$ if $k > 0$ since x_{t-k} only involves the a_j up to time $t-k$ - and these are uncorrelated with a_t)

Thus if $k > 0$

$$\gamma(k) = \phi_1 \gamma(k-1) + \dots + \phi_p \gamma(k-p) \quad (\forall k \leq p) \quad (6.10)$$

and if $k = 0$

$$\begin{aligned} \gamma(0) &= \sigma_x^2 \\ &= \phi_1 \gamma(-1) + \dots + \phi_p \gamma(-p) + E(a_t^2) \end{aligned}$$

But

$$\gamma(k) = \gamma(-k)$$

$$\begin{aligned} \text{and } E(a_t^2) &= E[a_t(\phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t)] \\ &= E(a_t^2) \\ &= \sigma_a^2 \end{aligned}$$

$$\begin{aligned}
 \gamma(0) &= \frac{\sigma_a^2}{\sigma_x^2} \\
 &= \frac{\sigma_a^2}{\sigma_a^2(1 + \theta_1^2 + \dots + \theta_p^2)} = \frac{1}{1 + \theta_1^2 + \dots + \theta_p^2} \\
 &= \frac{\sigma_a^2}{\sigma_x^2(1 + \theta_1^2 + \dots + \theta_p^2)} = \frac{\sigma_a^2}{\sigma_x^2}
 \end{aligned} \tag{6.11}$$

and hence

$$\sigma_a^2 = \sigma_x^2(1 - \theta_1^2 - \dots - \theta_p^2) \tag{6.12}$$

Thus if we know the value of p , we can use (6.10) and (6.12) to estimate $\sigma_a^2, \theta_1, \dots, \theta_p$.

The equations (6.10) for $k = 1, 2, \dots, p$ are known as the Yule-Walker (Y-W) equations.

[Note that AR processes are easier to work with than MA processes due to the linearity of the Y-W equations.]

6.3.2 THE DUALITY BETWEEN AN MA(1) AND AN AR(1)

Suppose $\{X_t\}$ is an MA(1) process then

$$X_t = a_t + \theta_1 a_{t-1}$$

and so

$$X_{t-1} = a_{t-1} + \theta_1 a_{t-2}$$

$$a_{t-1} = X_{t-1} - \theta_1 a_{t-2}$$

thus

$$X_t = a_t + \theta_1(X_{t-1} - \theta_1 a_{t-2})$$

$$= a_t + \theta_1 X_{t-1} - \theta_1^2 a_{t-2}$$

But

$$X_{t-2} = a_{t-2} + \theta_1 a_{t-3}$$

so

$$X_t = a_t + \theta_1 X_{t-1} - \theta_1^2 X_{t-2} + \theta_1^3 a_{t-3}$$

Now provided $\theta_1^k \rightarrow 0$, then we may write

$$\begin{aligned}
 x_t &= \theta_1 x_{t-1} + \theta_2 x_{t-2} + \theta_3 x_{t-3} + \dots \\
 &= \theta_1 x_{t-1} + \theta_2 x_{t-2} + \theta_3 x_{t-3} + \dots
 \end{aligned}$$

But $\theta_1^k \rightarrow 0$ only if $|\theta_1| < 1$ thus an MA(1) may be written as an AR(∞) provided $|\theta_1| < 1$. This condition is termed an invertibility condition (and this is the same as the stationarity condition for an AR(1)).

This is also true for the general case (by induction we may show that the stationarity conditions of an AR(2) are identical with the invertibility conditions of an MA(2) etc. Also see section 6.6.)

What are the implications of all this? An MA(1) may be approximated by a high order AR process and an AR(1) by an MA(q) (q large). Similarly for MA(2) and AR(2).

6.3.3 THE AR(2) PROCESS

Suppose X_t is an AR(2) process. Then

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$$

By section 6.3.1 (i.e. equation (6.10))

$$\phi_1(1) = \phi_1 + \phi_2(1) \quad (6.12a)$$

$$\phi_2(1) = \phi_1(1) + \phi_2 \quad (6.12b)$$

and

$$(2) = \phi_2 + \phi_1(1 - \phi_2)$$

or

$$\phi_1 = (1)(1 - \phi_2) / (1 - \phi_1^2) \quad (6.13)$$

and

$$\phi_2 = (1 - \phi_1^2) / (1 - \phi_1^2)$$

Now $|\phi(1)| \leq 1$

and $|\phi(2)| \leq 1$

Also $\gamma_x^2 = \gamma_a^2 / [1 - \phi(1)\phi_1 - \phi(2)\phi_2]$ (by (6.12)) (6.13a)

and substituting here for $\phi(1)$ and $\phi(2)$ from the equations (6.12b), after some algebra it may be shown that

$$\gamma_x^2 = (1 - \phi_2^2) \gamma_a^2 / \{(1 + \phi_2)(1 - \phi_1 + \phi_2)(1 + \phi_1 - \phi_2)\} \quad (6.13b)$$

Thus, since $\gamma_x^2 > 0$

$$\text{and } |\phi(1)| \leq 1 \quad \forall 1$$

then each factor in γ_x^2 must be greater than, or equal to, zero

hence

$$-1 < \phi_2 < 1$$

$$\phi_2 = \phi_1 - \phi_2 \quad (\text{from (6.12b)})$$

$$\phi_2 - \phi_1 < 1$$

These are the stationarity conditions for an AR(2) process, hence

the invertibility condition for an MA(2) will be identical

(provided ϕ_1 is replaced by θ_1 and ϕ_2 by θ_2).

The following figures show the form of the acrf for various values

of ϕ_1 and ϕ_2 in the AR(2) process

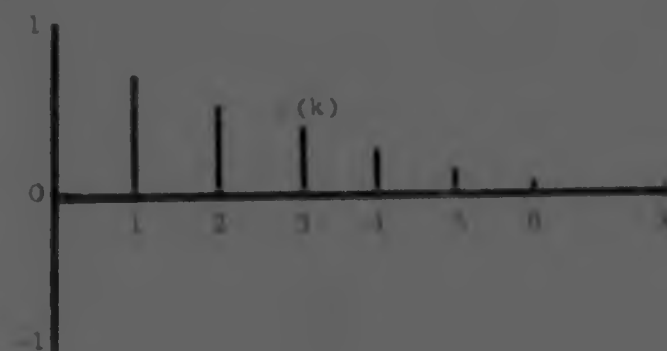
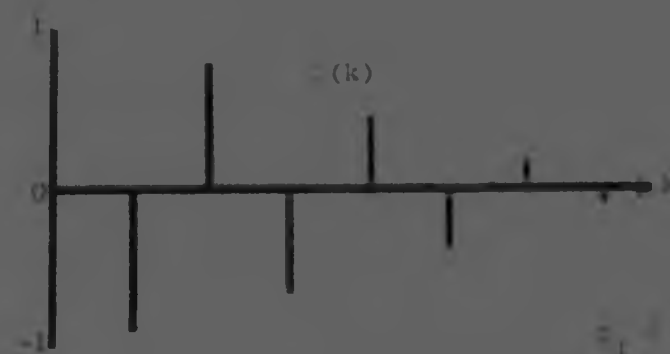


Fig 6.1

$$\phi_1 = 0$$

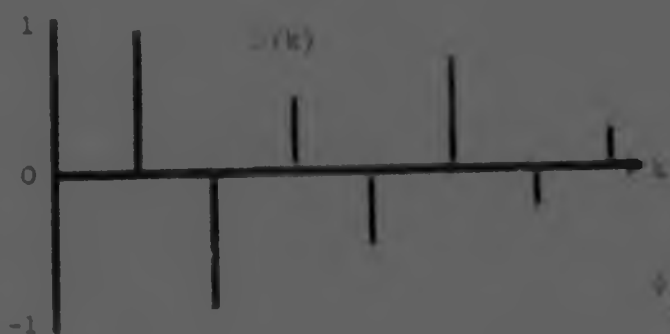
$$\phi_1^2 - 4\phi_2 > 0$$

For more detail see Box and Jenkins (1976), pg. 58

Fig 6.2
.....

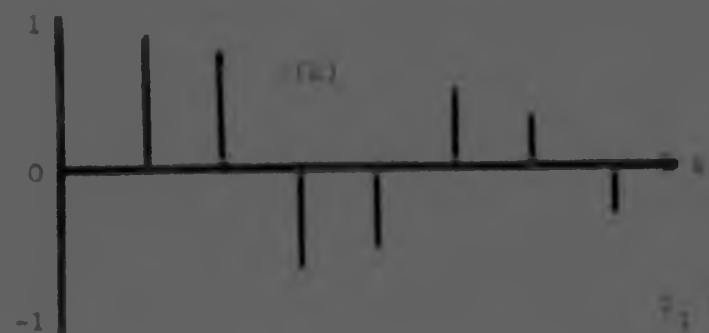
$$\phi_1 = 0$$

$$\phi_1^2 - 4\phi_2 = 0$$

Fig 6.3
.....

$$\phi_1 = 0$$

$$\phi_1^2 + 4\phi_2 = 0$$

Fig 6.4
.....

$$\phi_1 = 0$$

$$\phi_1^2 + 4\phi_2 = 0$$

Note that in all four cases here, the acf dies out (slowly).

Also note that in two cases the acf has cyclical form even though the AR(2) process itself is not cyclical!

6.4 THE PARTIAL AUTOCORRELATION FUNCTION (PACRF)

So far we have seen that it is possible to identify an MA(q) process from the acrf alone (if the acrf has cut-off after lag q). However this is not true for the AR(p) process. What we really require is some function which may be used in a similar way to identify p in an AR(p) process. Such a function exists and is known as the partial autocorrelation function (pacrf). This function completes the 'duality' between the MA and AR processes - the AR(p) process has cut-off after lag p and the MA(q) process dies out. The form of the pacrf and a proof that it can be used to identify p, is given below.

Consider the AR(1) process given in section 6.3. We know that

$$\rho(k) = \rho^k \quad (\text{see (6.9a)})$$

so observations one lag apart have a correlation of ρ , and observations two lags apart a correlation of ρ^2 i.e. observations x_{t+1} have a correlation with x_t . The question: "Is x_{t+3} independent of x_{t+1} , given x_{t+2} " can be answered by the partial correlation $\rho_{1,3|2}$ where $\rho_{1,3|2}$ is the correlation of x_{t+1} with x_{t+3} given x_{t+2} (or conditional on x_{t+2}). x_{t+3} is independent of x_{t+1} if $\rho_{1,3|2}$ is zero.

Now, after some algebra, it may be shown that

$$\rho_{1,3|2} = \frac{\rho_{1,3} - \rho_{1,2}\rho_{2,3}}{(1-\rho_{1,2}^2)(1-\rho_{2,3}^2)}$$

(see, for example, Gottman (1981))

where $\rho_{1,j} = \phi(1-j)$.

But $\rho_{1,3} = \rho(2) = \phi_1^2$

and $\rho_{1,2} = \rho(1) = \phi_1$

and thus

$$\rho_{1,3|2} = 0$$

i.e. there is no relationship between x_{t-1} and x_{t-3} once the intermediate association with x_{t-2} is removed. Similarly, all higher order partial correlations are also zero. So for an AR(1) process the partial correlations are zero after lag one

Consider the AR(2) process given in section 6.3.3. After algebraic manipulation it may be shown that

$$\rho_{1,3|2} = \frac{\phi(2) - \phi(1)^2}{1 - \phi(1)^2} = \phi_2$$

(see Gottman (1981))

and so, all higher order partial correlations (e.g. $\rho_{1,4|2,3}$) are zero (Box and Jenkins (1976)). Thus for an AR(2) process, the pacrf at lags 1 and 2 is nonzero, but is zero at all other lags.

Also note that the final nonzero partial correlation takes the value of the highest AR coefficient (i.e. ϕ_2) as in the AR(1) case. (This result holds in general for an AR(p) process - see, for example, Box and Jenkins (1976)).

Denote ϕ_i as the i^{th} AR coefficient in an AR(j) process. A plot of ϕ_i against i is termed the pacrf. The pacrf just tells us that if an AR(k-1) process has been fitted, then ϕ_{kk} is the 'gain' (or extra information) acquired by fitting an AR(k). Obviously if $\{X_t\}$ is really an AR(j) process then $\phi_{(j+1)(j+1)} = 0$ i.e. the cut-off point is 'lag j'.

Since any MA(q) process may be written as an AR(∞) process - see for instance Yule(1927) - then the pacrf of an MA(q) does not truncate after lag q (in fact it will have a pacrf of the same form as the acrf of an AR process).

Quenouille (1947) has shown that if $\{X_t\}$ is an AR(p) process then $\hat{\rho}_{jj}(j > p)$ has approximate variance of $1/N$ (where N is the number of observations in the time-series).

6.5 ARMA PROCESSES AND THE SDF OF BOX-JENKINS MODELS

6.5.1 ARMA PROCESSES

By definition 19, if $\{X_t\}$ is an ARMA(p,q) process then

$$\phi(B)x_t = \theta(B)a_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\text{and } \theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

Since $\{X_t\}$ has both an AR and an MA component, the acrf and pacrf are 'mixtures' of the functions for AR and MA processes.

The acvf satisfies

$$\begin{aligned} \gamma(k) &= \frac{1}{2}(\gamma(k-1) + \gamma(k-p) + E(x_{t-k} a_t) + E(x_{t-k-1} a_{t-1})) \\ &\quad + (1 + B(x_{t-k}^2) \theta_1 a_{t-1}) \end{aligned} \quad (6.14)$$

(see Box and Jenkins (1976) pg. 74)

Define

$$\gamma_{aa}(k) = E(a_{t-k} a_t) \quad (6.15)$$

= cross covariance function between

$$\{a_{t-k}\} \text{ and } \{a_t\}$$

then

$$\gamma_{aa}(k) = \begin{cases} 0 & k > 0 \\ \neq 0 & k \leq 0 \end{cases} \quad (6.16)$$

If $k > q + 1$, then (6.14) implies

$$\gamma_{aa}(k) = \phi_1 \gamma_{aa}(k-1) + \dots + \phi_p \gamma_{aa}(k-p)$$

and hence $\gamma_{aa}(k) = 0$

$$\gamma_{aa}(k) = 0 \quad (k > q) \quad (6.16a)$$

When $k = 0$, the variance of the process is

$$\begin{aligned} \gamma_{aa}(0) &= \sigma_a^2 \\ &= \phi_1^2 \gamma_{aa}(1) + \dots + \phi_p^2 \gamma_{aa}(p) + \sigma_a^2 \\ &= \phi_1^2 \gamma_{aa}(-1) + \dots + \phi_q^2 \gamma_{aa}(-q) \end{aligned} \quad (6.17)$$

6.3.2 THE SPECTRA OF AR, MA AND ARMA MODELS

Recall from section 5.4.1 that if

$$y_t = \sum_u a_u x_{t-u} \quad \forall t$$

then

$$\begin{aligned} h_y(\omega) &= \text{sdf of } Y_t \\ &= |A(\omega)|^2 h_x(\omega) \quad -\pi \leq \omega \leq \pi \end{aligned} \quad (6.18)$$

where

$$h_x(\omega) = \text{sdf of } X_t$$

$$\text{and } A(\omega) = \sum_u a_u e^{-i\omega u}$$

Now if X_t is an AR(p) process then

$$\phi(B)x_t = a_t \quad \forall t$$

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\{a_t\} = \text{white noise process}$$

$$a_t = \sum_{j=0}^p \phi_j x_{t-j} \quad \forall t \quad \text{and}$$

$$\phi_0 = 1 \text{ with } \phi_j = -\phi_j^* \quad j > 0$$

$$h_X(w) = |\phi(w)|^2 h_A(w) \quad -\pi \leq w \leq \pi \text{ by (6.18)}$$

where

$$\phi(w) = 1 - \sum_{j=1}^p \phi_j e^{-ijw}$$

i.e.

$$h_X(w) = \frac{h_A(w)}{|\phi(w)|^2}$$

but

$$h_A(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma_A(t) \cos(tw) \text{ by definition}$$

and

$$\gamma_A(t) = \begin{cases} 0 & t > 0 \\ \frac{1}{2} \sigma_a^2 & t = 0 \end{cases} \quad (\text{by definition of white noise})$$

$$\sigma_a^2 = \frac{1}{2\pi}$$

and

$$|\phi(w)|^2 = 1 - \sum_{j=1}^p \phi_j e^{-ijw} - \sum_{j=1}^p \phi_j^* e^{-ipw}$$

so the sdf of an AR(p) process is given by

$$h(w) = \frac{\frac{1}{2\pi}}{1 - \sum_{j=1}^p \phi_j e^{-ijw} - \sum_{j=1}^p \phi_j^* e^{-ipw}} \quad -\pi \leq w \leq \pi \quad (6.19)$$

If X_t is an MA(q) process

then

$$X_t = \theta(B) a_t \quad \forall t$$

where $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$

thus

$$h_X(w) = |\theta(w)|^2 h_a(w)$$

where

$$\theta(w) = 1 - \sum_{j=1}^q \theta_j e^{-ijw}$$

$$\text{i.e. } h_X(w) = \frac{\left| 1 - \sum_{j=1}^q \theta_j e^{-ijw} \right|^2}{2\pi} \quad -\pi \leq w \leq \pi \quad (6.20)$$

Similarly, if $\{X_t\}$ is an ARMA(p,q) process then

$$h_X(w) = \frac{\frac{2}{a} \left| 1 - \sum_{j=1}^q \theta_j e^{-ijw} \right|^2}{2\pi \left| 1 - \sum_{j=1}^p \phi_j e^{-ijw} \right|^2} \quad (6.21)$$

$$-\pi \leq w \leq \pi$$

So the form of the theoretical spectrum is easily calculated for any of these processes provided the model parameters are known. Even if the parameters are unknown, but we have estimates thereof then we may plot the spectrum of $\{X_t\}$ using the parameter estimates and then compare this spectrum with the spectrum (estimate) of the original data.

6.6 A GENERAL NOTE ON STATIONARITY AND INVERTIBILITY

Consider the general linear process $\{X_t\}$ defined by (6.8)

$$\text{i.e. } X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad \forall t, \psi_0 = 1$$

$$= \psi(B) a_t \quad \text{say} \quad (6.22)$$

where B = backshift operator

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

and $\{a_t\}$ = white noise process

Thus

$$\gamma_a(k) = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

and
$$a(k) = \begin{cases} 0 & k \neq 0 \\ 1 & k = 0. \end{cases}$$

Note that

$$\begin{aligned} x_t - \beta_1 x_{t-1} &= a_t - \beta_1 a_{t-1} + \beta_1 x_{t-1} - \beta_1 \sum_{j=1}^{\infty} \alpha_j a_{t-1-j} \\ &= a_t + (\alpha_1 - \beta_1) a_{t-1} + \sum_{j=2}^{\infty} \beta_1 \alpha_j a_{t-1-j} \end{aligned}$$

where $\alpha_j = x_j - \beta_1 x_{j-1}$

and choosing

$$\begin{aligned} \alpha_1 - \beta_1 &= 0 \\ \text{i.e. } \alpha_1 &= \beta_1 = \pi_1 \text{ say} \end{aligned}$$

thus

$$x_t - \pi_1 x_{t-1} = a_t + \sum_{j=2}^{\infty} \pi_1 \alpha_j a_{t-1-j} \quad \forall t$$

Similarly

$$\begin{aligned} x_t - \pi_1 x_{t-1} - \pi_2 x_{t-2} &= a_t + \sum_{j=2}^{\infty} \pi_1 \alpha_j a_{t-1-j} - \pi_2 a_{t-2} - \pi_2 \sum_{j=1}^{\infty} \alpha_j a_{t-2-j} \\ &= a_t + (\pi_2' - \pi_2) a_{t-2} + \sum_{j=3}^{\infty} \pi_1 \alpha_j a_{t-1-j} \text{ say.} \end{aligned}$$

Again we may choose $\pi_2' = \pi_2$

$$\text{i.e. } \pi_2' = \pi_2 \text{ say}$$

$$\pi_2' - \pi_2 = \pi_2' - \pi_2$$

Continuing in this manner we eventually obtain

$$x_t - \sum_{j=1}^{\infty} \pi_j x_{t-j} = a_t$$

or

$$x_t = a_t + \sum_{j=1}^{\infty} \pi_j x_{t-j}$$

Writing

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$$

thus $\pi(B)x_t = a_t$

But $x_t = \chi(B)a_t$

$$\therefore \pi(B) = (\chi(B))^{-1} \quad (6.23)$$

(B now a dummy variable)

Thus if $\{X_t\}$ is stationary, we may write an $MA(\infty)$ as an $AR(\infty)$.

Now $\gamma_X(k) = E(x_t x_{t-k}) \quad \forall t$

$$= E\left(\sum_{j=0}^{\infty} \alpha_j a_{t-j} \sum_{l=0}^{\infty} \alpha_l a_{t-k-l}\right)$$

$$= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \alpha_j \alpha_{j+k} \quad \alpha_0 = 1 \quad (6.24)$$

and for $k = 0$, $\gamma_X(0) = \sum_{j=0}^{\infty} \alpha_j^2$

$$= \sigma_a^2 \sum_{j=0}^{\infty} \alpha_j^2 \quad (\alpha_0 = 1) \quad (6.25)$$

Thus the α weights must decrease fast enough to ensure that $\{x_t\}$ has a finite variance, or more specifically, $\sum_{j=0}^{\infty} \alpha_j^2$ must converge

Define

$$\Gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_X(k) B^k \quad (6.26)$$

(called the autocovariance generating function)

then $\Gamma(B) = \sigma_a^2 \chi(B) \chi(F) \quad (6.27)$

where $F =$ forward shift operator.

((6.27) is obtained after substituting (6.24) in (6.26) - see Box and Jenkins (1976), pg.81).

$I_X(w)$, the sample periodogram of $\{X_t\}$ is defined by

$$I_X(w) = \frac{1}{2} \sum_{k=-N}^N c_k' \cos(wk) \quad -\pi \leq w \leq \pi \quad (6.28)$$

(see section 4.1)

and $\lim_{N \rightarrow \infty} E(I_X(w)) = h_X(w)$ the sdf

$$= \frac{1}{2\pi} f(x)$$

where $f(x)$ is the nsdf

Replacing B by e^{-1w} in (6.26) gives

$$\begin{aligned}\Gamma(e^{-1w}) &= \sum_{k=-\infty}^{\infty} \gamma_X(k) e^{-1wk} \\ &= \sigma_a^2 \lambda(e^{-1w}) \lambda(e^{1w}) \text{ since } F = B^{-1} \text{ (see defn. 15)}\end{aligned}$$

$$\begin{aligned}\text{i.e. } \alpha(e^{-1w}) \alpha(e^{1w}) &= \frac{1}{\sigma_a^2} \sum_{k=-\infty}^{\infty} \gamma_X(k) e^{-1wk} \\ &= \frac{2\pi}{\sigma_a^2} h_X(w)\end{aligned}$$

$$\text{thus } h_X(w) = \frac{\sigma_a^2}{2\pi} |\lambda(e^{-1w})|^2$$

Now

$$\begin{aligned}\gamma_X(0) &= \frac{\sigma_a^2}{2\pi} \\ &= \int_{-\pi}^{\pi} h(w) dw \text{ by definition}\end{aligned}$$

and so $\frac{\sigma_a^2}{2\pi}$ is finite only if $\lambda(B)$ converges - which occurs only when $|B| < 1$.

If $\lambda(B) = 1 - \sum_{q=1}^{\infty} B^q$, i.e. X_t is an MA(q) process, we still require $|B| < 1$ or equivalently we may ask that $\lambda(B)$ has weights λ_q such that $\lambda(B)$ converges on or within the unit circle. Similarly for $\pi(B)$.

Thus we have shown, more formally, the 'equivalence' of stationarity and invertibility conditions for AR and MA processes respectively (since AR and MA processes are just special cases of a general linear process and since $\lambda(B) = (\pi(B))^{-1}$ and vice versa).

Strictly speaking, B is only an operator, so $|B| < 1$ has no real meaning. B must be considered as a dummy variable.

6.7 NOTES ON THE APPENDICES A9-A11

Appendix A9 gives a procedure, suggested by Box and Jenkins (1976), for obtaining 'initial' estimates in an ARMA (p,q) process (given p and q). A10 provides a recursive method for calculating the pacrf.

In A11, the properties (i.e. variance, acrf, pacrf etc) of the 'main' BJ models are given. These 'main' models are the AR(1), AR(2), MA(1), MA(2) and ARMA(1,1) processes. Also provided are the covariance matrices.

Anderson (1977) has also derived an iterative procedure for calculating parameter estimates (given the order of the model i.e. p and q) in both the frequency and time-domains. These procedures are in matrix form and may be applied to any realization

$x_0, x_1, \dots, x_{N-1}$

CHAPTER 7: THE BOX-JENKINS STRATEGY

Box-Jenkins modelling, or the Box-Jenkins strategy, consists of five phases, namely

- (i) inducing stationarity
- (ii) identification of the model
- (iii) estimation of model parameters
- (iv) checking the fit of the model
- and (v) forecasting

In this chapter, the first four phases will be dealt with in detail and a note made of the fifth.

7.1 PHASE ONE: INDUCING STATIONARITY

If our original series was $\{Y_t\}$ (not necessarily stationary), then once the harmonic components (if any) have been removed, we are left with a set of residuals, $\{Z_t\}$ say. Now these $\{Z_t\}$ may be stationary or nonstationary. However, we saw in the last chapter that BJ models may only be applied to (second order) stationary processes. Thus the first stage of the BJ strategy consists of checking the condition of the residuals $\{Z_t\}$.

If $\{Z_t\}$ is a stationary process then we may proceed directly to phase two. However, if $\{Z_t\}$ is nonstationary, we would like to transform the data (i.e. $\{Z_t\}$) into a stationary process $\{X_t\}$.

Time-series, like any stochastic process, may be nonstationary in various ways. However, in section 6.6, when considering an ARMA (p,q) process $\{X_t\}$,

$$\text{i.e.} \quad \phi(B)x_t = \theta(B)a_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

$$\{a_t\} = \text{white noise process}$$

To ensure stationarity, we required that the roots of $\phi(B) = 0$ lie outside the unit circle. (Similarly to ensure invertibility, we required that the roots of $\theta(B) = 0$ lie outside the unit circle).

Thus any process which has roots of $\phi(B) = 0$ lying inside the unit circle, is nonstationary. The process will also be nonstationary if the roots lie on the unit circle. Since the behaviour of the process is 'explosive' if the roots fall (strictly) inside the unit circle (see, for instance, Box and Jenkins (1976), p.88), we shall only consider the case where roots fall either on or outside the unit circle.

So consider the model

$$I^d(B) z_t = \theta(B) a_t \quad \forall t \quad (7.1)$$

where $I^d(B) = 0$ has d roots, say, lying on the unit circle, and the rest outside.

We may rewrite (7.1) as

$$\begin{aligned} I^d(B) z_t &= (1-B)^d I(B) z_t \\ &= \theta(B) a_t \end{aligned} \quad (7.2)$$

where $I(B)$ has all roots outside the unit circle.

$$\begin{aligned} \text{But } (1-B)^d I(B) &= \phi(B) (1-B)^d \\ \text{and } \phi &= 1-B \quad (\text{see definition 16 of section 6.1}) \end{aligned}$$

i.e. the process is fluctuating wildly with no pattern

thus

$$\begin{aligned} (1-B)^d \phi(B) z_t &= \phi(B) (1-B)^d z_t \\ &= \phi(B) x_t \quad \text{say (where } x_t = (1-B)^d z_t \text{)}. \end{aligned}$$

Now $\{x_t\}$ is a stationary process since $\phi(B)$ has all its roots outside the unit circle and so it has an ARMA(p,q) representation. [$\phi(B)$ is called a stationary AR operator and $(1-B)^d$ a nonstationary AR operator].

Note that $x_t = (1-B)^d z_t$ is just the d^{th} successive 'difference' of z_t (see defn. 16, section 6.1).

Representation (7.2) is called an ARMA(p,d,q) process (AR integrated MA) if $d > 0$ (otherwise it is just an ARMA(p,q) process).

So, if we are given any process $\{z_t\}$, we must determine d in order to be sure that

$$(1-B)^d z_t = x_t$$

is stationary $\forall t$.

If d is made too large, we may overdifference the data and possibly introduce meaningless patterns into the transformed series. So care must be taken when deciding on the size of d .

What indication is there that $\{z_t\}$ may be nonstationary? In section 6.5 (specifically (6.16a)) it was shown that if $\{x_t\}$ is an ARMA(p,q) process then

$$\phi(B) \quad (k) = 0 \quad k > q \quad (7.3)$$

Now if $\phi(B)$ is a stationary AR operator (i.e. all roots of $\phi(B) = 0$ lie outside the unit circle) then we may write

$$\phi(B) = \prod_{i=1}^p (1 - s_i B) / \prod_{i=1}^q S_i$$

where s_i are the roots of $\phi(B) = 0$. Hence

$$\prod_{i=1}^p (1 - s_i B) = 0 \quad k = q$$

If the s_i are distinct (i.e. $s_i \neq s_j \forall i \neq j$) then the solution of the difference equation (7.3) is of the form

$$\phi(k) = \alpha_1 s_1^k + \dots + \alpha_p s_p^k \quad (k > q - p) \quad (7.4)$$

(by the theory of linear difference equations - see, for instance, Box and Jenkins (1976), pg.116). Since the roots of $\phi(B) = 0$ all lie outside the unit circle, then this implies that $|s_i| < 1 \forall i$, thus (7.4) tends rapidly to zero as k increases.

If just one of the roots, s_1^{-1} say, tends to unity (i.e. $s_1^{-1} = 1 - \epsilon$, $\epsilon > 0$) then if k is large

$$\begin{aligned} \phi(k) &= \alpha_1 \left(\frac{1}{1 - \epsilon} \right)^k + \text{terms tending rapidly to zero} \\ &\approx \alpha_1 (1 + k\epsilon) \end{aligned}$$

and so the acrf does not die out quickly.

Thus an indication of nonstationarity is the acrf not dying out quickly!

Using the sample acrf as an estimate of the theoretical acrf, we may ascertain the stationarity or nonstationarity of a process

(9)

If we conclude Z_t is non-stationary then it must be differenced d times (unless some of the roots lie inside the unit circle - i.e. the series shows explosive behaviour) i.e. find

$$\nabla^d Z_t = X_t$$

What value should d take?

Box and Jenkins (1976) suggest that d is invariably either 0 or 1 or 2 (see Box and Jenkins (1976) pg.173). So difference once, and if the acrf of the resulting series dies out 'quickly' we have

obtained stationarity. If not, carry on differencing until the acrf of the resulting series does in fact die out.

Gottman (1981, pg.262) has suggested that for each d , you calculate

$$D_d = \sum_{k=0}^{N^*} \rho^2(k,d) \quad (7.5)$$

where $\rho(k,d) = k^{\text{th}}$ acf of the d^{th} differenced series

and $N^* = N/6$ (this is in order to be able to compare the D_d across different values of d).

Estimate D_d by $\hat{D}_d$ (replace $\rho(k,d)$ by $r_{k,d}$) and then choose that value of d which gives a minimum value for $\hat{D}_d$.

Hill and Woodworth (1980) suggest using that value of d which minimizes $\chi^2(d)$

$$\text{i.e. } \min_d \chi^2(d) ; \chi_t^2 = \sum_{t=1}^n z_t^2 \quad (7.6)$$

Instead of using (7.5) or (7.6), it may be possible to determine when the (sample) acrf has effectively died out, by noting that

$$V(r_k) = \frac{1}{N} \sum_{t=-\infty}^{\infty} \left\{ r^2(t) + r(t+k)r(t-k) - 4r(t) \cdot r(k) \cdot r(t-k) + 2r^2(t)r^2(k) \right\} \quad (7.7)$$

if $\{X_t\}$ is a Gaussian process (see (A3.3), appendix A3).

If the acrf has died out after lag v say, then (7.7) may be rewritten as

The sample acf - see sections 2.4 and 2.5 (pages 33 and 37 resp.)

$$V(r'_k) \approx \frac{1}{N} \left(1 + 2 \sum_{i=1}^v c^2(i) \right) \quad k > v \quad (7.8)$$

thus

$$V(r'_k) \approx \frac{1}{N} \left(1 + 2 \sum_{i=1}^v r_i'^2 \right) \quad k > v \quad (7.9)$$

[So, for example, if $\{X_t\}$ is a (Gaussian) white noise process, then $v = 0$ (by definition of the acrf of a white noise process)

and

$$\sqrt{V(r'_k)} = \sigma(r'_k) \text{ say} \quad (7.10)$$

$$= \frac{1}{\sqrt{N}}$$

[= standard error]

Using (7.9), or (7.10), each lag in the (sample) acrf may be tested (successively) for a significant deviation from a zero value (since $H_0: \gamma(k) = 0 \quad k > v$)

and thus

$$\frac{r'_k - 0}{\sqrt{V(r'_k)}} \sim N(0,1) \text{ using the normal approximation to}$$

the distribution of r'_k : see Anderson (1942)). However the problem with this technique is the effect of multiple testing on the overall significance level.

7.2 PHASE TWO: IDENTIFICATION OF THE MODEL

Having induced stationarity by phase one, we are left with a (second-order) zero-mean stationary process $\{X_t\}$, and thus all the techniques from chapter six are available

We now wish to ascertain (i.e. estimate) the values of p and q in the ARMA(p, q) model

$$\phi(B)x_t = \theta(B)a_t.$$

If $p = q = 0$ then $x_t = a_t \quad \forall t$ (i.e. $\{x_t\}$ is a white noise process) and all we need now do is check the fit of the model (i.e. the 'whiteness' of the residuals).

Otherwise we must estimate p and q and hence the parameters in $\phi(B)$ and $\theta(B)$.

In chapter six, we saw some of the properties of the acrf and pacrf for AR, MA and ARMA processes. Recapping briefly, these were

- (a) the AR(p) has cut-off after lag p in the pacrf
- (b) the MA(q) has cut-off after lag q in the acrf.
- (c) if $\{x_t\}$ is an ARMA(p, q) (where p and q are both nonzero) then the acrf and pacrf do not have cut-off points
- (d) an AR(p) does not truncate in the acrf and neither does an MA(q) in the pacrf.

Thus if both the acrf and the pacrf tail off, an ARMA(p, q) is suggested, otherwise the form of the acrf or pacrf should suggest whether an AR(p) or MA(q) is likely. It is possible to ascertain a cut-off in the acrf by making use of the $V(r_k)$ (i.e. (7.9)).

Truncation in the pacrf is "determined" by noting that

$$\hat{V}(\hat{\phi}_{kk}) \approx \frac{1}{N} \quad k > p \quad (7.11)$$

if $\{x_t\}$ is an AR(p) process (see Quenouille (1949)), thus

$$\frac{\hat{\phi}_{kk}}{\hat{\sigma}_{kk}} \xrightarrow{d} N(0, 1)$$

where

$$\hat{\sigma}_{kk} = \sqrt{\hat{V}(\hat{\phi}_{kk})}$$

(see the final paragraph on pg. 114 for a definition of $\hat{\phi}_{kk}$)

If $\{X_t\}$ is an ARMA (p, q) process with both p and $q > 0$, then the number of anomalous (i.e. not a mixture of exponential decay and damped sine waves) terms in the acrf is $q - p$, and the number of anomalous terms in the pacrf is $p - q$ (see Box and Jenkins (1976), pg. 186). Note that $\rho(k) = \rho(-k)$, so the case when $p - q$ (or $q - p$) is negative is no problem. The number of anomalous terms only gives an idea of the difference between p and q and not about their actual sizes. In practise, both p and q are usually ≤ 3 .

7.2.1 THE INVERSE ACRF AND INVERSE PACRF

Another tool useful in the identification of p and q in an ARMA (p, q) process is the inverse autocorrelation function (iacrf) and analogously the inverse partial autocorrelation function (ipacrf) (or similarly the iacvf and ipacvf).

If X_t is stationary with acvf $\gamma(k)$ and acrf $\Gamma(k)$ ($k = 0, 1, \dots$), then the autocovariance generating function (acvgf) is defined by

$$\Gamma(Z) = \sum_{k=-\infty}^{\infty} \gamma(k) Z^k \quad (7.12)$$

and the autocorrelation generating function (acrgf) is defined by

$$P(Z) = \sum_{k=-\infty}^{\infty} \rho(k) Z^k \quad (7.13)$$

Then the inverse acvgf (iacvgf) is defined by $I(Z)$ where

$$\Gamma(Z) I(Z) = 1 \quad (7.14)$$

The coefficient in $I(Z)$ of Z^k is defined to be the inverse acv coefficient at lag k (and is denoted by $\gamma_I(k)$).

Similarly the inverse acrgf (or iacrgf) is defined by $P_I(Z)$ where

$$P(Z)P^*(Z) = 1 \quad (7.15)$$

and the coefficient of Z^k in $P^*(Z)$ is the inverse acf coefficient at lag k (and denoted $\gamma^*(k)$).

$$\text{Also} \quad \rho^*(k) = \gamma^*(k)/\gamma^*(0) \quad (7.16)$$

thus

$$P^*(Z) = \Gamma^*(Z)/\Gamma^*(0) \quad (7.17)$$

and

$$P(Z)P^*(Z) = [\gamma(0)\gamma^*(0)]^{-1} = 1 \quad (7.18)$$

(since $P(Z)P^*(Z) = [\gamma(0)\gamma^*(0)]^{-1}$ but $\gamma(0) = \gamma^*(0)$ and

$$\gamma^*(0) = \gamma(0))$$

where in general $\gamma^*(0) = (\gamma(0))^{-1}$

(and in fact $\gamma^*(k) = \gamma(k)^{-1}$ in general)

$$\text{Now if} \quad \phi(B)x_t = \theta(B)a_t \quad \forall t \quad (7.19)$$

then we may write (see, for example, section 6.6)

$$x_t = \psi(B)a_t \quad \forall t$$

where

$$\psi(B) = \phi^{-1}(B)\theta(B) \quad (7.20)$$

So, by section 6.6 (and specifically by (6.27)), it may be shown that the acvfg of x_t is

$$\Gamma^*(Z) = \psi^{-1}(Z)\gamma(Z)\psi(Z^{-1})$$

Thus the iacr^s of (7.19) are just the acf^s of the inverse model of (7.19) i.e. the acf^s of

$$\theta(B)x_t = \phi(B)a_t$$

(For more detail, see appendix A12.2).

So the iacr^f of an ARMA (p,q) process (or ARMA (p,d,q) process) is just the acrf of an ARMA (q,p) (or ARMA (q,d,p)) process!

Thus for an AR(p) process $\{x_t\}$, the iacf has cut-off after lag p (i.e. the iacf is just the same as the acrf of an MA(p) process)

So

$$\gamma(k) = \begin{cases} 1 & k = 0 \\ 0 & |k| > p \\ \frac{-\theta_k + \sum_{j=1}^{p-k} \theta_j \gamma(k+j)}{1 + \sum_{j=1}^p \theta_j^2} & \text{otherwise} \end{cases} \quad (7.21)$$

where the θ_j are the coefficients in an MA(p) process [by the definition of the acrf of an MA(p) process - see (6.9) in section 6.2 and substitute for γ^2 (see appendix A11)]

Similarly, for an MA(q) process, the iacf is just the same as the acrf of an AR(q) process. For an ARMA (p,q) process the iacf is the same as the acrf in an ARMA (q,p) process.

(See appendix A12.2)

Cleveland (1972) claims that the iacf is superior (for identification purposes) to the pacrf, as a result of the complicated form of the pacrf.

For an MA(q) process, we obtain inverse Yule-Walker equations of the form

$$\gamma(k) + \theta_1 \gamma(k-1) + \dots + \theta_q \gamma(k-q) = 0 \quad k = 1, \dots, q \quad (7.22)$$

In appendix A12, two techniques for estimating the iacf are given. The asymptotic behaviour of the iacf's is also treated there. [Note that as a result of the form of the iacf, an additional function, the inverse pacrf (ipacrf), may be defined in a manner analogous to the way in which we defined the pacrf (in section 6.4.)].

Thus, if the $iacf$ and $ipacf$ are estimated, we may also determine p and q , in the ARMA (p,q) process, from these two functions. If the $iacf$ truncates after lag p , this would indicate that $\{X_t\}$ is an AR(p) process. If the $ipacf$ truncates after lag q , then $\{X_t\}$ would seem to be an MA(q) process. Similarly if neither the $iacf$, nor the $ipacf$, truncate, then this indicates an ARMA (p,q) process ($p,q > 0$). Then analogous to the $acrf$ and $pacrf$, the number of anomalous terms in the $iacf$ and $ipacf$ are $p-q$ and $q-p$ respectively (see Hipel et al (1977)).

7.3 PHASE THREE: ESTIMATION OF MODEL PARAMETERS

Having tentatively identified a model (i.e. p,d and q in the ARMA (p,d,q) process) for the data, we now need to calculate efficient estimates of the parameters in the model

$$\Phi(B)x_t = (B)a_t$$

where $\{x_t\}$ is a zero-mean stationary process (and if d the degree of differencing, is greater than zero, then $x_t = \nabla^d z_t$).

In chapter six, we saw how to obtain estimates in an AR(p) process - by using the Yule-Walker equations. Techniques were also proposed for MA and ARMA processes (appendices 9 to 11). In this section, the estimation procedure will be tackled more formally.

If our realization of the stationary process $\{X_t\}$ is $x_1, x_2, \dots, x_N$, then it may be shown that the unconditional log-likelihood function of the process is given by

For computational convenience

$$L^* = L^*(\underline{\phi}, \underline{\theta}, \underline{\gamma}_a) = f(\underline{\phi}, \underline{\theta}) - N \ln \sigma_a^2 - S(\underline{\phi}, \underline{\theta}) / 2\sigma_a^2 \quad (7.23)$$

(see Box and Jenkins (1976) pg. 213)

where

$$S(\underline{\phi}, \underline{\theta}) = \sum_{t=-\infty}^N \{a_t | \underline{\phi}, \underline{\theta}, \underline{x}\}^2 \quad (7.24)$$

where the a_t 's are calculated recursively from $\underline{\phi}$, $\underline{\theta}$ and $\underline{x}$

and $\underline{\phi}' = (\phi_1, \dots, \phi_p)$ $f(\underline{\phi}, \underline{\theta})$ a function of $\underline{\phi}$, and $\underline{\theta}$

$$\underline{\theta}' = (\theta_1, \dots, \theta_q)$$

$$\underline{x}' = (x_1, x_2, \dots, x_N)$$

with $\{a_t | \underline{\phi}, \underline{\theta}, \underline{x}\} = E(a_t | \underline{\phi}, \underline{\theta}, \underline{x})$ (i.e. the expectation of a_t conditional on $\underline{\phi}$, $\underline{\theta}$ and $\underline{x}$)

$S(\underline{\phi}, \underline{\theta})$ is called the unconditional sum of squares function.

Usually $f(\underline{\phi}, \underline{\theta})$ in (7.23) is only of importance if N is small (Box and Jenkins (1976), pg. 213) while if N is large then L^* is dominated by $S(\underline{\phi}, \underline{\theta})$. If we find LSE of the parameters from $S(\underline{\phi}, \underline{\theta})$, then (if N is large) these estimates should be approximately equal to the maximum likelihood estimates (MLE^s) from L^* .

However, by the form of (7.24), we need to calculate $\{a_t | \underline{\phi}, \underline{\theta}, \underline{x}\}$ $\forall t$ (where from now on we shall denote $\{a_t | \underline{\phi}, \underline{\theta}, \underline{x}\}$ by $\{a_t\}$ i.e. the conditional expectation of the white noise variables (residuals)).

7.3.1 CALCULATING THE UNCONDITIONAL SUM OF SQUARES (USS)

In the calculation of the USS, the $\{a_t\}$ are computed recursively by taking the conditional expectations in

$$a_t = \phi(B)x_t + \theta(B)a_t \quad (t \leq N) \quad (7.25)$$

where $\phi(B)$ is as before but

$$\theta^*(B) = \theta_1 B + \dots + \theta_q B^q$$

(i.e. $\theta^*(B) = 1 - \theta(B)$)

By the form of (7.24), i.e. the USS, we need $\{a_t\}$ for $-\infty < t \leq N$ and so obviously values (or estimates) for $x_0, x_{-1}, x_{-2}, \dots$ (i.e. "backforecasts").

Since $\phi(B)x_t = \theta(B)a_t \quad \forall t$ then it is equally possible that the data could be generated by the model

$$\phi(F)x_t = \theta(F)\varepsilon_t \quad (7.26)$$

F = forward shift operator

$\{\varepsilon_t\}$ = white noise process

[Heuristic proof: If $\phi(B)x_t = \theta(B)a_t \quad \forall t$

and the zeroes of $\phi(B)$ and $\theta(B)$ lie outside the unit circle then we may write the model as

$$\prod_{i=1}^p (1 - \alpha_i B)x_t = \prod_{j=1}^q (1 - \beta_j B)a_t$$

where α_i^{-1} and β_j^{-1} are the roots of $\phi(B) = 0$ and $\theta(B) = 0$ resp.

Now if X_t has acvgf $\Gamma(B)$, then

$$\Gamma(B) = \frac{1}{a} \cdot \prod_{i=1}^p (1 - \alpha_i B)^{-1} (1 - \alpha_i F)^{-1} \cdot \prod_{j=1}^q (1 - \beta_j B)(1 - \beta_j F)$$

(see section 7.2.1)

But we may write (by multiplying out)

$$(1 - \beta_j B)(1 - \beta_j F) = \frac{1}{\beta_j} (1 - \beta_j^{-1} B)(1 - \beta_j^{-1} F),$$

so any of the 'models'

$$\prod_{i=1}^p (1 - \alpha_i B)x_t = \prod_{j=1}^q (1 - \beta_j^{-1} B)a_t$$

can have the same acvgf if k is appropriately chosen. But since all the α_j and β_j^s lie inside the unit circle, the acvgf is unique. Similarly, $\bar{\Gamma}(B)$ is unchanged if $(1 - \alpha_j B)$ is replaced by $(1 - \alpha_j F)$ or $(1 - \beta_j B)$ by $(1 - \beta_j F)$ and so all 'models' of the form

$$\sum_{i=1}^p (1 - \alpha_i B^{-1}) x_t = \sum_{j=1}^q (1 - \beta_j B^{-1}) a_t \quad (B^{-1} = F)$$

have the same covariance structure.]

(7.26) is referred to as the forward form of the process and

is used to generate the back forecasts $x_0, x_{-1}, \dots$.

(7.25) is then used to generate the a_t^s .

More formally, letting $\{e_t\} = E\{e_t | \underline{p}, \underline{\theta}, \underline{x}\}$:

(i) generate $\{e_t\}$ $t = N - p, \dots, 1$ and set $\{e_j\} = 0$ if $j \leq 0$

(ii) the $\{e_t\}$ in (i) are generated by using

$$\{e_t\} = \theta(F)x_t + \theta^*(F)\{e_t\}$$

$$\theta^*(F) = 1 - \theta(F), \quad t = N - p, N - p - 1, \dots, 1$$

$$\text{and } \{e_t\} = 0 \quad \text{if } t = N - p$$

$$(\text{since } E\{e_t\} = 0 \quad \forall t)$$

(iii) generate the back forecasts of x_t by using

$$x_t = \theta(F)\{e_t\} + \phi^*(F)x_t$$

$$(\text{where } \phi^*(F) = \phi_1 F + \dots + \phi_p F^p)$$

$$t = 0, -1, -2, \dots$$

(iv) do (iii) until the back forecasts become negligible (i.e. zero) - which occurs at R say (R a negative integer)

(v) define $\{a_t\} = 0$ if $t < R$

$$(\text{since } E\{a_t\} = 0 \quad \forall t)$$

(vi) use

$$\{a_t\} = \phi(B)x_t + \theta^*(B)\{a_t\}$$

to generate the $\{a_t\}^s$ for $t = R, R+1, \dots, -1, 0, 1, 2, \dots, N$

(vii) $S(\hat{\rho}, \hat{\theta})$ may now be calculated.

[(viii) if required, the whole procedure may be performed

iteratively, starting off each successive iteration by

using $\{a_N\}$ to calculate the forward forecasts of the

x_t^s using

$$x_t = \phi^*(B)x_t + \theta(B)\{a_t\};$$

the forward forecasts of x_t may then be used in

$$\{\varepsilon_t\} = \phi(F)x_t + \theta^*(F)\{\varepsilon_t\}$$

to obtain new back forecasts

However, we always need $\hat{\rho}$ and $\hat{\theta}$ to calculate $S(\hat{\rho}, \hat{\theta})$ and the reason we want $S(\hat{\rho}, \hat{\theta})$ is to obtain $\hat{\rho}$ and $\hat{\theta}$. Obviously, a grid-search may be performed to find the $\hat{\rho}$ and $\hat{\theta}$ to minimize $S(\hat{\rho}, \hat{\theta})$ although usually the techniques described in chapter 6 are used.

7.3.2 VARIANCES AND COVARIANCES OF MLE^s

Box and Jenkins (1976, pg. 226) state that if $\{X_t\}$ has an ARMA (p,q) representation, then over the relevant region of the parameter space, the log-likelihood function, L^* (as defined in (7.23)), will be approximately quadratic in the elements of $\hat{\rho}$ and $\hat{\theta}$. Thus

$$\begin{aligned} L^* &= L^*(\hat{\rho}, \hat{\theta}, \hat{\sigma}_a^2) \\ &= L^*(\hat{\rho}, \hat{\theta}) \quad \text{say} \\ &= L^*(\hat{\rho}, \hat{\theta}) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k L_{ij} (\hat{\rho}_i - \hat{\rho}_i) (\hat{\theta}_j - \hat{\theta}_j) \end{aligned} \quad (7.27)$$

where $k = p + q$

$$\text{and } L_{ij} = \frac{\partial^2 L(\underline{\Omega}, \underline{\theta}_a)}{\partial \Omega_i \partial \Omega_j} \quad (7.28)$$

are (approximately) constant.

Since

$$L^* = f(\underline{\phi}, \underline{\theta}) - N \log_a - \frac{S(\underline{\phi}, \underline{\theta})}{2\sigma_a^2} \quad (7.28a)$$

(by (7.23))

then for large N , $f(\underline{\phi}, \underline{\theta})$ can be ignored and so L^* is quadratic in $\underline{\Omega}$ if $S(\underline{\cdot})$ is quadratic in $\underline{\cdot}$.

Define

$$\| - E(L_{ij}) \| = I(\underline{\cdot}) \quad (7.29)$$

= k x k 'information' matrix (and expectation is over $\underline{x}$).

If σ_a^2 is given, then, from asymptotic likelihood theory

$$\Sigma(\hat{\underline{\Omega}}) = (-E(L_{ij}))^{-1} \quad (7.30)$$

(i.e. the variance-covariance matrix)

Defining

$$S_{ij} = \frac{\partial^2 S(\underline{\cdot})}{\partial \Omega_i \partial \Omega_j} \quad \forall i, j \quad (7.31)$$

then

$$L_{ij} = S_{ij} / 2\sigma_a^2 \quad \forall i, j \quad (7.32)$$

and

$$\Sigma(\underline{\cdot}) = 2\sigma_a^2 (S_{ij})^{-1} \quad (7.23)$$

(if we approximate L_{ij} and hence $\Sigma(\underline{\cdot})$ by their observed values)

Box and Jenkins (1976, appendices A7.4 and A7.5) have shown that

$$\sigma_a^2 = S(\underline{\underline{\theta}}) / N \quad (7.34)$$

by differentiating (7.28a) w.r.t. θ_a

$$\text{Cov}(\sigma_a^2, \underline{\underline{\theta}}) \approx 0 \quad (7.35)$$

$$\text{Cov}(\hat{\theta}_i, \hat{\theta}_j) \approx 2\sigma_a^2 S^{ij} \quad (7.36)$$

$$\text{where } (S^{ij}) = (S_{ij})^{-1}$$

(7.36) is derived directly from (7.30) to (7.32)).

It may also be shown that if $S(\underline{\underline{\theta}})$ is 'reasonably' quadratic in the model parameters, then a (1- α) confidence region is bounded by the contour on the SS surface, for which

$$S(\underline{\underline{\theta}}) \leq S(\hat{\underline{\underline{\theta}}}) \left\{ 1 + \frac{k}{N} \right\} \quad (7.37)$$

$$\text{where } k = p + q$$

(see, for instance, Box and Jenkins (1976) pg. 228 or use general linear least squares theory)

7.3.3 NONLINEAR ESTIMATION OF ϕ , θ AND σ_a^2

For most cases, the MLE's of ϕ , θ and σ_a^2 are closely approximated by the LSE which minimize

$$S(\underline{\underline{\theta}}) = \sum_{t=R}^N (a_t)^2 \quad (7.38)$$

(which may be approximated by

$$\sum_{t=R}^N (a_t)^2 \quad \text{where } R = \text{negative integer (defined on pg. 136)}$$

What we really need to calculate in (7.38) is

$$a_{t,t} = -2(a_t)/\sigma_a^2 \quad (7.39)$$

where λ_i is one of the 1^s or 2^s . Numerical estimates of (7.39) may be obtained by using any of the several nonlinear estimation techniques available. (Several optimization techniques which do not involve finding derivatives in the sum of squares, are also available.) Some of the more extensively applied algorithms include Gauss linearization (Draper and Smith, 1966, chapter 10), steepest descent (Fletcher and Powell, 1963), Marquardt's algorithm (Marquardt, 1963) and the method of conjugate directions (Powell, 1964a, 1964b).

7.4 PHASE FOUR : CHECKING THE FIT OF THE MODEL AND DETERMINING THE BEST MODEL

Especially in the case where $\{X_t\}$ is an ARMA process (with p and $q > 0$), we may formulate several models (due to the difficulty of deciding exactly how many terms in the acrf, pacrf, iacrf and ipacrf are anomalous). Obviously, when we have several models we would like some technique which may ascertain the 'best' model. We would also like to 'test' whether the residuals from the fitted model are sufficiently like a white noise process (and hence we may stop modelling).

7.4.1 MODEL INADEQUACY

Suppose we model $\{X_t\}$ by

$$\phi(B)x_t = \theta(B)a_t$$

then $a_t = \phi^{-1}(B)\theta(B)x_t \quad \forall t$

and so the residuals of the model are given by

$$\hat{a}_t = \hat{\theta}^{-1}(B)\phi(B)x_t \quad (7.40)$$

Now it is possible to show that

$$\hat{a}_t = a_t + O\left(\frac{1}{\sqrt{N}}\right) \quad (7.41)$$

if the model is adequate (Box and Jenkins, 1976, pg. 289) so as N increases, the estimated residuals become more and more like white noise.

Box (1942) has shown that if $\hat{\theta}$ and $\hat{\phi}$ are 'known', then the estimated acf's of the $\hat{a}_t$, are uncorrelated and independent and are approximately normally distributed as $N(0, \frac{1}{N})$ (see appendix A13).

If some pattern occurs in the residual acf, this probably indicates model inadequacy.

$$\text{Since } \hat{r}_{a,k} \sim N(0, \frac{1}{N})$$

(see appendix A13)

$$\text{then } \sqrt{N} \hat{r}_{a,k} \sim N(0, 1)$$

and hence

$$(\sqrt{N} \hat{r}_{a,k})^2 \sim \chi^2_1$$

Box and Jenkins (1976, pg. 290) have developed a test they term the portmanteau lack of fit test, and here they calculate

$$Q = N \sum_{i=1}^L \hat{r}_{a,i}^2 \quad (7.42)$$

$$\sim \chi^2_{L-p-q} \quad (L \text{ some integer } > 0)$$

If the model is 'appropriate' (i.e. the residuals are pure white noise) However Davies et al (1977) have shown that a test based on Q is unreliable. See also Ljung and Box (1978).

acrf (obtained from the diagonal entries of the matrix $\frac{1}{N} U$, see Hipel et al (1977)). If some of the residual acrf are significantly different from zero, this implies that the model is inadequate.

[The a_t may also be checked for normality. Normally distributed data should possess no significant skewness, and kurtosis equal to (approximately) three. In appendix A11, ARMA processes with nonnormal residuals are considered. If the residuals possess significant skewness and kurtosis, a transformation of the original data may remove this and thus make the residuals more normal.]

Alternatively, the cumulative periodogram of the residuals may be calculated.

$$I^*(w_j) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} c_{a,t}^* \cos(tw_j)$$

$$j = 0, \dots, \left[\frac{N}{2}\right] \quad w_j = 2^{-j}/N$$

= $h(w)$ an estimate of the spectrum (see section 4.1)

($c_{a,k}^*$ = estimate of the acv(at lag k) of the residuals).

If the residuals are pure white noise then

$$h(w) = \frac{1}{2\pi}$$

(since $h(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma_a(t) \cos(tw)$ by definition

but $\gamma_a(t) = 0$ if $|t| > 0$)

The cumulative spectrum, $H(w)$, may be defined by

$$H(w_j) = \int_0^{w_j} h(w) dw \quad 0 \leq w_j \leq \pi$$

+ This is not the definition of the true cumulative spectrum -

that is defined by $\int_{-\pi}^{w_j} h(w) dw$

acrf (obtained from the diagonal entries of the matrix $\frac{1}{N} U$, see Hipel et al (1977)). If some of the residual acf's are significantly different from zero, this implies that the model is inadequate.

[The $\hat{a}_t$ may also be checked for normality. Normally distributed data should possess no significant skewness, and kurtosis equal to (approximately) three. In appendix A14, ARMA processes with nonnormal residuals are considered. If the residuals possess significant skewness and kurtosis, a transformation of the original data may remove this and thus make the residuals more normal.]

Alternatively, the cumulative periodogram of the residuals may be calculated.

$$I^*(w_j) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \hat{c}_{a,k}^* \cos(kw_j)$$

$$j = 0, \dots, \left[\frac{N}{2}\right] \quad w_j = 2\pi j/N$$

$= h(w)$ an estimate of the spectrum (see section 4.1)

($\hat{c}_{a,k}^*$ = estimate of the acv(at lag k) of the residuals).

If the residuals are pure white noise then

$$h(w) = \frac{\sigma_a^2}{2\pi}$$

(since $h(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \hat{c}_a(t) \cos(tw)$ by definition

but $\hat{c}_a(t) = 0$ if $|t| > 0$)

The cumulative spectrum, $H(w)$, may be defined by

$$H(w_j) = \int_0^{w_j} h(w) dw \quad 0 \leq w_j \leq \pi^+$$

This is not the definition of the true cumulative spectrum -

that is defined by $\int_{-\pi}^{\pi} h(w) dw$

$$= \int_0^w \frac{a}{2\pi} d\psi$$

$$= \frac{a}{2\pi} (w_j)$$

$$\text{and } H(\pi) = \frac{a}{2}$$

thus $H(w)$ is a linear function of w over $(0, \pi)$ (if $\{a_t\}$ is a white noise process).

So define the cumulative periodogram by

$$CI(w_j) = \sum_{i=0}^j I^*(w_i)$$

and then use some goodness of fit test or nonparametric procedure.

7.4.2 SELECTING THE 'BEST' MODEL

As mentioned earlier, it may happen that we have several models for the data, and wish to choose the 'best' one. Several techniques are available for assessing this 'bestness'.

Akaike (1970) defined a criterion for assessing the best model by

$$\begin{aligned} \text{AIC} &= \text{Akaike's information criterion} \\ &= -2 \ln \hat{L}(\hat{\theta}) \quad (\text{maximum likelihood function of the process}) \\ &\quad + 2(\text{no. of parameters in the model}) \end{aligned}$$

or equivalently (for N large and hence $f(\cdot)$ negligible - see pg. 134)

$$\text{AIC} = N \ln \hat{\sigma}^2 + 2k \quad k = p + q$$

$$\hat{\sigma}^2 = S(\cdot)/N \quad (\text{see pg. 139})$$

since

$$\begin{aligned} \ln(\text{ML function}) &= L^* \\ &= -\frac{N}{2} \ln \frac{\sigma_a^2}{a} - \frac{1}{2\sigma_a^2} S(\underline{\Omega}) \quad (\text{for large } N) \end{aligned}$$

and hence $f(\underline{\Omega}, \underline{\Omega})$ negligible - see pg. 138)

where $\underline{\Omega} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$

and L^* maximized w.r.t. $\underline{\Omega}$ and σ_a^2 gives

$$\sigma_a^2 = \frac{1}{N} S(\underline{\Omega})$$

and the maximum of L^* is

$$L^* = -\frac{N}{2} \ln \frac{1}{N} S(\underline{\Omega}) - \frac{N}{2}$$

(and the second term can be ignored since it is independent of $\underline{\Omega}$ and k).

The model with the minimum AIC is defined to be 'best'.

Akaike (1970) also defined, for AR processes, a criterion for determining the 'best' model.

Define

FPE = final prediction error

$$= \frac{N-k}{N-k} \frac{\sigma_a^2}{a}$$

where $k = p$ and $\frac{\sigma_a^2}{a}$ is the MLE of $\frac{\sigma_a^2}{a}$.

The AR process with minimum FPE is the best fitting model.

[It may be shown that

$$AIC = N \cdot FPE$$

- Akaike (1970) or Priestley, 1981, pg. 374 .

More recently, Akaike (1979), has developed a new criterion for order determination in ARMA processes

$$BIC = N \left(\frac{\sigma_x^2}{N} - (N-k) \ln \left(1 - \frac{k}{N} \right) + k \ln N + k \ln \left(\frac{\sigma_x^2}{\sigma_a^2} - 1 \right) \right)$$

where σ_x^2 = raw sample variance of the observations

$$k = p + q$$

Parzen (1974), has also developed a technique for choosing the best fitting AR process using the CAT ("criterion for AR transfer functions") quantity defined by

$$CAT(p) = \begin{cases} \left(\frac{1}{N} \sum_{j=1}^p \frac{1}{\hat{\sigma}_j^2} \right) - \frac{1}{\hat{\sigma}_p^2} & p = 1, 2, 3, \dots \\ -(1 + \frac{1}{N}) & p = 0 \end{cases}$$

where $\hat{\sigma}_j^2$ is an (unbiased) estimate of the residual variance, σ_a^2 , when an AR(j) model is fitted. Choose the best fitting AR process when CAT is minimized.

A technique developed by McClave (1975, 1978, 1978a) for choosing the 'best' AR or MA model is described in appendix A15. This technique usually uses one of the above criteria (i.e. AIC, FPE, BIC or CAT). (This technique is mentioned here instead of in section 7.2., since it involves fitting several models and then choosing the best.)

Usually, if the model is deemed inadequate, either the residuals themselves may be modelled further or (if this is not possible) it may be necessary to transform the data (and hence start all over again). A class of transformations is presented in appendix A16.

7.5 PHASE FIVE : FORECASTING

Having deemed the model (models) to be satisfactory, we may wish to forecast estimates of future values. This is not particularly difficult, due to the form of the Box-Jenkins models, and is discussed more fully in appendix A17.

7.6 SEASONAL BOX-JENKINS MODELS

So far, the concept of a seasonal Box-Jenkins model has not been discussed since we have considered our data to be deseasonalized (by using periodogram or harmonic analysis). However it may so happen that either

- (a) we neglected to remove some periodic component
- or (b) we do not wish to use harmonic analysis.

In case (a), any periodicity in the data, which is not modelled, will probably be picked up in the residual acrf (or it may even be apparent in the residuals themselves). In (b), we must either deseasonalize the data using some other technique, or allow for the seasonal components by modelling them.

Let us assume the data contain some seasonal components and we wish to model them by somehow modifying the B-J techniques. Suppose the periodic behaviour has period S - i.e. similarities in the data appear every S basic time intervals.

Since observations S intervals apart are similar, it seems reasonable (assuming stationarity) that

$$B^S x_t = x_{t-S}$$

plays some role in the model. It is also possible that

$$\nabla_S^S x_t = (1 - B^S)x_t$$

$$= x_t - x_{t-S}$$

or even ∇_S plays some role.

[Note that $(1 - B^S)$ has S zeroes ($e^{i2\pi k/S}$, $k = 0, \dots, S-1$) evenly spaced on the unit circle.]

If the data are monthly say, then the two important time intervals correspond to months and years i.e. we expect relationships between successive months as well as relationships between months a year apart.

A model to link December say, to observations in previous Decembers, may be

$$\phi(B^S) \nabla_S^D x_t = \theta(B^S) e_t$$

S = period (in our case 12)

$\nabla_S = 1 - B^S$ D = degree of seasonal differencing

$\phi(B^S)$ = polynomial in B^S of degree P_S

$\theta(B^S)$ = polynomial in B^S of degree Q_S

and $\phi(B^S)$ and $\theta(B^S)$ satisfy stationarity and invertibility conditions.

The e_t will not usually be uncorrelated (i.e. the observation for December, while related to the previous December, may also bear some relationship to November, October etc). To take this into account, we model

$$\phi(B) \nabla^d e_t = \omega(B) a_t$$

where

$\{a_t\}$ = white noise process, and $\phi(B)$ and $\theta(B)$ are polynomials in B just as before.

So our overall model is

$$\phi(B)(1-B^S)^{-d-D}x_t = \theta(B)(1-B^S)a_t \quad (7.44)$$

- this is termed a $(p,d,q) = (P_S,D,Q_S)_S$ B-J model

If the data contain 2 or more seasonal components then we may model each component as in (7.44).

Our first job now is to determine d, D, p, q, Q_S and S . Firstly, for d and D , just apply $(1-B)$ and $(1-B^S)$ until the sample acf's of the differenced series die out quickly at high lags.

Hill and Woodworth (1980) have suggested finding

$$\min_{d,D,q} \left\{ \frac{1}{S} \sum_{k=1}^S \left| \hat{\gamma}_k - \frac{1}{S} \sum_{j=1}^S \hat{\gamma}_{k-jS} \right| \right\}$$

Once the acf of the differenced series dies out quick'y enough for lags $1, \dots, S$ and also at multiples of S , we may then attempt to estimate p, q, P_S and Q_S .

A summary of the properties of seasonal models is given below and this enables us to gain an idea about p, q, P_S and Q_S . (For a fuller account see Box and Jenkins (1976), Chapter 9).

7.6.1 THE ACRF

(1) If the process is $(0,q) = (0,Q_S)$, then $\hat{\gamma}_k$ truncates after

$$q = SQ_S$$

- (ii) If r_k' attenuates at lags which are multiples of S , then this implies the presence of a seasonal AR component; if the acrf fails to truncate at other lags, this may imply a nonseasonal AR component.

7.6.2 THE PACRF

- (i) If the process is $(p,0) + (P_S,0)$, then $\hat{r}_{kk}$ truncates after $p + SP_S$
- (ii) If $\hat{r}_{kk}$ damps out at lags which are multiples of S , this implies the presence of a seasonal MA component; the failure of the pacrf to truncate at other lags may imply a nonseasonal MA component.

7.6.3 THE IACRF

- (i) If the process is $(p,0) + (P_S,0)$, r_{1k}' truncates after $p + SP_S$
- (ii) If r_{1k}' decays at lags which are multiples of S , this implies a seasonal MA component; failure of the r_{1k}' to truncate at other lags implies a nonseasonal MA component.

7.6.4 THE IPACRF

- (i) If the process is $(0,q) + (0,Q_S)$, then $\hat{r}_{kk}$ truncates after $q + SQ_S$
- (ii) If $\hat{r}_{kk}$ decays at lags which are multiples of S , then this implies the presence of a seasonal AR component. Failure of the ipacrf to truncate at other lags implies a nonseasonal AR component.

[For further details, see Box and Jenkins (1976) and Hipel et al (1977).]

Having estimated p, q, P_s and Q_s , parameter estimation follows in a manner similar to the nonseasonal BJ models, and hence forecasts. Diagnostic checks on the fit of the model are also similar to those of the nonseasonal models.

CHAPTER 8: NONLINEAR MODELLING AND MISCELLANEOUS NOTES

In the previous two chapters, a special branch of time-series models has been considered - namely linear time-series models. In these, the (present) observation x_t is considered as a linear combination of past values of x_t as well as a linear function of some noise variables a_t . This assumption of linearity can be very restrictive and it also seems reasonable to consider nonlinear functions of the observations and noise variables

$$x_t = a_t + \sum_{i=1}^p a_{t-i} x_{t-i}$$

Recently a class of nonlinear time-series models called bilinear models has been proposed. These models seem to offer a wider extensive possibilities. A brief description of these models and their parameter estimates are given below.

BILINEAR TIME-SERIES MODELS

These models are of the form

$$x_t = \sum_{j=0}^p \phi_j x_{t-j} + \sum_{i=1}^p \sum_{j=1}^k \theta_{ij} x_{t-i} a_{t-j} \quad (8.1)$$

(ϕ_j , θ_{ij} and a_t parameters)

where x_t is a zero-mean stationary process and $\{a_t\}$ is a white noise series (and $\phi_0 = 1$).

The representation (8.1) may be written in the abbreviated form

$$BL(p, q, m, k) \quad (8.2)$$

where $0 \leq p < \infty$, $0 \leq q < \infty$, $0 \leq m < \infty$, $0 \leq k < \infty$ and $\theta_{ij} = 0$ for $i > m$ or $j > k$. If $\theta_{ij} = 0$ for $i > m$ or $j > k$ then

$$BL(p, q, 0, 0) = \text{ARMA}(p, q)$$

The bilinear models with the 'easiest' theory are those of the form
 $BL(p, 0, p, q)$ (due to the form of the model) i.e.

$$x_t + \phi_0 + \sum_{j=1}^p \phi_j x_{t-j} = \sum_{j=1}^q \delta_{ij} x_{t-i} a_{t-j} + a_t \quad (8.3)$$

(ϕ_0 not necessarily 1).

Assume the a_t in (8.3) are independent and are $\sim N(0, \sigma_a^2)$. Also assume (8.3) is an 'invertible' model (i.e. if x_t satisfies

$$x_t = [\text{function of } (x_{t-j}, a_{t-j}; j=1, \dots, p)] + a_t$$

and $\hat{a}_t$ are estimates of a_t then

$$\lim_{t \rightarrow \infty} E[a_t - \hat{a}_t]^2 = 0.$$

Suppose we have a realization $\{x_1, x_2, \dots, x_N\}$. Define

$$m = \max\{p, q\} + 1$$

then the joint distribution of

$$\underline{a}' = (a_m, a_{m+1}, \dots, a_N)$$

$$is \quad g(\underline{a}) = \left(\frac{1}{2\sigma_a^2} \right)^{\frac{N-m+1}{2}} \exp\left(-\frac{1}{2\sigma_a^2} \sum_{t=m}^N a_t^2\right) \quad (8.4)$$

(since the a_t are $\sim N(0, \sigma_a^2)$ independently).

Now the likelihood function of $\underline{x}' = (x_m, \dots, x_N)$ is the same as the joint pdf of $\underline{a}$ since the Jacobian of the transformation from $\underline{a}$ to $\underline{x}$ is 1. Thus maximizing the likelihood function, w.r.t. the model parameters, is equivalent to minimizing

$$S(\underline{a}) = \sum_{t=m}^N a_t^2 \quad (8.5)$$

where $\vec{\alpha} = (\alpha_0, \dots, \alpha_p, \alpha_1, \dots, \alpha_{pq})$
 $= (\alpha_1, \alpha_2, \dots, \alpha_R)$ say $R = p + pq = 1$

So

$$\frac{\partial S(\vec{\alpha})}{\partial \alpha_i} = 2 \frac{\partial a_i}{\partial \alpha_i} \quad i = 1, \dots, R \quad (8.6)$$

and

$$\frac{\partial^2 S(\vec{\alpha})}{\partial \alpha_i \partial \alpha_j} = 2 \frac{\partial^2 a_i}{\partial \alpha_i \partial \alpha_j} = 2 \frac{\partial^2 a_i}{\partial \alpha_j^2} \quad (8.7)$$

These partial derivatives satisfy

$$(1) \quad \frac{\partial a_i}{\partial \alpha_j} = \frac{q}{j+1} \binom{p}{j} (i) \frac{\partial a_{i-j}}{\partial \alpha_j} = \begin{cases} 1 & i=j \\ x_{i-j} & i \neq j \end{cases} \quad i=1, 2, \dots, p \quad (8.8)$$

$$(\text{and } \frac{\partial^2 a_i}{\partial \alpha_j^2} = \frac{p}{j+1} \binom{p}{j} x_{i-j})$$

$$(2) \quad \frac{\partial a_i}{\partial \alpha_j} + \frac{q}{k+1} \binom{p}{k} (i) \frac{\partial a_{i-k}}{\partial \alpha_j} = -x_{i-k} x_{i-j} \quad (8.9)$$

$$i=1, \dots, p \quad j=1, \dots, q$$

$$(3) \quad \frac{\partial^2 a_i}{\partial \alpha_j \partial \alpha_l} + \frac{q}{k+1} \binom{p}{k} (i) \frac{\partial^2 a_{i-k}}{\partial \alpha_j \partial \alpha_l} = 0 \quad (8.10)$$

$$i,j = 0, 1, \dots, p$$

$$(4) \quad \frac{\partial^2 a_i}{\partial \alpha_j \partial \alpha_l} + \frac{q}{k+1} \binom{p}{k} (i) \frac{\partial^2 a_{i-k}}{\partial \alpha_j \partial \alpha_l} + x_{i-k} \frac{\partial a_{i-j}}{\partial \alpha_l} = 0 \quad (8.11)$$

$$i = 0, \dots, p$$

$$j = 1, \dots, p \quad l = 1, \dots, q$$

$$(5) \quad \frac{\partial^2 a_i}{\partial \alpha_j \partial \alpha_{rs}} + \frac{q}{k+1} \binom{p}{k} (i) \frac{\partial^2 a_{i-k}}{\partial \alpha_j \partial \alpha_{rs}} + x_{i-k} \frac{\partial a_{i-j}}{\partial \alpha_{rs}} = -x_{i-j} \frac{\partial a_{i-k}}{\partial \alpha_{rs}} \quad (8.12)$$

$$i, r = 1, \dots, p$$

$$j, s = 1, \dots, q$$

(see Subba Rao (1981)).

Now assuming $a_t = 0$ ($t = 1, \dots, m$) and

$$\frac{\partial a_t}{\partial \Omega_i} = 0 \quad \frac{\partial^2 a_t}{\partial \Omega_i \partial \Omega_j} = 0 \quad (8.13)$$

$$t = 1, \dots, m$$

then together with (8.9), this implies the second-order derivatives w.r.t. Ω_i ($i = 1, \dots, p$) are zero.

Define

$$S'(\underline{\Omega}) = \left(\frac{\partial S(\underline{\Omega})}{\partial \Omega_1}, \dots, \frac{\partial S(\underline{\Omega})}{\partial \Omega_R} \right)$$

and

$$S''(\underline{\Omega}) = \left[\frac{\partial^2 S(\underline{\Omega})}{\partial \Omega_i \partial \Omega_j} \right]_{R \times R} \quad (\text{i.e. a matrix of second-order derivatives})$$

then expanding $S'(\underline{\Omega})$ in a Taylor series near $\underline{\Omega} = \underline{\Omega}^0$ gives

$$S'(\underline{\Omega}) = \underline{0} + S'(\underline{\Omega}^0) + S''(\underline{\Omega}^0) < (\underline{\Omega} - \underline{\Omega}^0)$$

thus

$$\underline{\Omega}^{(k+1)} = \underline{\Omega}^{(k)} - S''(\underline{\Omega}^{(k)})^{-1} S'(\underline{\Omega}^{(k)})$$

$$\text{where } S^I(\underline{\Omega}) = (S''(\underline{\Omega}))^{-1}$$

$$\text{and so } \underline{\Omega}^{(k+1)} = \underline{\Omega}^{(k)} - S''(\underline{\Omega}^{(k)})^{-1} S'(\underline{\Omega}^{(k)}) \quad (8.14)$$

where $\underline{\Omega}^{(k)}$ is the set of estimates at the k^{th} iteration. This leads to the standard Newton - Raphson iterative method of solving for $\underline{\Omega}$.

Thus letting $L(\underline{\Omega})$ denote the likelihood function of $\underline{x} = (x_1, \dots, x_N)'$

(and using the notation of Subba Rao (1981)) then

$$\frac{1}{N} \frac{\partial^2 \log L(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'} = - \frac{1}{N} \frac{\partial^2 S(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'}$$

Letting

$$\frac{1}{N} \frac{\partial^2 S(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'}$$

converge (element-wise stochastically) to $I(\underline{\theta})$ (the information matrix) then, as $N \rightarrow \infty$

$$- \frac{1}{N} \frac{\partial^2 \log L(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'} \longrightarrow \frac{1}{N} \frac{\partial^2 S(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'}$$

$$\text{and } \hat{\underline{\theta}} = \hat{\underline{\theta}} + N^{-1} I^{-1}(\hat{\underline{\theta}})$$

(see Subba Rao (1981))

Subba Rao (1981) suggests fitting an EM algorithm to the data (with a constant $\frac{1}{N}$ included) and obtaining estimates of $\underline{\theta} = (\theta_1, \dots, \theta_p)'$. Then use these $\hat{\theta}_i$ ($i = 1, \dots, p$) as initial values in the bilinear model; and set $\hat{\theta}_i = 0 \forall i$ as the initial values for θ_{i+1} .

Dane and Subba Rao (1981) have considered subset bilinear models. These models effectively take into account the situation where parameters in the 'full' bilinear model $BL(p, q, p, q)$ may be negligibly small. They have described an algorithm for fitting a subset bilinear model.

2.2 THRESHOLD AUTOREGRESSIVE MODELS

Another class of finite parameter nonlinear time series models is that known as threshold models (Tong and Lim (1980)). These are autoregressive models with parameters that vary according to the values of a finite number of past values

of x_t (or some associated process $\{Y_t\}$)

$$\text{e.g. (1)} \quad x_t = \begin{cases} \phi_1 x_{t-1} + a_{1t} & x_{t-1} \leq d \\ \phi_2 x_{t-1} + a_{2t} & x_{t-1} > d \end{cases}$$

where a_{1t} and a_{2t} are white noise processes and d is some real number

$$(2) \quad x_t = \phi_i x_{t-1} + a_{it} \quad \text{if } x_{t-1} \in R_i$$

and $R_1, \dots, R_k$ are given subsets of R (the real line).

$$(3) \quad x_t = \phi_{1i} x_{t-1} + \dots + \phi_{pi} x_{t-p} + a_{it} \quad (i=1, \dots, k)$$

if $(x_{t-1}, \dots, x_{t-p}) \in R^{(1)}$

where $R^{(1)}$ is a region of the Euclidean space R^p

This class of models is very difficult to apply (see Gabr and Subba Rao (1981)) however they seem to give better forecasts than ARMA or subset ARMA models in most cases. Gabr and Subba Rao (1981) claim that bilinear models (and in particular subset bilinear models) give the best forecast results. Note that these models are based on the idea of piece-wise linearization.

8.3 MISSING OBSERVATIONS

Often a data set is incomplete due to missing observations. This makes it difficult to estimate the spectrum and also any kind of model. Several authors have proposed techniques for estimating the spectrum (and hence the observations) - see, for example, Bloomfield (1970, 1976).

3.1 R AND S ARRAYS

The major problem, when using the Box-Jenkins technique, is determining p, d and q if $\{X_t\}$ is an ARMA (p, d, q) process. Gray, Kelly and McIntire (1978) have obtained an effective general method for simultaneously identifying p and q by analyzing two sets of ratios of Hankel determinants (Findley (1981)) of sample ac 's. These two sets of ratios are called R and S arrays. They have also suggested systematic procedures for identifying d .

For more details see Gray, Kelly and McIntire (1978) and Findley (1981).

CHAPTER 9: AN APPLICATION

In this chapter, an application of some of the techniques presented in earlier chapters is presented. The computer programs described in appendix A18 were used in this application.

9.1 THE DATA

The data consist of the 164 observations given in appendix A19 and plotted in figure 9.1. They represent weekly incoming-passenger totals for an airline at Jan Shuair Airport from January 1979 to February 1982. (The data are x_t as measured by PAS21X.)

I decided to use the first 120 observations for the process of modelling and the last 44 for forecasting purposes. The 'cut-off' point is shown in figure 9.1 ($x_{120} = 204$). From this plot, there seems to be some form of periodicity of approximately 36 weeks with the stretches of data in between not too well defined.

Since our data is weekly, it perhaps should not be too surprising if the data exhibit some form of seasonality with periods 12, 36 or 72 weeks (i.e. quarterly, half-yearly or annually). While there are no obvious periods.

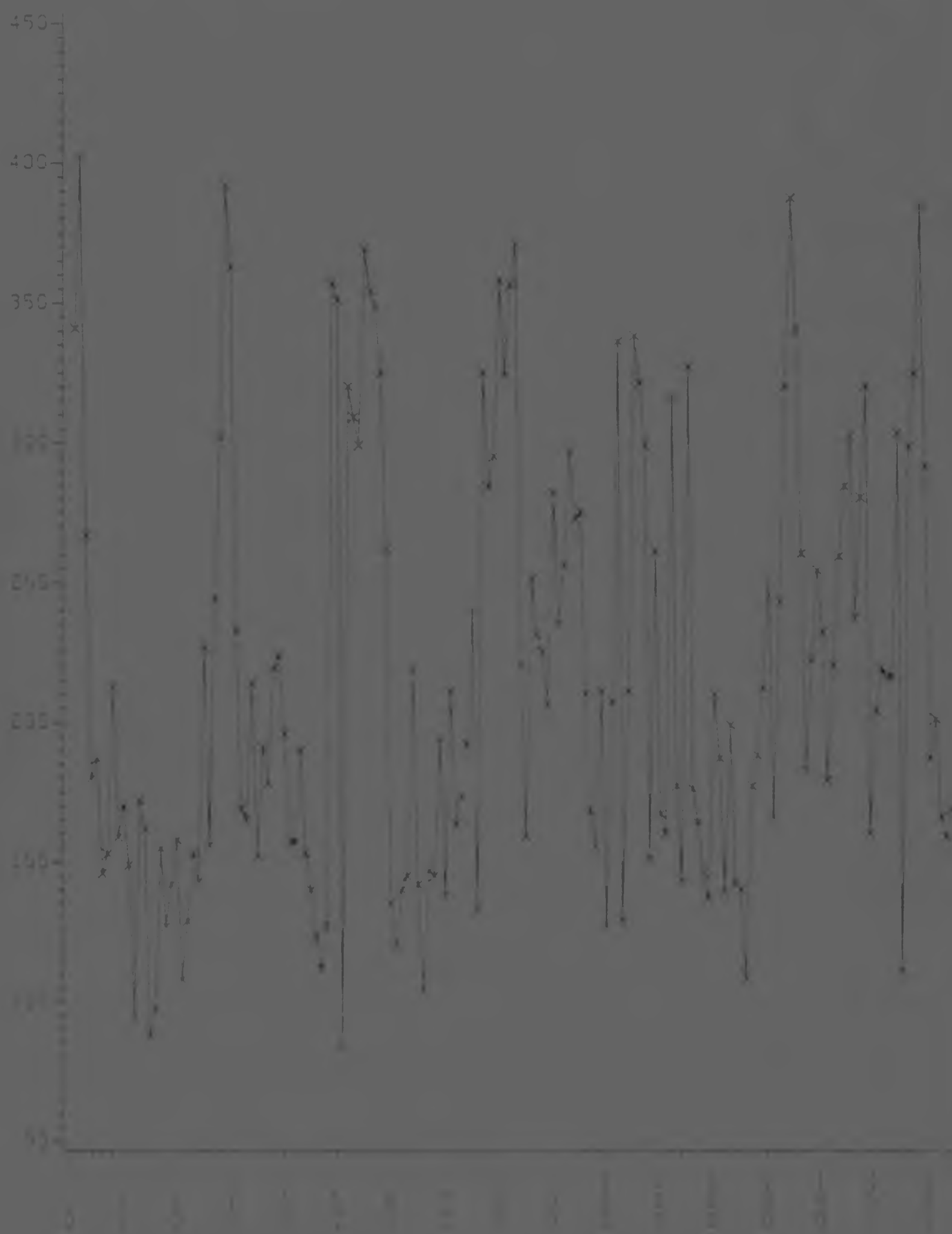
The data (or if necessary any transformation thereof) are to be modelled by

$$x_t = \text{deterministic cycles} + \text{stochastic process}$$

where the former part may contain a trend component (i.e. a constant) and consists of any sinusoids of fixed frequency, phase and amplitude.

Deterministic cycles (if present) are usually apparent in the

Fig. 9.1 A plot of the airline passenger data



data (i.e. a cyclic form). However, some nondeterministic cycles need not be cyclic but rather just meander around a particular frequency, phase and amplitude. Very often when some cyclicity or seasonality is apparent, a transformation of the data is made (see Chatfield (1980), pg. 13). For instance, if the data is seasonal and has escalating amplitude in the seasonal wave pattern, then a change in the variance over time, of the original data, would eventually be mirrored by heteroscedasticity in the residuals of the model fitted to the data. To rectify this situation (from the start) take natural logs of the data.

Also, the data are usually transformed if they have an asymmetrical appearance (see Campbell and Walker (1977)). Again the transformation to be used is a natural log transform (since this reduces the degree of asymmetry).

Parzen (1978) suggests a transformation without memory, or memoryless transformation, (i.e. independent of time), such as logarithm or square root if the data appear to have long tails or seem to be non-normally distributed. (Also see Parzen (1982)).

9.2 TRANSFORMING THE DATA

To ascertain whether a transformation is necessary, we shall use the technique of quantile - box plots suggested by Parzen (1978) and described in Parzen (1979).

This technique involves estimating the quantile function $Q(u)$ ($0 \leq u \leq 1$) defined by

See 'Comment' at the end of the chapter on page 214

$$Q(u) = F^{-1}(u)$$

where $F(x)$ is the cumulative density function (cdf) of the process from which the data is drawn.

$Q(u)$ is estimated by $\bar{F}^{-1}(u)$ where $\bar{F}(x)$ is the empirical distribution function (edf) of the sample.

Once $Q(u)$ has been plotted, H, E and D 'boxes' are superimposed on the plot and represent rectangles from the quantiles

$$\frac{1}{4} \rightarrow \frac{3}{4}, \quad \frac{1}{8} \rightarrow \frac{7}{8} \quad \text{and} \quad \frac{1}{16} \rightarrow \frac{15}{16}$$

resp. (see fig. 9.2). [For more detail see Parzen (1979)].

The quantile-box plot of the data is given in figure 9.2 with associated boxes. Parzen (1979) claims that dissimilar shapes within the boxes indicates asymmetrical data. If the data appear to be symmetric it is also possible to test whether the data are nonnormal and in particular if the distribution from which they are drawn possesses long tails (see Parzen (1979) - he also provides a statistic for determining skewness).

The box plots of the $\log_{10}$ and square root transformations are given in figures 9.3 and 9.4 resp.

Since memoryless transformations like $\log_{10}$ or square-root do not drastically change the basic shape of the distribution of the data (if all data points are > 1), a transformation of this kind is completely optional. As a result of this, instead of using the statistics suggested by Parzen (1979), I calculated the vertical distances between successive boxes - e.g. the distances a_i , b_i , c_i and d_i ($i=1,2$) given in figure 9.2 - for each quantile box plot. The results are given in table 9.1.

See 'Comment' at the end of the chapter on pg. 214

Fig. 9.2 Quantile-box plot of the airline passenger data

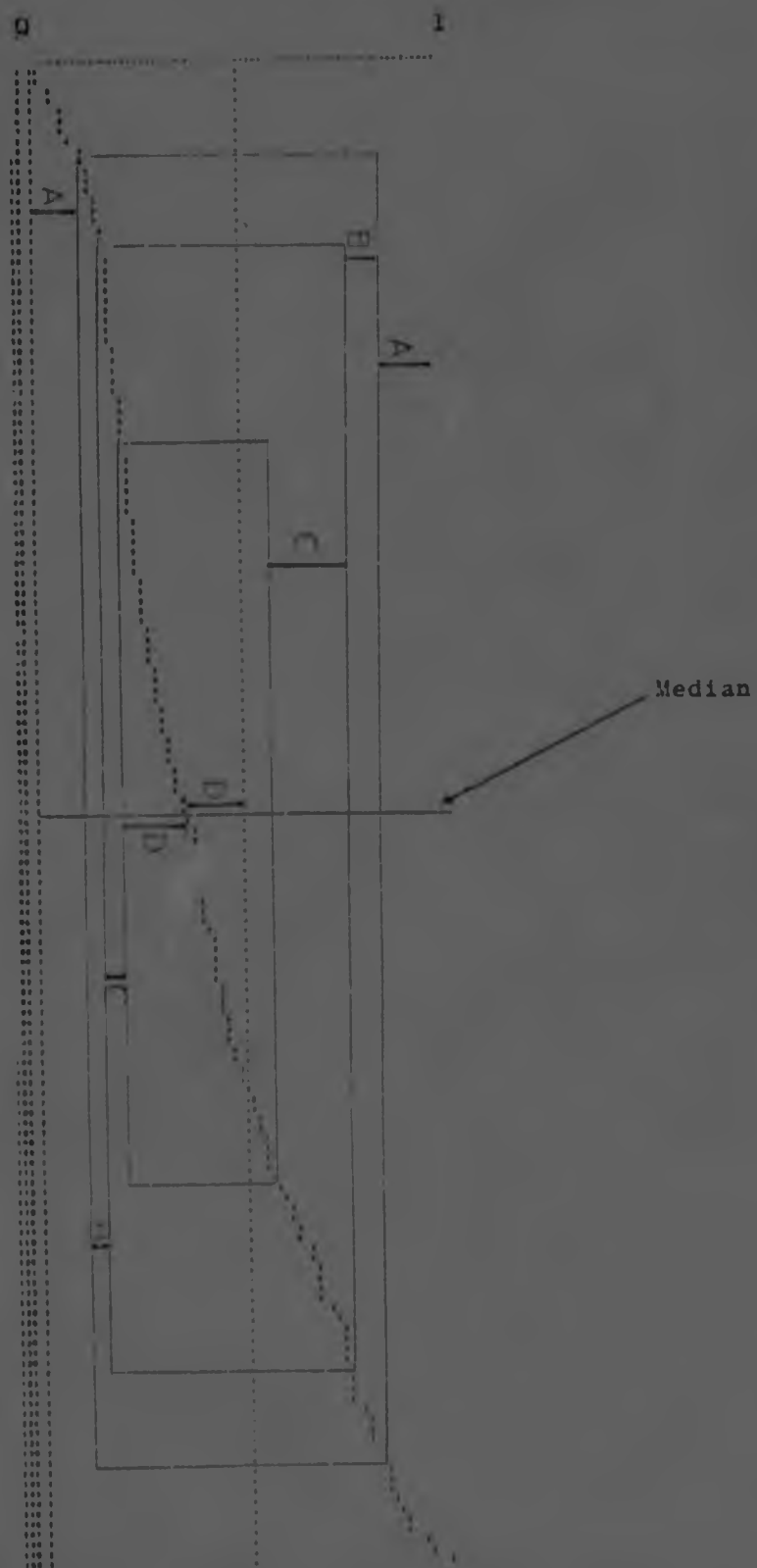


Fig. 9.3 Quantile-box plot of $\log_{10}$ of the airline passenger data



Fig. 9.4 Quantile-box plot of the square-root of the airline
passenger data



For the data to have a symmetric appearance, $x = \alpha_1/\alpha_2$ ($\alpha_1 = a, b, c, d$) should be approximately 1. The α values for each plot are given in table 9.1. Using c and d as criteria (since there is likely to be more variation in the tails of the data and hence the extreme quantiles), then since $\log_{10}$ gives minimum values for c and d, I decided to make use of this transformation⁺⁺

TABLE 9.1 A comparison of quantile-box plots

BOXES		ORIGINAL	LOG	SQRT
0 - 1/16	A	37,125	0,3664	1,8454
1/16 - 1/8	B	16,625	0,1289	0,7311
1/8 - 1/4	C	15,25	0,1047	0,6326
1/4 - M	D	43,0	0,2481	1,6307
M - 3/4	D	73,5	0,3184	2,4161
3/4 - 7/8	C	54,75	0,1852	1,5912
7/8 - 15/16	B	28,625	0,0843	0,7779
15/16 - 1	A	49,125	0,1340	1,2651

RATIOS (1)			
	O	L	S
A :	0,76	2,73	1,46
B :	0,58	1,53	0,94
C :	0,28	0,57	0,40
D :	0,59	0,78	0,68

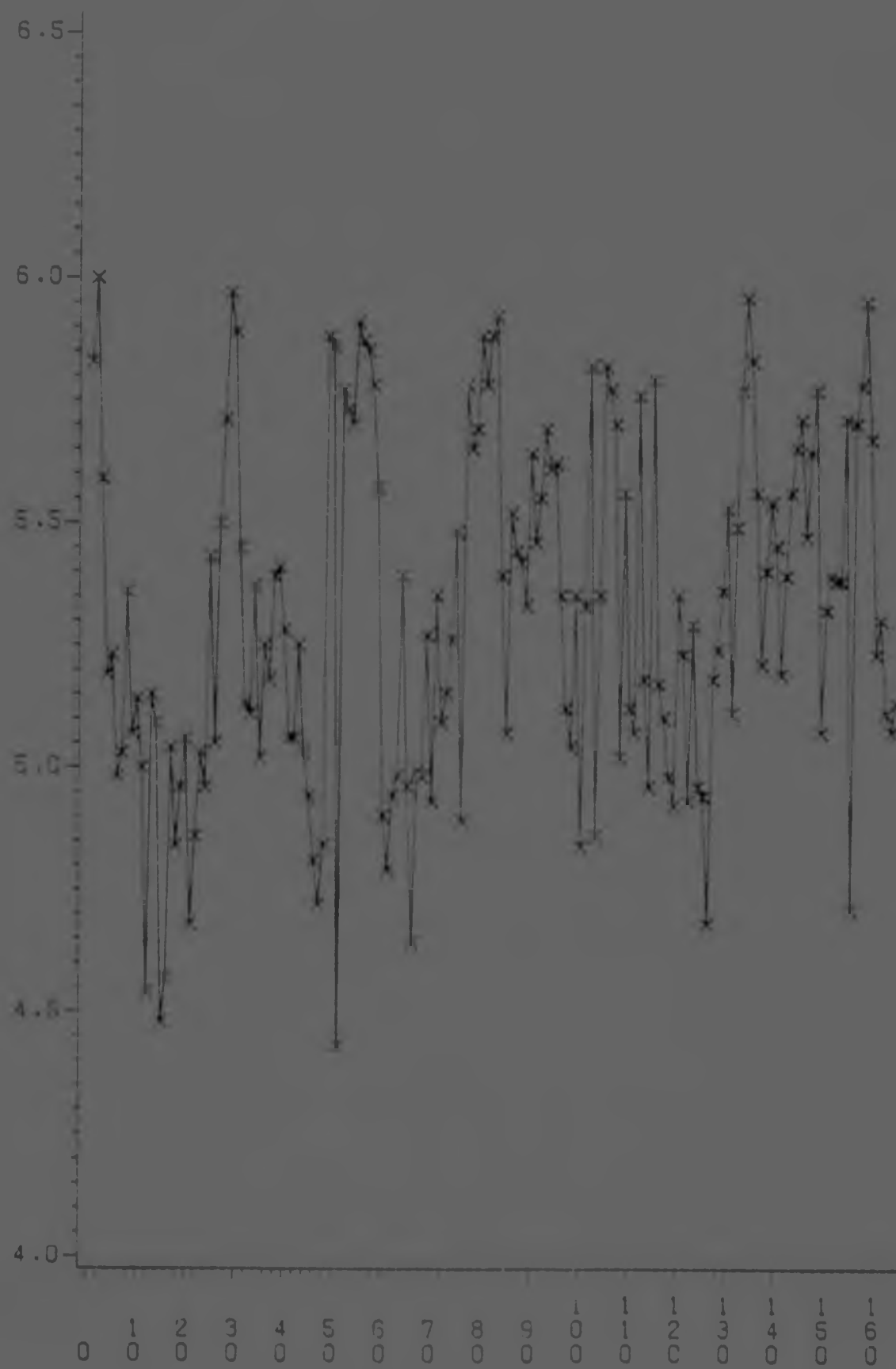
Since this transformation step is totally optional, I am quite satisfied that the very rough procedure I have used is accurate enough for any applications we may wish to make to the data. Nevertheless, the transformation does affect the shape of the distribution and hence the appropriate model.

A plot of $\log_{10} x_t$ ($t=1,2,\dots,164$) is given in figure 9.5 (and a listing of the data is given in table A19.2, appendix A19).

⁺⁺ By minimum, I mean values closest to 1.

See 'Comment' at the end of the chapter on page 214

Fig. 9.5 Plot of the transformed ($\log_{10}$) airline passenger data



[For interest's sake, figures A19.1 - A19.4 and tables A19.3 - A19.4, in appendix A19, contain information about PASSIN .

9.3 DETECTING ANY DETERMINISTIC CYCLES

The autocorrelation function of the (new) transformed data set is given in figure 9.6. Since the function is obviously not damping out at a fast rate, the data would seem to be nonstationary. This is backed up by the results given in tables A19.6 - A19.8 (appendix A19) where the time-series has been divided into the subseries lengths given and then the mean, variance and acrf of each subsection calculated.

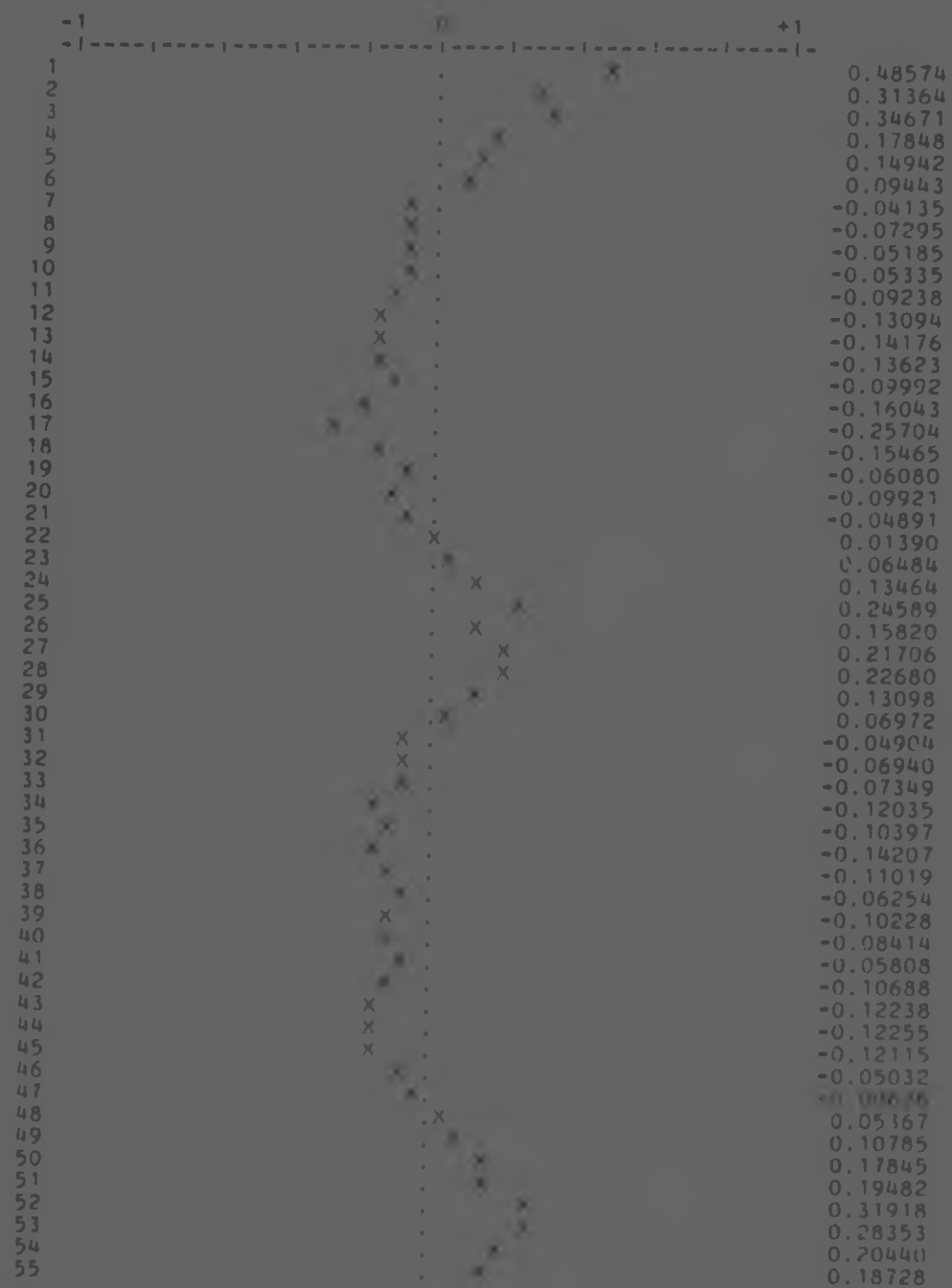
[Actually, for these tables, the first 156 (instead of 150) observations were used, in order to have segments consisting of the same number of observations (this should not affect the overall conclusion). The choice of subseries length was based upon the 'obvious' periods one may expect in weekly data (i.e. 13, 26 and 52) - any seasonality is 'more likely' to have periods corresponding to these values].

From these tables, we see that the series is at least nonstationary in its covariance structure since if the series were stationary, we would expect segments of equal length (from any point in time) to have approximately the same covariance structures. (See, for example, Gottman(1981), pg. 75).

Since the acrf of the $\log_{10}$ data seems to be cyclic with oscillations that do not die off quickly, we may conclude that any cycles in the data must be deterministic (since nondeterministic cycles damp down to zero quickly - Gottman (1981), pg 87).

Additional checks for distinguishing deterministic from

Fig. 9.6 Autocorrelogram of the transformed data



nondeterministic cycles consist of:

- (a) compute the periodogram with varying sample sizes; if the resulting plots show peaks that increase with sample size, this implies deterministic cycles (Gottman (1981), pg 197+)
- and
- (b) if a cycle is deterministic, then the (sample) sdf should peak sharply (approximately at one frequency); if the cycle is nondeterministic then the series repeats itself 'more or less' regularly and thus the sdf would be expected to peak across a band of frequencies.

Estimates of the sdf using the Parzen (b) window (see section 4.3.1, equation (4.25a)), with truncation points $M_1 = 0.3 \times N$ and $M_2 = 0.75 \times N$, are given in figures 9.7 and 9.8 resp. These plots seem to indicate sharp peaks thus implying deterministic cycles. (Full plots of the sample sdf are given in appendix A19 for truncation points 5%, 10%, 20%, 30% and 75% of N - see figures A19.5 - A19.10. Note how the spectrum estimate is initially smooth and as the truncation point increases tends to become 'rougher').

Now the sample sdf alone may be used to gauge at which frequencies there is 'significant' power and hence the frequencies of the deterministic cycles (and obviously the periods relating thereto). However I shall make use of the periodogram (an estimate of the sdf) to 'identify' the important frequencies (since this has been evaluated on a finer grid of frequencies).

The (natural) logarithms of the periodogram ordinates are given in fig. 9.9. The values on the vertical axis represent frequencies (e.g. $j = 6$ corresponds to $\omega_j = 2\pi \cdot 6/N$ where N' is the next power of 2 above 150 - the data to be modelled are $x_1, \dots, x_{150}$

Fig. 9.7 Spectrum estimate (using the Parzen (b) window (pg.80)
and truncation point 75) of the transformed ($\log_{10}$) data

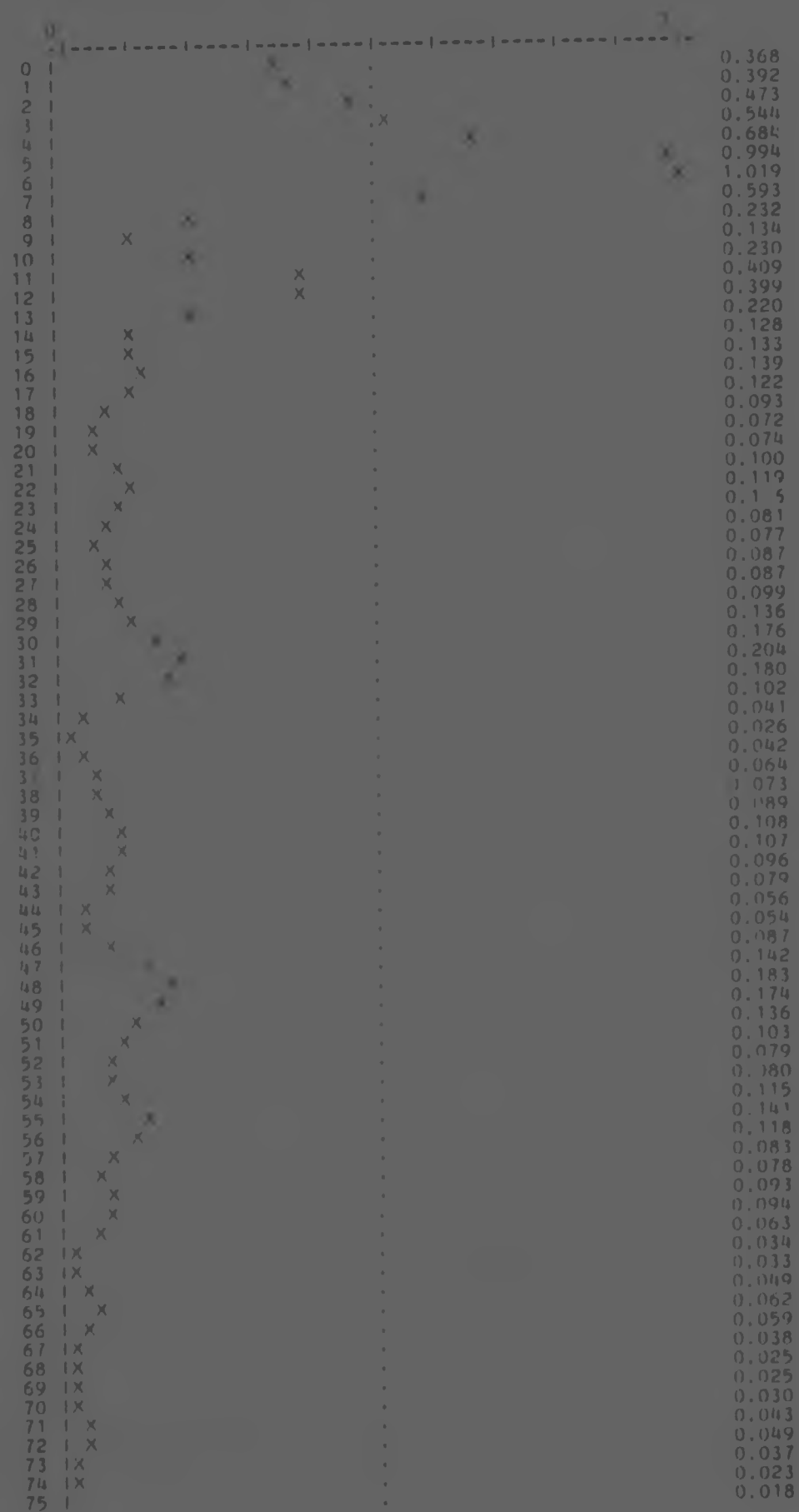
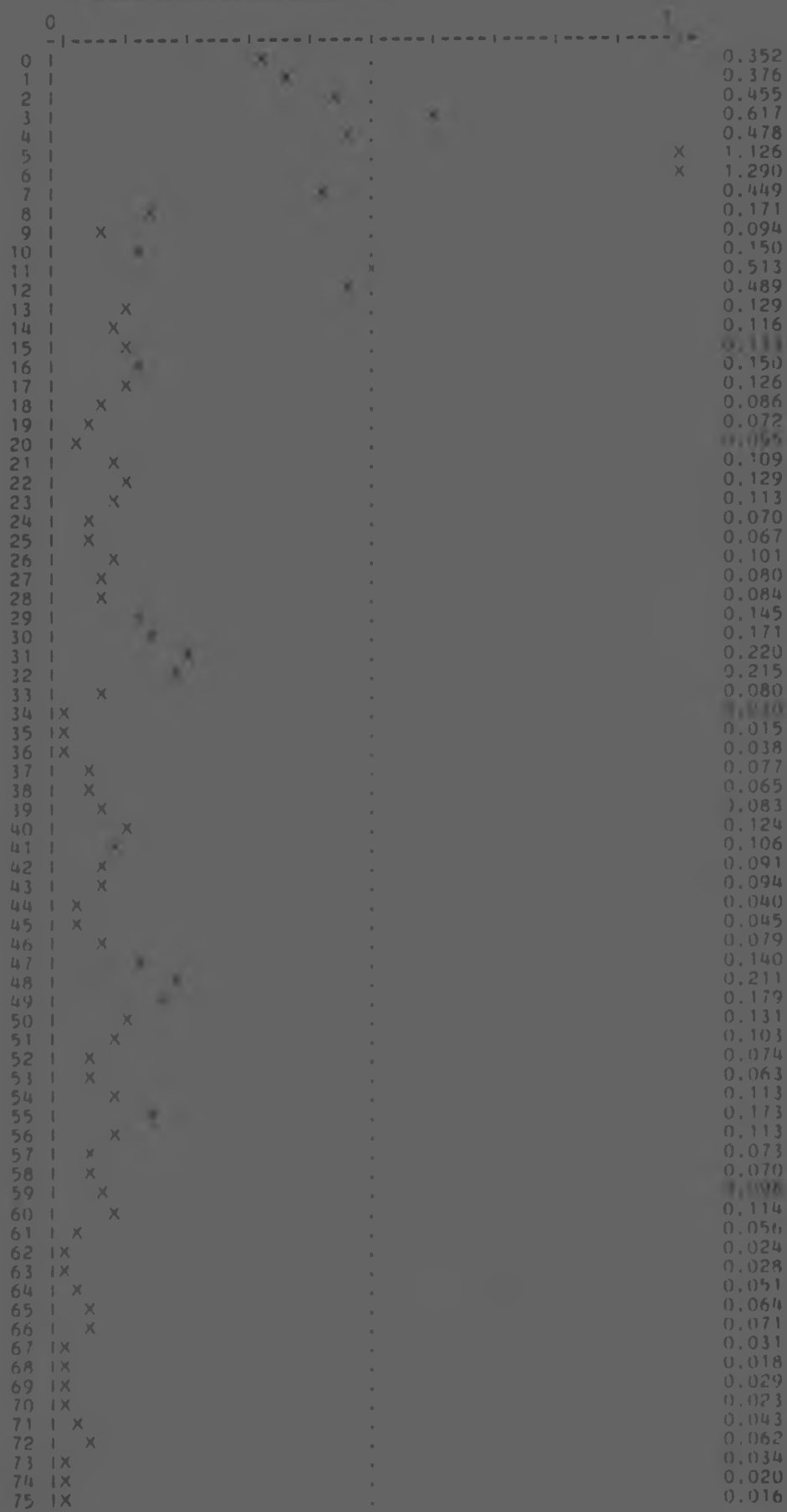


Fig. 9.8 Spectrum estimate (Parzen (b) window, truncation point 112)
of the transformed data



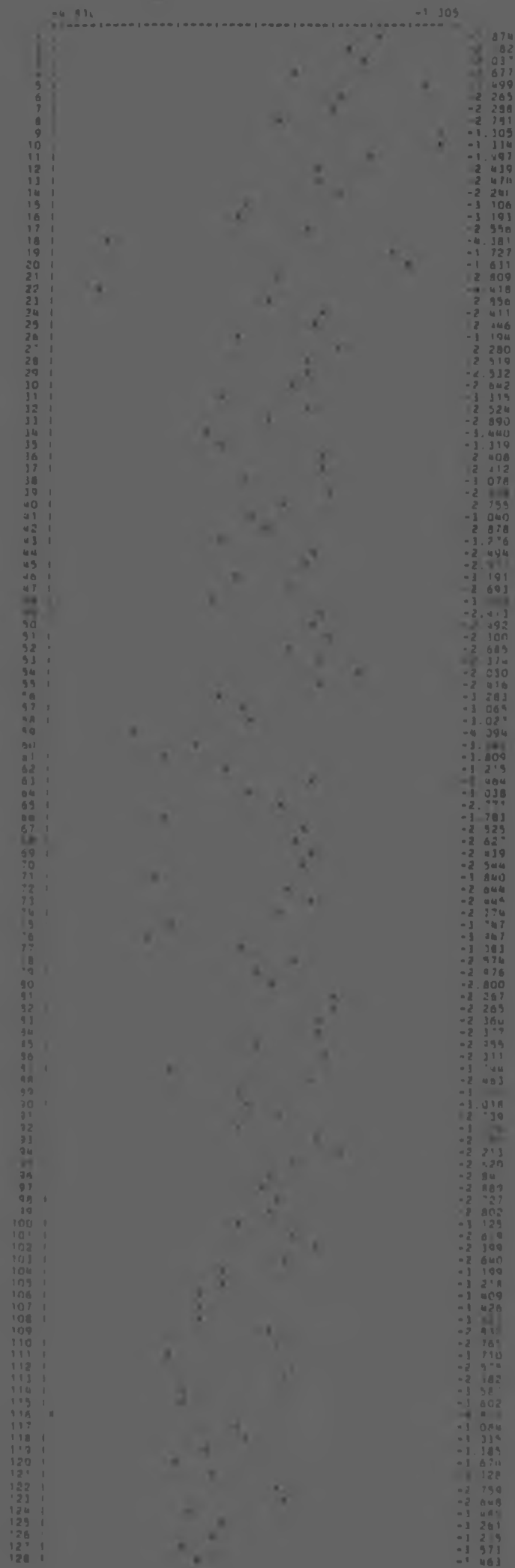


Fig. 9.9 $\log_{10}$ of the periodogram of the transformed data

- see program PGRAM in appendix A18). From this plot it appears that the 'major' peaks occur at w_5 , w_{10} and w_{20} . In figures 9.10 and 9.11 are the periodograms with 20% and 50% tapering (see section 5.4). These 'smoothed' periodograms make the peaks clearer.

The major peak or frequency would seem to correspond to w_{10} with possible subsidiary peaks at w_{20} and w_5 . However from section 3.4.1, note (vii), recall that if $\{x_t\}$ is periodic, with period h say, then the periodogram vanishes - i.e. is zero - at all frequencies except $w = 2\pi/h$ and its multiples - and w_{20} is a multiple of w_5 and w_{10} ; w_{10} is a multiple of w_5 .

However the sample sdf^5 (see either figs. 9.7 and 9.8 or figs. A19.5 - A19.10 in appendix A19) seem to indicate three 'distinct' peaks (not so much at w_5) thus possibly indicating a sum of three seasonal components. However, the procedure to be adopted here is that described in section 3.2 where the data will be detrended successively i.e. the major seasonal component removed from the data, the residuals then analysed for seasonality etc.

$$\begin{aligned}\text{Now } w_{10} &= 2\pi \times 10/256 \\ &= 0.245436926\end{aligned}$$

$$\begin{aligned}(\text{and the subscript } 10 \text{ in } w_{10} \text{ corresponds to a period of } \\ 2\pi/w_{10} = 26)\end{aligned}$$

and this will serve as an initial estimate for w_1^* in the model

$$x_t = \mu_1 + a_1 \cos(tw_1^*) + b_1 \sin(tw_1^*) + u_t$$

$$u_t = \text{stochastic process.}$$

w_{10} is used in the program HIDPI (which computes LSE's of all model parameters-see appendix A18) as an initial estimate of w_1^* . It serves to find a final estimate of w_1^* as well as a_1 and b_1 . $\bar{X}$, the sample mean, will serve as an estimate of μ_1 .

Fig. 9.10 $\log_{10}$ of the periodogram of the transformed data (20%

tapering)

Fig. 9.11 $\log_{10}$ of the periodogram of the transformed data (50%

tapering)

Defining $\hat{w}_1$, $\hat{a}_1$ and $\hat{b}_1$ as final estimates of w_1 , a_1 and b_1 , and $\hat{\mu}_1 (= \bar{X})$ as an estimate of μ_1 , we find that

$$\hat{\mu}_1 = 5,30486$$

$$\hat{w}_1 = 0,2305006$$

$$\hat{a}_1 = 0,246352$$

$$\hat{b}_1 = -0,00147534$$

with a residual variance of 0,1025 (a reduction of 22%)

A new time-series is now found by forming the residuals

$$\varepsilon_t = x_t - \hat{\mu}_1 - \hat{a}_1 \cos(\hat{w}_1 t) - \hat{b}_1 \sin(\hat{w}_1 t) \quad t = 0, 1, \dots, 149$$

The periodogram (tapered 20%) of the new series $\{\varepsilon_t\}$ is given in figure 9.12. The sample sdf (using a Parzen(b) window with truncation point $M = 60 = 40\%$ of N) is given in fig. 9.13

From both of these plots we now see that the angular frequency corresponding to w_{10} has been 'modelled out' of the data. (Note that w_{10} corresponds approximately to a period of 26)

The 'power' at frequencies w_8 and w_{20} now becomes more apparent, thus reinforcing the suspicion of further seasonal components. (Note that for the sample sdf, the 'numbered' axis represents the frequencies $w'_j = 2\pi j/N$ (i.e. not N') since $w'_{75} = w'_{N/2}$ corresponds to π in the periodogram, w'_{128} corresponds to 2π). This power is particularly evident in the sample sdf.

To remove any power at w_{20} in the periodogram (or w'_{11} in the sdf) use

$$\begin{aligned} w_{20} &= 2\pi \times 20/256 \\ &= 0,490873851 \quad (w_{20} \text{ corresponds to a period} \end{aligned}$$

of 13)

as an initial estimate of w_2 in the model



Fig. 9.13 $\log_{10}$ of the periodogram of the (once) de-signed transformed data (20% tapering)

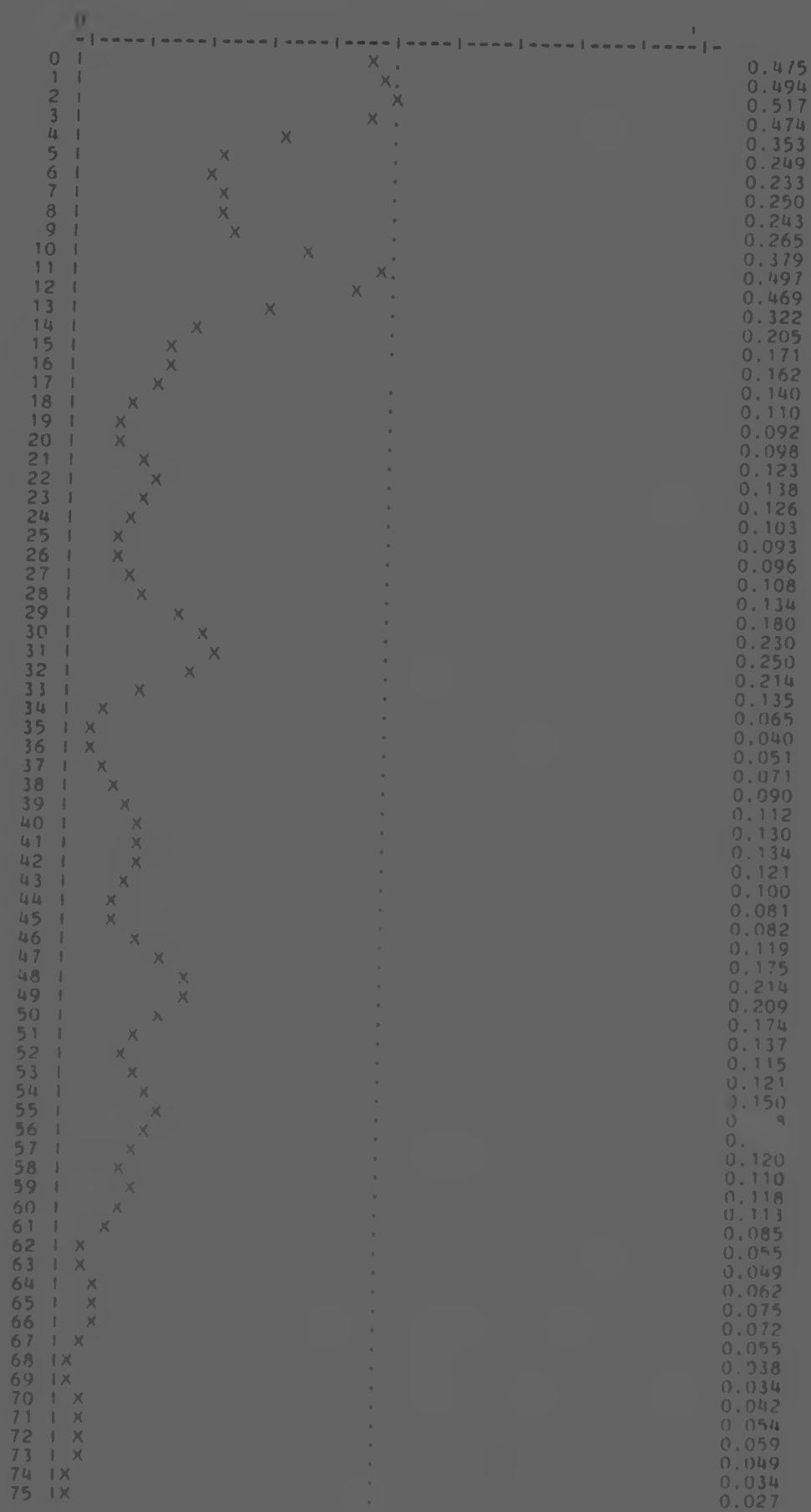


Fig. 9.13 Spectrum estimate of the (once) de-sined transformed data (Parzen (h) window, truncation point 60)

$$\epsilon_t = \mu_2 + a_2 \cos(w_2^* t) + b_2 \sin(w_2^* t) + v_t$$

v_t = stochastic process

Use w_{20} in the program HIDPER in the same manner that w_{10} was used.

Here we find that

$$\hat{\mu}_2 = -0,00308542 \approx 0$$

$$w_2 = 0,4776025$$

$$a_2 = 0,169991$$

$$b_2 = 0,0327472$$

with a residual variance of 0,0859 (a reduction (overall) of 35%)

Form the (new) time-series

$$x_t = \epsilon_t - \hat{\mu}_2 - \hat{a}_2 \cos(w_2 t) - \hat{b}_2 \sin(w_2 t) \quad t = 0, \dots, 149$$

The periodogram (tapered 20%) of the new series x_t as well as the sample sdf (using the Tukey-Hamming window - see section 4.3.1, equation (4.22) - and truncation point $M = 75 = 50\%$ of N) are given in figures 9.14 and 9.15 resp. (The use of different windows is no reflection of inadequacy, but rather just serves as an illustration of the different windows).

Of the three frequencies identified initially as having potential power, two have been removed - namely w_{10} and w_{20} (and in the sdf w_6 and w'_{11}). The final frequency noted in the periodogram given in fig. 9.9 (i.e the periodogram of the transformed data, no components removed) is w_3 (or w'_1 in the case of the sdf). It should not be too surprising to see then that the periodogram and sdf in figures 9.14 and 9.15 display power at this frequency. Thus we must remove the component associated with this frequency i.e. model

$$z_t = \mu_3 + a_3 \cos(w_3^* t) + b_3 \sin(w_3^* t) + v_t$$

z_t = stochastic process.

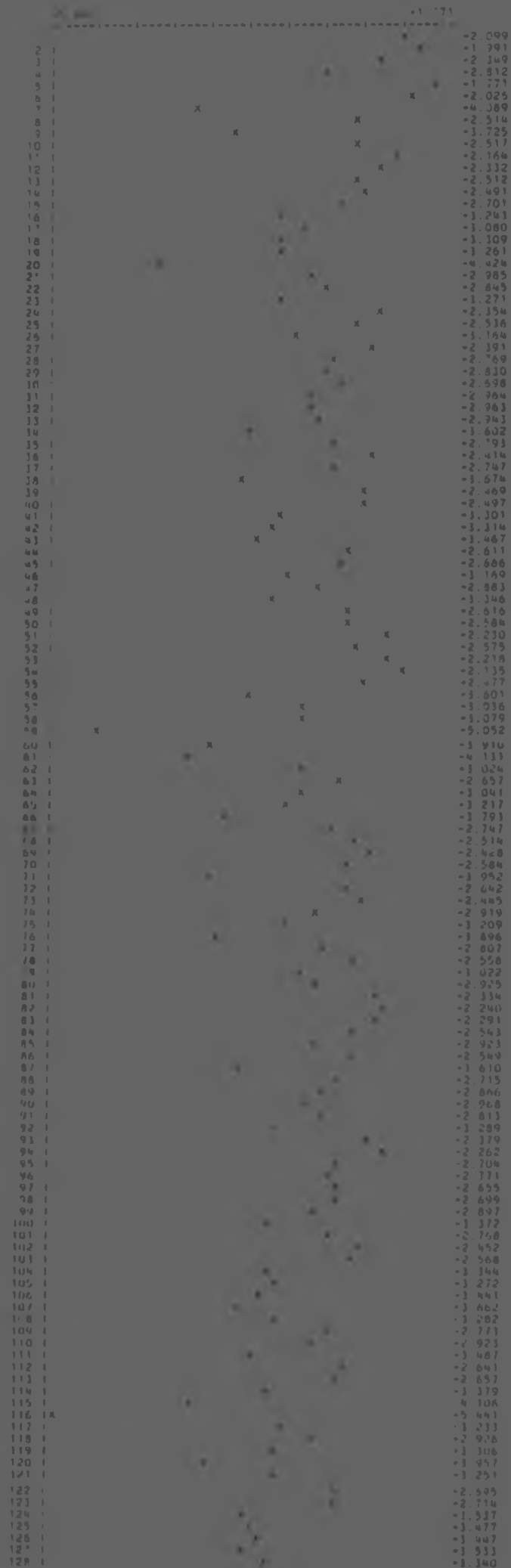


Fig. 9.14 $\log_{10}$ of the periodogram of the (twice) de-trended transformed data (20% tapering)

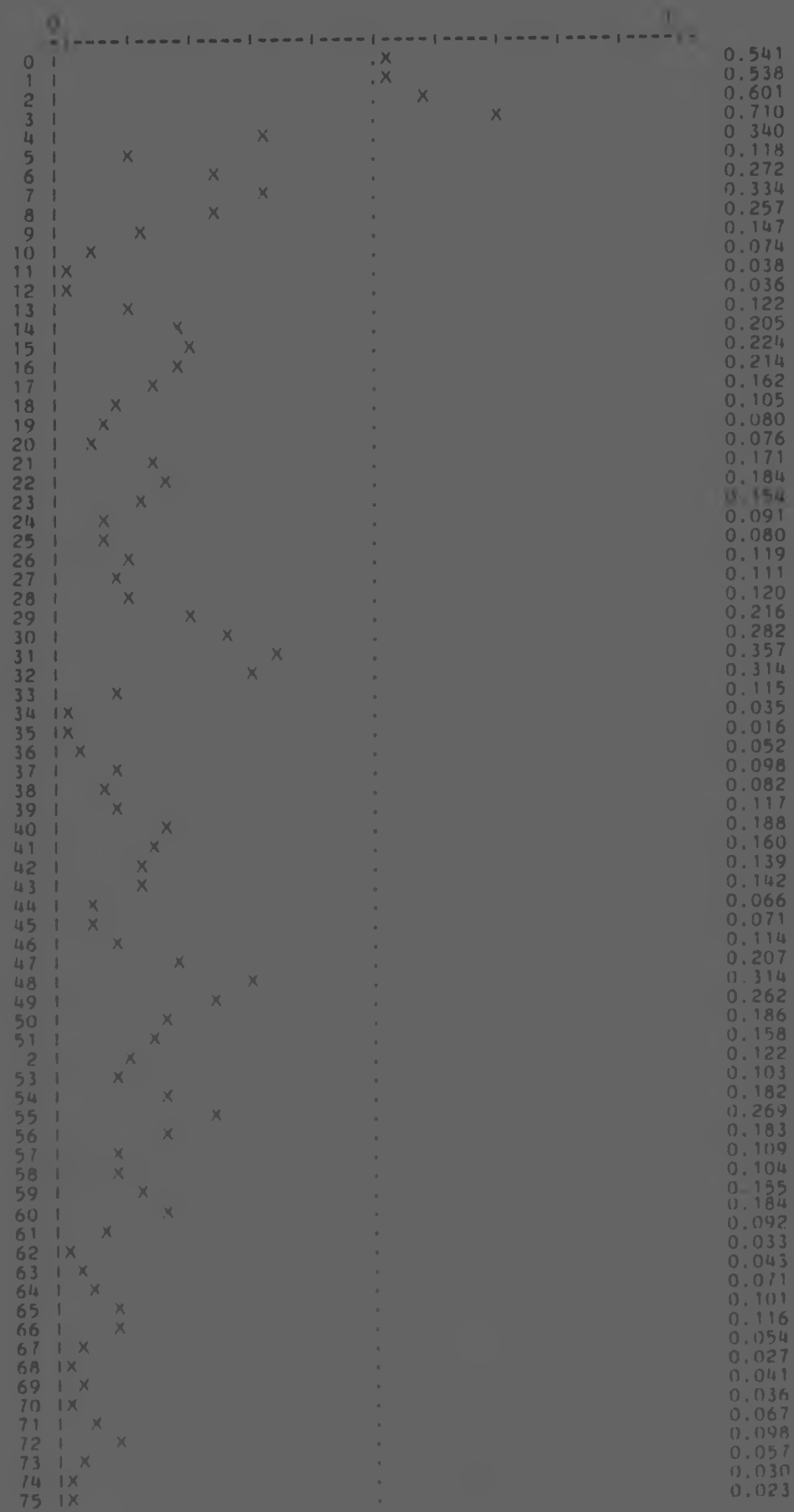


Fig. 9.15 Spectrum estimate of the (twice) de-sined transformed data (Tukey-Hanning window, truncation point 75)

Using $w_2 = 2\pi \times 5 / 256$
 $= 0,122718463$

as an initial estimate of w_3 in HIDPER, we find that

$$\hat{\mu}_3 = 0,00250543 \approx 0$$

$$\bar{w}_3 = 0,1251879$$

$$\bar{a}_3 = 0,0518517$$

$$\bar{b}_3 = -0,143498$$

[Note that w_2 and w_3 correspond approximately to a period of 52]

Create the new time-series $\{e_t\}$ by

$$e_t = x_t - \bar{\mu}_3 - \bar{a}_3 \cos(\bar{w}_3 t) - \bar{b}_3 \sin(\bar{w}_3 t)$$

$$t = 0, 1, \dots, 149.$$

$\{e_t\}$ is a time-series which has three seasonal components removed. These components correspond to periods of 13, 26 and 52. $\{e_t\}$ has a sample variance of 0,0741 a reduction of 44% from the original variance of the log-transformed series of 0,131476.

Instead of performing this third sequential removal, it is possible, using w_{10} , w_{20} and w_3 (as estimates of w_1^* , w_2^* and w_3^*) in HIDPER to obtain $\{e_t\}$ in one step
 i.e. model

$$x_t = \log_{10} \text{ of original data}$$

$$= \mu + \sum_{j=1}^3 \{a_j \cos(w_j^* t) + b_j \sin(w_j^* t)\} + \varepsilon_t$$

ε_t = residual error

From HIDPER, we find

$$\begin{aligned}
 \bar{\mu} &= 5,30387 \\
 w_1 &= 0,2331302 \\
 w_2 &= 0,4777789 \\
 w_3 &= 0,1239148 \\
 \bar{a}_1 &= 0,234516 \\
 \bar{b}_1 &= 0,027696 \\
 \bar{a}_2 &= 0,166452 \\
 \bar{b}_2 &= 0,0320354 \\
 \bar{a}_3 &= 0,0417698 \\
 \bar{b}_3 &= -0,150152
 \end{aligned} \tag{9.1}$$

and using these values in DETRND we find the sample variance of $\{e_0, \dots, e_{149}\}$ is 0,0738. Thus we have a reduction in variance of 44%.

This 'de-sining' procedure could have been stopped after removing the first component (corresponding to a period of approximately 26) which gave a reduction in variance of $\approx 23\%$. When three components are removed (corresponding to (approximate) periods of 26, 13 and 52) there is a variance reduction of $\approx 23\%$ i.e. the latter two components account for $\approx 23\%$ of the variability in the data. Obviously, this removal of cycles could be continued, however this may lead to the inclusion of an excessive number of model parameters (some possibly redundant). The analyst must decide at this stage how closely he/she wishes to follow the principle of parsimony.

The periodogram (no tapering) of the data (residuals) with 3 seasonal components removed simultaneously is given in figure 9.16. According to the values given in equation (9.1), we have

Fig. 9.16 $\log_{10}$ of the periodogram of the (thrice) de-sined trans-

formed data

$$\bar{w}_1 = 0,2331302$$

$$\bar{w}_2 = 0,4777789$$

$$\bar{w}_3 = 0,1239148$$

(corresponding to periods of 26.95, 13.15 and 50.71 resp.) and these will correspond to abscissae 9.50, 19.47 and 5.05 (resp.) in the periodogram i.e. $j = 9.50, 19.47$ and 5.05 in $w_j = 2\pi j/N'$ ($N'=256$) will yield $\bar{w}_1, \bar{w}_2$ and $\bar{w}_3$ (to 4 decimal places).

From the periodogram in Figure 9.16, we see that there appears to be 'holes' at the abscissae 5, 7 (=9) and 19 thus denoting the removal of w_1, w_2 and w_3 (see Bloomfield (1976), p.80). The peaks in this periodogram all occur (approximately) at multiples of 9.50, 19.47 and 5.03 thus reinforcing the assumption of additive seasonal components (see note (vii) section 3.4.1 or Bloomfield (1976), pg.80+).

However as a result of leakage and other factors (namely additional seasonal terms) some of these peaks may have magnitudes which are unacceptably large - for example additional seasonal terms may in fact have ordinates which correspond to a multiple of one of the ordinates of the seasonal components removed. Thus to be doubly certain no 'extra' (significant) seasonal terms are influencing the data (residuals), it is advisable to also examine the spectral estimate of the residuals. Spectral estimates (using the Tukey-Hanning window with truncation points $M = 30, 60$ and 75 (20%, 40% and 50% resp) - any values above or below these truncation points seemed to produce spectra which were either too 'smooth' or too 'rough' - are given in figures 9.17 - 9.19.

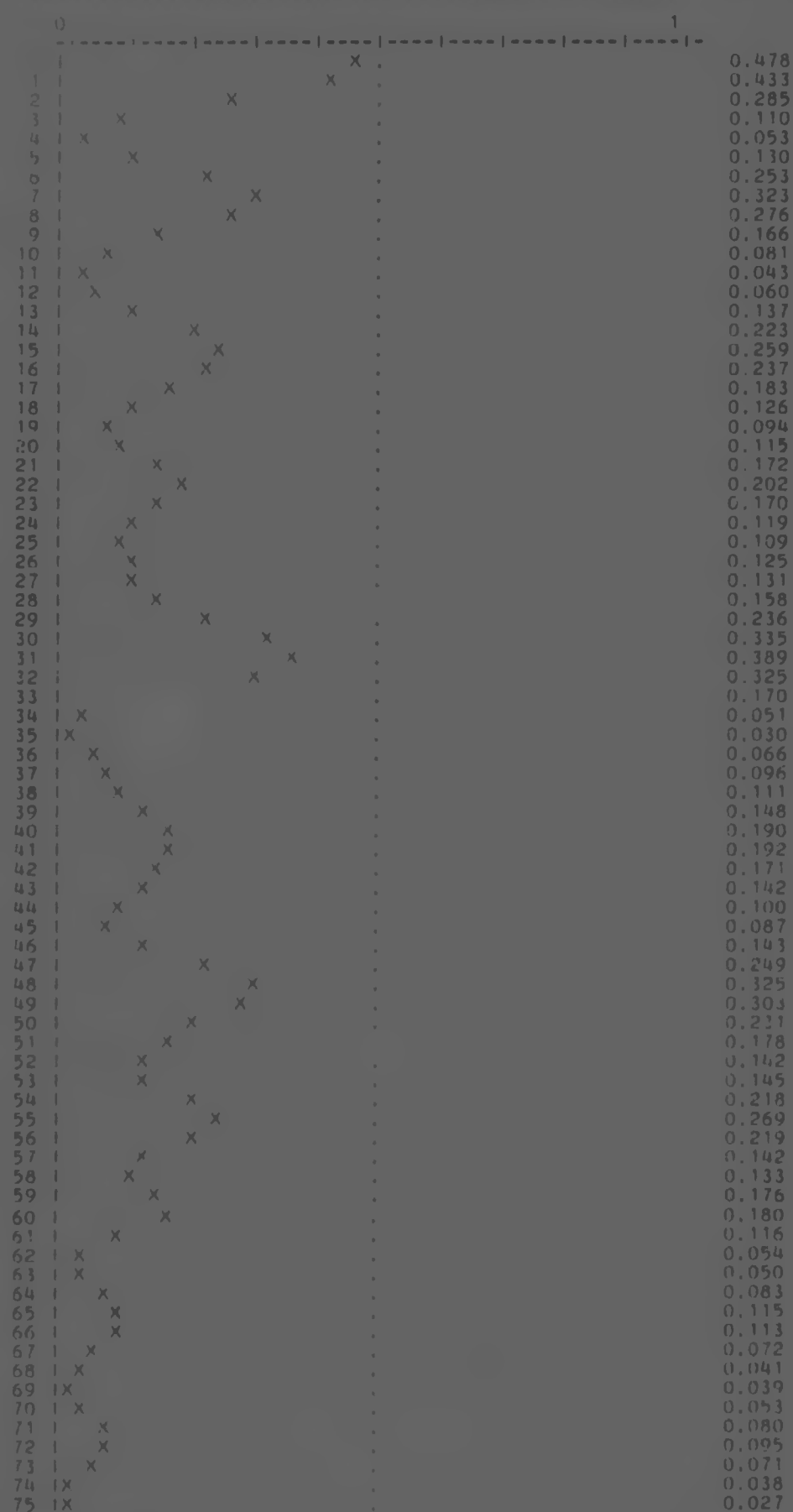
As the truncation point increases, peaks become more significant (pronounced). These peaks, although not as pronounced as

Fig. 9.17 Spectrum estimate of the (thrice) de-sined transformed
data (Tukey-Hanning window, truncation point 30)



Fig. 18 Spectrum estimate of the (thrice) de-sined transformed

data (Tukey-Hanning window, truncation point 60)



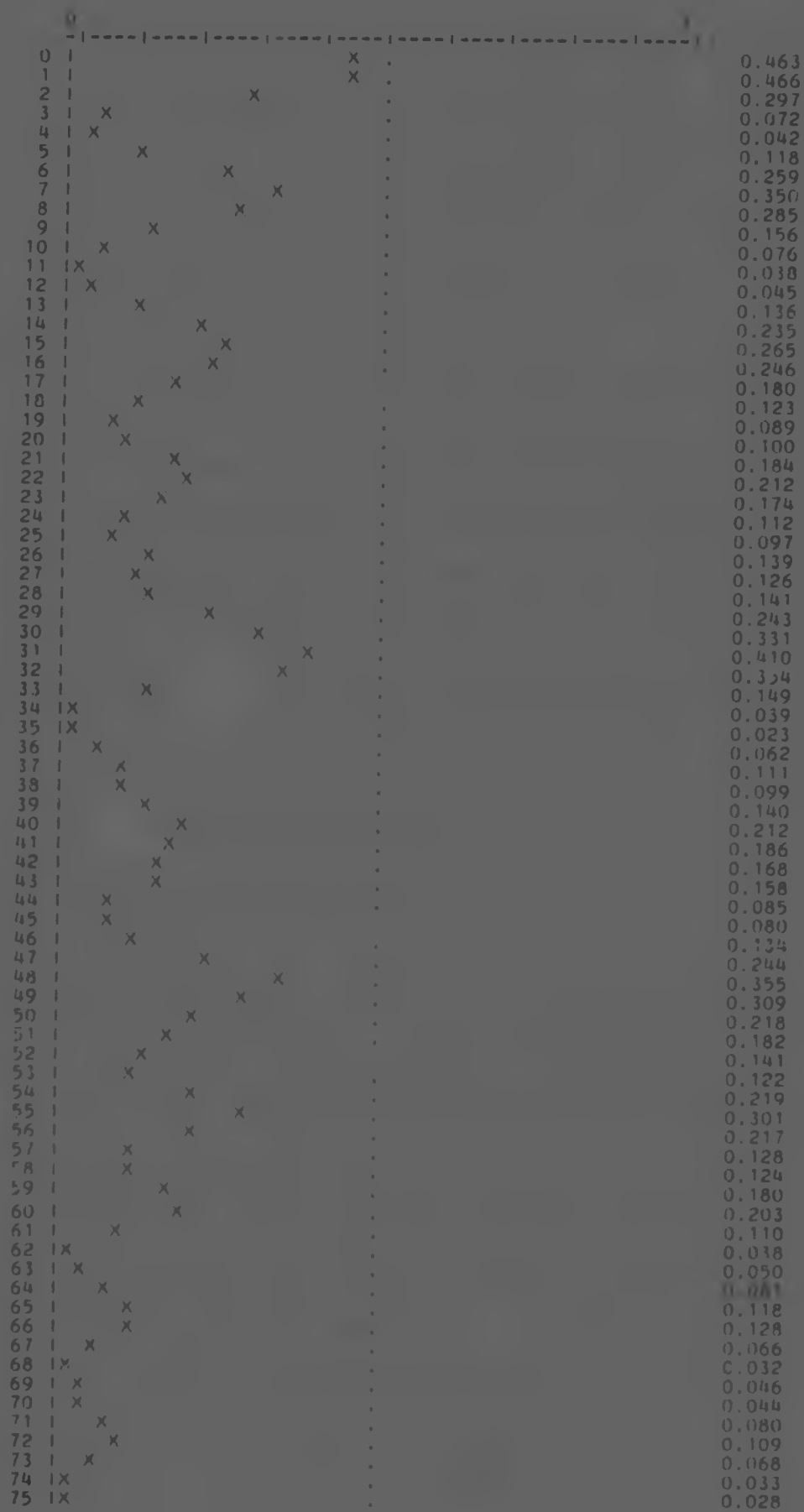


Fig. 9.19 Spectrum estimate of the thrice (de-sined transformed data) (Tukey-Hanning window, truncation point 75)

those corresponding to the peaks removed, are also visible in the spectral estimates of the $\log_{10}$ of the data (see figures A19.5 - A19.10) - however they will obviously account for less of the variability in the data (the peaks in figs. 9.17 - 9.19 have such large magnitudes now since they account for a 'large' amount of the variability in the residuals i.e. in the residual variance).

The peaks may be 'removed' from the spectrum of the residuals but as mentioned earlier, since one must decide on whether to adhere to the principle of parsimony, the choice is the analysts. In our case, we shall leave the data (i.e. the residuals) as they stand. The residuals will now be analysed using Box-Jenkins techniques.

9.4 THE BOX-JENKINS ANALYSIS

9.4.1 STATIONARITY

The first stage of the Box-Jenkins analysis consists of determining whether the residuals (referred to from here on as the data) are stationary. To do this, both the procedure of Hill and Woodworth (1980) - see section 7.1, equation (7.6)) - and a visual examination of the acrf of the residuals, will be given.

The first few results of the Hill-Woodworth method are given in table 9.2. (The maximum allowed degree of nonseasonal differencing was 2, of seasonal differencing, 1, and the maximum allowed seasonal base was 52). Recall that the degree of differencing, producing the minimum variance, is defined as 'optimum'.

* See 'Comment' at the end of the chapter on page 214

Table 9.2

DS = degree of seasonal differencing

DNS = degree of nonseasonal differencing

SB = seasonal base

<u>DNS</u>	<u>DS</u>	<u>SB</u>	<u>Variance</u>
0	0	0	0,673754
0	1	19	0,110626
0	1	52	0,123510
0	1	3	0,128368
1	0	0	0,128521
0	1	20	0,129580
0	1	6	0,132495

(the remaining results may be neglected - note how the variance is strictly increasing).

From Table 9.2 it is clear that no differencing is required to produce minimum variance. Note that the second-best result occurs when seasonal differencing (once) with a seasonal base of 19 is applied. A seasonal base of 19 (or period of 19) should result in a peak in the spectrum of the residuals near 7 and a peak in the periodogram near ordinate 13 - this is exactly what occurs. [To obtain these values recall that periods in the spectrum and periodogram are derived from

$$\text{period} = 2\pi/\omega$$

and

$$\hat{w}_j = \begin{cases} 2j/N & \text{spectrum} \\ 2j/N' & \text{periodogram} \end{cases}$$

(N' = next power of 2 above N)

$$\text{i.e. period} = \begin{cases} N/j \\ N^2/j \end{cases}$$

The residual acrf is given in figure 9.20.

The largest ordinate occurs at $j = 19$ (reinforcing the suspicion of some seasonality with period 19). Other than this lag, there seems to be no other lag which may hint at some nonstationarity and even so, the acf at lag 19 can just be reflecting the necessity for a seasonal component in the Box-Jenkins model.

Due to the general size of the acf's, one may suspect that in fact the residuals are pure white noise. To check this, note that

$$V(r'_k) = \frac{1}{N} \left\{ 1 + 2 \sum_{i=1}^q \rho^2(i) \right\} \quad k > q \quad (9.2)$$

q some integer

(see appendix A3 or Box and Jenkins (1976), pg.34).

If the whole series is white noise, then $q = 0$ in (9.2) and

$$V(r'_k) = \frac{1}{N} \quad \text{for all } k > 0$$

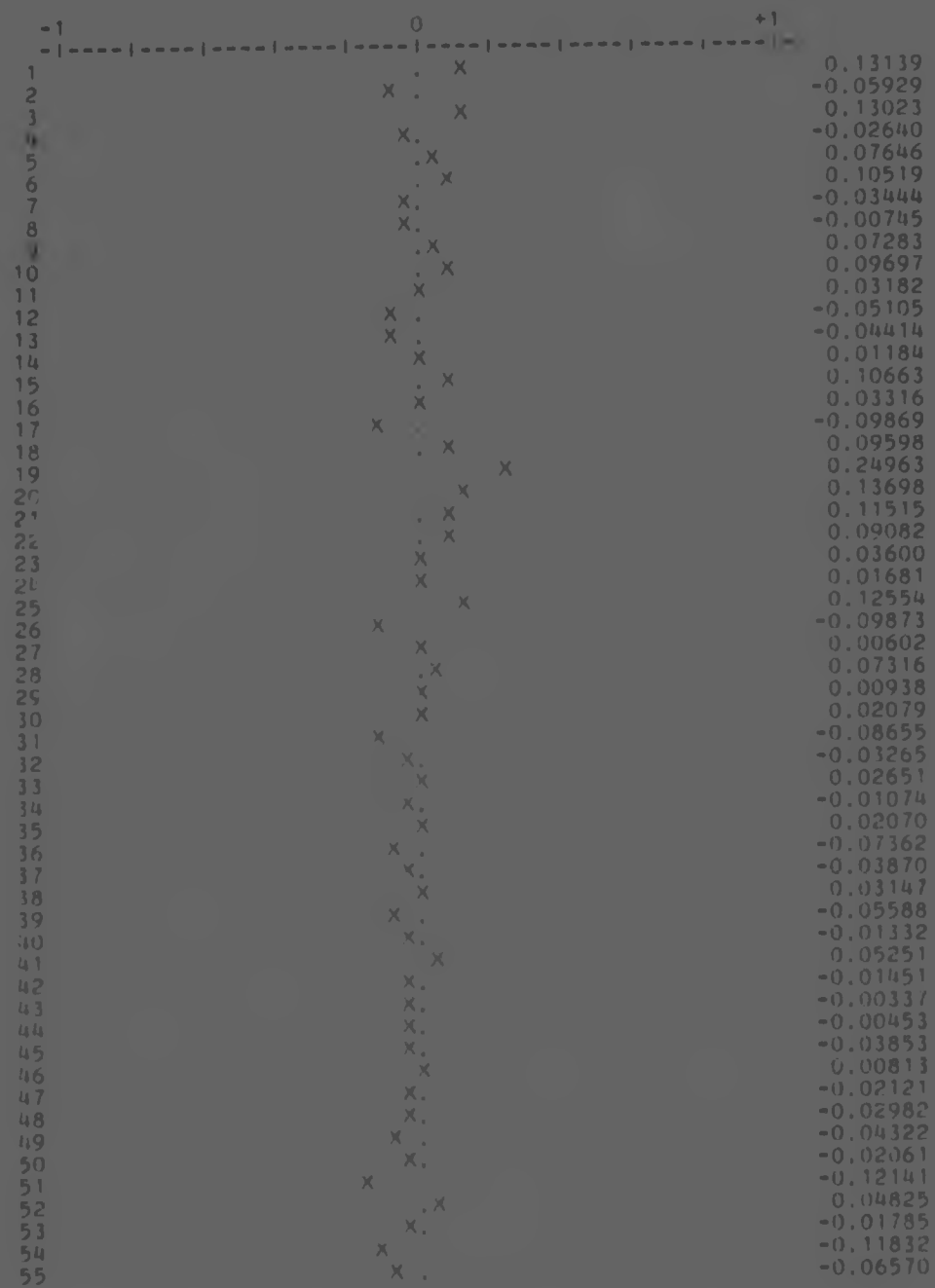
$$= 0.006 \quad \text{in our case}$$

The corresponding standard error is $0.08165 (= \sqrt{0.006})$.

Now if N is large then $r'_k \sim N(0, \frac{1}{N}) \quad \forall k$ (see appendix A13 or Anderson (1942)) when the data are white noise. Thus 95% confidence bounds for the r'_k would be $\pm 2/\sqrt{N}$ (approximately). In our case this results in 0.1633 and -0.1633.

Any r'_k falling outside these limits would seem to indicate a non-white noise process. The only value (lag) at which this occurs is

Fig. 9.20 Autocorrelogram of the (thrice, de-sined transformed data



at $j = 19$. Beyond lag 19, the acrf would appear to be negligible (i.e. $r_k = 0$ $k > 19$ (and in fact also for $k \leq 19$)).

In summary then, the data would appear to be stationary since

(a) no differencing (according to the method of Hill and Woodworth (1980)) is required

and

(b) the acrf appears to be damping out to zero, the only problem spot occurring at lag 19 (but this could indicate the necessity for a seasonal component in the model).

9.4.2 MODEL IDENTIFICATION AND VERIFICATION

A plot of the partial autocorrelation function is given in figure 9.21, and plots of the inverse acrf and inverse pacrf (i.e. iacrf and ipacrf) - see section 7.2.1 pg. 130 - are given in figures 9.22 and 9.23 resp. In all three plots, the dominant lag would appear to be 19. Besides this lag, there appears to be no power at any of the other lags.

The same technique for testing for whiteness of the acrf can be used in the pacrf - again only lag 19 appears to be significant.

Suppose for the moment, we ignore the dominant lag 19. By using the techniques described in section 7.2 or by using an automatic model selection procedure (on the computer), it may be possible to identify any ARMA(p,q) models ($p, q \leq 5$). This could be done by generating all possible ARMA (p,q) models ($p, q \leq 5$) and choosing the model with the minimum residual variance as 'best-fitting' - Akaike's AIC will take the number of parameters in the model into account (see section 7.4.2)).

The 'best' model then turns out to be an AR(3) model, which has

Fig. 9.21 Partial autocorrelogram of the (thrice) de-sined trans-
formed data

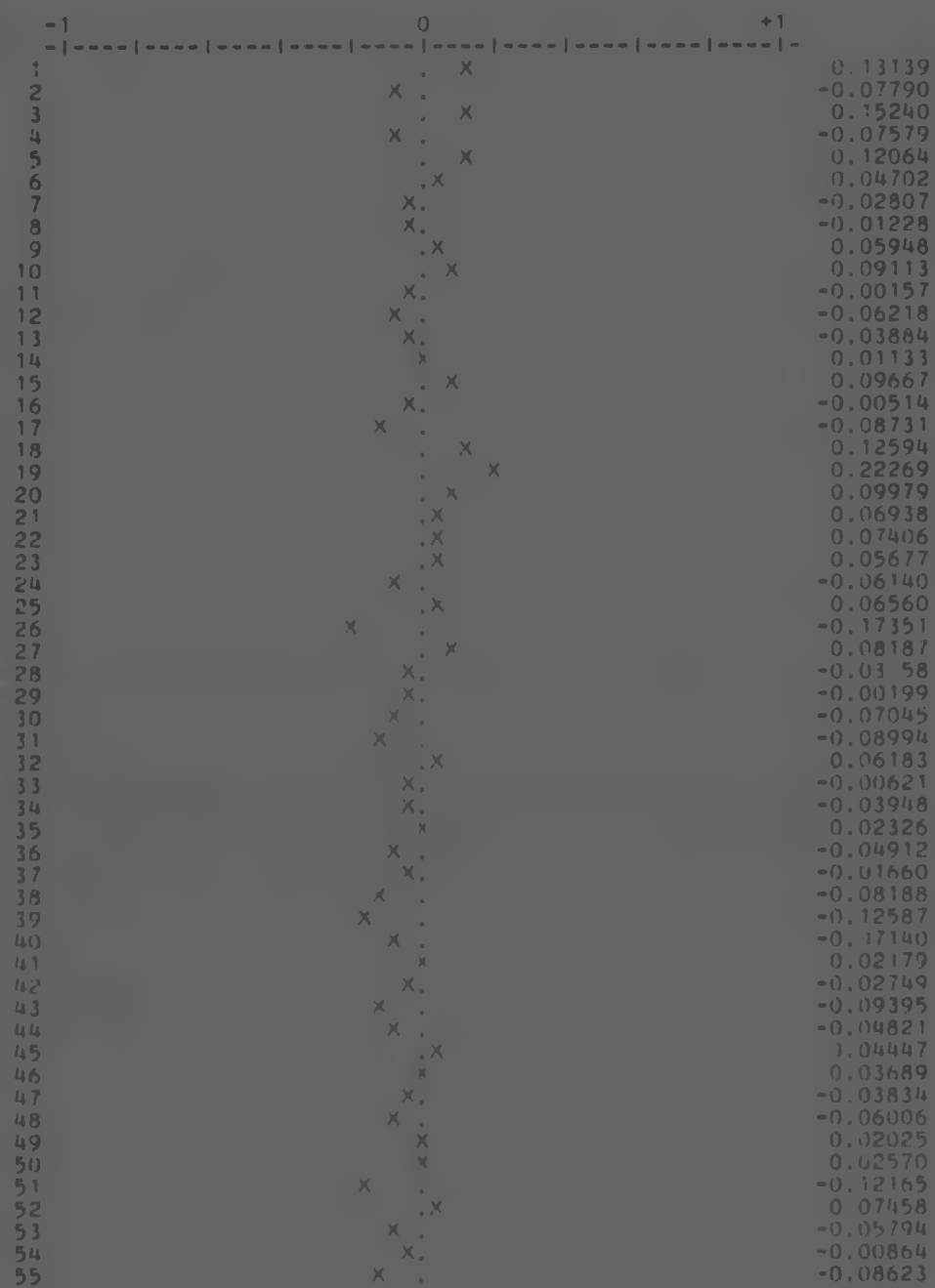




Fig. 9.23 Inverse partial autocorrelation of the

(thrice de-sined transformed data

Fig. 9.22 Inverse autocorrelation of the (thrice) de-sined

transformed data

parameters

$$\begin{aligned}\phi_1 &= 0.1578 & \text{S.E.} &= 0.0817 \\ \phi_2 &= -0.1046 & &= 0.0824 \\ \phi_3 &= 0.1616 & &= 0.0826 \\ \sigma^2 &= 0.07017\end{aligned}$$

A plot of the residual autocorrelation function is given in figure 9.24 with associated 99% confidence levels (the first column of data represents the residual acr's and the second column the upper asymptotic confidence levels - the lower levels follow by symmetry [see, for example, Hipel et al (1977)]).

All the acr's except the one at lag 19, fall within these significance levels - thus indicating 'power' at lag 19. This lag could be removed by 'overfitting' (see, for example, Hipel et al (1977)) i.e. instead of fitting an AR(3) model fit an AR model which incorporates ϕ_1, ϕ_2, ϕ_3 and ϕ_{19} [this overfitting might be attempted if the partial acrf of the 'data' had significant, but decreasing lags at 1, 2, 3 and 19]. However, note that a period of 19, besides being doubtful, makes no intuitive sense (as do 13, 26 and 52) and is difficult to explain. When attempted, this model (i.e. nonzero coefficients at ϕ_1, ϕ_2, ϕ_3 and ϕ_{19}) only produced a white noise variance estimate of $\sigma^2 = 0.066705$.

Several other over-fitting models were attempted and these (and their white noise variance estimates) are listed in table 9.3.

Fig. 9.24 Residual autocorrelogram with 99% confidence limits

COL.1- RESIDUAL ACR'S
COL.2- UPPER 99% CONFIDENCE LIMITS

	-1	0	+1		
1		0		0.0127	0.0340
2		0		-0.0251	0.0378
3		0X0		0.0102	0.0465
4	0 X	.	0	-0.0805	0.2041
5	0	.	X 0	0.0940	0.2074
6	0	.	X 0	0.0729	0.2083
7	0 X	.	0	-0.0484	0.2098
8	0	X.	0	-0.0062	0.2102
9	0	.	X 0	0.0537	0.2102
10	0	.	X 0	0.0907	0.2102
11	0	X	0	0.0306	0.2102
12	0 X	.	0	-0.0704	0.2102
13	0 X	.	0	-0.0455	0.2102
14	0	X	0	0.0058	0.2102
15	0	.	X 0	0.0827	0.2102
16	0	X	0	0.0547	0.2102
17	0 X	.	0	-0.1105	0.2102
18	0	.	X 0	0.0616	0.2102
19	0	.	X	0.2130	0.2102
20	0	.	X 0	0.1110	0.2102
21	0	.	X 0	0.1055	0.2102
22	0	X	0	0.0285	0.2102
23	0	.	X 0	0.0453	0.2102
24	0	X.	0	-0.0320	0.2102
25	0	.	X 0	0.1304	0.2102
26	0X	.	0	-0.1289	0.2102
27	0	X	0	0.0220	0.2102
28	0	X	0	0.0654	0.2102
29	0	X	0	0.0008	0.2102
30	0	X	0	0.0413	0.2102
31	0 X	.	0	-0.0978	0.2102
32	0	X.	0	-0.0717	0.2102
33	0	X	0	0.0559	0.2102
34	0	X.	0	-0.0141	0.2102
35	0	X	0	0.0367	0.2102
36	0 X	.	0	-0.0633	0.2102
37	0	X.	0	-0.0391	0.2102
38	0	X	0	0.0315	0.2102
39	0 X	.	0	-0.0431	0.2102
40	0	X.	0	-0.0082	0.2102

Table 9.3 : Overfitting in AR models

<u>nonzero lags</u>	<u>$\hat{\sigma}^2$</u>
1,2,3,19,25,26	0,064949
1,2,3,4,5,19	0,065148
1,3,5,19,26	0,065782

Another way of removing the 'power' at lag 19 is to model a seasonal parameter with period 19. It is unlikely that any periodicity or seasonality would entail more than one AR or MA parameter so I limited P_s and Q_s to maximum values of one and tried several models. The 'best' models are listed in table 9.4.

Table 9.4 : SARIMA models

<u>model</u>	<u>$\hat{\sigma}^2$</u>	<u>AIC</u>
(0,0,3)(1,0,1)19	0,06058	-410,54
(3,0,0)(1,0,1)19	0,05988	-412,31
(0,0,4)(1,0,1)19	0,06040	-409,02
(4,0,0)(1,0,1)19	0,06003	-409,94
(0,0,1)(1,0,1)19	0,06337	-407,82
(0,0,0)(1,0,1)19	0,06402	-408,28

(the above notation represents

$$(P,D,Q)(P_s,D_s,Q_s)SB$$

SB = seasonal base

D = degree of nonseasonal differencing

D_s = degree of seasonal differencing

AIC = Akaike's criterion

P, P_s = nonseasonal and seasonal AR orders

Q, Q_s = nonseasonal and seasonal MA orders)

Table 9.3 Overfitting in AR models

<u>nonzero lags</u>	<u>$\hat{\sigma}^2$</u>
1,2,3,19,25,26	0,064949
1,2,3,4,5,19	0,065148
1,3,5,19,26	0,065782

Another way of removing the 'power' at lag 19 is to model a seasonal parameter with period 19. It is unlikely that any periodicity or seasonality would entail more than one AR or MA parameter so I limited P_S and Q_S to maximum values of one and tried several models. The 'best' models are listed in table 9.4.

Table 9.4 SARIMA models

<u>model</u>	<u>$\hat{\sigma}_\varepsilon$</u>	<u>AIC</u>
(0,0,3)(1,0,1)19 *	0,06058	-410,54
(3,0,0)(1,0,1)19 *	0,05988	-412,31
(0,0,4)(1,0,1)19	0,06040	-409,02
(4,0,0)(1,0,1)19	0,06003	-409,94
(0,0,1)(1,0,1)19 *	0,06337	-407,82
(0,0,0)(1,0,1)19 *	0,06402	-408,28

The above notation represents

$$(P,D,Q)(P_S,D_S,Q_S)^{SB}$$

SB = seasonal base

D = degree of nonseasonal differencing

D_s = degree of seasonal differencing

AIC = Akaike criterion

P, P_s = nonseasonal and seasonal AR orders

Q, Q_s = nonseasonal and seasonal MA orders)

The residual autocorrelograms of the four models marked * are shown in figures 9.25 - 9.28 resp. Since for all 4 models, all the residual acf's fall inside the 1% significance limits, all the models are 'acceptable'.

The best model, i.e. the model with minimum AIC, would appear to be (3,0,0)(1,0,1)₁₉. The one-step-ahead forecasts of the de-sined data, based on this model, are given in table 9.5. As a result of the large amount of unexplained variability in the model (the de-sining procedure accounts for about 44% and the B-J model for an additional 11% i.e., only about 55% of the variability in the data can be modelled - the rest is due to white noise) these forecasts are not very accurate.

An estimate of the spectrum (using the Parzen (t) window - see section 4.3.1, equation (4.25) - and truncation point $M = 60 = 0.4 \times N$) of the final residuals (i.e. $\log_{10}$ of data minus sinusoids minus Box-Jenkins model) is given in figure 9.29.

This plot seems to be dominated by several peaks and troughs, the 'major' peak occurring at lag 31 (which corresponds, assuming seasonality, to a period of 4 or 5 i.e. monthly periodicity). However even this 'peak' does not seem to have any overall dominance in the spectrum. It does not seem worthwhile trying to model the data any further.

Thus it would appear that the best model is

$$Z_t = \log_{10} x_t$$

with

$$Z_t = \text{constant} + 3 \text{ sinusoids} +$$

Fig. 9.25 Residual autocorrelogram of (0,0,3)(1,0,1)19 BJ model with
99% confidence limits

COL.1- RESIDUAL ACR'S
COL.2- UPPER 99% CONFIDENCE LIMITS

	-1	0	+1		
1		0		0.0023	0.3344
2		OXO		0.0133	0.0450
3		OXO		0.0169	0.0603
4	0	X	0	-0.0468	0.2044
5	0	.	X	0.0922	0.2054
6	0	.	X	0.0898	0.2062
7	0	X	.	-0.0487	0.2095
8	0	X	.	-0.0110	0.2098
9	0	X	.	0.0288	0.2100
10	0	.	X	0.0790	0.2102
11	0	X	.	0.0349	0.2102
12	0	X	.	-0.0379	0.2102
13	0	X	.	-0.0581	0.2102
14	0	X	.	-0.0095	0.2102
15	0	.	X	0.1171	0.2102
16	0	X	.	-0.0206	0.2102
17	0	X	.	-0.0700	0.2102
18	0	.	X	0.0597	0.2102
19	0	.	X	0.0710	0.1502
20	0	.	X	0.1057	0.1818
21	0	.	X	0.0961	0.1959
22	0	X	.	-0.0041	0.2024
23	0	.	X	0.1034	0.2054
24	0	X	.	-0.0484	0.2066
25	0	.	X	0.1169	0.2070
26	0	X	.	-0.1060	0.2069
27	0	.	X	0.0130	0.2066
28	0	.	X	0.0486	0.2062
29	0	X	.	0.0279	0.2058
30	0	X	.	-0.0012	0.2053
31	0	X	.	-0.0987	0.2048
32	0	.	X	0.0029	0.2043
33	0	.	X	0.0325	0.2038
34	0	.	X	-0.0315	0.2032
35	0	.	X	0.0578	0.2027
36	0	X	.	-0.0754	0.2020
37	0	X	.	-0.0205	0.2014
38	0	.	X	0.0993	0.2007
39	0	X	.	-0.0386	0.1999
40	0	X	.	0.0134	0.1991

Fig. 9.26 Residual autocorrelogram of (3,0,0)(1,0,1)19 BJ model
with 99% confidence limits

COL.1- RESIDUAL ACR'S
COL.2- UPPER 99% CONFIDENCE LIMITS

	-1	0	+1		
1		0		0.0223	0.1338
2		0		-0.0269	0.0381
3		0X0		0.0251	0.0424
4		0 X . 0		-0.0767	0.2053
5		0 . X 0		0.1376	0.2072
6		0 . X 0		0.0679	0.2082
7		0 X . 0		-0.0302	0.2099
8		0 X . 0		-0.0284	0.2102
9		0 X . 0		-0.0057	0.2102
10		0 . X 0		0.1047	0.2102
11		0 . X 0		0.0513	0.2102
12		0 X . 0		-0.0582	0.2102
13		0 X . 0		-0.0560	0.2102
14		0 X . 0		-0.0139	0.2102
15		0 . X 0		0.0946	0.2102
16		0 X . 0		-0.0102	0.2102
17		0 X . 0		-0.0584	0.2102
18		0 . X 0		0.0716	0.2102
19		0 . X 0		0.0496	0.1475
20		0 . X 0		0.1173	0.1819
21		0 . X 0		0.1038	0.1965
22		0 X . 0		0.0191	0.2032
23		0 . X 0		0.0764	0.2061
24		0 X . 0		-0.0453	0.2073
25		0 . X 0		0.1071	0.2076
26		0 X . 0		-0.1137	0.2076
27		0 . X 0		0.0155	0.2073
28		0 . X 0		0.0810	0.2069
29		0 X . 0		0.0007	0.2065
30		0 X . 0		-0.0065	0.2065
31		0 X . 0		-0.0963	0.2054
32		0 X . 0		-0.0034	0.2048
33		0 X . 0		0.0358	0.2041
34		0 X . 0		-0.0109	0.2034
35		0 . X 0		0.0561	0.2025
36		0 X . 0		-0.0942	0.2016
37		0 X . 0		-0.0260	0.2006
38		0 . X 0		0.1009	0.1995
39		0 X . 0		-0.0343	0.1982
40		0 X . 0		-0.0083	0.1968

Fig. 9.27 Residual autocorrelogram of (0,0,1)(1,0,1)19 BJ model
with 99% confidence limits

COL.1- RESIDUAL ACR'S		COL.2- UPPER 99% CONFIDENCE LIMITS	
-1	0	-1	+1
1	0	-0.0052	0.0258
2	0 x . 0	-0.0676	0.2087
3	0 . x0	0.1748	0.2102
4	0 x . 0	-0.0782	0.2102
5	0 . x 0	0.0956	0.2102
6	0 . x 0	0.0966	0.2102
7	0 x . 0	-0.0567	0.2102
8	0 x . 0	-0.0111	0.2102
9	0 . x 0	0.0421	0.2102
10	0 . x 0	0.0689	0.2102
11	0 x . 0	0.0297	0.2102
12	0 x . 0	-0.1279	0.2102
13	0 x . 0	-0.0601	0.2102
14	0 x . 0	-0.0069	0.2102
15	0 . x 0	0.1300	0.2102
16	0 x . 0	-0.0189	0.2102
17	0 x . 0	-0.0758	0.2102
18	0 . x 0	0.0810	0.2102
19	0 . x 0	0.0792	0.1556
20	0 . x 0	0.1087	0.1821
21	0 . x 0	0.0730	0.1951
22	0 x . 0	0.0341	0.2016
23	0 . x 0	0.0711	0.2049
24	0 x . 0	-0.0198	0.2065
25	0 . x 0	0.1158	0.2071
26	0 x . 0	-0.0883	0.2072
27	0 x . 0	-0.0034	0.2070
28	0 x . 0	0.0575	0.2066
29	0 x . 0	0.0175	0.2061
30	0 x . 0	0.0041	0.2056
31	0 x . 0	-0.1038	0.2051
32	0 x . 0	0.0183	0.2045
33	0 x . 0	0.0251	0.2039
34	0 x . 0	-0.0461	0.2032
35	0 . x 0	0.0746	0.2025
36	0 x . 0	-0.0924	0.2017
37	0 x . 0	-0.0173	0.2008
38	0 . x 0	0.1116	0.1999
39	0 x . 0	-0.0640	0.1988
40	0 x . 0	0.0114	0.1977

Fig. 9.28 Residual autocorrelogram of (0,0,0)(1,0,1)19 BJ model
with 99% confidence limits

COL.1- RESIDUAL ACR'S
COL.2- UPPER 99% CONFIDENCE LIMITS

	-1	0	+1		
	----- ----- ----- ----- ----- -----				
1		0 . X 0		0.0895	0.2102
2		0 X . 0		-0.0505	0.2102
3		0 . X 0		0.1555	0.2102
4		0 X . 0		-0.0548	0.2102
5		0 . X 0		0.0955	0.2102
6		0 . X 0		0.0941	0.2102
7		0 X . 0		-0.0350	0.2102
8		0 X . 0		-0.0125	0.2102
9		0 X . 0		0.0376	0.2102
10		0 . X 0		0.0855	0.2102
11		0 X . 0		0.0313	0.2102
12		0 X . 0		-0.0265	0.2102
13		0 X . 0		-0.0651	0.2102
14		0 X . 0		0.0003	0.2102
15		0 . X 0		0.1322	0.2102
16		0 X . 0		-0.0217	0.2102
17		0 X . 0		-0.0609	0.2102
18		0 . X 0		0.0859	0.2102
19		0 . X 0		0.0730	0.1397
20		0 . X 0		0.1331	0.1820
21		0 . X 0		0.0910	0.1974
22		0 . X 0		0.0471	0.2032
23		0 . X 0		0.0763	0.2052
24		0 X . 0		0.0016	0.2057
25		0 . X 0		0.1055	0.2057
26		0 X . 0		-0.0819	0.2053
27		0 X . 0		-0.0078	0.2050
28		0 . X 0		0.0638	0.2047
29		0 X . 0		0.0210	0.2045
30		0 X . 0		-0.0076	0.2043
31		0 X . 0		-0.1011	0.2041
32		0 X . 0		0.0170	0.2040
33		0 . X 0		0.2010	0.2039
34		0 X . 0		-0.0406	0.2038
35		0 . X 0		0.0663	0.2037
36		0 X . 0		-0.0898	0.2036
37		0 X . 0		-0.0173	0.2036
38		0 . X 0		0.1053	0.2035
39		0 X . 0		-0.0552	0.2034
40		0 X . 0		0.0112	0.2034

Table 9.5 One-step-ahead forecasts

ORIG = POINT FROM WHICH FORECAST IS MADE
 OBS = OBSERVATION WHICH IS FORECAST
 ACT = RESIDUALS OF THE TRANSFORMED SERIES MINUS 3 SINUSOIDAL COMPONENTS

ORIG	OBS	ACT	FCS	U LIM	L LIM	ERROR

LEAD TIME = 1						
150	151	0.3520	0.4750	0.9547	-0.0046	-0.1230397
151	152	0.3750	-0.1710	0.3086	-0.6506	0.5460034
152	153	0.3650	0.3186	0.7983	-0.1610	0.0463570
153	154	0.6480	0.1033	0.5830	-0.3763	0.5446674
154	155	-0.4480	0.1366	0.6162	-0.3430	-0.5846208
155	156	0.4260	0.1131	0.5928	-0.3665	0.3128607
156	157	0.3970	0.1585	0.6382	-0.3211	0.2384598
157	158	0.4740	0.1089	0.5885	-0.3707	0.3650810
158	159	0.1380	0.4305	0.9101	-0.0492	-0.2924682
159	160	-0.3180	-0.3863	0.0933	-0.8659	0.0683070
160	161	-0.2160	-0.4498	0.0298	-0.9294	0.2338156
161	162	-0.1610	0.1870	0.6666	-0.2927	-0.5299582
162	163	-0.3030	0.0635	0.5431	-0.4161	-0.3665286
163	164	-0.1570	0.2835	0.7631	-0.1961	-0.4404879



Fig. 9.29 Spectrum estimate of the residuals (Parzen (b) window, truncation

point 60)

$$= \mu + \sum_{j=1}^3 \left\{ a_j \cos(tw_j) + b_j \sin(tw_j) \right\} + \varepsilon_t$$

(see equations (9.1) for μ , a , b , w) where ε_t is such that

$$(1 - \phi_{19} B^{19})(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3) \varepsilon_t = (1 - \theta_{19} B^{19}) a_t$$

and a_t are white noise variables with $V(a_t) = 0,05988$

and

$$\phi_{19} = -0,71$$

$$\theta_{19} = -1,0553$$

$$\phi_1 = 0,1144$$

$$\phi_2 = -0,102$$

$$\phi_3 = 0,1609$$

(with associated standard errors

$$0,1253$$

$$0,0786$$

$$0,0952$$

$$0,0340$$

$$0,0849)$$

[All these numerical values were obtained from the program BJES described in appendix A18, which is a program for obtaining parameter estimates in Box-Jenkins models].

9.4.3 SIMULATION

In figures 9.30 - 9.33 are the spectral estimates (using the Parzen (b) window and truncation point $M = 67$) for the stretch of data $x_{31}, \dots, x_{104}$ in the four cases

(i) original data

(ii) sinusoid simulation

Fig. 9.30 Spectrum estimate of the residuals (Parzen (b) window,
truncation point 67)

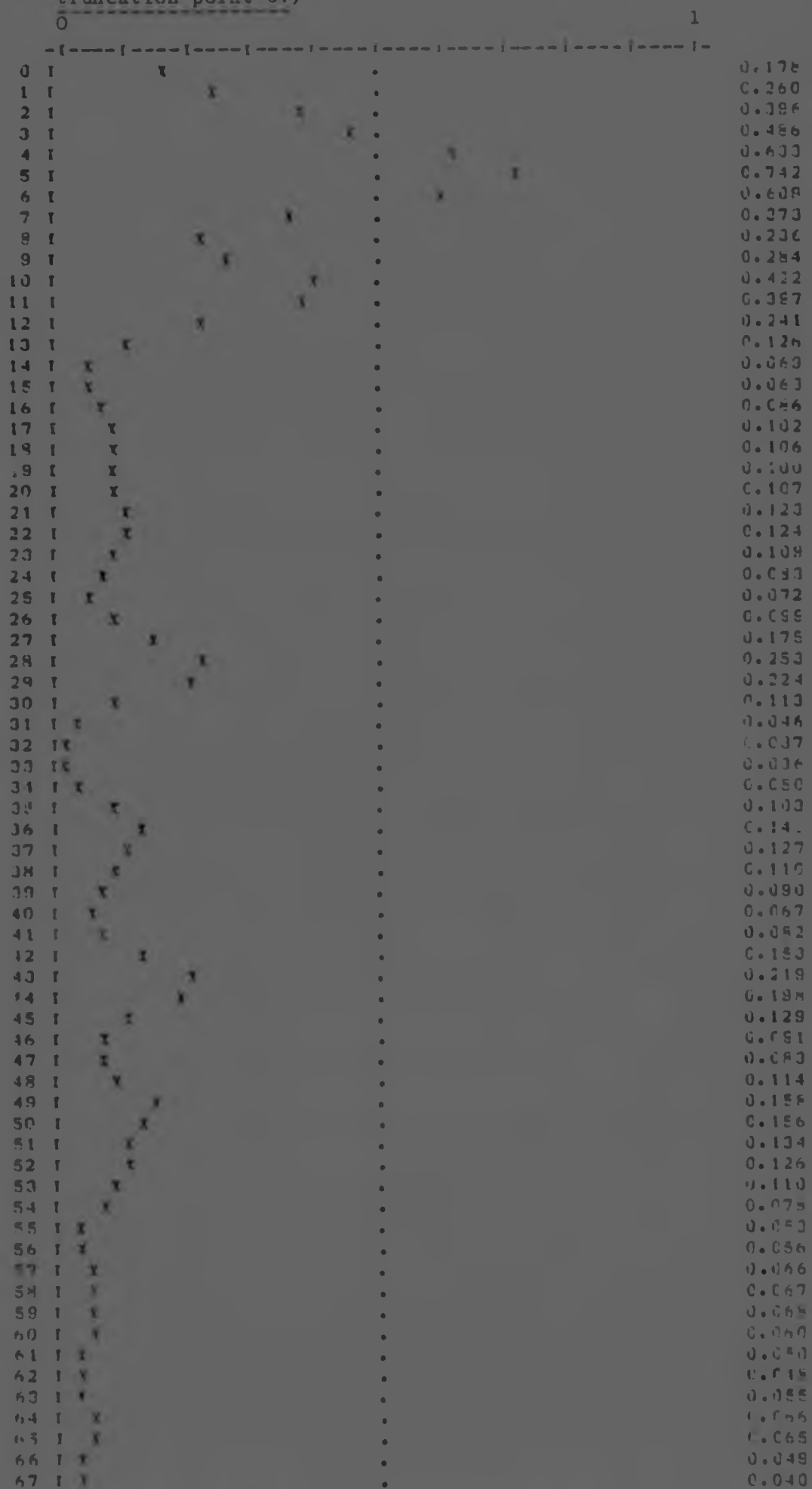
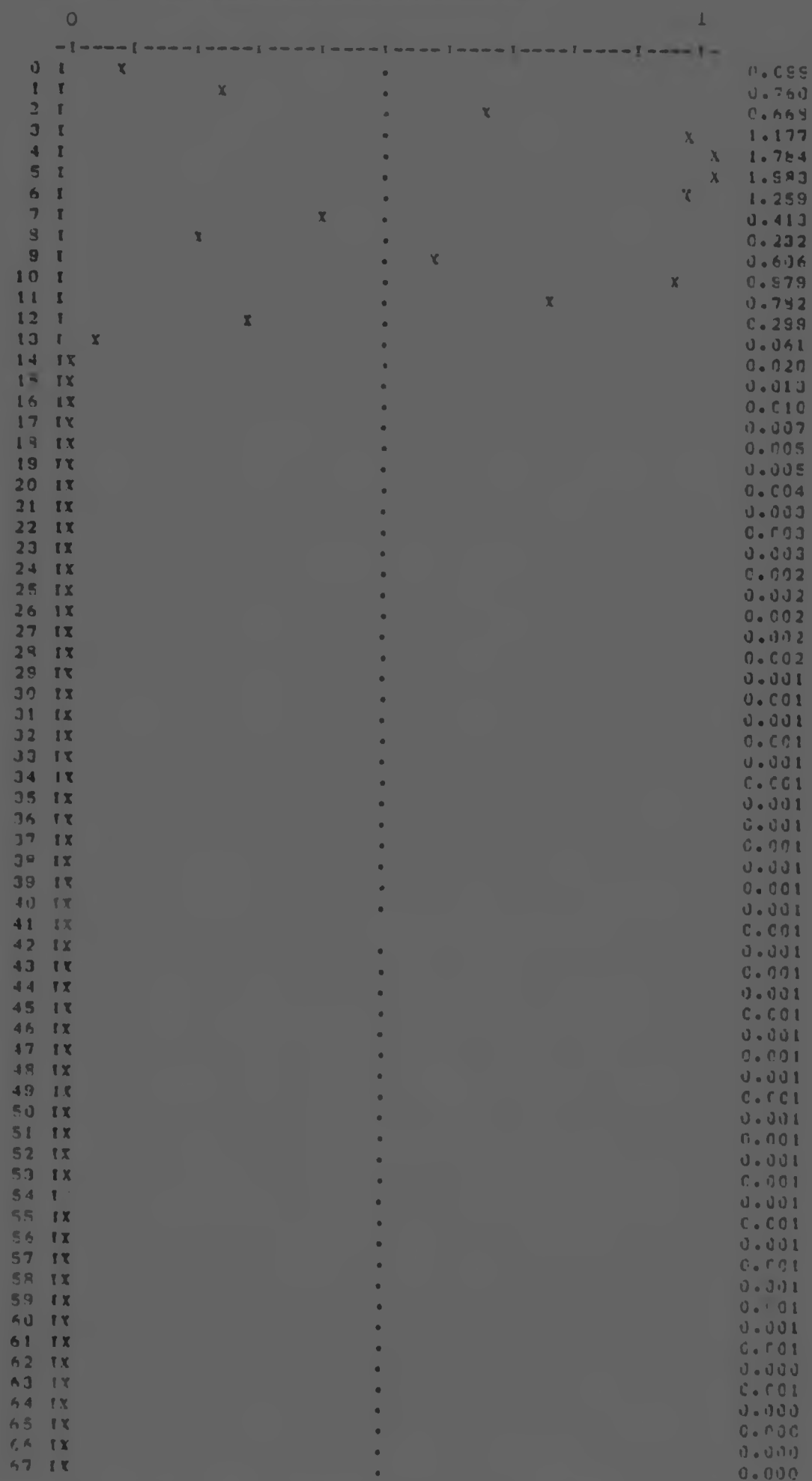


Fig. 9.31 Spectrum estimate of the transformed data minus 3 sinusoids
(Parzen (b) window, truncation point 67)



minus BJ one-step-ahead forecasts (Parzen (b) window, truncation point 67).

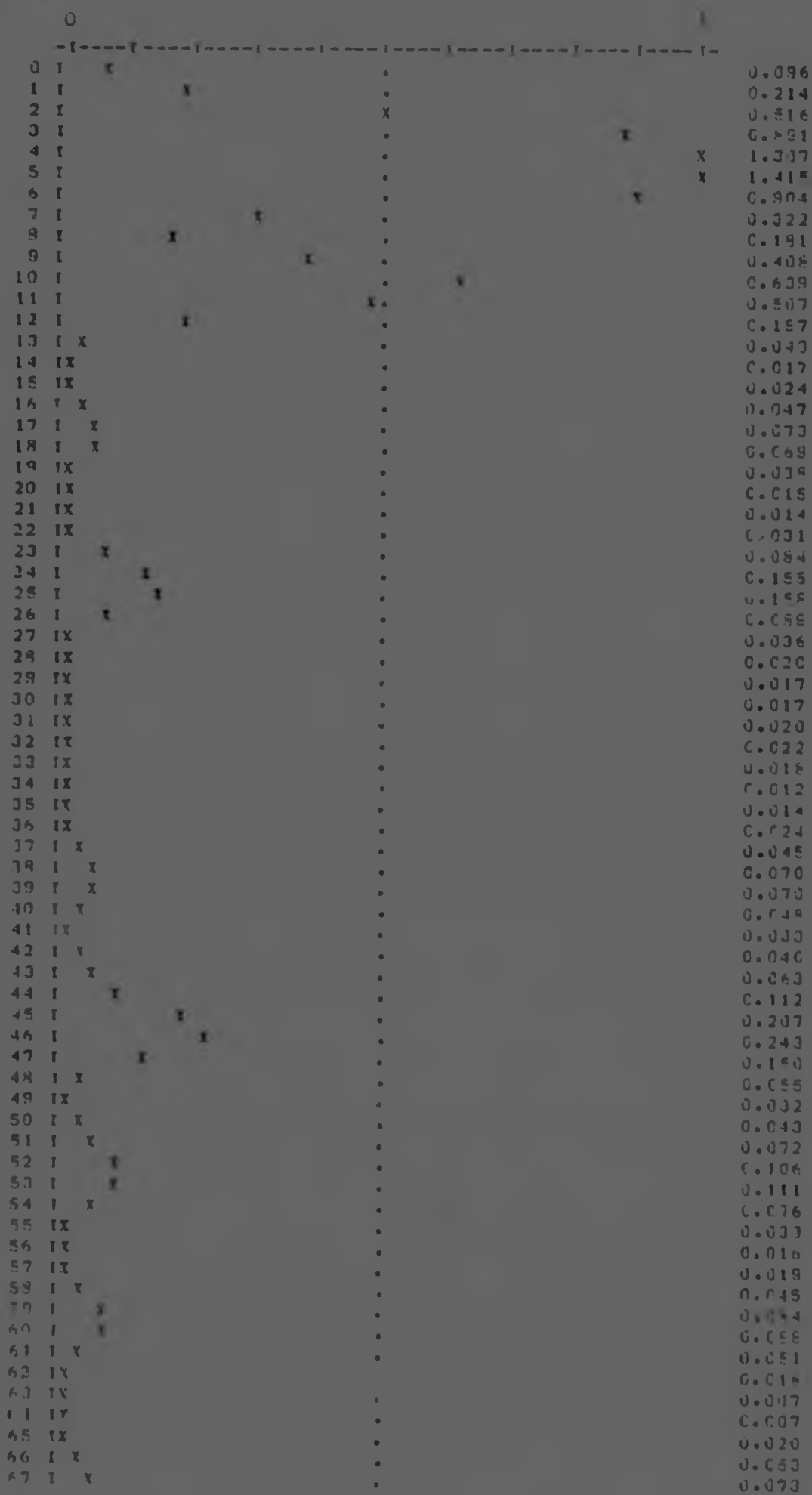
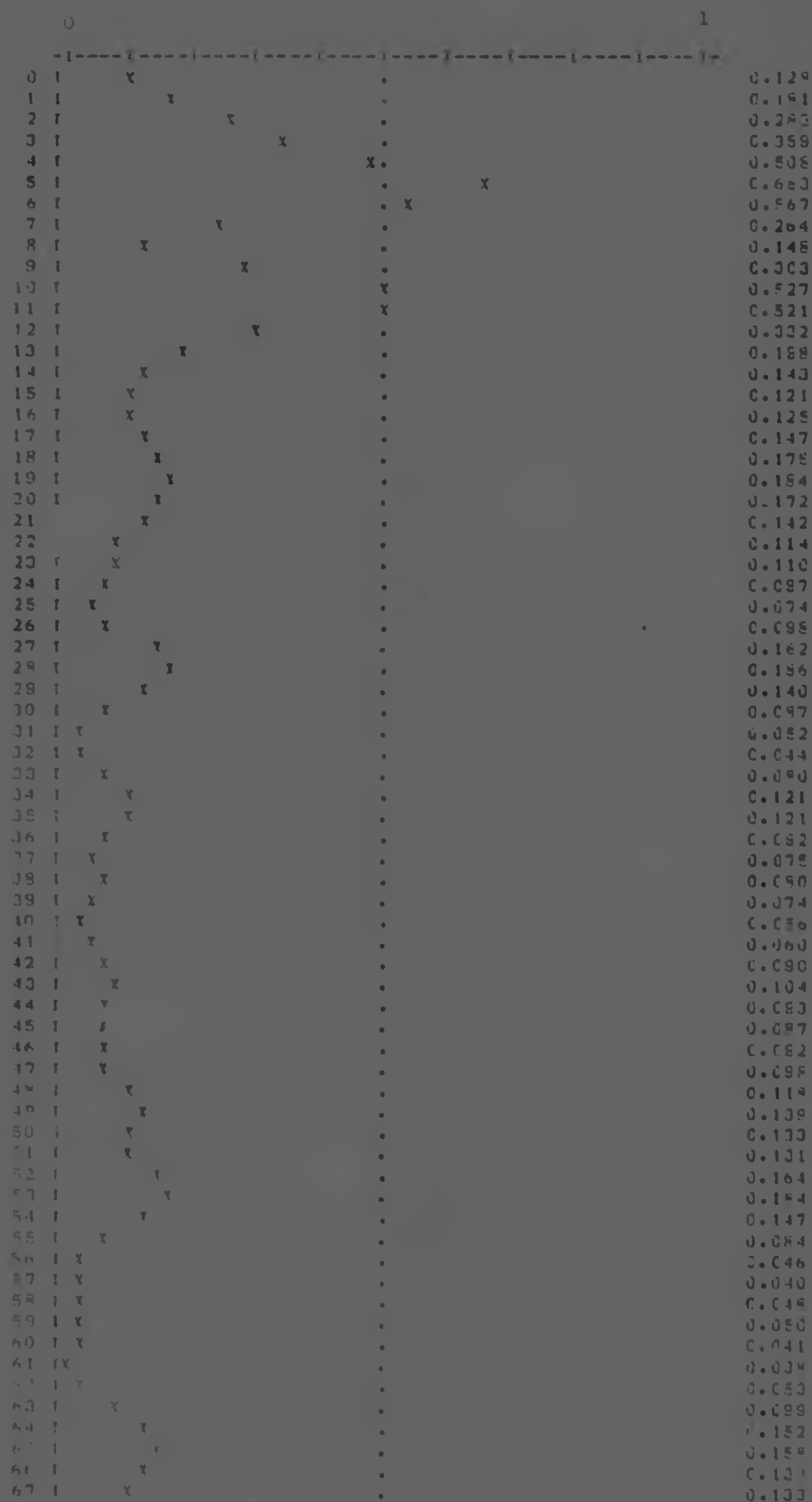


Fig. 4.33 Spectrum estimate of the transformed data minus 3 sinusoids
minus the simulated BJ series (Parzen (b) window, truncation point 67)



- (iii) sinusoid + 1-step-ahead forecasts
- (iv) sinusoid + simulated $N(0, \sigma_a^2)$ variables.

Case (ii) above, is just a simulated data set involving μ and the three sinusoidal components modelled in the data. Cases (iii) and (iv) take this simulated sinusoid and add the stochastic element, modelled by the Box-Jenkins process $(3,0,0)(1,0,1)I_9$, obtained from either the 1-step-ahead forecasts or just a simulation of the process.

Except for the sinusoid by itself (i.e. (ii)) the remaining spectra have a relatively strong resemblance to one another. However due to the large amount of unexplained variability in the data, this resemblance cannot be strengthened. It is interesting to note that the simulated stochastic element has the closest resemblance to the spectrum of the original data, (more so than the forecasted element) and it may be wiser to use a simulated stochastic process (with $N(0, \sigma_a^2)$ random errors) as a forecasting model.

9.4.4 AR MODELLING AND AR SUBSET MODELS

Due to the low degree of variability explained by the model fitted up to this point, I have tried using McClave's chi-squared technique (see appendix A15) for fitting AR models to the $\log_{10}$ of the data. The results, for a maximum model size of AR(30), are given in table 9.6 (by using the programs AUTO and AUTOAR described in appendix A18). For each AR(k) model, $k = 1, \dots, 30$, the values of the CAT and AIC statistics are given (see section 7.4.2 pg 144) (RV = residual variance, and is obtained using backforecasting).

Table 9.6 AR modelling (using backforecasting)

ORDER OF AR(P)	RESIDUAL VARIANCE	AIC	CAT
1	0.0992563	-342.5073	-10.0077562
2	0.0985795	-341.5337	-10.0092993
3	0.0943434	-346.1221	-10.3941202
4	0.0937431	-345.0793	-10.3908777
5	0.0942543	-342.2637	-10.2622871
6	0.0942513	-340.2683	-10.1918898
7	0.0931962	-339.9570	-10.2404757
8	0.0934135	-337.6077	-10.1441460
9	0.0940214	-334.6348	-10.0040312
10	0.0945394	-331.8105	-9.87523460
11	0.0949264	-329.1980	-9.76188564
12	0.0945764	-327.7520	-9.73037434
13	0.0947852	-325.4211	-9.63674831
14	0.0953175	-322.5813	-9.50789165
15	0.0960436	-319.4429	-9.35915756
16	0.0945473	-319.7981	-9.45342731
17	0.0911782	-323.2407	-9.77112675
18	0.0917165	-320.3579	-9.63407803
19	0.0912215	-319.1694	-9.62015533
20	0.0917032	-316.3794	-9.48987198
21	0.0923973	-313.2483	-9.33580208
22	0.0931508	-310.0303	-9.17669201
23	0.0932844	-307.8152	-9.08984756
24	0.0939752	-304.7036	-8.93976974
25	0.0914246	-306.8357	-9.16405773
26	0.0916952	-304.3926	-9.05907440
27	0.0886908	-307.3896	-9.35334110
28	0.0894342	-304.1377	-9.18507195
29	0.0900660	-301.0818	-9.03262997
30	0.0896607	-299.7249	-9.00599289

From these results we obtain the following

CAT best model : AR(3)

AIC best model : AR(3)

(using the CAT criterion, one may choose a 'best' model, or 'better' model, as occurring at any relative minimum - the above model, i.e. AR(3), is for the overall minimum CAT value).

Thus it would seem that an AR(3) process is the best AR model to fit to the $\log_{10}$ of the data. However this doesn't explain very much of the variability in the data. (It is also pointless trying to fit subset AR models since p is so small).

Similar results are obtained by applying this procedure to the $\log_{10}$ data minus sinusoid (1) (corresponding to a period of 26) - i.e. an AR(3) is the best fitting AR model (based on Akaike's criteria). In this case, the residual variance (of the AR(3) model) is 0.0363.

9.5 COMMENT

9.5.1 MIXED SPECTRA

Suppose we have a process x_t which we intend modelling as

$$\begin{aligned} x_t &= \text{deterministic cycles} + \text{stochastic process} \\ &= \text{harmonic component} + \text{general linear process} \\ &= y_t + z_t \\ &= \sum_{i=1}^m A_i \cos(w_i t + \phi_i) + \sum_{u=1}^L g_u \epsilon_{t-u} \end{aligned}$$

i.e. the best model when harmonic analysis is not used

where $E(\epsilon_t) = 0$ (independent $\forall t$)

$$E(\epsilon_t^2) = 1$$

$$E(\epsilon_t^4) < \infty$$

and $E(\epsilon_t^2) < \infty$

$$E(\epsilon_t^4) < \infty$$

Suppose $\{y_t\}$ has sdf $h_y(w)$ and $\{z_t\}$, $h_z(w)$. If $\{y_t\}$ and $\{z_t\}$ are uncorrelated processes, then

$$h_x(w) = h_y(w) + h_z(w)$$

and $\{x_t\}$ is said to have a mixed spectrum if $h_z(w)$ is not a constant.

All the sampling properties of the sdf and periodogram were derived for the special case where $\gamma_z(r) = 0 \forall r \neq 0$. For instance, in chapter three, when searching for harmonic components, z_t was assumed to be white noise (and hence has a constant spectrum). The null hypothesis of no harmonic components present was put forward and then m in

$$y_t = \sum_{i=1}^m \lambda_i \cos(\omega_i t + \phi_i)$$

was determined by testing the periodogram ordinates (in order of magnitude) until we failed to get a significant result.

Now it may happen that $\{z_t\}$ does not have a constant spectrum and might only have power at one frequency, ω_0 , say. If $\{y_t\}$, or more correctly, $h_y(w)$, does not have power at ω_0 , a periodogram test (or spectrum test) may incorrectly indicate an harmonic component at ω_0 . So if $\{z_t\}$ is not a white noise process, then any test for periodogram ordinates (see appendix A5 for example) is not strictly applicable - thus the procedure, and conclusions, of the analysis of the data used in this chapter, may be incorrect.

What is needed is a test (under the assumption that we have a mixed spectrum) which will separate, or determine the existence of, the discrete spectrum.

If a discrete spectral component is detected, its parameters must be estimated (i.e. the amplitude and frequency). The contribution of this harmonic term must then be removed from the data and a test applied for another harmonic component in the residuals. This must be continued until no further harmonic components are indicated.

Several tests for the existence of the discrete spectrum are given in Priestley, 1981, pg. 617+. The one I have used is termed the $P(\cdot)$ test and is described fully in Bhansali (1979) and Priestley (1962a,b). It is summarized here without proof.

The $P(\cdot)$ test

Suppose we model

$$x_t = y_t + z_t$$

$$y_t = \sum_{i=1}^m A_i \cos(\omega_i t + \phi_i)$$

$$\phi_i \sim U(-\pi, \pi) \text{ (i.e. uniform distribution on } (-\pi, \pi))$$

z_t = general linear process

H_0 : all the A_i are zero.

If H_0 is true, the acrf of $|x_t|$ should eventually damp out to zero. However, if one (or more) of the A_i are nonzero, then the acrf will not damp out and in fact its tail will behave like a linear combination of cosine terms of the same (or similar) frequency as those of the harmonic components.

Let $h_1(w)$ and $h_2(w)$ be two spectral estimates of $h_x(t)$ where i and j denote the truncation points of the spectral windows and $j > 2i$. The point i is such that y_t has acvf with

$$\gamma_y(t) \approx 0 \quad |t| > i$$

If we are searching for harmonic components, we may do this by performing an harmonic analysis of the tail of the acvf so examine

$$P(\lambda) = \hat{h}_1(\lambda) - \hat{h}_2(\lambda)$$

$$\lambda = \lambda_p$$

$$= 2\pi p/N$$

$$p = 0, 1, 2, \dots, \left[\frac{N}{2}\right]$$

$$N > j > 2i$$

If some of the λ_p 's are nonzero then a plot of $P(\lambda)$ will exhibit several well defined peaks. Testing each peak in order of frequency, and supposing the first peak is at λ_0 , form

$$J_q = \left(\frac{N}{4\lambda(1,j)} \right) \sum_{s=0}^q P^* \left(\frac{2\pi s}{1} + \delta \right) g^* \quad q = 0, \dots, \frac{1}{2}$$

where $P^*(\cdot) = P(\cdot)/c_0$

c_0 = sample variance

$$g^* = \left\{ \frac{1}{8\pi} \left\{ 2 \sum_{|u| \leq i} r_u^2 - \sum_{|u| \leq 2i} r_u^2 \right\} \right\}$$

r_u = acf at lag u

δ is such that $0 \leq \delta < \frac{1}{N}$

and $\frac{2\pi s}{1} + \delta = \lambda_0$ some integer s

$\lambda(1,j)$ depends on the form of the spectral windows $h_1(w)$ and $h_2(w)$

For the special case where h_i and h_j are both Bartlett spectrum estimates given by equation (4.18)

$$\Lambda(i,j) = \frac{2}{3} j - \frac{4}{3} i + \frac{2i^2}{3j} \quad (\text{Priestley, 1962a})$$

If $\max_q (J_q) > \alpha_0$ where α_0 is the upper 100% point of a $N(0,1)$ distribution, then the peak at k_0 is significant at the 1% level of significance.

[For more details on the theory behind the test, see Priestley (1962a,b)].

Application of the $P(\lambda)$ test

Priestley's $P(\lambda)$ test was applied to the logarithms of the data over the section of interest (i.e. $x_1, \dots, x_{150}$).

Several truncation points were used, namely $i = 7, 17, 22$ and 58 and these represented the lags at which I felt the acf's were ≈ 0 . For each i (the lower truncation point), several j 's (the upper truncation points) were used and these were such that $j > 2i$ and represented the end of a 'cycle' in the acf.

e.g. $i = 22 \quad j = 48, 53, 59$
 $i = 58 \quad j = 120$

The plots of $P^*(\lambda)$ for $i = 22$ and $j = 53, 59$ are given in figures 9.34 and 9.35 resp. In both cases, the peak at $k = 6$ was tested and I obtained

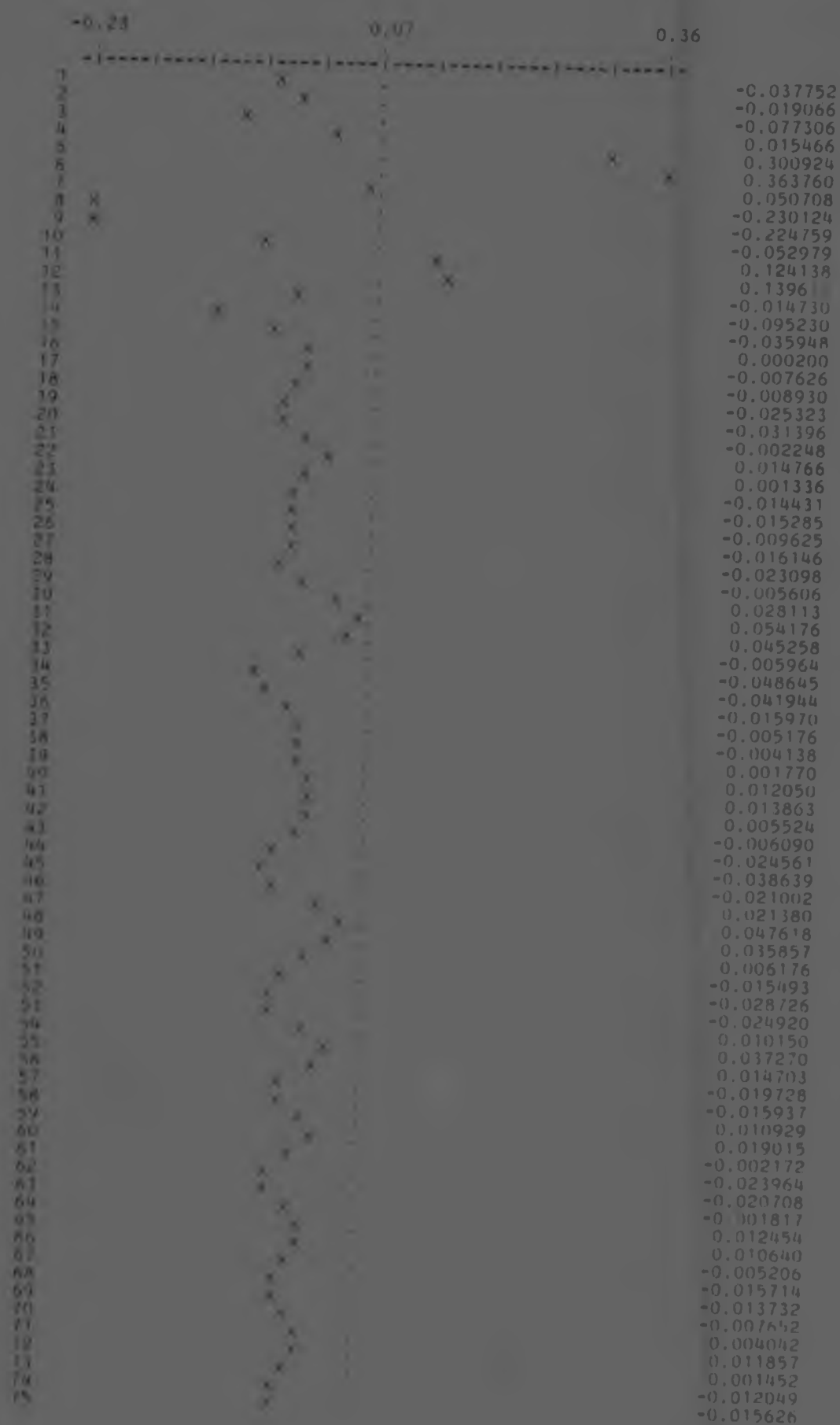
$$J_q(22, 53) = 2.94$$

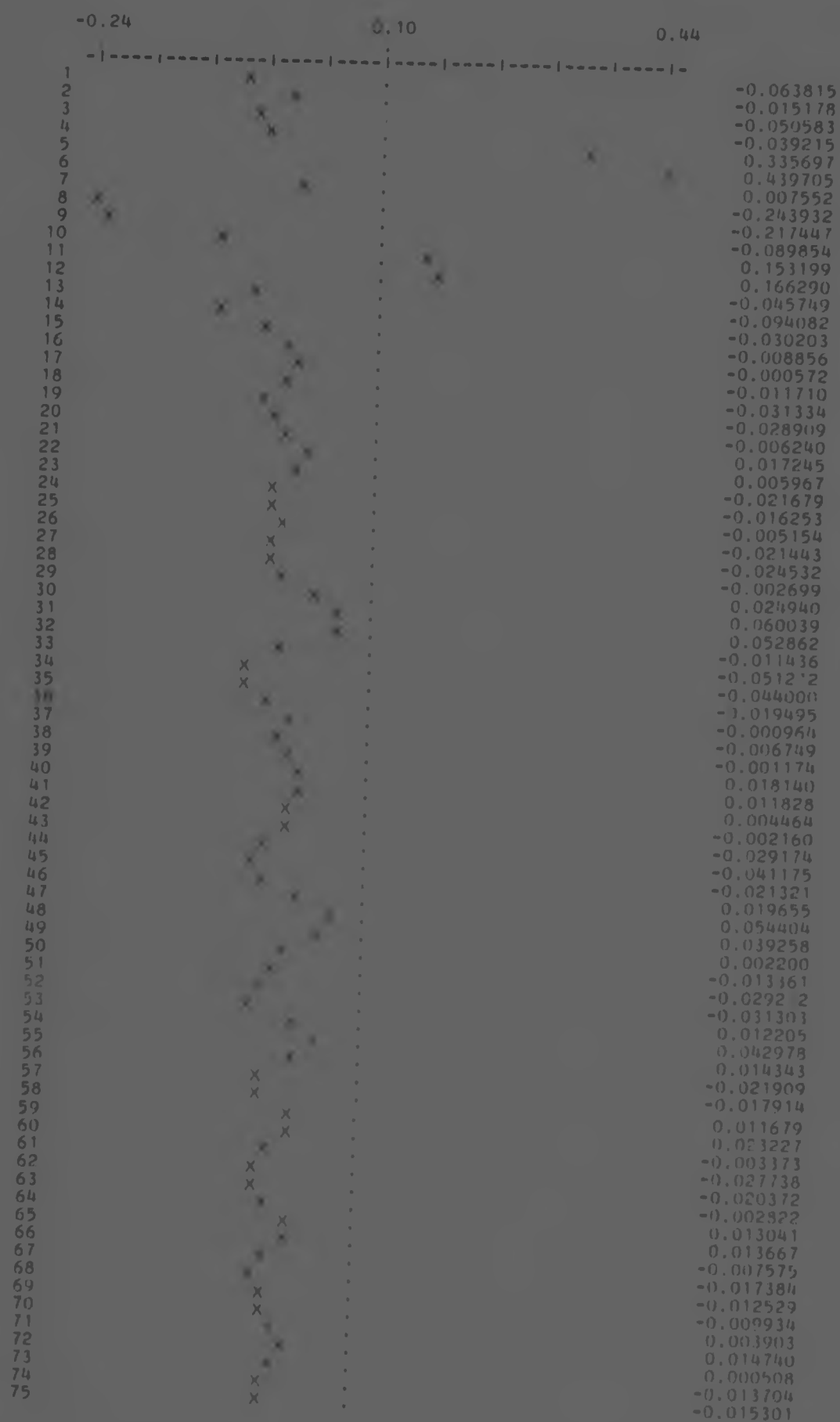
$$J_q(22, 59) = 3.01$$

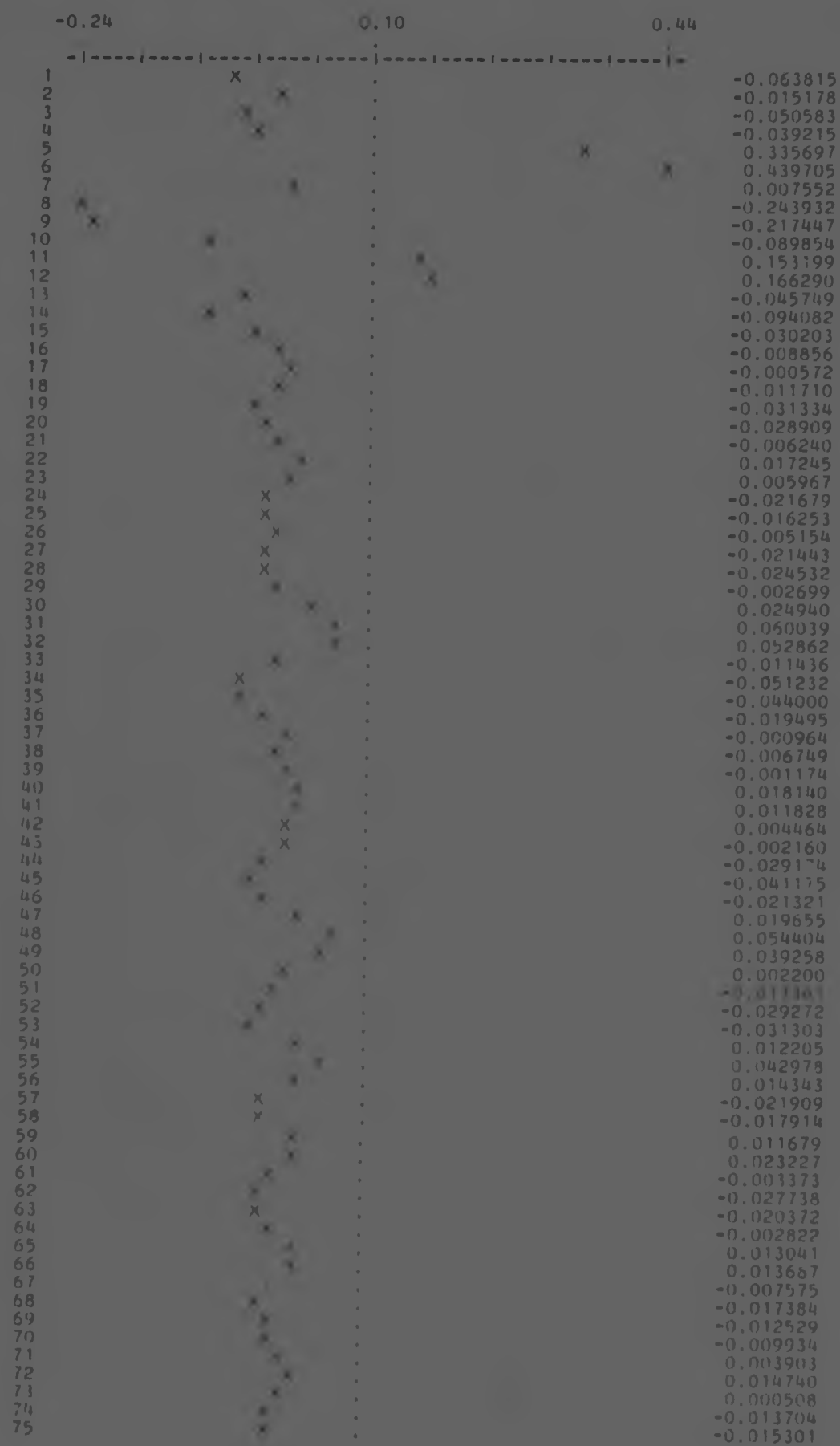
i.e. significant at the 1% level. Similarly, with $i = 58$ and $j = 120$, a value of

$$J_q(58, 120) = 2.09$$

was obtained for the peak at $k = 6$ - i.e. significant at the 5% level. (In this case, a peak at $k = 3$ was the first major peak



Fig. 9.35 $P(\lambda)$ with lower truncation point 22 and upper truncation point 59

Fig. 9.35 $P(\lambda)$ with lower truncation point 22 and upper truncation point 59

and this had a J_q value of 2.16 - however if we ignore this component and test $k = 0$ we obtain the above J_q value).

A peak therefore, at $k = 0$, seems to be present in the data, and this corresponds to a frequency of $\omega_0 = 2\pi \cdot 0.125 = 0.25$ (and corresponds to a period of 20 or 28 - exactly the same result as obtained using periodogram analysis).

$$J_q(0, 100) = 2.73$$

and corresponds to a period of 20 or 28 - exactly the same result as obtained using periodogram analysis).

Removing this harmonic component from the data, it was now necessary to perform the whole test over again.

Consider a (lower) truncation point of 50 (-4) and upper value of 100 (4). The choice of 50 seems 'reasonable' from the appearance of the plot, of this new data set, given in Figure 9.36.

In this case, the first dominant peak appears at lag $k = 3$.

Testing for H_0 , I obtained

$$J_q(3, 100) = 2.73$$

- again highly significant.

Similarly, using $i = 70$ and $j = 145$, I obtained

$$J_q(70, 145) = 2.07$$

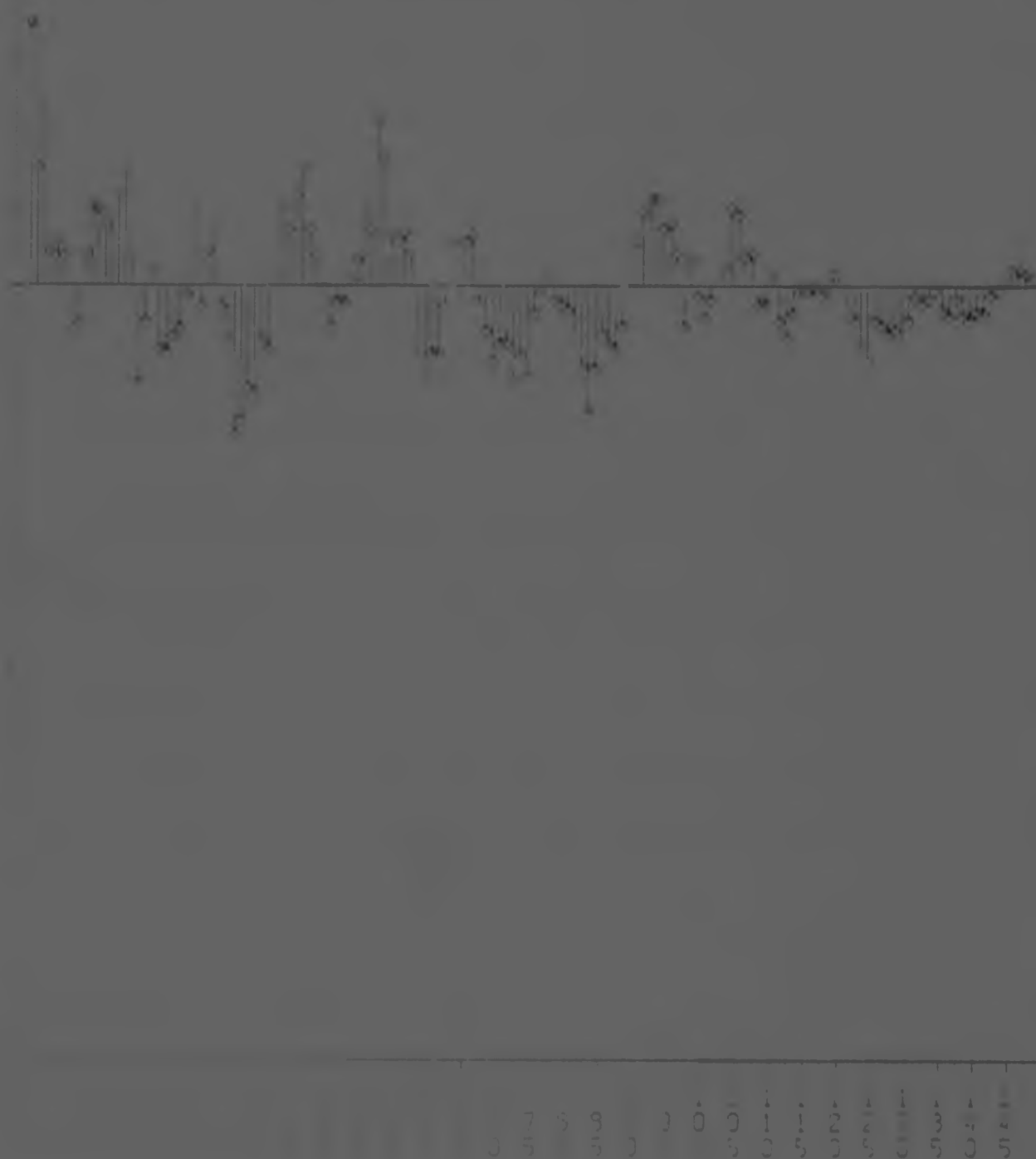
(still significant at the 5% level).

Thus it would seem that a harmonic component corresponding to a frequency of $\omega_0 = 2\pi \cdot 0.125 = 0.25$ (and corresponding to a period of approximately 20) is evident in the data. (This is also suggested by the failure of the test to drop out - it seems to be oscillating fairly regularly).

Removing this component, and performing the $P(k)$ test again, a peak at lag $k = 11$ is dominant and when tested, is also significant.

Fig. 9.16 Autocorrelogram (up to lag 148)

of the transformed data minus the harmonic
component detected by the $P(\lambda)$ test



The peak at $k = 11$ corresponds to an harmonic component with frequency $11 \cdot 2^{-150} = 0,461$ and approximate period of 13. No more peaks were found to be significant.

Thus the $P(\lambda)$ test seems to support the conclusions (and strictly speaking, the incorrect procedure) of the harmonic analysis

However, I still view these conclusions with suspicion since

- (a) Priestley's $P(\lambda)$ test was designed so that peaks are tested in order of frequency and this should prevent one obtaining spurious significant results (cf Priestley (1962 a,b)) - however Bhansali (1979) has shown that it is possible to get spurious results

and

- (b) due to the large variability in the data, it is not really possible to assess whether the rises and falls in the data are equal - Galbraith (1977) notes that if the data are asymmetrical as far as the rises and falls go (for example the data may always require about 10 values to reach a peak from a trough, but may only have about 5 values from a peak to a trough and if this asymmetry is not incorporated into the model, then this asymmetry will show up as minor peaks (in the sdf) at frequencies which are multiples of the fundamental frequency.

I am fairly confident that an harmonic term corresponding to a period of 26 is evident in the data - but if the data are 'asymmetrical' as described above, then peaks in the sdf will occur at multiples of approximately $w = 0,25$ - this may account for peaks at $w = 0,46$ and $w = 0,13$.

9.5.2 LOGARITHMIC TRANSFORMATIONS

It is unreasonable to assume that our process $\{x_t\}$ can always be modelled by an additive form

$$\text{i.e. } x_t = y_t + z_t.$$

Another obvious possibility is a multiplicative model

$$\text{i.e. } x_t = y_t z_t.$$

By taking logarithms, any multiplicative model is converted into an additive form. However we may incorrectly assume the data have a multiplicative form, when they really have an additive form, and then a logarithmic transformation is not really applicable. However the logarithmic data should still have a form which is approximately additive.

Parzen (1978) states that if the data do not appear to be normally distributed, one may perform a log transformation (or any other memoryless transformation) since such transformations often do not change the overall shape of the data. When this is true, it then seems reasonable to assume that if the data originally had an additive form, by taking logarithms this should not change things dramatically (obviously one of the exceptions being data which are ≈ 0 , negative or in the interval $(0,1)$).

Thus I am tempted to suggest that generally the form of the model will not change if logarithms of the original data are taken.

9.6 CONCLUSION

The model I finally decided upon was the following:

$$\hat{x}_t = \mu + \sum_{j=1}^3 \left(a_j \cos(tw_j) + b_j \sin(tw_j) \right) + e_t$$

$$t = 0, \dots, 163$$

(see equation (9.1) for a_j, b_j, w_j)

where e_t is such that

$$(1 - \frac{1}{19}B^{19})(1 - \frac{1}{2}B - \frac{1}{2}B^2 - \frac{1}{3}B^3)e_t = (1 - \frac{1}{19}B^{19})a_t$$

(see pg. 207 for these parameter values) and

$$\text{and } Z_t = \log_{10} x_t \quad \forall t$$

$\{x_t\}$ = original data set

$\{a_t\}$ is a white noise series with $V(a_t) \approx 0.05988$

$$\text{Cov}(a_t, a_s) = 0 \quad \forall t \neq s$$

$$E(a_t) = 0 \quad \forall t$$

The model turned out to be very disappointing in terms of variance reduction. This does not detract from the power of the techniques presented - it may be that the data used consisted mainly of white noise (this will account for the small reduction in variance i.e. only 55%).

It is probable that the residuals may be modelled further, however this will only lead to an increase in the number of parameters in a model already containing a large number. By including more parameters, the variance will not be reduced substantially (see, for example, section 8.4.4) and we may also have the problem of overfitting (if this had not already been done in the final model given).

A more successful application of this de-sining/BJ modelling procedure has been performed on the Canadian lynx data i.e. the annual trapping of lynx - this data is given in Campbell and Walker (1977) and the application may be found in Bhansali (1979). For an application of subset autoregression models to the same data see Gabr and Subba Rao (1981). This data set shows how successful this modelling procedure can be and atones for the data analysed in this chapter.

If this airline data consists mainly of white noise, as I believe it does, then it is unlikely that any model, of the types described, will achieve a variance reduction (i.e. fit) of over 70%.

APPENDICES

APPENDIX A1: STOCHASTIC LIMITING OPERATIONS

1. Let $X_1, X_2, \dots$ be a sequence of r.v.^s. The sequence $\{X_i\}$ converges in mean square iff $\exists$ a random variable X such that

$$\lim_{i \rightarrow \infty} E[(X_i - X)^2] = 0$$

This is written as

$$\text{l.i.m.}_{i \rightarrow \infty} X_i = X.$$

2. $X(t)$ is stochastically continuous at $t = t_0$ if

$$\text{l.i.m.}_{t \rightarrow t_0} X(t) = X(t_0)$$

$$\text{i.e. iff } \lim_{t \rightarrow t_0} E[(X(t) - X(t_0))^2] = 0$$

3. If $\{X(t)\}$ is stationary with variance σ_X^2 and acrf $\rho(\tau)$, then

$$\begin{aligned} E[(X(t) - X(t_0))^2] &= V(X(t)) + V(X(t_0)) \\ &\quad - 2 \text{Cov}[X(t), X(t_0)] \\ &= \sigma_X^2 [1 - \rho(t - t_0)] \end{aligned}$$

so $X(t)$ is stochastically continuous at $t = t_0$ iff

$$\lim_{t \rightarrow t_0} \rho(t - t_0) = 1$$

$$\text{i.e. iff } \lim_{\tau \rightarrow 0} \rho(\tau) = 1$$

$$\text{i.e. iff } \rho(\tau) \text{ is continuous at } \tau = 0.$$

4. If $\rho(\tau)$ is continuous at $\tau = 0$, it is continuous everywhere (consider $E[X(t - \tau)X(t + h) - X(t)]$ and use the Cauchy-Schwarz inequality; then assume $\rho(\tau)$ is continuous at $\tau = 0$).

5. $X(t)$ is stochastically differentiable at $t = t_0$ with derivative $X'(t_0)$ iff

$$\begin{aligned} & \text{l.i.m.}_{h \rightarrow 0} \left\{ \frac{X(t_0 + h) - X(t_0)}{h} \right\} \text{ exists (and equals } X'(t_0)) \\ \text{i.e. iff } & \lim_{h \rightarrow 0} E \left[\left(\frac{X(t_0 + h) - X(t_0)}{h} \right)^2 \right] \text{ exists} \\ \text{But } E \left[\left(\frac{X(t_0 + h) - X(t_0)}{h} \right)^2 \right] &= \frac{2\sigma_X^2}{h^2} \{1 - \rho(h)\} \end{aligned}$$

for $X(t)$ to be stochastically differentiable, we require

$$\lim_{h \rightarrow 0} \frac{1 - \rho(h)}{h^2} = 0$$

i.e. that $\rho'(0)$ exists and is zero.

6. Consider $\int_a^b g(t)X(t)dt$ where $g(t)$ is a deterministic function.

Now consider an arbitrary (finite) set of points $(t_0, \dots, t_n)$ in (a, b) and define the Riemann integral as

$$\text{l.i.m.}_{n \rightarrow \infty} \left\{ \sum_{i=1}^n g(t_i)X(t_i)(t_i - t_{i-1}) \right\} \text{ as } \max(t_i - t_{i-1}) \rightarrow 0$$

Now this limit exists in mean square iff $\int_a^b \int_a^b g(t)g(s) \cdot (t-s)dsdt$

exists as a double Riemann integral.

In a similar manner Riemann-Stieltjes integrals involving stochastic processes may be defined.

For two deterministic functions $g(t)$ and $F(t)$, the Riemann-Stieltjes integral

$$\int_a^b g(t)dF(t)$$

is defined as the limiting value of

$$\sum_{i=1}^n g(t_i) \{F(t_i) - F(t_{i-1})\}$$

as $\max (t_i - t_{i-1}) \rightarrow 0$

So a Riemann-Stieltjes integral of the form

$$\int_a^b g(t) dX(t)$$

may now be defined as the mean square limit of

$$\sum_{i=1}^n g(t_i) \{X(t_i) - X(t_{i-1})\}$$

APPENDIX A2: THE VARIANCE AND COVARIANCE OF ACVF ESTIMATES

Suppose we have a (second-order) stationary process $\{X_t\}$ and w.l.o.g, assume μ , the mean of X_t , is 0 i.e. $E(X_t) = 0 \quad \forall t$

So c'_r as defined by (2.28), reduces to

$$c'_r = \frac{1}{N} \sum_{t=1}^{N-|r|} x_t x_{t+|r|}$$

$$\text{and } E(c'_r) = (1 - \frac{|r|}{N}) \gamma(r) \quad \forall r \quad (\text{by (2.29)})$$

If $r \geq 0$ and $r+k \geq 0$, (and both are $\leq N-1$)

$$\text{Cov}(c'_r, c'_{r+k}) = E(c'_r c'_{r+k}) - E(c'_r)E(c'_{r+k})$$

$$= \frac{1}{N^2} E \left(\sum_{t=1}^{N-r} \sum_{s=1}^{N-r-k} x_t x_{t+r} x_s x_{s+r+k} \right) - (1 - \frac{r}{N})(1 - \frac{r+k}{N}) \gamma(r) \gamma(r+k) \quad (\text{a2.1})$$

Now a standard result for quartic forms states that if $E(X_t) = 0 \quad \forall t$, then (by Isserlis (1918))

$$\begin{aligned} E(x_t x_{t+r} x_s x_{s+r+k}) &= E(x_t x_{t+r}) E(x_s x_{s+r+k}) \\ &\quad + E(x_t x_s) E(x_{t+r} x_{s+r+k}) \\ &\quad + E(x_t x_{s+r+k}) E(x_{t+r} x_s) + K_4(s-t, r, k) \end{aligned}$$

(where $K_4(s-t, r, k)$ is the coefficient of

$$\frac{(it_1)^t (it_2)^{r+t} (it_3)^s (it_4)^{s+r+k}}{t!(r+t)! s! (s+r+k)!}$$

in $\phi(t_1, t_2, t_3, t_4)$, the multivariate characteristic function, if an expansion in power series exists)

Using this result, (a2.1) reduces to

$$\begin{aligned} \text{Cov}(c'_r, c'_{r+k}) &= \frac{1}{N^2} \sum_{t=1}^{N-r} \sum_{s=1}^{N-r-k} (s-t) \gamma(s+k-t) + \gamma(s+r+k-t) \gamma(s-t-r) \\ &\quad + K_4(s-t, r, k) \end{aligned} \quad (\text{a2.2})$$

APPENDIX A2: THE VARIANCE AND COVARIANCE OF ACVF ESTIMATES

Suppose we have a (second-order) stationary process $\{X_t\}$ and w.l.o.g, assume μ , the mean of $\{X_t\}$, is 0 i.e. $E(X_t) = 0 \quad \forall t$

So c'_r as defined by (2.28), reduces to

$$c'_r = \frac{1}{N} \sum_{t=1}^{N-r} x_t x_{t+r}$$

$$\text{and } E(c'_r) = (1 - \frac{r}{N}) \gamma(r) \quad \forall r \quad (\text{by (2.29)})$$

If $r > 0$ and $r+k > 0$, (and both are $\leq N-1$)

$$\begin{aligned} \text{Cov} \{c'_r, c'_{r+k}\} &= E(c'_r c'_{r+k}) - E(c'_r) E(c'_{r+k}) \\ &= \frac{1}{N^2} E \left\{ \sum_{t=1}^{N-r} \sum_{s=1}^{N-r-k} x_t x_{t+r} x_s x_{s+r+k} \right\} - (1 - \frac{r}{N})(1 - \frac{r+k}{N}) \gamma(r) \gamma(r+k) \end{aligned} \quad (\text{a2.1})$$

Now a standard result for quartic forms states that if $E(X_t) = 0 \quad \forall t$, then (by Isserlis (1918))

$$\begin{aligned} E(x_t x_{t+r} x_s x_{s+r+k}) &= E(x_t x_{t+r}) E(x_s x_{s+r+k}) \\ &\quad + E(x_t x_s) E(x_{t+r} x_{s+r+k}) \\ &\quad + E(x_t x_{s+r+k}) E(x_{t+r} x_s) + K_4(s-t, r, k) \end{aligned}$$

(where $K_4(s-t, r, k)$ is the coefficient of

$$\frac{(it_1)^t (it_2)^{r+t} (it_3)^s (it_4)^{s+r+k}}{t! (r+t)! s! (s+r+k)!}$$

in $\ln \phi(t_1, t_2, t_3, t_4)$, the multivariate characteristic function, if an expansion in power series exists)

Using this result, (a2.1) reduces to

$$\begin{aligned} \text{Cov}(c'_r, c'_{r+k}) &= \frac{1}{N^2} \sum_{t=1}^{N-r} \sum_{s=1}^{N-r-k} \{ \gamma(s-t) \gamma(s+k-t) + \gamma(s+r+k-t) \gamma(s-t-r) \\ &\quad + K_4(s-t, r, k) \} \end{aligned} \quad (\text{a2.2})$$

$$= \frac{1}{N} \sum_{m=a}^b \left\{ 1 - \frac{\gamma(m)+r+k}{N} \right\} \left\{ \gamma(m)\gamma(m+k) + \gamma(m+r+k)\gamma(m-r) \right. \\ \left. + K_4(m, r, k) \right\}$$

$$\text{where } a = -(N-r) + 1 \quad b = N - r - k - 1$$

$$\text{and } v(m) = \begin{cases} m & m > 0 \\ 0 & -k \leq m \leq 0 \\ -(m+k) & a \leq m < -k \end{cases}$$

Now if $\{X_t\}$ is Gaussian, then $K_4(m, r, k)$ is zero. Letting $k = 0$,

$$V(c_r) = \frac{1}{N} \sum_{m=a}^b \left\{ 1 + \frac{\gamma(m)+r}{N} \right\} \left\{ \gamma^2(m) + \gamma(m+r)\gamma(m-r) \right\}$$

For large N ,

$$\text{Cov}(c'_r, c'_{r+k}) = \frac{1}{N} \sum_{m=-\infty}^{\infty} \left\{ \gamma(m)\gamma(m+k) + \gamma(m+r+k)\gamma(m-r) \right\} \\ = 0 \text{ if } \gamma(m) = 0 \quad \forall m \neq 0 \quad (\text{a2.3})$$

$$\text{and } V(c'_r) = \frac{1}{N} \sum_{m=-\infty}^{\infty} \left\{ \gamma^2(m) + \gamma(m+r)\gamma(m-r) \right\} \\ = \frac{\gamma(0)}{N} \text{ if } \gamma(m) = 0 \quad \forall m \neq 0 \quad (\text{a2.4})$$

i.e. $V(c'_r)$ is of order $\frac{1}{N}$ for large N .

Note

If $\{X_t\}$ has a continuous spectrum then we may express (a2.3) and (a2.4) i.t.o. this continuous sdf, $h(w)$ say.

$$\text{Cov}(c'_r, c'_{r+k}) = \frac{2\pi}{N} \int_{-\pi}^{\pi} \left(e^{i\omega k} + e^{i\omega(2r+k)} \right) h^2(\omega) d\omega$$

and

$$V(c'_r) = \frac{2\pi}{N} \int_{-\pi}^{\pi} (1 + e^{2i\omega r}) h^2(\omega) d\omega$$

So for $r = 0$,

$$\begin{aligned} s_0^2 &= s_X^2 \\ &= \frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2 \end{aligned}$$

and so

$$V(c'_0) = \frac{2}{N} \sum_{m=-\infty}^{\infty} \gamma^2(m)$$

$$c^2(m)]$$

Since $c'_r = \frac{N-r}{N} c_r$, the variance and covariance terms for c_r can be found in a similar fashion.

In particular, for $r > 0$

$$V(c_r) = \frac{1}{N-r} \sum_{m=-\infty}^{\infty} \gamma^2(m) + ((m-r)\gamma(m-r)) \quad (\text{proof not shown})$$

thus $V(c_r)$ is of order $\frac{1}{N-r}$

APPENDIX A3 THE VARIANCE AND COVARIANCE OF ACRF ESTIMATES

Suppose we have a realization $x_1, \dots, x_N$ from a (second-order) zero-mean stationary process $\{x_t\}$. Define r'_k as in (2.33) and define a function $\delta(x)$ as

$$\delta(x) = x - E(x)$$

$$\begin{aligned}\delta(r'_k) &= \delta_k \quad \text{say} \\ &= r'_k - E(r'_k)\end{aligned}$$

If we assume δ_k and δ_{k+v} are small compared to $E(r'_k)$ and $E(r'_{k+v})$ then

$$\begin{aligned}E(r'_k) &= \frac{E(c'_k)}{E(c'_0)} \\ &= (1 - \frac{|k|}{N}) \rho(k)\end{aligned}$$

and

$$\begin{aligned}\delta_k &= \frac{c'_k - c'_0}{c'_0} \\ &= \frac{c'_k}{c'_0} - \frac{c'_k c'_0}{(c'_0)^2} \\ &= \frac{\delta(c'_k)}{\gamma(0)} - \frac{\gamma(k) \delta(c'_0)}{\gamma^2(0)}\end{aligned}$$

to a first-order approximation.

Hence

$$\begin{aligned}\delta_k \delta_{k+v} &= \frac{\delta(c'_k) \delta(c'_{k+v})}{\gamma^2(0)} - \frac{\gamma(k+v) \delta(c'_0) \delta(c'_k)}{\gamma^3(0)} \\ &\quad - \frac{\gamma(k) \delta(c'_0) \delta(c'_{k+v})}{\gamma^3(0)} + \frac{\gamma(k) \gamma(k+v) \delta(c'_0)^2}{\gamma^4(0)}\end{aligned}$$

Taking expectations gives

$$\begin{aligned} \text{Cov}(r'_k, r'_{k+v}) &= \frac{1}{4} \left\{ \text{Cov}(c'_k, c'_{k+v}) - \rho(k+v) \text{Cov}(c'_0, c'_k) - \right. \\ &\quad \left. \rho(k) \text{Cov}(c'_0, c'_{k+v}) + \rho(k) \rho(k+v) V(c'_0) \right\} \quad (\text{a3.1}) \end{aligned}$$

Assuming $\{X_t\}$ is Gaussian and using (a2.3), then if $k > 0$ and $k + v > 0$,

$$\begin{aligned} \text{Cov}(r'_k, r'_{k+v}) &= \frac{1}{N} \sum_{m=-\infty}^{\infty} \left\{ \rho(m) \rho(m+v) + \rho(m+k+v) \rho(m-k) \right. \\ &\quad \left. + 2\rho(k) \rho(k+v) \rho^2(m) - 2\rho(k) \rho(m) \rho(m-k-v) - 2\rho(k+v) \rho(m) \rho(m-k) \right\} \quad (\text{a3.2}) \end{aligned}$$

$$= 0 \text{ if } \rho(m) = C \quad \forall m \neq 0$$

So if $v = 0$ and $k > 0$,

$$\begin{aligned} V(r'_k) &= \frac{1}{N} \sum_{m=-\infty}^{\infty} \left\{ \rho^2(m) + \rho(m+k) \rho(m-k) \right. \\ &\quad \left. + 2\rho^2(m) - 4\rho(k) \rho(m) \rho(m-k) \right\} \quad (\text{a3.3}) \end{aligned}$$

If $\rho(m) = C \quad \forall m \neq 0$ then

$$V(r'_k) = \frac{1}{N}$$

[For further details see Bartlett (1946)]

APPENDIX A4: THE VARIANCE AND COVARIANCE OF THE PERIODOGRAM

The periodogram, $I(w)$, can be defined by

$$I(w) = \sum_{h=-(N-1)}^{(N-1)} c'_h \cos(hw) \quad -\pi \leq w \leq \pi$$

c'_h = sample acvf

(see (3.51) and section 2.4)

For any w_j, w_l (where $w_j = 2^{-1}j/N$),

$$\text{Cov}(I(w_j), I(w_l)) = 4 \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} a(s)a(r) \cos(sw_j) \cos(rw_l) \text{Cov}(c'_s, c'_r)$$

where $a(0) = 1$

and $a(s) = \frac{1}{2} \quad s \geq 1$

If we assume the x_t 's are independent,

then

$$K_4(s-t, k, v) = \begin{cases} E(X_t^4) & s-t = k = v = 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\gamma(s) = \begin{cases} \frac{2}{x} & s = 0 \\ 0 & \text{otherwise} \end{cases}$$

(i.e. $x_t = \epsilon_t \quad \forall t$ where $\{\epsilon_t\}$ is a purely random process).

Then $\text{Cov}(c'_s, c'_r) = 0 \quad r \neq s$

(see appendix A2)

$$\text{and} \quad V(c'_s) = \frac{N-s}{s^2} \sigma_X^4 (1 + \gamma_s) + \gamma_s \frac{K_4}{N} \quad (s \geq 0)$$

where $K_4 = E(X_t^4)$

$$\text{and} \quad \gamma_s = \begin{cases} 1 & s = 0 \\ 0 & s \neq 0 \end{cases}$$

Hence

$$\text{Cov}(I(w_j), I(w_l))$$

$$= \frac{4K_1}{N} + \frac{4C_X}{N} \left\{ 2 + 4 \sum_{a=1}^{N-1} \left(1 - \frac{s}{N}\right) \times \cos(sw_j) \cos(sw_l) \right\}$$

$$= \frac{4K_1}{N} + \frac{4C_X}{N} \left\{ \frac{(N-1)}{h=-(N-1)} \left(1 - \frac{|h|}{N}\right) (a+b) \right\}$$

$$(\text{where } a = \cos(h(w_j + w_l)))$$

$$b = \cos(h(w_j - w_l))$$

- proof not shown, but is similar to that used to find the $E(I(w))$

i.e. (3.53)

$$= \frac{4K_1}{N} + \frac{4C_X}{N} \{c+d\}$$

where

$$c = \frac{\sin^2 \left(\frac{N(w_j + w_l)}{2} \right)}{\sin^2 \left(\frac{w_j + w_l}{2} \right)}$$

and

$$d = \frac{\sin^2 \left(\frac{N(w_j - w_l)}{2} \right)}{\sin^2 \left(\frac{w_j - w_l}{2} \right)}$$

Hence

$$V(I(w_j)) = \frac{4K_1}{N} + \frac{4C_X}{N} \quad w_j = 0$$

APPENDIX A5: TESTING THE SIGNIFICANCE OF PERIODOGRAM ORDINATES

A plot of $I(w_j)$, the sample periodogram, against j (or w_j), will yield a series of ordinates some of which may be approximately zero. Those which are (nearly) zero will indicate that our model

$$x_t = \sum_{i=0}^{\lfloor \frac{N}{2} \rfloor} (a_i \cos(w_i t) + b_i \sin(w_i t)) + \varepsilon_t$$

$$t = 0, \dots, N-1$$

$$w_i = 2^{-1} i/N$$

does not contain either cosine or sine components at these particular frequencies. However we need an objective test to decide which of the ordinates are not significantly different from zero.

If the X_t are Gaussian, then by the theory of sections 3.5, 3.6, and 3.7,

$$I(w_j) \sim \frac{\sigma_X^2}{2} \chi^2_2 \quad j \neq 0 \text{ and } N \text{ odd}$$

So $I(w_j) / \frac{\sigma_X^2}{2}$ has pdf

$$f(x) = \frac{1}{2} e^{-x/2} \quad 0 \leq x < \infty \quad (\text{i.e. exponential with parameter } \lambda = \frac{1}{2})$$

$$P(I(w_j) / \frac{\sigma_X^2}{2} \leq k) = 1 - e^{-k/2} \quad (\text{a5.1})$$

Let

$$\beta = \max_{1 \leq j \leq \lfloor \frac{N}{2} \rfloor} \{ I(w_j) / \frac{\sigma_X^2}{2} \} \quad (N \text{ odd}) \quad (\text{a5.2})$$

$$\text{then } P(\beta > k) = 1 - (1 - e^{-k/2})^{\lfloor \frac{N}{2} \rfloor} \quad (\text{since the } I(w_j) \text{ are independent}).$$

Usually σ_X^2 is unknown and we have to estimate it from the data.

From (3.49),

$$E(I(w_j)) = \frac{\sigma_X^2}{2} \quad 0 \leq j \leq \lfloor \frac{N}{2} \rfloor \quad (N \text{ odd})$$

$$E \left\{ \sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I(w_j) \right\} = 2 \lfloor \frac{N}{2} \rfloor \cdot \frac{2}{N}$$

$$s_X^2 = s_N^2 = \frac{1}{2 \lfloor \frac{N}{2} \rfloor} \sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I(w_j) \quad (a5.3)$$

(s_X^2 is an unbiased estimator!)

Replacing σ_X^2 by s_X^2 in (a5.2), we obtain the test statistic

$$\hat{\beta} = \frac{\max_j I(w_j)}{\frac{1}{2 \lfloor \frac{N}{2} \rfloor} \sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I(w_j)} \quad (a5.4)$$

[Note that $I(w_0)$ need not be considered since $E(x_t) = 0 \quad \forall t$,

$$\text{and } I(w_0) = a_0^2$$

$$= (\bar{x})^2$$

$$= 0 \quad]$$

Ignoring sampling fluctuations in the denominator of $\hat{\beta}$ as given in (a5.3), then, asymptotically,

$$\begin{aligned} P(\hat{\beta} > k) &\approx 1 - (1 - e^{-\frac{k}{2} \lfloor \frac{N}{2} \rfloor}) \\ &\approx \lfloor \frac{N}{2} \rfloor e^{-k/2} \end{aligned} \quad (a5.5)$$

(for small $e^{-k/2}$ i.e. large k and N not too large)

So our test procedure for the maximum ordinate in the periodogram, consists simply of choosing our significance level α and hence finding k such that

$$P(\hat{\beta} > k) = \alpha$$

If $\hat{\beta} > k$ we conclude that

$$\max_j I(w_j) \neq 0$$

Suppose we conclude $\max_j I(w_j) \neq 0$.

What about $\max_{j \neq k} \{I(w_j)\}$ where $I(w_j) = \max \{I(w_j)\}$?

Omit $I(w_k)$ from $\sum_j I(w_j)$ in the denominator of $\bar{B}$ and replace N'

by $N'-1$ (where $N' = [\frac{N}{\alpha}]$). We may then test the second largest

ordinate. The major problem with this procedure is that our

overall significance level will not be α (in fact it will be α^2).

APPENDIX A6: A GUIDE FOR DETECTING PERIODICITY

Suppose we have a time series and we assume it has a period of T time units. A simple way for detecting this period T was devised by Buys-Ballot in 1847 (see, for example Whittaker and Robinson (1924) or Buys-Ballot (1847)). Write out all the data in the time series in groups of m ($m=2,3,\dots$)

1	2	3		m
x_1	x_2	x_3		x_m
x_{m+1}	x_{m+2}	x_{m+3}		x_{2m}
$\vdots$				
x_{km+1}	x_{km+2}	x_{km+3}		x_{km+m}
$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$		$\bar{x}_m$

Calculate the mean of each column (as above) and calculate the variance of these means. Compute the ratio η or η^2 where

$$\eta = \frac{\sigma_m}{\sigma}$$

and σ_m = std. dev. of means

σ = std. dev. of all the data.

Whittaker and Robinson (1924) have shown that η plotted against m is very similar to the periodogram and has much the same properties.

Excess x_i 's are just discarded

APPENDIX A7: SAMPLING PROPERTIES OF SPECTRAL ESTIMATES

As in chapter four, define an estimate of the sdf $h(w)$ to be of the form

$$\hat{h}(w) = \frac{1}{2\pi} \sum_{|t| < N} \lambda_t \bar{c}_t e^{-itw} \quad (\text{a7.1})$$

where λ_t = lag window

$\bar{c}_t$ = sample acvf

and our sample estimate is based on a realization $x_0, x_1, \dots, x_{N-1}$ from a stationary (second-order) process $\{X_t\}$.

Define

$$w_N(\theta) = \frac{1}{2\pi} \sum_{|t| < N} \lambda_t e^{it\theta} \quad (\text{a7.2})$$

= spectral window.

Assume $\lambda_t = \lambda_{-t}$ i.e. $\{\lambda_t\}$ is an even function, thus $w_N(\theta)$ is a real-valued function of θ and is continuous for all θ (since it is a finite sum of cosine functions).

$$w_N(\theta) \geq 0 \quad \forall N, \theta \quad (\text{by a7.2}) \quad (\text{a7.3})$$

If we assume $\lambda_0 = 1$, then

$$\int_{-\pi}^{\pi} w_N(\theta) d\theta = 1 \quad (\text{a7.4})$$

Furthermore, assume $\{\lambda_t\}$ is such that

$$\int_{-\pi}^{\pi} w_N^2(\theta) d\theta < \infty \quad \text{for finite } N \text{ and} \quad (\text{a7.5})$$

$$w_N(\theta) \rightarrow 0 \text{ uniformly as } N \rightarrow \infty \text{ for } |\theta| > \epsilon \quad (\epsilon > 0) \quad (\text{a7.6})$$

Also note that

$$\frac{\sum_{j=1}^N \frac{g_j(\theta)}{N} \frac{1}{\epsilon}}{\sum_{j=1}^N \frac{1}{\epsilon}} \longrightarrow 0 \quad \text{as } N \rightarrow \infty \quad (\text{a7.7})$$

and

$$\int_{-\pi}^{\pi} W_N(\theta) d\theta \rightarrow \infty \quad \text{as } N \rightarrow \infty \quad (\text{a7.8})$$

(by (a7.3), (a7.4) and (a7.5))

The above assumptions and notes make it easier to calculate asymptotic statistical properties of $h(w)$, which are given here, but with no details.

Theorem A

If $g_1(w)$ and $g_2(w)$ are 2 real-valued functions on $(-\pi, \pi)$, each with a finite number of discontinuities (possibly zero) and if both are absolutely integrable and square integrable

$$\begin{aligned} \text{i.e.} \quad & \left. \begin{aligned} \int_{-\pi}^{\pi} |g_i(\theta)| d\theta < \infty \\ \int_{-\pi}^{\pi} |g_i(\theta)|^2 d\theta < \infty \end{aligned} \right\} \quad i = 1, 2 \\ \text{and} \quad & \end{aligned}$$

then if $\{X_t\}$ is a process of the form

$$X_t = \sum_{u=-\infty}^{\infty} a_u \epsilon_{t-u} \quad \forall t$$

where $\{\epsilon_t\}$ is a white noise process with $E(\epsilon_t^2) = 1$ and

$$\sum_{u=-\infty}^{\infty} |a_u| |u|^k < \infty \quad (\forall k > 0)$$

where a_u is of order $|u|^{-k}$ ($k > 1$) and writing

$$I_k = \int_{-\pi}^{\pi} g_1(w) I^*(w) dw$$

$I^*(w)$ as in (4.5)

$$\text{and } \hat{\beta}_i = \int_{-\pi}^{\pi} g_i(w) h(w) dw \quad i = 1, 2$$

then

$$(a) \lim_{N \rightarrow \infty} E(\hat{\beta}_i) = \beta_i \quad i = 1, 2$$

$$(b) \lim_{N \rightarrow \infty} N \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = e_{12} + 4\pi \int_{-\pi}^{\pi} g_1(w) g_2^*(w) h^2(w) dw$$

$$\text{where } e = E(\varepsilon_t^4) - 3 \quad (a7.9)$$

$$\text{and } g_2^*(w) = \frac{1}{2}(g_2(w) + g_2(-w))$$

$$(c) \lim_{N \rightarrow \infty} NV(\hat{\beta}_i) = e\sigma_i^2 + 4\pi \int_{-\pi}^{\pi} g_i(w) g_i^*(w) h^2(w) dw \quad i = 1, 2$$

Proof: see, for example, Priestly (1981), p. 427.

Result 1

Replacing $g_1(w)$ and $g_2(w)$ by $w_N(w_1 - w)$ and $w_N(w_2 - w)$ (resp.),

then provided

$$x_t = \sum_u a_u \varepsilon_{t-u} \quad \forall t \quad \text{as in the theorem, then}$$

$$E(h(w)) \approx \int_{-\pi}^{\pi} h(\theta) W_N(w - \theta) d\theta = h^*(w) \text{ say} \quad (a7.10)$$

and

$$E(I^*(w)) \approx \int_{-\pi}^{\pi} h(\theta) D_N(\theta - w) d\theta \quad (a7.11)$$

$$\text{where } D_N(\cdot) = \frac{1}{2\pi} \frac{\sin^2(\frac{N\theta}{2})}{N \sin^2(\frac{\theta}{2})}$$

provided $h(w)$ has bounded first derivative ((a7.10) follows from theorem A and (a7.11) was shown by Fejer (1910)).

Result 2

In fact result 1 can be extended so that

$$E(h(w)) = \int_{-\pi}^{\pi} h(\theta) W_N(w - \theta) d\theta + O\left(\frac{\log N}{N}\right) \quad (a7.12)$$

Result 3

If we assume $h(w)$ is continuous for all w , then using (a7.4) and (a7.6), it may be shown that

$$\lim_{N \rightarrow \infty} E(h(w)) = h(w) \quad \forall w \quad (a7.13)$$

Result 4

By theorem A,

$$(a) \quad N V(h(w)) \approx e h''(w)^2 + 2 \int_{-\pi}^{\pi} h^2(\theta) W_N(w - \theta) [W_N(w - \theta) + W_N(w + \theta)] d\theta$$

where e and W_N are defined by (a7.10) and (a7.9)

$$(b) \quad N \text{Cov}(h(w_1), h(w_2)) \approx e h''(w_1) h''(w_2) + 2 \int_{-\pi}^{\pi} h''(\theta) W_N(w_1 - \theta) W_N(w_2 - \theta) + W_N(w_1 - \theta) W_N(w_2 + \theta) d\theta$$

Result 5

From (a7.6), as $N \rightarrow \infty$, $W_N(\cdot)$ has the limiting form of a Dirac delta function. The first term in result 4(a) is of order 1 and so negligible when compared with the second term as $N \rightarrow \infty$.

Now

$$W_N(w - \theta) \rightarrow \text{Dir}(w - \theta)$$

$$W_N(w + \theta) \rightarrow \text{Dir}(w + \theta)$$

(where Dir is the Dirac delta function)

so

$$\int_{-\pi}^{\pi} h^2(\theta) w_N(w-\theta) w_N(w+\theta) d\theta \rightarrow 0 \text{ as } N \rightarrow \infty$$

$w \neq 0 \text{ or } \pi$

In fact

$$V(h(w)) \approx (1+\delta) \frac{2\pi}{N} \int_{-\pi}^{\pi} h^2(\theta) w_N^2(w-\theta) d\theta$$

$$\delta = \begin{cases} 1 & w = 0 \text{ or } \pi \\ 0 & \text{otherwise} \end{cases}$$

But

$$\begin{aligned} \int_{-\pi}^{\pi} h^2(\theta) w_N^2(w-\theta) d\theta &= h^2(w) \int_{-\pi}^{\pi} w_N^2(w-\theta) d\theta \\ &= h^2(w) \int_{-\pi}^{\pi} w_N^2(\theta) d\theta \end{aligned}$$

(since $w_N(\cdot)$ has period 2π)

so

$$V(h(w)) \approx (1+\delta) \frac{2\pi}{N} h^2(w) \int_{-\pi}^{\pi} w_N^2(\theta) d\theta$$

$$\text{However } \int_{-\pi}^{\pi} w_N^2(\theta) d\theta = \sum_{|t| < N} \frac{2}{t}$$

$$\circ V(h(w)) \approx (1+\delta) h^2(w) \frac{1}{N} \left[\sum_{|t| < N} \frac{2}{t} \right]$$

Result 6

The first term of result 4(b) is always of order 1 and the second term involves the integral of products of functions, so if w_1 and w_2 are fixed (and $w_1 \neq w_2$), the second term tends to zero and

$$\lim_{N \rightarrow \infty} \text{Cov}(h(w_1), h(w_2)) = 0 \quad w_1 \neq w_2$$

Considering N to be large but finite, the first term of 4(b) is still of order 1, but the second term tends to zero if w_1 and w_2 are close, thus

$$\text{Cov}(h(w_1), h(w_2)) = \frac{2\pi}{N} \int_{-\pi}^{\pi} h^2(\theta) w_N(w_1 - \theta) w_N(w_2 - \theta) d\theta$$

Result 7

Let

$$\begin{aligned} b(w) &= E\{h(w)\} - h(w) \\ &= \text{bias in } h(w) \text{ at } w \end{aligned}$$

Now

$$E(h(w)) = \int_{-\pi}^{\pi} h(\theta) w_N(w - \theta) d\theta + O\left(\frac{\log N}{N}\right) \quad (\text{result 2})$$

so

$$b(w) = \int_{-\pi}^{\pi} [h(\theta) - h(w)] w_N(w - \theta) d\theta + O\left(\frac{\log N}{N}\right)$$

Ignoring the second term,

$$b(w) = \int_{-\pi}^{\pi} [h(w - \theta) - h(w)] w_N(\theta) d\theta$$

If $h(w)$ is twice differentiable (with bounded second derivative) and for θ small,

$$h(w - \theta) = h(w) - \theta h'(w) + \frac{\theta^2}{2} h''(w) + O(\theta^2)$$

$$\therefore b(w) = \frac{1}{2} h''(w) \int_{-\pi}^{\pi} \theta^2 w_N(\theta) d\theta$$

Result 8

$\hat{h}(w)$ is asymptotically unbiased (result 3) so $\hat{h}(w)$ is consistent if

$$\lim_{N \rightarrow \infty} \left\{ \frac{1}{N} \int_{-\pi}^{\pi} W_N^2(\theta) d\theta \right\} = 0 \quad (\text{by result 5})$$

i.e. if $\lim_{N \rightarrow \infty} \left\{ \frac{1}{N} \sum \lambda_{\epsilon}^2 \right\} = 0$

Result 9

(Proving $\hat{I}_X(w)$ is not a consistent estimate of $h_X(w)$)

$$\hat{h}(w) = \int_{-\pi}^{\pi} \hat{I}^*(w) W_N(w-\theta) d\theta \quad (\text{from section 4.1})$$

Suppose we approximate this integral by a sum over $\theta_j = 2\pi j/N$

$$\hat{h}(w) \approx \frac{2\pi}{N} \sum_j W_N(w-\theta_j) \hat{I}^*(w_j)$$

Suppose also that

$$x_t = \sum_u a_u x_{t-u} \quad \forall t$$

where $x_t \sim N(0, \sigma^2)$ independently

and $E(x_t^4) < \infty$

with $\sum_u |a_u| |u|^\alpha < \infty$ some $\alpha > 0$

i.e. suppose x_t is a moving average process

Then

$$\hat{I}_X(w) = |A(w)|^2 \hat{I}^*(w)$$

$$\text{where } |A(w)|^2 = \sum_u a_u e^{-i u w}$$

Since the x_t are uncorrelated, then

$$h_X(w) = \text{sdf of } \{x_t\} \\ = \frac{\sigma^2}{2}$$

$$\text{thus } h_X(w) = \frac{1}{2\pi} |A(w)|^2$$

$$\text{hence } I_X^*(w) = |A(w)|^2 I_X^*(w)$$

$$= \frac{2\pi}{2} h_X(w) I_X^*(w)$$

By the distribution of $I_X(w_j)$, (asymptotically), the

$$I_X(w_j)/2\pi h_X(w_j) \quad (j = 0, \dots, \lfloor \frac{N}{2} \rfloor)$$

are independently distributed, and in fact

$$I_X(w_j)^2 \sim \begin{cases} 4\pi h_X(w_j)^2 & j = 0 \text{ or } \lfloor \frac{N}{2} \rfloor \\ 2\pi h_X(w_j)^2 & \text{otherwise} \end{cases}$$

thus

$$E(I_X(w_j)) \approx 4\pi h_X(w_j) \quad j = 0 \text{ or } \lfloor \frac{N}{2} \rfloor$$

$$V(I_X(w_j)) = 16\pi^2 h_X(w_j)^2$$

$$\text{But } I^*(w) = \frac{1}{4\pi} I(w)$$

$$\text{so } E(I^*(w_j)) = h_X(w_j)$$

$$V(I^*(w_j)) = h_X(w_j)^2$$

$$j = 0 \text{ or } \lfloor \frac{N}{2} \rfloor$$

Thus $I_X^*(w)$ is not a consistent estimate of $h_X(w)$.

Result 10

Consider the case where λ_t is of the form

$$\lambda_t = k\left(\frac{t}{M}\right)$$

where $k(x)$ is independent of N .

Then $W_N(\beta)$ is termed a scale parameter window. Assume

(i) $k(x)$ is continuous

(ii) $k(x)$ is an even function

(iii) $k(0) = 1$

(iv) $\int_{-\infty}^{\infty} k^2(x) dx < \infty$

If we make the above assumptions, then (a7.3) - (a7.8) are satisfied provided that

$$\frac{M}{N} \rightarrow 0 \text{ as } M, N \rightarrow \infty$$

So

$$\frac{1}{M} \sum_{|t| \leq N} \frac{t^2}{t} = \frac{1}{M} \sum k^2\left(\frac{t}{M}\right) \rightarrow \int_{-\infty}^{\infty} k^2(x) dx \text{ as } M \rightarrow \infty$$

Instead of result 5, we have

$$(a) \lim_{N \rightarrow \infty} \left\{ \frac{N}{M} V(h(w)) \right\} = (1+\delta) h^2(w) \int_{-\infty}^{\infty} k^2(x) dx$$

$$(b) \lim_{N \rightarrow \infty} \left\{ \frac{N}{M} \text{Cov}(h(w_1), h(w_2)) \right\} = 0 \quad w_1 \neq w_2 \quad \text{instead of result 6}$$

(c) replacing $\frac{t}{M}$ by $k(\frac{t}{M})$ in $h(w)$ and taking expectations gives

$$\begin{aligned} b(w) &= \frac{1}{2\pi} \sum_{|t| < N} \left[k\left(\frac{t}{M}\right) \left(1 - \frac{|t|}{N}\right) - 1 \right] \gamma(t) e^{-itw} \\ &\quad - \frac{1}{2\pi} \sum_{|t| \geq N} \gamma(t) e^{-itw} \\ &= \frac{1}{2\pi} \sum_{|t| < N} \left[k\left(\frac{t}{M}\right) - 1 \right] \gamma(t) e^{-itw} \\ &\quad - \frac{1}{2\pi N} \sum_{|t| < N} |t| k\left(\frac{t}{M}\right) \gamma(t) e^{-itw} \\ &\quad - \frac{1}{2\pi} \sum_{|t| \geq N} \gamma(t) e^{-itw} \end{aligned}$$

Let r be the largest integer (≥ 0) such that

$$k^{(r)} = \lim_{x \rightarrow 0} \frac{1 - k(x)}{x^r} \quad (x \neq 0 \text{ and } < \infty)$$

Suppose

$$\sum_{t=-\infty}^{\infty} |t|^s |\gamma(t)| < \infty \quad s \leq r$$

and $\frac{M^r}{M^r} \rightarrow \infty$ as $N \rightarrow \infty$

then taking the limit as $M \rightarrow \infty$ of $b(w)$, results in the second and third terms in $b(w)$ tending to zero (as $O(M^{-r})$) uniformly in w , whilst the first term is (asymptotically) $-k^{(r)} M^{-r} h^{(r)}(w)$ where

$$h^{(r)}(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} |t|^r \gamma(t) e^{-itw}$$

$$\lim_{N \rightarrow \infty} \{M^r b(w)\} = -k^{(r)} h^{(r)}(w)$$

i.e. $b(w) \sim -M^{-r} k^{(r)} h^{(r)}(w)$ as $N \rightarrow \infty$

[For further details of all these results see Priestley (1981)]

Note: Motivation behind the technique of comparing lag windows by using negated derivatives

As in chapter 4 define

$h(w)$ = spectral density function (sdf)

$$= \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma(t) \cos(tw) \quad -\pi \leq w \leq \pi$$

where $\gamma(t)$ = acvf at lag t

Now $h(w)$ may be estimated by either

$$\hat{h}(w) = \frac{1}{2\pi} \sum_{t < N} \lambda_{1,t} c_t' \cos(tw)$$

or by

$$\hat{\hat{h}}(w) = \frac{1}{2\pi} \sum_{t < N} \lambda_{2,t} c_t \cos(tw)$$

where

$$\lambda_{i,t} = \text{lag window} \quad (i=1,2)$$

$$c_t^* = (\text{biased}) \text{ estimate of } \gamma(t) \quad (\text{see section 2.4})$$

$$c_t = (\text{unbiased}) \text{ estimate of } \gamma(t) \quad (\text{see section 2.4})$$

$$N = \text{size of the observed time-series}$$

Define

$$k_i\left(\frac{t}{M}\right) = \lambda_{i,t} \quad \forall t \quad i = 1, 2 \quad (M = \text{truncation point})$$

thus

$$k_1(0) = 1 = \lambda_{1,0}$$

$$k_1(\theta) = k_1(-\theta) \quad 0 \leq \theta \leq 1$$

$$k_1(\theta) = 0 \quad \theta > 1 \quad (\text{i.e. } t > M)$$

[Note that

$$h(\cdot) = h(\cdot) \quad \text{iff}$$

$$\lambda_{2,t} = \left(1 - \frac{|t|}{N}\right) \lambda_{1,t} \quad \forall t]$$

Define another estimate of the sdf by

$$S_t(w) = \frac{1}{2\pi} c_0 + \frac{1}{\pi} \sum_{u=1}^t c_u \cos(uw) \quad (t \geq 1)$$

$$\text{with } S_0(w) = c_0.$$

(Note that if $t = N-1$, $S_{N-1}(w)$ is very similar to the periodogram)

Since

$$\frac{1}{\pi} c_t \cos(tw) = S_t(w) - S_{t-1}(w) \quad (t > 0)$$

then

$$\begin{aligned} h(w) &= \frac{1}{2\pi} c_0 + \frac{1}{\pi} \sum_{t=1}^M \lambda_{2,t} c_t \cos(tw) \\ &= \frac{1}{2\pi} S_0(w) + \sum_{t=1}^M \lambda_{2,t} \{S_t(w) - S_{t-1}(w)\} \\ &= \sum_{t=0}^{M-1} \left[\lambda_{2,t} - \lambda_{2,t+1} \right] S_t(w) + \lambda_{2,M} S_M(w) \end{aligned}$$

$$= \sum_{t=0}^{M-1} \left\{ k_2\left(\frac{t}{M}\right) - k_2\left(\frac{t+1}{M}\right) \right\} S_t(w) + k_2(1) S_M(w)$$

From the mean value theorem, for some $\theta \in (0,1)$

$$k_2\left(\frac{t+1}{M}\right) - k_2\left(\frac{t}{M}\right) = \frac{1}{M} k_2'\left(\frac{t+\theta}{M}\right) \approx \frac{1}{M} k_2'\left(\frac{t}{M}\right)$$

if M is not too small.

Thus the weight applied to $S_t(w)$ in $\hat{h}(\cdot)$ is approximately proportional to $-k_2'\left(\frac{t}{M}\right)$.

At those values of t for which $S_t(\cdot)$ is likely to be a good estimate of $h(\cdot)$, the negated derivative, $-k_2'\left(\frac{t}{M}\right)$, should be (comparatively) large, while for other values it should be small.

So an indicator of the properties of the spectrum estimate $\hat{h}(\cdot)$ (or $h(\cdot)$) using $S_{2,t}$ (or $S_{1,t}$), is the (negated) derivative of $k_2(\cdot)$ (or $k_1(\cdot)$)

Neave (1972) states that if the spectrum is not too complicated and has no excessively high and narrow peaks, then the weighting factor (i.e. $-k_2'\left(\frac{t}{M}\right)$) should be small for t close to zero (since then $S_t(\cdot)$ has large bias) and rise to a maximum somewhere near $t \approx 20$. (For t large, $S_t(\cdot)$ is also a bad estimate since it has a large variance - see Neave(1972)).

$$= \sum_{t=0}^{M-1} \left(k_2\left(\frac{t}{M}\right) - k_2\left(\frac{t+1}{M}\right) \right) S_t(w) + k_2(1) S_M(w)$$

From the mean value theorem, for some $\xi \in (0,1)$

$$k_2\left(\frac{t+1}{M}\right) - k_2\left(\frac{t}{M}\right) = \frac{1}{M} k_2'\left(\frac{t+\xi}{M}\right) \approx \frac{1}{M} k_2'\left(\frac{t}{M}\right)$$

if M is not too small.

Thus the weight applied to $S_t(w)$ in $h(\cdot)$ is approximately proportional to $-k_2'\left(\frac{t}{M}\right)$.

At those values of t for which $S_t(\cdot)$ is likely to be a good estimate of $h(\cdot)$, the negated derivative, $-k_2'\left(\frac{t}{M}\right)$, should be (comparatively) large, while for other values it should be small. So an indicator of the properties of the spectrum estimate $h(\cdot)$ (or $S_t(\cdot)$) using $k_{2,t}$ (or $k_{1,t}$), is the (negated) derivative of $k_2(\cdot)$ (or $k_1(\cdot)$)

Neave (1972) states that if the spectrum is not too complicated and has no excessively high and narrow peaks, then the weighting factor (i.e. $-k_1'\left(\frac{t}{M}\right)$) should be small for t close to zero (since then $S_t(\cdot)$ has large bias) and rise to a maximum somewhere near $t \approx 20$. (For t large, $S_t(\cdot)$ is also a bad estimate since it has a large variance - see Neave(1972)).

APPENDIX A8: APPROXIMATE VARIANCE IN SPECTRUM ESTIMATES

$$I^*(w) = \frac{1}{2\pi} \sum_{t=-N}^N x_t e^{-itw} = \text{periodogram}$$

$h(w)$ = estimate of sdf $h(w)$

$$= \int_{-\pi}^{\pi} I^*(\theta) W_N(w-\theta) d\theta$$

$$\text{where } W_N(w-\theta) = \frac{1}{2\pi} \sum_{t=-N}^N e^{-it(w-\theta)}$$

= spectral window

If the time series $\{X_t\}$ is Gaussian, then the periodogram ordinates $I^*(w_j)$ ($w_j = 2\pi j/N$, $j = 0, 1, \dots, [N/2]$) are independently distributed as $\frac{1}{2} \chi^2_2$ (if $j \neq 0$ and $j \neq [N/2]$)

As $N \rightarrow \infty$

$$(1) \quad E(I^*(w)) \rightarrow h(w)$$

$$\text{so } E(h(w)) = \int_{-\pi}^{\pi} h(\theta) W_N(w-\theta) d\theta$$

$$(\text{since } E(h(w)) = E \left\{ \int_{-\pi}^{\pi} I^*(\theta) W_N(w-\theta) d\theta \right\})$$

$$= \int_{-\pi}^{\pi} E(I^*(\theta)) W_N(w-\theta) d\theta$$

$$= \int_{-\pi}^{\pi} h(\theta) W_N(w-\theta) d\theta$$

$$(11) \quad V(h(w)) = \frac{2}{N} \int_{-\pi}^{\pi} h^2(\theta) W_N^2(w-\theta) d\theta \quad (w \neq 0, \pm\pi)$$

$$= \frac{2}{N} h^2(w) \int_{-\pi}^{\pi} W_N^2(w-\theta) d\theta \quad (w \neq 0, \pm\pi)$$

(since $h(\theta)$ is continuous and for large N , $W_N^2(w-\theta)$ is highly concentrated about $\theta = w$ - see, for example, Priestley (1981))

If we approximate $h(w) = \int_{-\pi}^{\pi} I^*(\theta) w_N(w-\theta) d\theta$

by a discrete sum over the ordinates $w_j = \frac{2\pi j}{N}$,

then $\hat{h}(w) = \left(\frac{2\pi}{N}\right) \sum_j w_N(w-w_j) I^*(w_j)$

and hence $\hat{h}(w)$ is (approximately) a weighted sum of χ^2 r.v.'s and so it seems quite reasonable to approximate its distribution by a χ^2 distribution where

$$f = \frac{N}{\pi \int_{-\pi}^{\pi} w_N^2(w-\theta) d\theta} = \frac{2E(h(w))^2}{V(h(w))}$$

The (approximate) variances of some spectral estimates are given on the next page in the form

$$\frac{V(h(w))}{h^2(w)} \cdot \frac{N}{2\pi} = V^* \quad (w \neq 0, \pi)$$

[Since $2\pi \int_{-\pi}^{\pi} w_N^2(\theta) d\theta = \frac{2\pi}{N} \sum_{t < N} \frac{1}{t}$

$$\Rightarrow V(h(w)) \approx h^2(w) \frac{1}{N} \left\{ \frac{2\pi}{t} \right\} \quad w \neq 0, \pi$$

and thus we have an alternative method of calculating the approximate variance]

The bias of these windows may be obtained from, for example, Priestley (1981), pg. 163.

Window	V^* (defined on previous page)
Truncated periodogram	2
Bartlett (a)	2/3
Bartlett (b)	$2 + \frac{2}{3}(2a+b^2) + \frac{2}{3}a^2$ where $a = \frac{M}{N}$ and $b = \frac{M+N}{N}$
Daniell	1
Blackman-Tukey	$2(1 - 4a + 6a^2)$ $0 < a < 1$
Tukey-Hamming	0.7948
Tukey-Hanning	3/4
Parzen (a)	16/15
Parzen (b)	151/280 = 0.5393
Bartlett-Priestley	6/5
Bohman	$4/3 + \frac{1}{2^{1/2}} = 1.5866$
Tukey-Parzen	$2(1-2a)^2 + 4a^2$ and $a = 0.282$

Notes:

(i) For the case $w = 0$ or π , each of the above V^* must be doubled

(ii) Each of the above V^* was calculated by considering the windows as scale parameter windows i.e. writing

$\gamma_t = \gamma(\frac{t}{M})$ where M = truncation point. Since

$$\frac{1}{M} \sum_{t < |N|} \gamma_t^2 = \int_{-\infty}^{\infty} k^2(x) dx \quad \text{as } N \rightarrow \infty$$

$$\lim_{N \rightarrow \infty} \left\{ \frac{N}{M} V(h(w)) \right\} = h^2(w) \int_{-\infty}^{\infty} k^2(x) dx \quad w \neq 0, \pi$$

and hence for each window only $\int_{-\infty}^{\infty} k^2(x) dx$ need be calculated.

(iii) Choice of window is not only dependent on the (asymptotic) variance but also on the (asymptotic) bias. So before defining any of the above windows as best, the bias for each window must be calculated and then some technique decided upon in order to maximize (or minimize $P_1(w)$ or $P_2(w)$ (as defined in section 5.3)).

APPENDIX A9 : OBTAINING INITIAL ESTIMATES IN AN ARMA PROCESS

Suppose $\{X_t\}$ is zero-mean (second order) stationary and follows an ARMA (p,q) process i.e.

$$\phi(B)x_t = \theta(B)a_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

$$\{a_t\} = \text{white noise}$$

Box and Jenkins (1976) have suggested the following procedure for calculating initial estimates in an ARMA (p,q) model.

- (a) Calculate $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$ from $c'_{-p+1}, \dots, c'_{q+p}$ by using the Y-W equations. (See equation (6.16a) to see why $c'_0, \dots, c'_{q+1}$ are not used)

- (b) Find the autocovariances (q + 1 of them) of the new series

$$x_t^* = x_t - \sum_{i=1}^p \hat{\phi}_i x_{t-i} \quad \forall t.$$

- (c) Calculate $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ by using the following quadratically convergent procedure

- (i) Denote $\underline{\tau}' = (\tau_0, \tau_1, \dots, \tau_q)$ and let

$$\hat{\theta}_j = -\tau_j / \tau_0 \quad j = 1, \dots, q \quad \tau_0 = \sigma_a^2$$

- (ii) Let $\underline{\tau}_{(1)}$ denote the $\underline{\tau}$ obtained at iteration 1

then

$$\underline{\tau}_{(i+1)} = \underline{\tau}_{(i)} - (\underline{\tau}_{(i)}')^{-1} \underline{f}_{(i)}$$

$$\text{where } \underline{f}' = (f_0, f_1, \dots, f_q)$$

Strictly speaking, the Yule-Walker equations were defined i.t.o. the acf^s. However, multiplying the Y-W equations (through-out) by $\gamma(0)$ yields a set of equations i.t.o. the acv^s.

with

$$x_j = \sum_{i=0}^{q-j} \tau_i \tau_{i+j} = a_j''$$

and $c_j'' = j^{\text{th}}$ acv of (x_t'')

and

$$T = \begin{bmatrix} \tau_0 & \tau_1 & \dots & \tau_q \\ 0 & \tau_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} + \begin{bmatrix} \tau_0 & \tau_1 & \dots & \tau_q \\ 0 & \tau_0 & \dots & \tau_{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \tau_0 \end{bmatrix}$$

For more details, see Box and Jenkins (1976) pg.201

APPENDIX A10 : A RECURSIVE METHOD FOR CALCULATING ESTIMATES OF AR PARAMETERS

Here we will show how Yule-Walker estimates for the parameters of an AR(p+1) process may be obtained from the estimates of an AR(p) process, and used to approximate the pacrf.

$$r'_j = \phi_{k1} r'_{j-1} + \dots + \phi_{k,k-1} r'_{j-k+1} + \phi_{k,k} r'_{j-k} \quad (j \leq p)$$

(by the Yule-Walker equations - see section 6.3.1 and in particular equation (6.10)) where ϕ_{ki} ($i=1, \dots, k$) refer to the coefficients of an AR(k) process

In particular, when $k = 2$

$$r'_2 = \phi_{21} r'_1 + \phi_{22}$$

$$r'_1 = \phi_{21} + \phi_{22} r'_1$$

and when $k = 3$

$$r'_3 = \phi_{31} r'_2 + \phi_{32} r'_1 + \phi_{33} \quad (\text{a10.1})$$

$$r'_2 = \phi_{31} r'_1 + \phi_{32} + \phi_{33} r'_1 \quad (\text{a10.2})$$

$$r'_1 = \phi_{31} + \phi_{32} r'_1 + \phi_{33} r'_2$$

It may be shown, by writing (a10.2) in matrix form and solving for ϕ_{31} and ϕ_{32} that

$$\phi_{31} = \phi_{21} - \phi_{33} \phi_{22}$$

$$\phi_{32} = \phi_{22} - \phi_{33} \phi_{21}$$

Substituting these two equations into (a10.1) gives

$$\phi_{33} = \frac{r'_3 - \phi_{21} r'_2 - \phi_{22} r'_1}{1 - \phi_{21} r'_1 - \phi_{22} r'_2}$$

In general, (Durban (1960)) it may be shown that

$$\hat{\gamma}_{p+1,j} = \hat{\gamma}_{pj} - \hat{\gamma}_{p+1,p+1} \hat{\gamma}_{p,p-j+1} \quad j = 1, \dots, p$$

$$\hat{\gamma}_{p+1,p+1} = \frac{r'_{p+1} - \sum_{j=1}^p \hat{\gamma}_{pj} r'_{p+1-j}}{1 - \sum_{j=1}^p \hat{\gamma}_{pj} r'_j}$$

(Also see Box and Jenkins (1976) pg. 92)

APPENDIX A11 : PARAMETER ESTIMATES AND PROPERTIES OF SPECIFIC BJ MODELS

AR(1)

stationarity condition

$$|\phi_1| < 1$$

invertibility condition

acrf

$$\rho(k) = \phi_1 \rho(k-1) \quad k > 0$$

variance

$$\sigma_x^2 = \sigma_a^2 / (1 - \phi_1 \rho(1))$$

spectrum

$$h(w) = \frac{\sigma_a^2}{2\pi [1 + \phi_1^2 - 2\phi_1 \cos(w)]}$$

ϕ_1

$$\phi_1 = r_1'$$

behaviour of acrf

decays slowly

behaviour of pacrf

only ϕ_{11} nonzero

σ_a^2 (residual variance)

$$\sigma_a^2 = \gamma(0) (1 - \phi_1^2)$$

AR(2)

stationarity condition

$$\begin{cases} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1 \end{cases}$$

invertibility condition

acrf

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) \quad k > 0$$

variance

$$\sigma_x^2 = \sigma_a^2 / (1 - \sum_{i=1}^2 \phi_i \rho(i))$$

spectrum

$$h(w) = \frac{\sigma_a^2 (2\pi)^{-1}}{1 + \phi_1^2 + \phi_2^2 + 2\phi_1(1 - \phi_2) \cos(w) - 2\phi_2 \cos(2w)}$$

ϕ_1

$$\phi_1 = r_1' (1 - r_2') / (1 - r_1'^2)$$

ϕ_2

$$\phi_2 = (r_2' - r_1'^2) / (1 - r_1'^2)$$

behaviour of acrf

a mix of exponentials or damped
sine waves

behaviour of pacrf

only ϕ_{11} and ϕ_{22} nonzero

σ_x^2 (residual variance)

$$\sigma_x^2 = \sigma_a^2(1 - \phi_1 \rho(1) - \phi_2 \rho(2))$$

MA(1)

stationarity condition

invertibility condition

acrf

$$|\theta_1| < 1$$

$$\rho(1) = -\theta_1 / (1 + \theta_1^2)$$

$$\rho(k) = 0 \quad k > 1$$

variance

$$\sigma_x^2 = (1 + \theta_1^2) \sigma_a^2$$

spectrum

$$h(w) = \sigma_a^2 (1 + \theta_1^2 - 2\theta_1 \cos(w))^{-1}$$

$$\hat{\theta}_1$$

$$r_1' = -\theta_1 / (1 + \theta_1^2)$$

(and solve for θ_1)

behaviour of acrf

only $\rho(1)$ nonzero

behaviour of pacrf

exponential dominates the decay

σ_a^2 (residual variance)

$$\sigma_a^2 = \sigma_x^2 / (1 + \theta_1^2)$$

MA(2)

stationarity condition

invertibility condition

acrf

$$\theta_2 + \theta_1 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$-1 < \theta_2 < 1$$

$$\rho(1) = -\theta_1(1 - \theta_2) / (1 + \theta_1^2 + \theta_2^2)$$

$$\rho(2) = -\theta_2 / (1 + \theta_1^2 + \theta_2^2)$$

$$\rho(k) = 0 \quad k > 2$$

variance

$$\sigma_x^2 = \sigma_a^2 (1 + \theta_1^2 + \theta_2^2)$$

spectrum

$$h(w) = 2\sigma_a^2 \left(1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1-\theta_2)\cos(w) - 2\theta_2\cos(2w) \right)$$

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}$$

$$r'_1 = -\hat{\theta}_1(1-\hat{\theta}_2)/(1+\hat{\theta}_1^2+\hat{\theta}_2^2)$$

$$r'_2 = -\hat{\theta}_2/(1+\hat{\theta}_1^2+\hat{\theta}_2^2)$$

(Solve for $\hat{\theta}_1$ and $\hat{\theta}_2$)

behaviour of acrf

only $\phi(1)$ and $\psi(2)$ nonzero

behaviour of pacrf

mix of exponentials and sine waves

σ_a^2 (residual variance)

$$\sigma_a^2 = \sigma_X^2 / (1 + \theta_1^2 + \theta_2^2)$$

ARMA (1,1)

stationarity condition

$$|\phi_1| < 1$$

invertibility condition

$$|\theta_1| < 1$$

acvf

$$\gamma(1) = \frac{\sigma_a^2(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}$$

variance

$$\gamma(k) = \phi_1^{k-1} \quad k \geq 2$$

$$\sigma_X^2 = (1+\theta_1^2-2\phi_1\theta_1)\sigma_a^2/(1-\phi_1^2)$$

spectrum

$$h(w) = \frac{\sigma_a^2(1+\theta_1^2-2\theta_1\cos(w))}{2(1+\phi_1^2-2\phi_1\cos(w))}$$

$\hat{\phi}_1$ and $\hat{\theta}_1$

$$r'_2 = r'_1 \hat{\phi}_1$$

$$r'_1 = \frac{(1-\hat{\phi}_1\hat{\theta}_1)(\hat{\phi}_1-\hat{\theta}_1)}{1+\hat{\phi}_1^2-2\hat{\phi}_1\hat{\theta}_1}$$

(Solve for $\hat{\phi}_1$ and $\hat{\theta}_1$)

behaviour of acrf

decays exponentially

behaviour of pacrf

dominated by an exponential

σ_a^2 (residual variance)

$$\sigma_a^2 = \frac{\sigma_X^2(1-\phi_1^2)}{1+\phi_1^2-2\phi_1\theta_1}$$

Notes

(i) The Yule-Walker equations may be expressed in the form

$$\underline{\hat{\phi}} = P^{-1} \underline{\phi}$$

where $\underline{\phi}' = (\phi(1), \dots, \phi(p))$

$$\text{and } P = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & & \rho(p-2) \\ \vdots & & \ddots & \vdots \\ \rho(p-1) & \dots & \dots & 1 \end{bmatrix}$$

with $\underline{\phi} = (\phi_1, \phi_2, \dots, \phi_p)$

So

$$\underline{\hat{\phi}} = \underline{\phi}^{-1} \underline{\Gamma}$$

and $V(\underline{\hat{\phi}}) = \frac{1}{N} \sigma_a^2 \underline{\Gamma}^{-1} = \frac{1}{N} (1 - \underline{\phi}' \underline{\phi}) P^{-1}$

since $\sigma_a^2 = \gamma(0) (1 - \underline{\phi}' \underline{\phi})$

where $\underline{\Gamma}$ and P are the acv and acr matrices of an AR(p) process

(ii) In particular, from (i) if we have an AR(2) process then

$$V(\hat{\phi}_1, \hat{\phi}_2) = \frac{1}{N} \begin{bmatrix} 1 - \hat{\phi}_2^2 & -\hat{\phi}_1(1 + \hat{\phi}_2) \\ -\hat{\phi}_1(1 + \hat{\phi}_2) & 1 - \hat{\phi}_1^2 \end{bmatrix}$$

(iii) For an MA(2) process

$$V(\hat{\theta}_1, \hat{\theta}_2) = \frac{1}{N} \begin{bmatrix} 1 - \hat{\theta}_2^2 & -\hat{\theta}_1(1 + \hat{\theta}_2) \\ -\hat{\theta}_1(1 + \hat{\theta}_2) & 1 - \hat{\theta}_1^2 \end{bmatrix}$$

(iv) For an ARMA(1,1) process

$$V(\hat{\phi}, \hat{\theta}) = \frac{1}{N} \begin{bmatrix} (1 - \hat{\phi}^2)(1 - \hat{\theta}^2) & (1 - \hat{\phi}^2)(1 - \hat{\theta}^2) \\ (1 - \hat{\phi}^2)(1 - \hat{\theta}^2) & (1 - \hat{\phi}^2)(1 - \hat{\theta}^2) \end{bmatrix}$$

(See Box and Jenkins (1976))

APPENDIX A12 . ESTIMATION OF THE IACRF AND THE ASYMPTOTIC BEHAVIOUR OF THE IACR^S

ESTIMATION OF THE IAC

There are two techniques for calculating estimates for the iacrf - either by using the inverse model, or by using the periodogram estimate. The former is just like the procedures discussed in the For-Jenkins modelling sections (i.e. chapter six). The latter is as follows.

(i) Suppose we have $x_0, x_1, \dots, x_{N-1}$

(ii) Define the periodogram as in earlier chapters i.e.

$$I(w) = \frac{1}{2\pi} \left| \sum_{t=0}^{N-1} x_t e^{-itw} \right|^2$$

(and w.l.o.g assume $E(x_t) = 0 \quad \forall t$)

(iii) An estimate of the sdf is given by

$$h(w) = \frac{1}{M} \sum_{k=1}^M I_k(w)$$

(see Bloomfield (1976), pg 163 - all that we do here is divide the data into M segments each of length k say (where $N = kM$) and let $I_j(w)$ be the periodogram of the j^{th} segment).

Any other estimate of the spectrum may be used.

(iv) Define the 'inverse' spectral density function by

$$h_1(w) = 1/h(w)$$

$$\therefore h_1(w) = 1/h(w)$$

Here $\{x_0, \dots, x_{N-1}\}$ is used instead of $\{x_1, \dots, x_N\}$ for notational convenience.

(v) The $\text{iacv}^s \gamma I(k)$ are estimated by

$$\gamma I(k) = \int_{-\pi}^{\pi} \hat{h}_1(w) e^{ikw} dw$$

(see Cleveland (1972))

[Chatfield (1979) has also proposed a technique for estimating the iacr^s and this involves fitting a large order AR process to the data i.e.

$$x_t = a_t + \sum_{i=1}^v \phi_i x_{t-i}$$

$$\therefore \rho I(k) = r_{1k} \quad \text{say}$$

$$r_{1k} = \frac{\sum_{i=1}^{v-k} \hat{\phi}_i \hat{\phi}_{i+k}}{1 + \sum_{i=1}^v \hat{\phi}_i^2}$$

(which is the form of $\rho(k)$ for an MA(v) process).

Hipel et al (1977) suggest choosing about four values of v between 10 and 40 (where $v \propto \frac{N}{1}$) and then using the 'most representative' graph].

2 THE ASYMPTOTIC DISTRIBUTION OF THE IACR^s

Suppose $\{x_t\}$ is zero-mean stationary with acvgf and iacvgf $\gamma(z)$ and $\Gamma(z)$ resp. Let the realization of the process be $x_0, x_1, \dots, x_{N-1}$

Then

$$\sqrt{N}(r_k^s - \gamma(k)) \xrightarrow{d} \text{Normal}(0, \text{cov})$$

where

$$\text{cov} = N(V(r_k^s))$$

$$V(r_k^s) = \frac{1}{N} \sum_{m=-\infty}^{\infty} \gamma(m) \gamma(m+k) \gamma(m-k) + 2 \gamma(k) \gamma(m) \gamma(m-k)$$

(See appendix A3, (a3.3))

Now suppose

$$(B)x_t = \theta(B) a_t \quad \forall t$$

then

$$h_X(w) = \text{sdf of } \{X_t\}$$

$$= \frac{a^2}{2\pi} \frac{1 - \sum_{j=1}^q \theta_j e^{-ijw}}{1 - \sum_{j=1}^p \phi_j e^{-ijw}} \quad \text{by (6.21)} \quad (\text{a12.1})$$

Now

$$h_1(w) = 1/h(w)$$

$$h_1(w) = \left(\frac{2\pi}{a} \right)^2 \frac{1 - \sum_{j=1}^p \phi_j e^{-ijw}}{1 - \sum_{j=1}^q \theta_j e^{-ijw}} \quad (\text{a12.2})$$

and

$$I(k) = \int_{-\pi}^{\pi} h_1(w) e^{ikw} dw$$

hence so

$$I(k) = \left(\frac{2\pi}{a} \right)^2 \int_{-\pi}^{\pi} \frac{1 - \sum_{j=1}^p \phi_j e^{-ijw}}{1 - \sum_{j=1}^q \theta_j e^{-ijw}} e^{ikw} dw \quad (\text{a12.3})$$

$$= \left(\frac{2\pi}{a} \right)^2 \int_{-\pi}^{\pi} \frac{1 - \sum_{j=1}^p \hat{\phi}_j e^{-ijw}}{1 - \sum_{j=1}^q \theta_j e^{-ijw}} e^{ikw} dw$$

$$= \left(\frac{2\pi}{a} \right)^2 G(w) \text{ say}$$

But $G(w) = h_y(w)$

where $h_y(w)$ is the sdf of the process $\{Y_t\}$ where

$$\phi(B)a_t = \theta(B)y_t \quad \text{i.e. an ARMA (q,p) process.}$$

Alternatively

$$\phi(B)x_t = \theta(B)a_t \quad \forall t$$

then write

$$x_t = \alpha(B)a_t \quad \forall t$$

$$\text{where } \alpha(B) = \phi^{-1}(B)\theta(B)$$

and define

$$\Gamma_1(Z) = \text{acvfg of } \{X_t\}$$

$$= \sigma_a^2 \alpha(Z)\alpha(Z^{-1}) \quad (\text{see section 6.6})$$

If

$$\theta(B)x_t = \phi(B)\varepsilon_t \quad \forall t$$

then

$$x_t = \phi^{-1}(B)\theta(B)\varepsilon_t$$

$$= \alpha^{-1}(B)\varepsilon_t$$

$$\text{since } \alpha^{-1}(B) = \phi(B)\theta^{-1}(B)$$

So

$$\Gamma_2(Z) = \sigma_\varepsilon^2 \alpha^{-1}(Z)\alpha^{-1}(Z^{-1})$$

But $\Gamma_1(Z)\Gamma_2(Z) = 1$ (by section 7.2.1)

$$\Gamma_1(Z) = 1/\Gamma_2(Z)$$

$$= (\sigma_a^2)^{-1} \alpha^{-1}(Z)\alpha^{-1}(Z^{-1})$$

$$= (\Gamma_2(Z))(\sigma_a^2)^{-1}/\sigma_\varepsilon^2$$

$$\sigma_a^2 \Gamma_1(Z) = (\Gamma_2(Z)) \quad \text{say}$$

i.e. $\Gamma_1(Z) = \Gamma_2(Z)$ if $\sigma_a^2 = \sigma_\varepsilon^2$

Similarly

$$PI_1(Z) \approx P_2(Z)$$

$$\text{and } P_1(Z) \approx PI_2(Z)$$

Getting back to the original argument,

$$\text{if } \sqrt{n} (r'_k - \rho(k)) \xrightarrow{d} N(0, \text{cov})$$

then surely

$$\sqrt{n} (r'_k - \rho I(k)) \xrightarrow{d} N(0, \text{cov})?$$

Bhansali (1980) has shown that, asymptotically, the sampling distribution of the $iacr^s$, from an ARMA(p,q) process, is the same as that of the acr^s from an ARMA(q,p) process. (Also see Hosking (1980))

APPENDIX A13 : THE DISTRIBUTION OF RESIDUAL AUTOCORRELATIONS

1. Consider

$$\phi(B)x_t = \theta(B)a_t$$

where $\{x_t\}$ is a zero-mean stationary process.

Define

$r'_{a,k}$ = estimated acr of the residuals a_t , at lag k

$$r'_{a,k} = \frac{\sum_{t=k}^N a_t a_{t-k}}{\sum_{t=k}^N a_t^2} \quad (\text{a13.1})$$

$$t = 0, \dots, N-1$$

Now

$$V(r'_{a,k}) = \frac{N-k}{N(N+2)} \quad (\text{Anderson (1942) or Priestley (1981), pg. 339})$$

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

so if N is large

$$\sum_{k=1}^m \frac{1}{N-k} r_{a,k}^2 \sim \chi_m^2 \quad (\text{a13.2})$$

assuming the $r'_{a,k}$ are independent for different k

(since $r'_{a,k} \sim N(0, \frac{N-k}{N(N+2)}) \quad \forall k$ (Anderson (1942)))

2. Let $\beta = (\underline{\phi}, \underline{\theta})$

then

$$\underline{\beta} \sim N(\underline{\beta}; I^{-1}/N)$$

where

$$I = \begin{pmatrix} \frac{\partial^2 \log L}{\partial \phi \partial \phi} & \frac{\partial^2 \log L}{\partial \phi \partial \theta} \\ \frac{\partial^2 \log L}{\partial \theta \partial \phi} & \frac{\partial^2 \log L}{\partial \theta \partial \theta} \end{pmatrix}$$

and

$$\phi(B)v_t = -a_t$$

$$\theta(B)u_t = a_t$$

$$\gamma_{vv}(k) = E(v_t v_{t+k})$$

$$\gamma_{uu}(k) = E(u_t u_{t+k})$$

$$\gamma_{vu}(k) = E(v_t u_{t+k})$$

$$\gamma_{uv}(k) = \gamma_{vu}(-k)$$

(see Hannan (1970))

and

$$v_{t-1} = \partial \hat{a}_t / \partial \phi_1$$

$$u_{t-1} = \partial \hat{a}_t / \partial \theta_1$$

where $\hat{a}_t$ are the residuals obtained by replacing β with a known $\hat{\beta}$.

Hannan (1970) has shown that the large sample distribution of $\underline{r} = (r'_{a,1}, \dots, r'_{a,m})'$ is normal with mean 0 and covariance matrix

$$(I - XI^{-1}X')/N$$

I as on the previous page

$I = m \times m$ identity matrix

$X = \begin{pmatrix} -\phi_1 & \theta_1 \\ \vdots & \vdots \end{pmatrix}$ (a partitioned matrix,

with each partition showing it's ij^{th} element)

where ϕ_j and θ_j are defined by

$$1/\phi(B) = \sum_{j=0}^{\infty} \phi_j B^j$$

and

$$1/\theta(B) = \sum_{j=0}^{\infty} \theta_j B^j$$

and $\phi_j = \theta_j = 0$ if $j < 0$.

(see McLeod (1978))

APPENDIX A14 : ARMA PROCESSES WITH NON-NORMAL RESIDUALS

Invariably in the section on Box-Jenkins models, when finding the asymptotic distribution of acr^s , residual acr^s etc, the a_t^s (white noise) in the model

$$\phi(B)x_t = \psi(B)a_t$$

have been assumed to be independently distributed $N(0, \sigma_a^2)$. Davies, Spedding and Watson (1980) considered the case where $\{a_t\}$ is a zero-mean white noise process with variance σ_a^2 , but with nonzero moment coefficient of skewness $\gamma_1(a)$ and moment coefficient of kurtosis $\beta_2(a) \neq 3$.

Since any stationary ARMA process can be written i.t.o an $MA(\infty)$ process i.e.

$$x_t = \alpha(B)a_t$$

where $\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \dots$

(and $\phi(B)\alpha(B) = \psi(B)$)

then

$$\beta_1(x) = \frac{E(x_t^3)}{\{E(x_t^2)\}^{3/2}} = \frac{\sum_{i=0}^{\infty} \alpha_i^3 \gamma_1(a)}{\left\{ \sum_{i=0}^{\infty} \alpha_i^2 \right\}^{3/2}} \quad (\alpha_0 = 1) \quad (\text{a14.1})$$

and

$$\beta_2(x) = \frac{E(x_t^4)}{\{E(x_t^2)\}^2} = \frac{\sum_{i=0}^{\infty} \alpha_i^4}{\left\{ \sum_{i=0}^{\infty} \alpha_i^2 \right\}^2} \beta_2(a) + \frac{6 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \alpha_i^2 \alpha_j^2}{\sum_{i=0}^{\infty} \alpha_i^2} \quad (\text{a14.2})$$

Davies, Watson and Spedding (1980) claim that usually only α_1 $1 \leq i \leq 30$ is required (provided $p, q \leq 2$).

(a14.1) and (a14.2) can be used to approximate the appropriate skewness and kurtosis that need be imposed on the noise process $\{a_t\}$ to generate $\{X_t\}$ with given skewness and kurtosis.

[Also see Lomnicki (1961), Gabr and Subba Rao (1981), Subba Rao and Gabr (1981)]

APPENDIX A15 : ESTIMATING THE ORDER OF AR AND MA MODELS - THE MAX

 χ^2 METHOD

McClave (1975, 1978, 1978a) has developed a technique for determining the order of AR and MA models. Only the method for AR models will be given here since that of MA models follows from the AR procedure.

Suppose $\{X_t\}$ is stationary AR(p) process i.e.

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t \quad \forall t$$

where $\{a_t\}$ is a zero-mean white noise process with variance σ_a^2 .

If we have an AR(10), say, model, it is feasible that some subsets of the AR coefficients may provide models that fit equally well

$$\text{e.g. } x_t = \phi'_1 x_{t-1} + \phi'_5 x_{t-5} + \phi'_{10} x_{t-10} + a_t$$

(analogous to subset regression models).

A general subset model will be denoted $g \max p^*$ with lags

$\{j_1, j_2, \dots, j_g\}$ (where $j_g = p$) - so the model in our example is an AR model of order $3 \max 10$ with lags $\{1, 5, 10\}$.

If $\underline{1}$ is the parameter set for the 'full' AR(p) model, then we may write the subset parameter vector $\underline{1}^*$ as the union of two disjoint sets namely

$$\underline{1}^* = \underline{1}_1 \cup \underline{1}_2$$

where $\underline{1}_1 = \{g : j_1, j_2, \dots, j_g\}$

and $\underline{1}_2 = \{\phi_{j_1}, \phi_{j_2}, \dots, \phi_{j_g}\}$

Note that $\underline{1}_1$ must be 'known' before $\underline{1}_2$

Consider a realization of the AR(p) process (i.e. $x_0, x_1, \dots, x_{N-1}$) and suppose this can be generated 'equally well' by the AR model of order $g \leq p$ with lags $\{j_1, j_2, \dots, j_g\}$. (The purpose of this technique is to find the best subset AR model of order $g \leq p$.)

Let $P_g = g \times g$ matrix with rs th element $\rho(j_r - j_s)$
 $(r, s = 1, 2, \dots, g)$

and $\underline{r}_g = (\rho(j_1), \dots, \rho(j_g))$

hence we may estimate these by

$$\hat{R}_g = (\hat{r}_{j_r - j_s})$$

and $\hat{\underline{r}}_g = (\hat{r}_{j_1}, \hat{r}_{j_2}, \dots, \hat{r}_{j_g})$

Define $\hat{\phi}_2$ to be an estimate of ϕ_2 (if we are given $\hat{\phi}_1$ (or $\hat{\phi}_1$))

then

$$\hat{\phi}_2 = \hat{R}_g^{-1} \hat{\underline{r}}_g \quad (\text{called the modified Yule-Walker equations})$$

and $\hat{\phi}_a = \hat{\phi}_0 (1 - \hat{\underline{r}}_g' \hat{R}_g^{-1} \hat{\underline{r}}_g)$

To estimate $\hat{\phi}_1$, we need the 'best' AR model of each order $g \leq p$ ($g = 1, 2, \dots, p$).

McClave (1975, 1978) gives the following procedure for determining the best AR model for a fixed g :

- (i) compute $\delta(i)$ ($i=1, 2, \dots, p$) where $\delta(i)$ is the increase in the estimated residual variance when lag i is removed from the AR model of order p
- (ii) order the $\delta(i)$ such that

$$\delta(i_1) \leq \delta(i_2) \leq \dots \leq \delta(i_p)$$

- (iii) let $q = p - g$
- (iv) remove the lags $i_1, \dots, i_q$ from the AR(p) model and compare the increase in residual variance with $\delta(i_{q+1})$; if this increase is $\delta(i_{q+1})$, the AR model of order $g \max p$ with smallest residual variance among all $\binom{p}{g}$ such models, is obtained by removing lags $i_1, \dots, i_q$; if this increase $> \delta(i_{q+1})$ proceed to the next step
- (v) find the minimum increase in residual variance for the set of $\binom{q+1}{q}$ models obtained by removing each of the subsets of q lags in $(i_1, i_2, \dots, i_{q+1})$; compare this minimum increase with $\delta(i_{q+2})$ - if it is $< \delta(i_{q+2})$ stop, otherwise proceed to the next step
- (vi) find the minimum increase in residual variance for the set of $\binom{q+2}{q}$ models obtained by removing each of the subsets of q lags in $(i_1, i_2, \dots, i_{q+2})$; if this increase is $< \delta(i_{q+3})$ we have found the 'best' subset model of order $g \max p$, otherwise continue in this manner

Doing this for each g ($g=1, \dots, p$) we eventually obtain a set of vectors $\hat{p}_1(1), \hat{p}_1(2), \dots, \hat{p}_1(p)$ where $\hat{p}_1(i)$ is an estimate of $\hat{p}_1 = \{g : j_1, j_2, \dots, j_g\}$ for the case when $g = 1$.

We now need to estimate the 'best' $g(1, 2, \dots, p)$ and hence use $\hat{p}_1(\hat{g})$ to estimate $\hat{p}_1$

McClave (1975, 1978) suggests the following

- (1) compute $\{g_0, \dots, g_{p-1}$

This technique may cause one to 'stop' too early, however the result should be close enough (Mc Clave 1978)

$$\text{where } g_k = \frac{(N-k-s-1)(\hat{\sigma}_k^2 - \hat{\sigma}_{k+1}^2)}{\hat{\sigma}_{k+1}^2}$$

$$(k=0, \dots, p-1)$$

and $\hat{\sigma}_k^2$ = residual variance corresponding to the set of nonzero lags in $\hat{\Phi}_1(k)$

s = no. of functionally independent parameters in the estimated model (usually k)

(ii) for a given choice of α , find d_k such that

$$P(g_k^* > d_k) = \alpha$$

where $g_{k,1}^*$ is the maximum order statistic in a sequence of $p-k$ independent χ_1^2 random variables

(iii) choose the estimated order $\hat{g}$ such that

$$\min k : g_k \leq d_k, \quad 0 \leq k < p$$

$$\text{then } \hat{\Phi}_1 = \hat{\Phi}_1(g)$$

Thus $\hat{\Phi}_1$ may be estimated.

Example. to fit an MA subset model of order $g \leq q$ with nonzero lags $j_1, j_2, \dots, j_g$, it is sufficient to find estimates of $I(k)$ the inverse acf coefficients - which is equivalent to having estimates of the acf of an AR model of order $g \leq q$ with non-zero lags $(j_1, j_2, \dots, j_g)$. Exactly the same procedure is used as for the AR method except that all acf are replaced by the iacf.

APPENDIX A16 : TRANSFORMED SERIES

Often, before or after we apply Box-Jenkins models, a (preliminary) transformation is required. A family of transformations presented by Box and Cox (1964) is given by

$$x_{\lambda,t} = \begin{cases} \frac{x_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases}$$

where $\{x_t\}$ is a positive nonstationary series and λ = transformation parameter. $\{x_{\lambda,t}\}$ = transformed series.

In the case where some of the $x_t < 0$

$$x_{\lambda,t} = \begin{cases} \frac{(x_t + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \lambda_1 \neq 0 \\ \log(x_t + \lambda_2) & \lambda_1 = 0 \end{cases}$$

This family of transformations holds for $x_t > -\lambda_2 \forall t$ (see Box and Cox (1964)).

Ansley, Spivey and Wroblewski (1977) have designed an algorithm to find nonlinear LSEs of λ and the other model parameters. (Also see Vera and Guerrero (1981)).

Granger and Newbold (1976) considered the following problem : given a time series $\{X_t\}$ and a set of forecasts $x_t(l)$ (l = lead time), how do you find forecasts for a transformation of $\{X_t\}$? This is related to the next appendix, i.e. A17].

Suppose $x_{\lambda,t} = T(x_t) \forall t$ where

T = some well-behaved function, and $\{X_t\}$ is a stationary Gaussian series with mean μ and variance σ^2 (and acrf $\gamma(k)$).

Let $y_t = \frac{x_t}{\sigma}$

y_t and y_{t-k} are jointly distributed as a bivariate normal with mean zero, unit variances and correlations $\rho(k)$.

Consider $x_{\lambda,t} = T(y_t)$ with the expansion

$$T(Z) = \sum_{j=0}^M \alpha_j H_j(Z) \quad \text{where}$$

$H_j(Z)$ = Hermite polynomial such that

$$H_n(x) = e^{x^2/2} \left(-\frac{d}{dx} \right)^n e^{-x^2/2}$$

or $H_n(x) = (-1)^n f^{(n)}(x)/f(x)$ where f is a std. normal pdf.

Explicitly,

$$H_n(x) = n! \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{(-1)^m}{2^m m!} (n-2m)! \cdot x^{n-2m}$$

and $\lfloor k \rfloor$ = largest integer $\leq k$.

Then $E(y_t x_{\lambda,t-\tau}) = \alpha_1 \rho(\tau)$

$$\text{Cov}(y_t x_{\lambda,t-\tau}) = \alpha_1 \rho(\tau) \sigma^2$$

$$\text{Cov}(x_{\lambda,t} x_{\lambda,t-\tau}) = \sum_{j=1}^M \alpha_j^2 j! \rho(\tau)^j$$

$$V(x_{\lambda,t}) = \sum_{j=1}^M \alpha_j^2 j!$$

so the varf of $x_{\lambda,t}$ is $V_{\lambda} = \sum_{j=1}^M \alpha_j^2 j! \rho(1)^j / \sum_{j=1}^M j!$

If $\rho(1) = 0$ for $t = 0$

then $\rho(t) = 0$ for $m = 1$ (i.e. in the expansion i.t.o.

Hermite polynomials).

If at time n , we have $x_n(h)$ = forecast of x_{n+h}

let

$$\begin{aligned} x_{\lambda, n+h} &= T(y_{n+h}) \\ &= T\left(\frac{y_{n+h} - \mu}{\sigma}\right) \end{aligned}$$

Granger and Newbold (1976) claim that

$$x_{\lambda, n}(h) = T\left(\frac{x_n(h) - \mu}{\sigma}\right) \text{ is invariably the most}$$

practical forecast of $x_{\lambda, n+h}$

APPENDIX A17 : FORECASTING

Suppose the data are generated by the model $\Phi(B)x_t = \Theta(B)a_t$, i.e. an ARMA (p,q) model (and $\{x_t\}$ is a stationary process).

x_{t+l} can be represented in three ways.

(a) In terms of a difference equation

$$x_{t+l} = \phi_1 x_{t+l-1} + \dots + \phi_p x_{t+l-p} - \theta_1 a_{t+l-1} - \dots - \theta_q a_{t+l-q} + a_{t+l}$$

(b) In terms of values (past and present) of the white noise $\{a_t\}$

$$\text{i.e. } x_{t+l} = \sum_{j=-\infty}^{t+l} \psi_{t+l-j} a_j = \sum_{j=0}^{\infty} \psi_{t+l-j} a_{t+l-j} \quad (\text{a17.1})$$

and $\psi_0 = 1$

(c) In terms of previous x values and noise i.e.

$$x_{t+l} = \sum_{j=1}^l \psi_{t+l-j} x_{t+l-j} + a_{t+l} \quad (\text{a17.2})$$

[see Box and Jenkins (1976), pg. 94; also see section 6.6 for the relationship between the ψ weights and $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$.]
Suppose we are at origin t and wish to forecast $x_{t+1}, x_{t+2}, \dots$
i.e. $\hat{x}_t(1), \hat{x}_t(2)$, etc. and suppose the best forecast of x_{t+l} at time t , using (a17.1), is

$$\hat{x}_t(l) = \alpha_l^* a_t + \alpha_{l+1}^* a_{t-1} + \dots \text{some } \alpha_j^*, \quad j = l, l+1, \dots$$

The MSE of the forecast is

$$E(x_{t+l} - \hat{x}_t(l))^2 = (1 + \alpha_l^2 + \dots + \alpha_{l-1}^2) \sigma_a^2 + \sum_{j=0}^{\infty} (\alpha_{l+j} - \alpha_{l+j}^*)^2 \sigma_a^2$$

(using (a17.1)).

This is minimized when $\alpha_{l+j}^* = \alpha_{l+j}$

$$\therefore x_{t+l} = (a_{t+l} + \alpha_1^* a_{t+l-1} + \dots + \alpha_{l-1}^* a_{t+1}) + (\alpha_l^* a_t + \alpha_{l+1}^* a_{t-1} + \dots)$$

$$= e_t(l) + \bar{x}_t(l) \quad \text{say} \quad (a17.3)$$

Note that (i) $e_t(1) = x_{t+1} - \bar{x}_t(1) = a_{t+1}$

= one step ahead forecast error (which is
uncorrelated with any other one step ahead
forecast error)

$$(ii) \quad \bar{x}_t(l) = \sum_{j=1}^l E(x_{t+j} / x_t, x_{t-1}, \dots) \\ + E(a_{t+j} / a_t, a_{t-1}, \dots)$$

(by using (a17.2))

e.g. if $x_{t+l} = \pi_1 x_{t+l-1} + \pi_2 x_{t+l-2} + a_{t+l}$

then

$$\bar{x}_t(1) = \pi_1 x_t + \pi_2 x_{t-1}$$

$$\bar{x}_t(2) = \pi_1 \bar{x}_t(1) + \pi_2 x_t$$

$$\bar{x}_t(l) = \pi_1 \bar{x}_t(l-1) + \pi_2 \bar{x}_t(l-2) \quad l = 3, 4, 5, \dots$$

e.g. if $x_{t+l} = \phi_1 x_{t+l-1} + \phi_2 x_{t+l-2} + a_{t+l} - \theta_1 a_{t+l-1} - \theta_2 a_{t+l-2}$

$$\hat{x}_t(1) = \phi_1 x_t + \phi_2 x_{t-1} - \theta_1 a_t - \theta_2 a_{t-1}$$

$$\bar{x}_t(2) = \phi_1 \hat{x}_t(1) + \phi_2 x_t - \theta_2 a_t$$

$$\bar{x}_t(l) = \phi_1 \bar{x}_t(l-1) + \phi_2 \hat{x}_t(l-2) \quad l = 3, 4, \dots$$

Suppose forecasts at lead times $1, 2, \dots, L$ are required. If we are
to use an updating procedure, $\bar{x}_t, \dots, \bar{x}_{t-1}$ are required. These
can be obtained by solving

$$\Phi(B)\bar{x}(B) = \Theta(B)$$

$$\text{i.e. } (1 - \phi_1 B - \dots - \phi_p B^p)(1 + \pi_1 B + \pi_2 B^2 + \dots) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

This results in

$$\begin{aligned} \alpha_1 &= \phi_1 - \theta_1 & \alpha_0 &= 1 \\ \alpha_2 &= \phi_1 \alpha_1 + \phi_2 - \theta_2 & \alpha_j &= 0 \quad j < 0 \\ & & \theta_j &= 0 \quad j > q \\ \alpha_j &= \phi_1 \alpha_{j-1} + \dots + \phi_p \alpha_{j-p} - \theta_j \end{aligned}$$

Knowledge of $\alpha_1, \dots, \alpha_q$ and $\phi_1, \dots, \phi_p$ implies knowledge of the α 's.

Since $x_{t+1}(l) = \alpha_{l+1} a_t + \dots$

and $\hat{x}_t(l+1) = \alpha_{l+1} a_t + \alpha_{l+2} a_{t-1} + \dots$

then $\hat{x}_{t+1}(l) = \hat{x}_t(l+1) + \alpha_l a_{t+1}$

By (A17.3), $e_t(l) = x_{t+1}(l) - \hat{x}_t(l)$

so

$$V(e_t(l)) = (1 + \alpha_1^2 + \alpha_2^2 + \dots + \alpha_{l-1}^2) \sigma_a^2$$

If the a_t 's are independent and normally distributed (with mean zero),

then the approximate 1 - α probability limits for x_{t+1} are

$$x_t(l) \pm Z_{\frac{1-\alpha}{2}} \left(1 + \sum_{j=1}^{l-1} \alpha_j^2 \right)^{\frac{1}{2}} s_a$$

where s_a = estimate of σ_a

$Z_{\frac{1-\alpha}{2}}$ = deviate exceeded by a proportion $\frac{1-\alpha}{2}$ in $N(0,1)$ distribution.

E.g. suppose $\{X_t\}$ is generated by an ARMA(1,1) model ($\{X_t\}$ is stationary) i.e.

$$(1 - \phi B)x_t = (1 - \theta B)a_t$$

then

$$\hat{x}_t(l) = \phi \hat{x}_t(l-1) + \theta a_t$$

$$x_t(l) = \phi x_t(l-1) \quad l \geq 2$$

and so

$$\bar{x}_{t+1}(\ell) = x_t(\ell+1) + (\phi - \theta) \phi^{-1} a_{t+1}$$

Also, since $\alpha_j = (\phi - \theta) \theta^{j-1} \quad j = 1, 2, \dots$

and $\beta_j = (\phi - \theta) \phi^{j-1} \quad j = 1, 2, \dots$

$$V(e_t(\ell)) = \frac{1}{a} \left\{ 1 + (\phi - \theta)^2 \left(\frac{1 - \phi^{2\ell-2}}{1 - \phi^2} \right) \right\}$$

Correlation between forecast errors

Since $e_t(\ell) = a_{t+1} + \alpha_1 a_{t+2-1} + \dots + \alpha_{\ell-1} a_{t+\ell}$

and $e_{t-j}(\ell) = a_{t-j+1} + \alpha_1 a_{t-j+2-1} + \dots + \alpha_{\ell-1} a_{t-j+\ell} \quad j < \ell$

then

$$\text{Cov}(e_t(\ell), e_{t-j}(\ell)) = \frac{1}{a} \sum_{i=0}^{\ell-1} \alpha_i \alpha_{i-j} \quad (\alpha_0 = 1)$$

* acvf of the forecast errors

Similarly, the acf's are

$$\rho(e_t(\ell), e_{t-j}(\ell)) = \begin{cases} \frac{\sum_{i=0}^{\ell-1} \alpha_i \alpha_{i-j}}{\sum_{i=0}^{\ell-1} \alpha_i^2} & 0 \leq j < \ell \\ 0 & j \geq \ell \end{cases}$$

In a similar manner it can be shown that

$$\rho(e_t(\ell), e_{t+\ell+j}) = \frac{\sum_{i=0}^{\ell-1} \alpha_i \alpha_{i+j}}{\sum_{h=0}^{\ell-1} \alpha_h^2} \quad 2 \leq j$$

[for more detail see Box and Jenkins, 1976, chapter 5]

APPENDIX A18 : DESCRIPTION OF COMPUTER PROGRAMS

Due to the large amount of computation needed for the effective analysis of a time-series, it is generally necessary to develop a suite of computer programs. Such a suite was developed during the preparation of this dissertation, and a description of the programs is given in this appendix

All programs were written in Fortran IV in conversational mode, and implemented on the IBM SYSTEM/370.

The programs used were

- | | |
|-------------|-------------|
| (1) PGRAM | (12) PITEST |
| (2) HIDPER | (13) QBPLOT |
| (3) MINVAR | |
| (4) AUTOAR | |
| (5) BJID | |
| (6) BJES | |
| (7) BJFO | |
| (8) AUTO | |
| (9) SPEC | |
| (10) TRANS | |
| (11) BUYBAL | |

In the foreseeable future it is intended that all these programs will be condensed into one package and will allow the user either no control (i.e. fully-automatic) or selective control.

(1) PGRAM (Periodogram computation)

This program is adapted from Bloomfield (1976, pg. 114) and computes the discrete Fourier transform of a time series, and hence the periodogram. The transform is computed by using the Tukey-Cooley fast Fourier algorithm (Cooley and Tukey (1965)).

This basically just involves extending the time-series by zeroes to the next power of 2 above N (the length of the series) and calculating the discrete Fourier transform $J(w)$ for $w_j = 2\pi j/N'$ (N' = next power of 2 above N).

Input (besides the time-series)

- (a) length of time series (i.e. N)
- (b) ordinates of the section of data of interest
- (c) $K = 0$ or 1 - the degree of the LS polynomial to be removed from section of interest
- (d) the degree of tapering required (see section 5.4.2)

Only the section of data given by (b) is used. The next power of 2 (i.e. N') is computed automatically.

Output

- (a) the degree of the polynomial removed
- (b) the degree of tapering
- (c) a plot of $\log_{10} \bar{R}(w)^2$

where

$$\bar{R}(w)^2 = \bar{A}(w)^2 + \bar{B}(w)^2$$

$$\bar{A}(w) = \frac{1}{N'} \sum (x_t - \bar{x}) \cos(wt)$$

$$\bar{B}(w) = \frac{1}{N'} \sum (x_t - \bar{x}) \sin(wt)$$

$\bar{R}(w)^2$ differs from $I(w)$, the periodogram by a factor of $8^{-1} N'$. It is more useful in harmonic analysis since it represents the LS estimate of the squared amplitude of a sinusoid of frequency w .

$\bar{R}(w)^2$ is plotted against j where $w_j = 2\pi j/N'$ and

$$j = 1, \dots, \frac{N'}{2} \quad (\text{i.e. } 0 < w_j \leq \pi)$$

The logarithm is used due to the variation in the order of magnitude of the periodogram between different frequencies. If only the periodogram is plotted (i.e. not the $\log_{10}$) so that the largest value is on the graph, then details at higher frequencies may be obscured (i.e. the $\log_{10}$ plot just reduces the range).

(2) HIDPER (Search for hidden periodicities)

This program is adapted from Bloomfield (1976, pg. 33) and fits the model

$$x_t = \sum_{k=1}^m \{ a_k \cos(w_k t) + b_k \sin(w_k t) \} + \epsilon_t$$

to the data.

Input (besides the time-series)

- (a) length of the time-series
- (b) ordinates of the section of the data of interest
- (c) m (the no. of frequencies)
- (d) initial estimates of $w_1, w_2, \dots, w_m$ (gauged from PGRAM)
- (e) initial estimates of $a_1, \dots, a_m$ and $b_1, \dots, b_m$ (usually zero)
- (f) $L = \begin{cases} 0 & \text{if exact LS method is required} \\ 1 & \text{if approximate method is required} \end{cases}$

(see section 3.2)

The program then computes either exact LS estimates or approximate LS estimates of w, a, b using subprogram LOCALM from Brent(1973) [this finds a local minimum of a multi-parameter function f in interval (u,v)].

Output

- (a) whether exact/approximate LS method is used
- (b) initial values of frequencies and sine and cosine coefficients
- (c) optimum values of frequencies and sine cosine coefficients
- (d) fitted constant (i.e. $\hat{\mu}$)
- (e) the (exact or approximate) residual sum of squares
- (3) MINVAR (Inducing stationarity)

This program is based on the procedure used by Hill and Woodworth (1980) for determining d and D - the degrees of nonseasonal and seasonal differencing resp.

i.e. calculate

$$\min_{d,D,S} \left\{ \sigma_Z^2(d,D,S) \right\}$$

$$\text{where } Z_t = \frac{d,D}{S} x_t \quad \forall t$$

S = seasonal base

Input (besides the time-series)

- (a) maximum values of d,D,S
- (b) length of the time-series
- (c) ordinates of the section of interest.

The program now tries all combinations of values (for d,D and S). The variances of the differenced series are sorted into ascending order.

Output

- (a) the variances of the differenced series (in ascending order) with associated d,D,S, as well as the coefficient of variation associated with each series (this just provides another criterion for determining (d,D,S)).

(4) AUTOAR (Automatic AR model fitting)

This program fits successive autoregressive models up to some specified maximum order i.e. AR(1), AR(2), ..., AR(p).

Input (besides the time-series)

- (a) length of series
- (b) ordinates of section of interest
- (c) degree of seasonal differencing and seasonal base
- (d) degree of nonseasonal differencing
- (e) maximum value of p (i.e. the highest order AR model required).

If differencing is required, the program first computes this. Then the acrf (and acvf) is calculated upto a lag greater than the p given in (e). The Yule-Walker equations are then solved i.e. solve $\underline{\hat{\phi}} = \underline{R}^{-1} \underline{r}$

where $\underline{\hat{\phi}} = (\hat{\phi}_1, \dots, \hat{\phi}_k)'$

$\underline{r} = (r'_1, \dots, r'_k)'$

$$\underline{R} = \begin{pmatrix} 1 & r'_1 & \dots & r'_{k-1} \\ r'_1 & 1 & & r'_{k-2} \\ & & \ddots & \\ r'_{k-1} & r'_{k-2} & & 1 \end{pmatrix}$$

for each $k = 1, 2, \dots, p$

Calculate

$$\hat{\sigma}_{a,k}^2 = c'_0 - \sum_{i=1}^k \hat{\phi}_i c'_i \quad \text{for each } k$$

Output

- (a) the parameter estimates for each AR model
- (b) the associated white noise variance estimate $\hat{\sigma}_{a,k}^2$

(c) AIC and CAT (as defined in section 7.4.2.) for each AR model - this enables the user to decide which AR model gives the best fit.

(5) BJID (Box-Jenkins identification)

This program is a modification of that described in Box and Jenkins (1976), pg. 496 and helps in the initial stages of model identification. It consists of two sections - an editing phase and an identification phase. The editing phase allows one to transform the data, replace certain values, plot the data etc.

The identification phase provides plots of the *acrf*, *pacrf*, *iacrf* and *ipacrf*.

Input (besides the time-series)

- (a) length of the series
- (b) ordinates of section of interest
- (c) any changes to be made in the editing phase
- (d) degree of seasonal differencing and seasonal base
- (e) degree of nonseasonal differencing
- (f) maximum lag of the *acrf*
- (g) maximum lag of the *pacrf*
- (h) 4 values (between 10 and 50) which denote the maximum AR model to be fitted and hence the *iacrf* and *ipacrf* (see appendix A12).

Firstly, any desired changes to be made by the editing phase are made. If differencing is required, this is carried out.

Using the value given in (f), the Yule-Walker equations are solved (as in program (3), AUTOAR).

Using the technique described in appendix A10, and the value in (g), the pacrf is plotted. A technique, described by Chatfield(1979), for estimating the iacrf, is given in appendix A12. This technique is used for each of the 4 values given in (h). To find the ipacrf, the 'inverse' Yule-Walker equations must be solved i.e.

$$\begin{bmatrix} 1 & \rho I(1) & \dots & \rho I(k-1) \\ \rho I(1) & 1 & & \\ \vdots & & \ddots & \\ \rho I(k-1) & \rho I(k-2) & & 1 \end{bmatrix} \begin{bmatrix} \rho I_{k1} \\ \rho I_{k2} \\ \vdots \\ \rho I_{kk} \end{bmatrix} = \begin{bmatrix} \rho I(1) \\ \vdots \\ \rho I(k) \end{bmatrix}$$

where $\rho I(k)$ is one of the 4 values given in (h).

Replace $\rho I(2)$ by $\rho I(1)$ and solve for ρI_{kj} ($j = 1, \dots, k$)

ρI_{kk} is called the sample ipacrf.

The user is left to decide which of the 4 iacrf/ipacrf representations is appropriate.

The user may now also wish to 'fit' some BJ models and the only input required are the values for p and q in the desired ARMA (p,q) process. Model parameters are estimated by

(i) if $p > 0$, solving

$$A \underline{\phi} = C$$

where

$$A_{ij} = c_{|q+1-j|}^i$$

$$C = (c_1^1, \dots, c_p^1)'$$

and

(ii) if $q = 0$, use the procedure described in appendix A9.

(iii) the overall constant in the model is estimated by

$$\hat{\delta} = \begin{cases} \bar{x}(1 - \sum_{i=1}^p \hat{b}_i) & p > 0 \\ \bar{x} & p = 0 \end{cases}$$

and the white noise variance by

$$\hat{\sigma}_0^2 = \begin{cases} \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2 & q = 0 \\ \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2 - \sum_{i=1}^p \hat{b}_i^2 \hat{\sigma}_0^2 & q = 1 \end{cases}$$

($\hat{\sigma}_0^2$ from appendix A9 and $\bar{x}$ the mean of the series)

Output

- (a) mean of (differenced) series
- (b) variance of (differenced) series
- (c) degrees of differencing
- (d) acrf
- (e) pacrf
- (f) iacrf and ipacrf for each of the 4 values desired
- (g) estimates of model parameters (if so required), overall constant in the model and white noise variance.

This program is 'recursive' i.e. it allows several successive runs of the program for different editing phases, degrees of differencing, models etc.

The initial estimates (of model parameters) required in the next program are given in (g).

[Note that the Box-Jenkins programs 1 and 2 (Box and Jenkins (1976) pg. 496+) do not calculate the iacrf or ipacrf.]

(6) BJES (Box-Jenkins model estimation).

This program is an adaptation of the program (3) described in Box and Jenkins (1976), pg. 500, and uses the initial estimates of the nonseasonal parameters from program (5) (BJID) (and if seasonal parameters are required then a technique described by Box and Jenkins (1976), chapter 9, provides initial estimates of the seasonal parameters) this program calculates LSE's for μ , ϕ , θ , ϕ^s , θ^s and ψ the seasonal parameter vectors) in the model

$$\phi(B)(x_t - \mu) = \theta(B)a_t$$

$$\text{or } \phi(B)\phi(B^s)(x_t - \mu) = \theta(B)\theta^s(B^s)a_t$$

where $\{x_t\}$ is a stationary series.

Input (besides the time series)

- (a) length of series, N
- (b) ordinates of section of interest
- (c) degrees of differencing
- (d) initial estimates of all parameters (i.e. p, q, P, Q, S and hence parameter values)

The program calculates backforecasts of the x_t 's (see section 7.3.1) and hence

$$S(\mu, \hat{\phi}, \hat{\theta}, \hat{\phi}^s, \hat{\theta}^s) = \sum_{t=k}^N \hat{a}_t^2.$$

The values which minimize S are obtained by the constrained optimization technique of Marquardt (1963) (which is described fully in Box and Jenkins, 1976, pg. 504). These values are the LSE's.

An estimate of the residual variance is then obtained using these LSE's, and the covariance matrix of the (optimum)

estimates is given by

$$\hat{\Sigma} = (X'X)^{-1} \hat{\sigma}_a^2$$

(where X is the regression matrix in the linearized model). The standard errors of the estimates are given by

$$\hat{\sigma}_1 = \sqrt{\hat{\sigma}_{11}} \quad (\hat{\Sigma} = (\hat{\sigma}_{ij}))$$

An estimate of the overall constant in the model is given by

$$\hat{\delta} = \bar{u}L$$

where

$$L = (1 - \sum_{i=1}^p \hat{\phi}_i)$$

(or if the model is seasonal

$$L = (1 - \sum_{i=1}^p \hat{\phi}_i)(1 - \sum_{j=1}^q \hat{\phi}_j))$$

The residuals a_t corresponding to the LSE's are calculated and hence the acrf of the residuals (for some set lag). The mean and variance of the residuals are also given. The acrf of the residuals is plotted as well as 1% significance levels (using the technique of Hipel et al (1977)).

Output

- (a) initial and final estimates of the parameters
- (b) white noise variance estimate
- (c) standard errors and covariance matrix of the estimates
- (d) $\hat{\delta}$, the overall constant
- (e) residual acrf (a plot) and 1% significance levels

(7) BJFO (Box-Jenkins forecasting)

Using the LSE's from program 6, (BJES), this program forecasts future values of the series $\{X_t\}$ using the techniques described

in appendix A17. This program is described in Box and Jenkins (1976), pg. 505.

(8) AUTO (Subset autoregression technique)

Using the procedure of McClave (1975, 1978, 1978a) - see appendix A15 - the user specifies a maximum order AR model and the program then finds the best-fitting AR model (using AIC and CAT criteria). Having found the 'optimum' AR order, the program now finds the best subset autoregression for each different subset size (again using the above-mentioned criteria). For each 'full' AR model, parameters are given, and the parameters are also given for each best subset in the optimum 'full' model. For more details, see appendix A15. (Also see Gabr and Subba Rao (1981)).

(9) SPEC (A spectrum estimate)

This program calculates the sample acrf (acvf) and based on this, computes an estimate of the nsdf (see equation (4.9), section 4.2 - divide both sides by $\frac{1}{N}$).

Input (besides the time series)

- (a) length of time series, N
- (b) ordinates of section of interest
- (c) the number of spectral windows to be used - 10 are offered - and the specific windows to be used.

The program automatically computes the sample acrf (up to lag $0.75 \times N$). The ten windows offered are:

Truncated periodogram, Bartlett A, Bartlett B, Daniell, Tukey-Hamming, Tukey-Hanning, Parzen A, Parzen B, Bartlett-Priestley and Tukey-Parzen. The truncation points M (for each window)

are 5%, 10%, 20%, 40%, 50%, and 75% of N (i.e. for each window, six spectrum estimates are given). The spectra are evaluated at w_j where $0 \leq w_j \leq \pi$ (and $j = 0, 1, \dots, [\frac{N}{2}]$) and $w_j = 2\pi j/N$.

Output (for each desired window)

- (a) 6 spectral estimates (each at the associated truncation point)

(10) TRANS (Box-Cox transformation)

In appendix A16, a family of transformations is presented.

These transformations take the form

$$x_{\lambda,t} = \begin{cases} \frac{x_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln x_t & \lambda = 0 \end{cases}$$

(assuming all the $x_t^s > 0$).

What value should λ take?

Vera and Guerrero (1981) have proposed a method for finding λ in the following class of transformation (a special case of the above).

$$x_{\lambda,t} = \begin{cases} x_t^{1-\lambda} & \lambda \neq 1 \\ \ln x_t & \lambda = 1 \end{cases}$$

Divide the data into subseries of some set length. For each subset calculate the mean Z , standard deviation S and the ratio S^*/Z^* (for each $\lambda \in [a,b]$ say) where $S^* = 1000 \times S$ and $Z^* = (1000 \times Z)^\lambda$.

Supposing there are M subseries of length L , the program calculates the mean, $\bar{Z}_\lambda$, of the S^*/Z^* (for each λ), as well as the standard deviation, S_λ , of the S^*/Z^* . The coefficient of variation is defined by

$$CV = S_\lambda / \bar{Z}_\lambda$$

The value of λ giving a minimum value to CV is defined to be the transformation parameter. Successive runs with different lengths of subseries are possible.

Input (besides the time-series)

- (a) length of series, N
- (b) ordinates of section of interest
- (c) length of the subseries, L

The program divides the data into an integer number of subseries of length L and calculates Z^* and S^* for each subseries.

- (d) lower bound, upper bound and increment of λ

Output

- (a) the coefficient of variation and associated λ (and obviously the length of the subseries)

(11) BUYBAL (Buys-Ballot test for period)

A procedure for testing for an assumed period is described in appendix A6. This program performs that procedure.

Input (besides the time-series)

- (a) length of the time-series
- (b) ordinates of section of interest
- (c) minimum and maximum values of the 'assumed' periodicity.

Output

- (a) a plot of the values τ (see appendix A6) for each (integer) value between the minimum and maximum given in
- (c) above

(12) PITEST

This program tests for the existence of the discrete spectrum using Priestley's $P(\lambda)$ test as described on pg.216(or see Priestley (1962a,b)) using the Bartlett (a) window (see pg. 77).

Input (besides the time-series)

- (a) length of the series, N
- (b) ordinates of section of interest
- (c) lower truncation point
- (d) upper truncation point
- (e) lag at which the peak occurs

Output (after the entry of (a) - (d))

- (a) a plot of the acrf
- (b) a listing of the acr's
- (c) a plot of $P(\cdot)$
- (d) a listing of $P(\lambda)$ ordinates

Upon entry of (e) (in the input), the peak is tested for significance.

(13) QBPLOT

This program plots the quantile-box plot of the data using the technique described in section 9.2 and Parzen (1979).

[Note: In each program, if the ordinates of the section of interest are not 1 and N (representing x_1 and x_N) then the program considers the series between these ordinates to be the (new) series and the length of this section the new N].

Listings of two of the programs (SPEC (9) and PITEST (12)) follow and serve as exemplars of the suite.

PROGRAM SPEC

```

C
C
C
PROGRAM TO CALCULATE ACRF AND PERFORM A SPECTRAL ANALYSIS
OF THE DATA

DOUBLE PRECISION X(360),YY(360),Y(360),R(150,2)
REAL*8 A,V
INTEGER JUMP,I(6),N2(10)
LOGICAL*1 MAP(3)
COMMON JUMP,MAP
EXTERNAL W1,W2,W3,W4,W5,W6,W7,W8,W9,W10
WRITE(6,1)
1  FORMAT(T2,'ENTER NC. OF OBSERVATIONS')
   READ(9,*)N
   READ(5,*)(X(I),I=1,N)
   READ(5,66)(MAP(I),I=1,3)
66  FORMAT(3A1)
   WRITE(6,67)
67  FORMAT(T2,'ENTER ORDINATES OF SECTION OF DATA TO',
C' BE ANALYSED')
   READ(9,*)NO1,NO2
   J=0
   DO 68 I=NO1,NO2
   J=J+1
68  YY(J)=X(I)
   N=NO2-NO1+1
   PI=3.141593
   A=0.0
   V=0.0
   CON=0.0
   L(1)=(0.05*N)+0.4999
   L(2)=(0.10*N)+0.4999
   L(3)=(0.20*N)+0.4999
   L(4)=(0.40*N)+0.4999
   L(5)=(0.50*N)+0.4999
   L(6)=(0.75*N)+0.4999
   DO 2 J=1,N
   Y(J)=YY(J)
   IF(Y(J).LT.CON)CON=Y(J)
2  CONTINUE
   IF(CON.GE.0.0)GOTO 3
   DO 3 J=1,N
   Y(J)=Y(J)-CON
3  CONTINUE
   DO 4 J=1,N
   A=A+Y(J)
   A=A/FLOAT(N)
   DO 5 J=1,N
   Y(J)=Y(J)-A
5  CONTINUE
   DO 6 J=1,N
   V=V+Y(J)*Y(J)
   V=V/FLOAT(N)
   L6=L(6)
   DO 7 J=1,L6
   R(J,2)=0.0
   NJ=N-J
   DO 8 I=1,NJ
8  R(J,2)=R(J,2)+Y(I)*Y(I+J)
   R(J,2)=R(J,2)/FLOAT(N)
7  R(J,1)=R(J,2)/V
   WRITE(6,15)
15  FORMAT(/T2,'THE FOLLOWING ARE THE AVAILABLE WINDOWS',///)
   WRITE(6,9)
9  FORMAT(T5,'(1) TRUNCATED PERIODOGRAM',/,T5,'(2) BARTLETT A',/,
CT5,'(3) BARTLETT B',/,T5,'(4) DANIELL',/,T5,'(5) TUKEY-HAMMING',
C/,T5,'(6) TUKEY-HANNING',/,T5,'(7) PARZEN A',/,T5,
C'(8) PARZEN B',/,T5,'(9) BARTLETT-PRIESTLEY',/,T4,
C'(10) TUKEY-PARZEN',///)
   WRITE(6,10)
10  FORMAT(T2,'IF FURTHER DETAILS OF THE ABOVE WINDOWS ARE REQUIRED',
C' ENTER 1, OTHERWISE 0',/)

```

```

      READ(9,*)N1
      IF(N1.EQ.1)CALL WINDOW
      WRITE(6,11)
11  FORMAT(/T4,'ENTER THE NO. OF SPECTRAL ESTIMATES REQUIRED',
C' I.E. THE NO. OF DIFFERENT WINDOWS TO BE USED',/)
      READ(9,*)N1
      WRITE(6,12)
12  FORMAT(T2,'ENTER THE NO'S OF THE ASSOCIATED WINDOWS ON ONE LINE')
      READ(9,*)(N2(I),I=1,N1)
      WRITE(6,750)
750 FORMAT(T2,'ENTER THE "JUMP" TO BE USED IN THE PLOT OF H(W) I.E.',
C' 1 IF EVERY POINT MUST BE PLOTTED, 2 IF EVERY 2ND POINT ETC')
      READ(9,*)JUMP
      DO 13 I=1,N1
      IF(N2(I).NE.1)GOTO 59
C
C   TRUNCATED PERIODOGRAM
C
      WRITE(7,21)
21  FORMAT(/T2,'TRUNCATED PERIODOGRAM',/T2,21('-'),/T2,21('-'),/)
      DO 41 I1=1,6
41  CALL WIND(L(I1),W1,R,V,N)
      GOTO 13
59  IF(N2(I).NE.2)GOTO 51
C
C   BARTLETT A
C
      WRITE(7,22)
22  FORMAT(/T2,'BARTLETT A',/T2,10('-'),/T2,10('-'),/)
      DO 42 I1=1,6
42  CALL WIND(L(I1),W2,R,V,N)
      GOTO 13
51  IF(N2(I).NE.3)GOTO 52
C
C   BARTLETT B
C
      WRITE(7,23)
23  FORMAT(/T2,'BARTLETT B',/T2,10('-'),/T2,10('-'),/)
      DO 43 I1=1,6
43  CALL WIND(L(I1),W3,R,V,N)
      GOTO 13
52  IF(N2(I).NE.4)GOTO 53
C
C   DANIELL
C
      WRITE(7,24)
24  FORMAT(/T2,'DANIELL',/T2,7('-'),/T2,7('-'),/)
      DO 44 I1=1,6
44  CALL WIND(L(I1),W4,R,V,N)
      GOTO 13
53  IF(N2(I).NE.5)GOTO 54
C
C   TUKEY-HAMMING
C
      WRITE(7,25)
25  FORMAT(/T2,'TUKEY-HAMMING',/T2,13('-'),/T2,13('-'),/)
      DO 45 I1=1,6
45  CALL WIND(L(I1),W5,R,V,N)
      GOTO 13
54  IF(N2(I).NE.6)GOTO 55
C
C   TUKEY-HANNING
C
      WRITE(7,26)
26  FORMAT(/T2,'TUKEY-HANNING',/T2,13('-'),/T2,13('-'),/)
      DO 46 I1=1,6
46  CALL WIND(L(I1),W6,R,V,N)
      GOTO 13
55  IF(N2(I).NE.7)GOTO 56
C

```

```

C      PARZEN A
C
      WRITE(7,27)
27     FORMAT(/T2,'PARZEN A',/T2,8(' '),/T2,8(' '),/)
      DO 47 I1=1,6
47     CALL WIND(L(I1),W7,R,V,N)
      GOTO 13
56     IF(N2(I).NE.8)GOTO 57
C
C      PARZEN B
C
      WRITE(7,28)
28     FORMAT(/T2,'PARZEN B',/T2,8(' '),/T2,8(' '),/)
      DO 48 I1=1,6
48     CALL WIND(L(I1),W8,R,V,N)
      GOTO 13
57     IF(N2(I).NE.9)GOTO 58
C
C      BARTLETT-PRIESTLEY
C
      WRITE(7,29)
29     FORMAT(/T2,'BARTLETT-PRIESTLEY',/T2,18(' '),/T2,18(' '),/)
      DO 49 I1=1,6
49     CALL WIND(L(I1),W9,R,V,N)
      GOTO 13
58     IF(N2(I).NE.10)GOTO 13
C
C      TUKEY-PARZEN
C
      WRITE(7,30)
30     FORMAT(/T2,'TUKEY-PARZEN',/T2,12(' '),/T2,12(' '),/)
      DO 50 I1=1,6
50     CALL WIND(L(I1),W10,R,V,N)
13     CONTINUE
      STOP
      END
      SUBROUTINE WIND(M,W,R,V,N)
      DOUBLE PRECISION R(150),W(150),OM1(180)
      LOGICAL*1 MAP(3)
      INTEGER JUMP
      COMMON JUMP,MAP
      PI=3.141593
      NM=N/2.
      NM=NM+1
      DO 1 II=1,M
1      WW(II)=W(II,M,N)
      DO 2 I=1,NM
      III=II-1
      OM1(II)=0.0
      OMEG=2.*PI*FLOAT(III)/FLOAT(N)
      DO 3 JJ=1,M
      OMEGA=OMEG*FLOAT(JJ)
3      OM1(II)=OM1(II)+WW(JJ)*R(JJ,1)*COS(OMEGA)
      OM1(II)=2.*OM1(II)+1.
2      OM1(II)=OM1(II)/(2.*PI)
      WRITE(7,6)M
      WRITE(7,4)
4      FORMAT(T6,'-I',10('----I'),'--')
      DO 5 JJ=1,NM,JUMP
      LL=DINT(OM1(JJ)*51.)
6      FORMAT(/T2,'TRUNC. PT. =',I4,/)
      VL=OM1(JJ)
      IF(LL.GT.51)LL=51
      IF(LL.LE.1)LL=1
5      CALL PLOT(LL,JJ,VL)
      RETURN
      END
      SUBROUTINE PLOT(LL,J,VL)
      LOGICAL*1 MAP(3),HEAD(51)
      COMMON JUMP,MAP

```

```

      JJJ=J-1
      DO 10 K=1,51
10     HEAD(K)=MAP(1)
        JA=LL
        HEAD(26)=MAP(3)
        HEAD(JA)=MAP(2)
        WRITE(7,11)JJJ,(HEAD(I),I=1,51),VL
11     FORMAT(T2,I3,1X,'I' 51A1,2X,F6.3)
        RETURN
      END

C
C     W1 TRUNCATED PERIODOGRAM
C
      FUNCTION W1(IJ,M,N)
        W1=1.
        RETURN
      END

C
C     W2 BARTLETT A
C
      FUNCTION W2(IJ,M,N)
        W2=1.-(FLOAT(IJ)/FLOAT(M))
        RETURN
      END

C
C     W3 BARTLETT B
C
      FUNCTION W3(IJ,M,N)
        W3=(1.-(FLOAT(IJ)/FLOAT(M)))*(1.-(FLOAT(IJ)/FLOAT(N)))
        RETURN
      END

C
C     W4 DANIELL
C
      FUNCTION W4(IJ,M,N)
        PI=3.141593
        W4=SIN(PI*FLOAT(IJ)/FLOAT(M))*PI*FLOAT(IJ)/FLOAT(M)
        RETURN
      END

C
C     W5 TUKEY-HAMMING
C
      FUNCTION W5(IJ,M,N)
        PI=3.141593
        W5=0.54+0.46*COS(PI*FLOAT(IJ)/FLOAT(M))
        RETURN
      END

C
C     W6 TUKEY-HANNING
C
      FUNCTION W6(IJ,M,N)
        PI=3.141593
        W6=0.5*(1.+COS(PI*FLOAT(IJ)/FLOAT(M)))
        RETURN
      END

C
C     W7 PARZEN A
C
      FUNCTION W7(IJ,M,N)
        W7=1.-(FLOAT(IJ)/FLOAT(M))**2
        RETURN
      END

C
C     W8 PARZEN B
C
      FUNCTION W8(IJ,M,N)
        MN=(M/2)+1
        IF(IJ.GT.MN)GOTO 1
        W8=1.-6.*((FLOAT(IJ)/FLOAT(M))**2)+6.*((FLOAT(IJ)/FLOAT(M))**3)
        GOTO 2

```

```

1      W8=2.*((1.-(FLOAT(IJ)/FLOAT(M))**.3.)
2      RETURN
      END
C
C      W9 BARTLETT-PRIESTLEY
C
      FUNCTION W9(IJ,M,N)
      PI=3.141593
      W9=(SIN(PI*FLOAT(IJ)/FLOAT(M)) PI*FLOAT(IJ)/FLOAT(M))
      W9=W9-COS(PI*FLOAT(IJ)/FLOAT(M))
      W9=W9*3.*(FLOAT(M)/PI*FLOAT(IJ))**.2
      RETURN
      END
C
C      W10 TUKEY-PARZEN
C
      FUNCTION W10(IJ,M,N)
      A=0.282
      PI=3.141593
      W10=1.-2.*A+2.*A*COS(PI*FLOAT(IJ)/FLOAT(M))
      RETURN
      END
      SUBROUTINE WINDOW
      WRITE(6,1)
1      FORMAT(T5,'(1) TRUNCATED PERIODOGRAM',//)
      WRITE(6,2)
2      FORMAT(T2,'W(I)=1      -(M+1)<I<(M+1)')
      WRITE(6,3)
3      FORMAT(/T5,'(2) BARTLETT A',//)
      WRITE(6,4)
4      FORMAT(T2,'W(I)= 1-(MOD(I)/M)      -(M+1)<I<(M+1)')
      WRITE(6,5)
5      FORMAT(/T5,'(3) BARTLETT B',//)
      WRITE(6,6)
6      FORMAT(T2,'W(I)=(1-(MOD(I)/M))(1-(MOD(I)/N))      -(M+1)<I<(M+1)')
      WRITE(6,7)
7      FORMAT(/T5,'(4) DANIELL',//)
      WRITE(6,8)
8      FORMAT(T2,'W(I)=SIN(PI*I/M)/(PI*I/M)      -(M+1)<I<(M+1)')
      WRITE(6,9)
9      FORMAT(/T5,'(5) TUKEY-HAMMING',//)
      WRITE(6,10)
10     FORMAT(T2,'W(I)=0.54+0.46*COS(PI*I/M)      -(M+1)<I<(M+1)')
      WRITE(6,11)
11     FORMAT(/T5,'(6) TUKEY-HANNING',//)
      WRITE(6,12)
12     FORMAT(T2,'W(I)=0.5*(1+COS(PI*I/M))      -(M+1)<I<(M+1)')
      WRITE(6,13)
13     FORMAT(/T5,'(7) PARZEN A',//)
      WRITE(6,14)
14     FORMAT(T2,'W(I)=1-(I/M)**2      -(M+1)<I<(M+1)')
      WRITE(6,15)
15     FORMAT(/T5,'(8) PARZEN B',//)
      WRITE(6,16)
16     FORMAT(T2,'W(I)= 1-6*(I/M)**2+6*(MOD(I)/M)**3      MOD(I)<(M/2)+1',
      C/T6,'= 2*(1-(MOD(I)/M))**3      OTHERWISE')
      WRITE(6,17)
17     FORMAT(/T5,'(9) BARTLETT-PRIESTLEY',//)
      WRITE(6,18)
18     FORMAT(T2,'W(I)=((SQRT(3)-PI/2)*(SIN(PI*I/M)/(PI*I/M))',
      C/'-COS(PI*I/M))      -(M+1)<I<(M+1)')
      WRITE(6,19)
19     FORMAT(/T4,'(10) TUKEY-PARZEN',//)
      WRITE(6,20)
20     FORMAT(T2,'W(I)=1-2A+2A*COS(PI*I/M)      -(M+1)<I<(M+1) A=0.282')
      RETURN
      END

```

A18.19

```

1  W8=2*((1.-(FLOAT(IJ)/FLOAT(M)))**3.)
2  RETURN
   END
C
C  W9 BARTLETT-PRIESTLEY
C
   FUNCTION W9(IJ,M,N)
   PI=3.141593
   W9=(SIN(PI*FLOAT(IJ)/FLOAT(M)))/PI*FLOAT(IJ)/FLOAT(M)
   W9=W9-COS(PI*FLOAT(IJ)/FLOAT(M))
   W9=W9*3.*(FLOAT(M)/PI*FLOAT(IJ))**2
   RETURN
   END

C
C  W10 TUKEY-PARZEN
C
   FUNCTION W10(IJ,M,N)
   A=0.282
   PI=3.141593
   W10=1.-2.*A+2.*A*COS(PI*FLOAT(IJ)/FLOAT(M))
   RETURN
   END

SUBROUTINE WINDOW
WRITE(6,1)
1  FORMAT(T3,'(1) UNMODIFIED PERIODOGRAM',//)
WRITE(6,2)
2  FORMAT(T2,'W(I)=1-(MOD(I)/M) - (M+1)<I<(M+1)')
WRITE(6,3)
3  FORMAT(T5,'(2) BARTLETT A',//)
WRITE(6,4)
4  FORMAT(T2,'W(I)=1-(MOD(I)/M) - (M+1)<I<(M+1)')
WRITE(6,5)
5  FORMAT(T5,'(3) BARTLETT B',//)
WRITE(6,6)
6  FORMAT(T2,'W(I)=(1-(MOD(I)/M))(1-(MOD(I)/N)) - (M+1)<I<(M+1)')
WRITE(6,7)
7  FORMAT(T5,'(4) DANIELL',//)
WRITE(6,8)
8  FORMAT(T2,'W(I)=SIN(PI*I/M)/(PI*I/M) - (M+1)<I<(M+1)')
WRITE(6,9)
9  FORMAT(T5,'(5) TUKEY-HAMMING',//)
WRITE(6,10)
10  FORMAT(T2,'W(I)=0.54+0.46*COS(PI*I/M) - (M+1)<I<(M+1)')
WRITE(6,11)
11  FORMAT(T5,'(6) TUKEY-HANNING',//)
WRITE(6,12)
12  FORMAT(T2,'W(I)=0.5*(1+COS(PI*I/M)) - (M+1)<I<(M+1)')
WRITE(6,13)
13  FORMAT(T5,'(7) PARZEN A',//)
WRITE(6,14)
14  FORMAT(T2,'W(I)=1-(I/M)**2 - (M+1)<I<(M+1)')
WRITE(6,15)
15  FORMAT(T5,'(8) PARZEN B',//)
WRITE(6,16)
16  FORMAT(T2,'W(I)=1-2*(I/M)**2+2*(MOD(I)/M)**2 - MOD(I)*MOD(I/2)+1',
C/T6,'= 2*(1-(MOD(I)/M))**3 OTHERWISE')
WRITE(6,17)
17  FORMAT(T5,'(9) BARTLETT-PRIESTLEY',//)
WRITE(6,18)
18  FORMAT(T2,'W(I)=((SQRT(3)-1/PI*I)**2)*(SIN(PI*I/M)/(PI*I/M))',
C'-COS(PI*I/M) - (M+1)<I<(M+1)')
WRITE(6,19)
19  FORMAT(T4,'(10) TUKEY-PARZEN',//)
WRITE(6,20)
20  FORMAT(T2,'W(I)=1-2A+2A*COS(PI*I/M) - (M+1)<I<(M+1) A=0.282')
RETURN
END

```

```

DIMENSION X(360),ARR(360),Y(360),PILAM(360)
DIMENSION PIL(360),QJ(180),R(360)
COMMON MAP(3)
PI=3.141592653
WRITE(6,1)
1  FORMAT(/T2,'ENTER NO. OF OBSNS IN SERIES')
   READ(9,*)N
   READ(5,*)(X(I),I=1,N)
   READ(5,444)(MAP(I),I=1,3)
444  FORMAT(3A1)
   WRITE(6,2)
2  FORMAT(/T2,'ENTER ORDINATES OF SECTION OF INTEREST')
   READ(9,*)N1,N2
   II=0
   CON=0.0
   DO 3 J=N1,N2
     II=II+1
     IF(X(J).LT.CON)CON=X(J)
3  Y(II)=X(J)
   N=N2-N1+1
   IF(CON.GE.0.0)GOTO 95
   DO 6 J=1,N
     Y(J)=Y(J)-CON+1.
6  XM1=0.0
95  DO 97 J1=1,N
     XM1=XM1+Y(J1)
97  CONTINUE
     XM1=XM1/N
85  DO 99 J1=1,N
     Y(J1)=Y(J1)-XM1
99  CONTINUE
110 V1=0.0
     DO 111 J1=1,N
       B1=Y(J1)
       IF(B1+1)83,111,83
83  V1=V1+B1*B1
111 CONTINUE
     V1=V1/(N-1)
     NA=N-1
408 DO 120 J1=1,NA
     R(J1)=0.0
     NJ=N-J1
     DO 81 J2=1,NJ
       R(J1)=R(J1)+Y(J2)*Y(J1+J2)
81  CONTINUE
     R(J1)=R(J1)/N
120 CONTINUE
     WRITE(6,7)
     WRITE(1,7)
7  FORMAT(/T2,'PLOT OF ACRF',/T2,12(' '),/T2)
     DO 14 J=1,NA
       ARR(J)=R(J)/V1
       CALL PLOT(ARR,NA,1)
       WRITE(6,46)
       WRITE(1,46)
46  FORMAT(/T2,'LISTING OF ACR'S :',/T2)
       WRITE(6,37)(ARR(J),J=1,NA)
       WRITE(1,37)(ARR(J),J=1,NA)
37  FORMAT(T2,8F8.4)
555 WRITE(6,8)
8  FORMAT(/T2,'ENTER THAT LAG IN THE ACVF BEYOND WHICH THE
C/T2,'ACV'S APPEAR TO SETTLE DOWN TO STEADY OSCILLATIONS
C/T2,'WITH CONSTANT AMPLITUDES')
   READ(9,*)M
   WRITE(6,51)
51  FORMAT(/T2,'ENTER AN UPPER TRUNCATION POINT, LN,'
C/T2,'WHERE LN > 2M')
   READ(9,*)LN
   IF(LN.GT.NA)LN=NA
   IF(LN.LT.10)LN=20

```

```

M1=M+1
M22=2*M
N2=N/2
N21=N2+1
777 WRITE(6,54)M, LN
    WRITE(1,54)M, LN
54  FORMAT(/T2, 'ACRF TRUNCATION PT =', I3,
C/T2, 'UPPER TRUNCATION PT =', I3)
    CALL PIT(R, M, LN, 0.0, PILAM, N)
    DO 13 J=2, N21
        J1=J-1
13  ARR(J1)=PILAM(J)/V1
    CALL PLOT(ARR, N2, 0)
    WRITE(6,49)
    WRITE(1,49)
49  FORMAT(/T2, 'LISTING OF P(L) ORDINATES : ', /T2)
    WRITE(6,47) ARR(J), J=1, N2)
    WRITE(1,47) (ARR(J), J=1, N2)
47  FORMAT(T2, 8F8.3)
    WRITE(6,15)
15  FORMAT(T2, 'ENTER THE LAG AT WHICH THE PEAK OCCURS')
    READ(9,*)JW
    OMEGA=2.*PI*JW/FLOAT(N)
    M2=M/2
    TPM=2.*PI/FLOAT(M)
    M21=M2+1
    DO 17 J=1, M21
        J1=J-1
        DEL=OMEGA-(2.*PI*J1/FLOAT(M))
        IF(DEL.GE.0.0.AND.DEL.LE.TPM)GOTO 67
17  CONTINUE
    WRITE(6,18)
18  FOPMAT(/T2, 'NO CONSTANT VALUE POSSIBLE')
    STOP
67  WRITE(6,68)DEL, J1
    WRITE(1,68)DEL, J1
68  FORMAT(/T2, 'CONSTANT=', F15.6, ' INTEGRAL VALUE=', I3)
    CALL PIT(R, M, LN, DEL, PIL, M)
    U1=0.0
    U2=0.0
    DO 21 J=1, M
21  U1=U1+R(J)*R(J)
    U1=2.*(2.*U1+V1*V1)/(V1*V1)
    DO 22 J=1, M22
22  U2=U2+R(J)*R(J)
    U2=2.*(U2+V1*V1)/(V1*V1)
    GSTAR=(U1-U2)/(4.*PI)
    VNM=LN*2./3.-(4.*FLOAT(M)/3.)+(2.*M*M/(3.*FLOAT(LN)))
    VNM=FLOAT(M)*VNM
    A1=SQRT(FLOAT(N)/VNM)
    GST=GSTAR/(2.*PI)
    GST=1./SQRT(GST)
    CON=-9999999.
    DO 23 J=1, M21
        QJ(J)=0.0
        DO 24 JJ=1, J
24  QJ(J)=QJ(J)+(PIL(JJ)/V1)
        QJ(J)=GST*A1*QJ(J)
        IF(ABS(QJ(J)).GT.CON)CON=ABS(QJ(J))
23  CONTINUE
    WRITE(6,26)CON
    WRITE(1,26)CON
26  FORMAT(T2, 'THE ABSOLUTE MAXIMUM OF THE JQ"S =', F15.6)
    WRITE(6,25)
25  FORMAT(/T2, 'ENTER 1 IF THIS IS CONSIDERED SIGNIFICANT',
C/T2, 'ENTER 0 TO CONTINUE RUN',
C/T2, 'ENTER 2 TO CONTINUE RUN BUT WITH A DIFFERENT M',
C/T2, 'ENTER 3 TO EXIT')
    READ(9,*)NI
    IF(NI.NE.1)GOTO 888

```


APPENDIX A19 : RESULTS RELEVANT TO THE CHAPTER 9 AIRLINE DATA

This appendix contains several subsidiary results relevant to chapter 9. These results consist of the following tables and figures.

Tables

A19.1 Listing of the original data (read across)

A19.2 Listing of $\log_{10}$ of the data given in table A19.1

A19.3 The first 156 observation in the original data set are divided into subsets of size 13, and for each subset the mean, variance and acrf are calculated (this enables one to roughly ascertain whether the data is stationary or nonstationary).

A19.4 The same as table A19.3, except a subset length of 26 is used

A19.5 The same as table A19.3, except a subset length of 52 is used.

A19.6-A19.8 A similar situation to tables A19.3-A19.5 except that the $\log_{10}$ of the data is used

Figures

A19.1 The acrf of the first 150 observations in the original series.

A19.2 A spectral estimate (using the Parzen (b) window) of the first 150 observations, in the original series, with a truncation point of $M = 75 = 50\% \text{ of } N (N=150)$.

A19.3 Same as A19.2 but with a truncation point of 75

A19.4 Output of UYBAL (see appendix A18) applied to the first 150 data points in the original series

APPENDIX A19 : RESULTS RELEVANT TO THE CHAPTER 9 AIRLINE DATA

This appendix contains several subsidiary results relevant to chapter 9. These results consist of the following tables and figures.

Tables

A19.1 Listing of the original data (read across)

A19.2 Listing of $\log_{10}$ of the data given in table A19.1

A19.3 The first 156 observation in the original data set are divided into subsets of size 13, and for each subset the mean, variance and acrf are calculated (this enables one to roughly ascertain whether the data is stationary or nonstationary).

A19.4 The same as table A19.3, except a subset length of 26 is used

A19.5 The same as table A19.3, except a subset length of 52 is used

A19.6-A19.8 A similar situation to tables A19.3-A19.5 except that the $\log_{10}$ of the data is used

Figures

A19.1 The acrf of the first 150 observations in the original series.

A19.2 A spectral estimate (using the Parzen (b) window) of the first 150 observations, in the original series, with a truncation point of $M = 75 = 50\%$ of N ($N=150$).

A19.3 Same as A19.2 but with a truncation point of 75

A19.4 Output of BUYBAL (see appendix A18) applied to the first 150 data points in the original series.

- A19.5-A19.10 The (Parzen (b)) spectral estimates of the first 150 data points in the $\log_{10}$ transformed series for truncation points 7, 15, 30, 60, 75 and 112 (i.e. 5%, 10%, 20%, 40%, 50% and 75% of $N=150$).
- A19.11 Autocorrelogram (up to lag 148) of the transformed data.

Table A19.1 Listing of original data

341.0	402.0	267.0	180.0	187.0	146.0	153.0	213.0	159.0	170.0
149.0	94.0	172.0	162.0	88.0	97.0	155.0	127.0	142.0	158.0
108.0	129.0	153.0	143.0	227.0	156.0	244.0	302.0	392.0	363.0
233.0	169.0	165.0	214.0	152.0	191.0	178.0	219.0	224.0	196.0
157.0	157.0	190.0	153.0	140.0	123.0	112.0	127.0	357.0	351.0
84.0	320.0	309.0	299.0	369.0	354.0	349.0	325.0	262.0	135.0
120.0	140.0	145.0	219.0	142.0	104.0	147.0	145.0	194.0	138.0
211.0	163.0	173.0	192.0	240.0	133.0	325.0	284.0	295.0	358.0
324.0	356.0	371.0	220.0	159.0	251.0	231.0	225.0	206.0	282.0
235.0	256.0	297.0	272.0	275.0	210.0	168.0	155.0	211.0	127.0
207.0	336.0	129.0	211.0	338.0	321.0	299.0	151.0	261.0	167.0
160.0	316.0	177.0	143.0	327.0	176.0	164.0	146.0	127.0	210.0
187.0	139.0	199.0	142.0	140.0	108.0	177.0	188.0	212.0	252.0
166.0	243.0	320.0	387.0	340.0	260.0	183.0	222.0	254.0	232.0
179.0	220.0	259.0	284.0	302.0	237.0	280.0	320.0	160.0	204.0
219.0	216.0	216.0	303.0	111.0	298.0	324.0	384.0	291.0	187.0
201.0	166.0	159.0	169.0						

Table A19.2 Listing of transformed ($\log_{10}$) data

5.832	5.996	5.587	5.193	5.231	4.984	5.030	5.361	5.069	5.136
5.004	4.543	5.147	5.088	4.477	4.575	5.043	4.844	4.956	5.063
4.682	4.360	5.030	4.963	5.425	5.050	5.497	5.710	5.971	5.894
5.451	5.130	5.106	5.366	5.024	5.252	5.182	5.389	5.412	5.278
5.056	5.056	5.247	5.030	4.942	4.812	4.718	4.844	5.878	5.861
4.431	5.768	5.733	5.700	5.911	5.869	5.855	5.784	5.568	4.905
4.787	4.942	4.977	5.389	4.956	4.644	4.990	4.977	5.268	4.927
5.352	5.094	5.153	5.257	5.481	4.890	5.784	5.649	5.687	5.881
5.781	5.875	5.916	5.394	5.069	5.525	5.442	5.416	5.328	5.642
5.460	5.545	5.694	5.606	5.617	5.347	5.124	5.043	5.352	4.844
5.333	5.817	4.860	5.352	5.823	5.771	5.700	5.017	5.565	5.118
5.075	5.756	5.176	4.963	5.790	5.170	5.100	4.984	4.920	5.347
5.231	4.934	5.293	4.956	4.942	4.682	5.176	5.236	5.357	5.529
5.112	5.493	5.768	5.958	5.829	5.561	5.209	5.403	5.537	5.447
5.187	5.394	5.557	5.649	5.710	5.469	5.635	5.768	5.075	5.318
5.389	5.375	5.375	5.714	4.710	5.697	5.781	5.951	5.673	5.231
5.303	5.112	5.069	5.130						

Table A19.3 Stationarity 'test' (with subseries length 13) on
the original data

SUBSERIES LENGTH = 13

	MEAN	VARIANCE	LAGS	1	2	ACR 3	4	5	6
SUBSET 1	202.5385	0.734624E+04		0.57	0.15	0.00	-0.12	-0.08	0.02
SUBSET 2	141.9231	0.123824E+04		0.10	-0.05	0.06	-0.27	0.08	0.19
SUBSET 3	234.3077	0.564572E+04		0.63	0.21	-0.09	-0.14	-0.19	-0.26
SUBSET 4	189.7692	0.854752E+04		0.02	-0.07	0.12	-0.12	-0.12	-0.05
SUBSET 5	243.6923	0.931555E+04		0.71	0.52	0.22	-0.05	-0.24	-0.38
SUBSET 6	188.3846	0.401909E+04		0.12	0.14	0.12	0.10	-0.17	0.12
SUBSET 7	270.2307	0.442286E+04		0.47	0.16	0.11	-0.05	-0.30	-0.24
SUBSET 8	219.5385	0.422477E+04		0.08	-0.01	0.15	-0.23	-0.23	-0.08
SUBSET 9	230.7692	0.626868E+04		-0.07	-0.13	0.30	-0.22	-0.17	0.18
SUBSET 10	172.0769	0.164074E+04		0.21	-0.00	-0.10	-0.30	-0.25	-0.11
SUBSET 11	251.1538	0.421597E+04		0.49	-0.11	-0.37	-0.19	-0.04	-0.07
SUBSET 12	242.3077	0.391523E+04		-0.26	0.08	0.14	-0.06	-0.19	0.19

Table A19.4 Stationarity 'test' (with subseries length 26) on the original data

SUBSERIES LENGTH = 26

	MEAN	VARIANCE	LAGS	1	2	3	4	5	6
				7	8	9	10	11	12
SUBSET 1	172.2308	0.507583E+04		0.57	0.23	0.16	0.04	0.07	0.15
				0.06	-0.00	-0.04	-0.16	-0.09	-0.05
SUBSET 2	212.0385	0.732848E+04		0.33	0.08	0.08	-0.06	-0.08	-0.07
				-0.02	-0.11	0.03	0.06	0.01	-0.07
SUBSET 3	216.0385	0.719593E+04		0.64	0.51	0.32	0.15	0.01	-0.03
				-0.15	-0.18	-0.21	-0.22	-0.26	-0.30
SUBSET 4	244.8846	0.481396E+04		0.38	0.17	0.25	-0.02	-0.17	-0.10
				-0.02	-0.08	-0.00	0.00	0.11	0.22
SUBSET 5	201.4231	0.469215E+04		0.18	0.09	0.25	-0.10	0.03	0.15
				-0.01	0.14	0.03	0.02	-0.13	-0.19
SUBSET 6	246.7308	0.392331E+04		0.11	0.01	-0.16	-0.16	-0.11	0.00
				-0.23	-0.09	0.15	0.09	0.11	0.13

Table A19.5 Stationarity 'test' (with subseries length 52) on the original data

SUBSERIES LENGTH = 52

	MEAN	VARIANCE	LAGS	1	2	3	4	5	6
				7	8	9	10	11	12
SUBSET 1	192.1346	0.648448E+04		0.46	0.21	0.15	0.00	-0.02	-0.01
				-0.07	-0.12	-0.04	-0.07	-0.07	-0.13
SUBSET 2	230.4615	0.610175E+04		0.58	0.45	0.36	0.14	0.01	-0.03
				-0.14	-0.16	-0.13	-0.20	-0.16	-0.15
SUBSET 3	224.0769	0.474652E+04		0.27	0.19	0.21	0.02	0.08	0.13
				-0.14	0.00	0.04	0.01	-0.02	-0.03

Table A19.6 Stationarity 'test' (with subseries length 13) on
transformed data

SUBSERIES LENGTH = 13

	MEAN	VARIANCE	LAGS	1	2	ACR 3	4	5	6
SUBSET 1	5.2395	0.146598E+00		0.51	0.16	0.03	-0.14	-0.03	0.56
SUBSET 2	4.9274	0.609630E-01		0.12	-0.10	0.08	-0.28	0.07	0.25
SUBSET 3	5.4141	0.872618E-01		0.61	0.26	-0.07	-0.15	-0.23	-0.30
SUBSET 4	5.1478	0.202496E+00		-0.12	-0.16	0.09	-0.09	-0.09	-0.02
SUBSET 5	5.4135	0.189086E+00		0.67	0.48	0.18	-0.05	-0.21	-0.36
SUBSET 6	5.1897	0.103188E+00		0.06	0.14	0.11	0.15	-0.19	0.15
SUBSET 7	5.5705	0.639682E-01		0.43	0.11	0.07	-0.02	-0.32	-0.23
SUBSET 8	5.3488	0.960851E-01		0.05	0.01	0.22	-0.24	-0.19	-0.07
SUBSET 9	5.3865	0.118844E+00		-0.10	-0.13	0.33	-0.21	-0.17	0.20
SUBSET 10	5.1221	0.564400E-01		0.19	-0.01	-0.15	-0.29	-0.25	-0.11
SUBSET 11	5.4965	0.628924E-01		0.42	-0.15	-0.35	-0.17	-0.05	-0.03
SUBSET 12	5.4525	0.911762E-01		-0.28	0.07	0.11	-0.04	-0.14	0.23

Table A19.7 Stationarity 'test' (with subseries length 26) on transformed data

SUBSERIES LENGTH = 26

	MEAN	VARIANCE	LAGS	ACR					
				1	2	3	4	5	6
				7	8	9	10	11	12
SUBSET 1	5.0834	0.124951E+00		0.50	0.21	0.24	0.07	0.11	0.22
				-0.04	-0.03	-0.04	-0.20	-0.14	-0.06
SUBSET 2	5.2810	0.157533E+00		0.22	0.05	0.11	-0.01	-0.05	-0.03
				-0.05	-0.11	0.05	0.09	0.03	-0.04
SUBSET 3	5.3016	0.153319E+00		0.57	0.43	0.28	0.15	0.03	0.00
				-0.17	-0.21	-0.23	-0.21	-0.27	-0.32
SUBSET 4	5.4596	0.896037E-01		0.32	0.15	0.28	-0.02	-0.14	-0.07
				-0.03	-0.09	-0.00	-0.04	0.10	0.20
SUBSET 5	5.2543	0.102310E+00		0.17	0.09	0.21	-0.10	0.02	0.12
				-0.01	0.16	0.02	-0.02	-0.14	-0.17
SUBSET 6	5.4745	0.744561E-01		0.03	-0.01	-0.10	-0.12	-0.09	0.08
				-0.30	-0.07	0.13	0.05	0.07	0.11

Table A19.8 Stationarity 'test' (with subseries length 52) on transformed data)

SUBSERIES LENGTH = 52

	MEAN	VARIANCE	LAGS	ACR					
				1	2	3	4	5	6
				7	8	9	10	11	12
SUBSET 1	5.1822	0.148419E+00		0.38	0.20	0.22	0.07	0.04	0.06
				-0.05	-0.12	-0.01	-0.07	-0.08	-0.14
SUBSET 2	5.3806	0.125443E+00		0.53	0.42	0.35	0.16	0.06	0.03
				-0.11	-0.14	-0.11	-0.18	-0.16	-0.17
SUBSET 3	5.3644	0.990174E-01		0.23	0.21	0.25	0.07	0.13	0.18
				-0.10	0.04	0.03	-0.02	-0.04	-0.03

Fig. A19.1 Autocorrelogram of the first 150 observations
(in the original form)

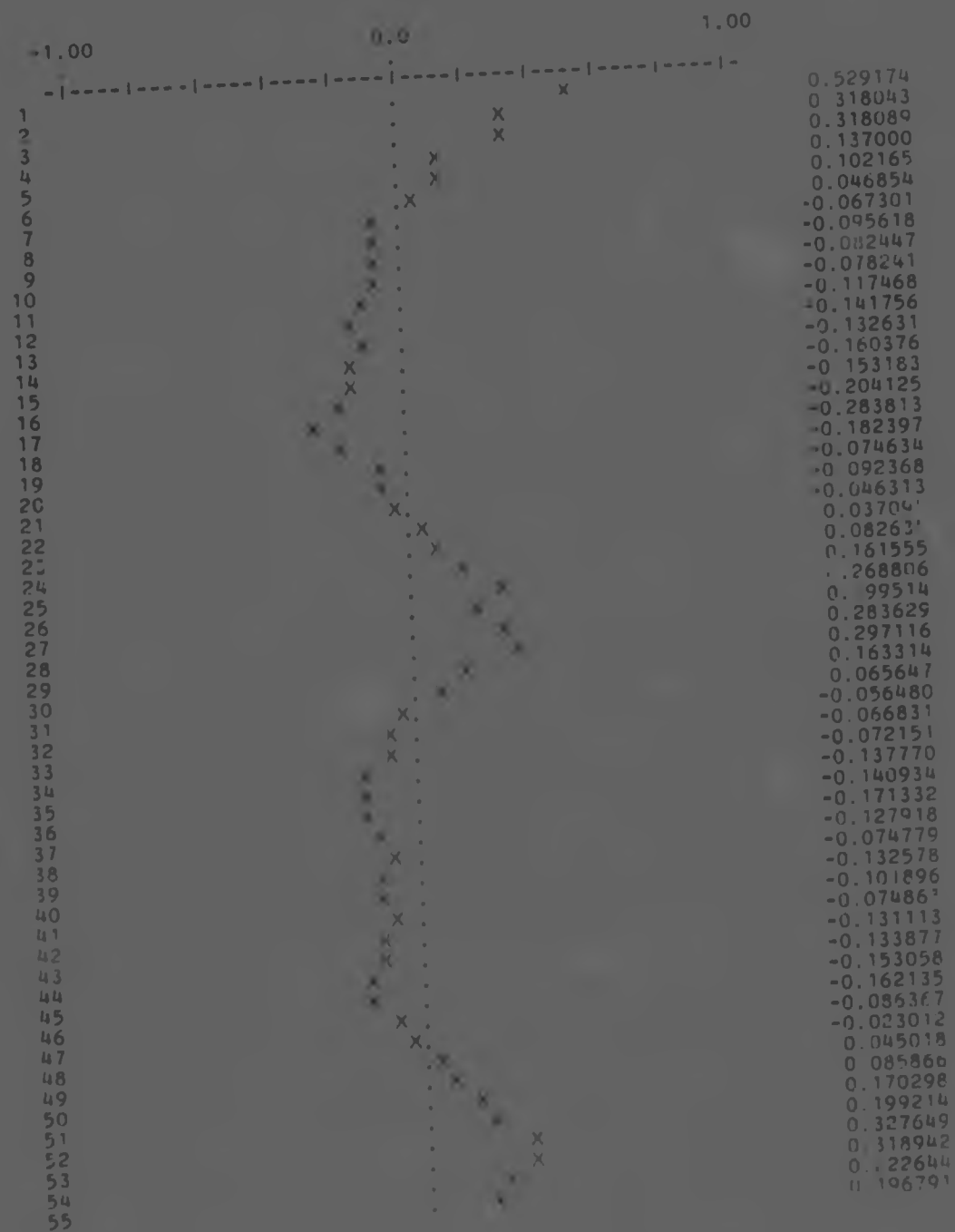
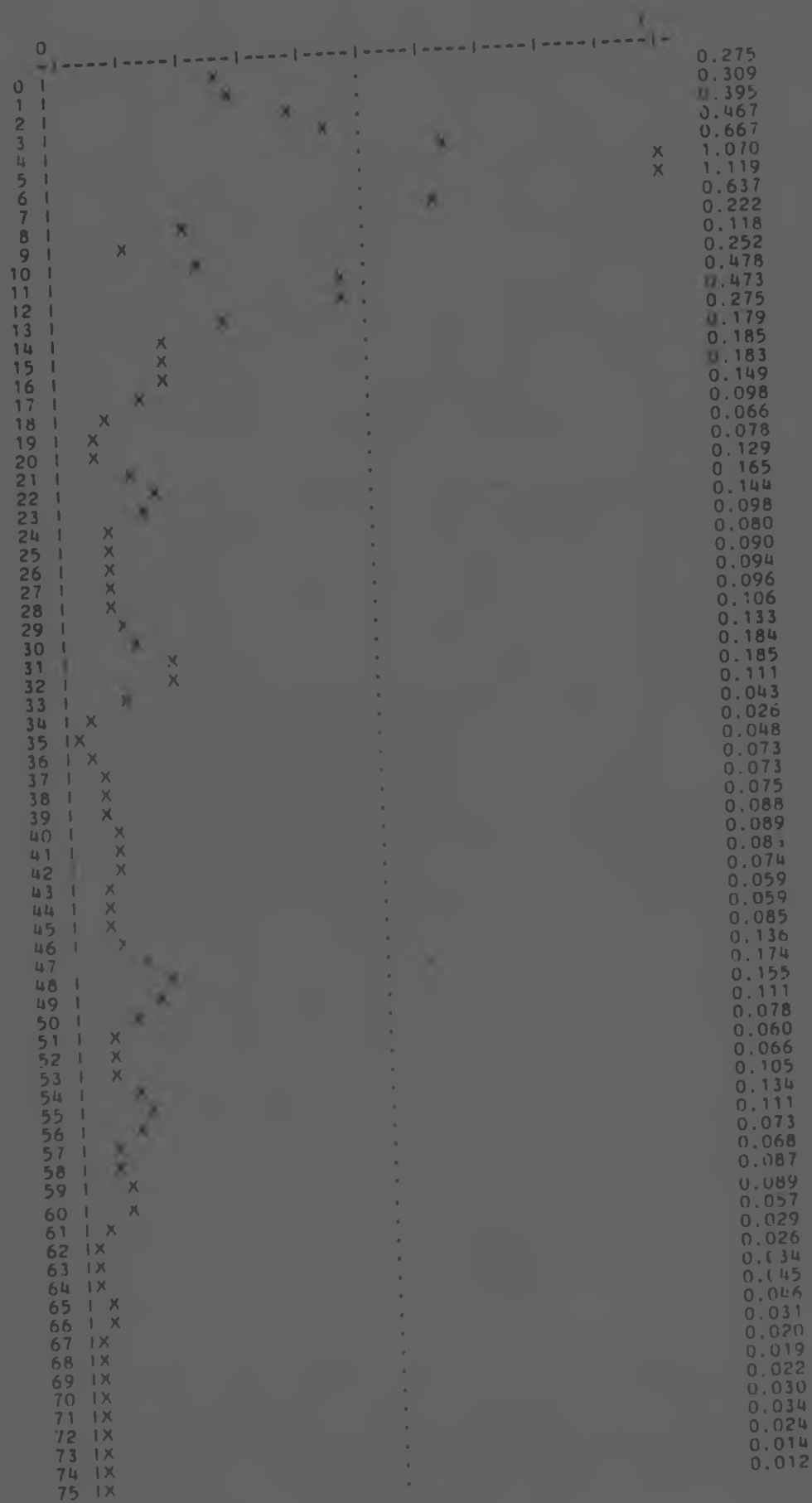


Fig. A19.2 Spectrum estimate for the original data

A19.9

(Parzen (b) window, truncation point 75)



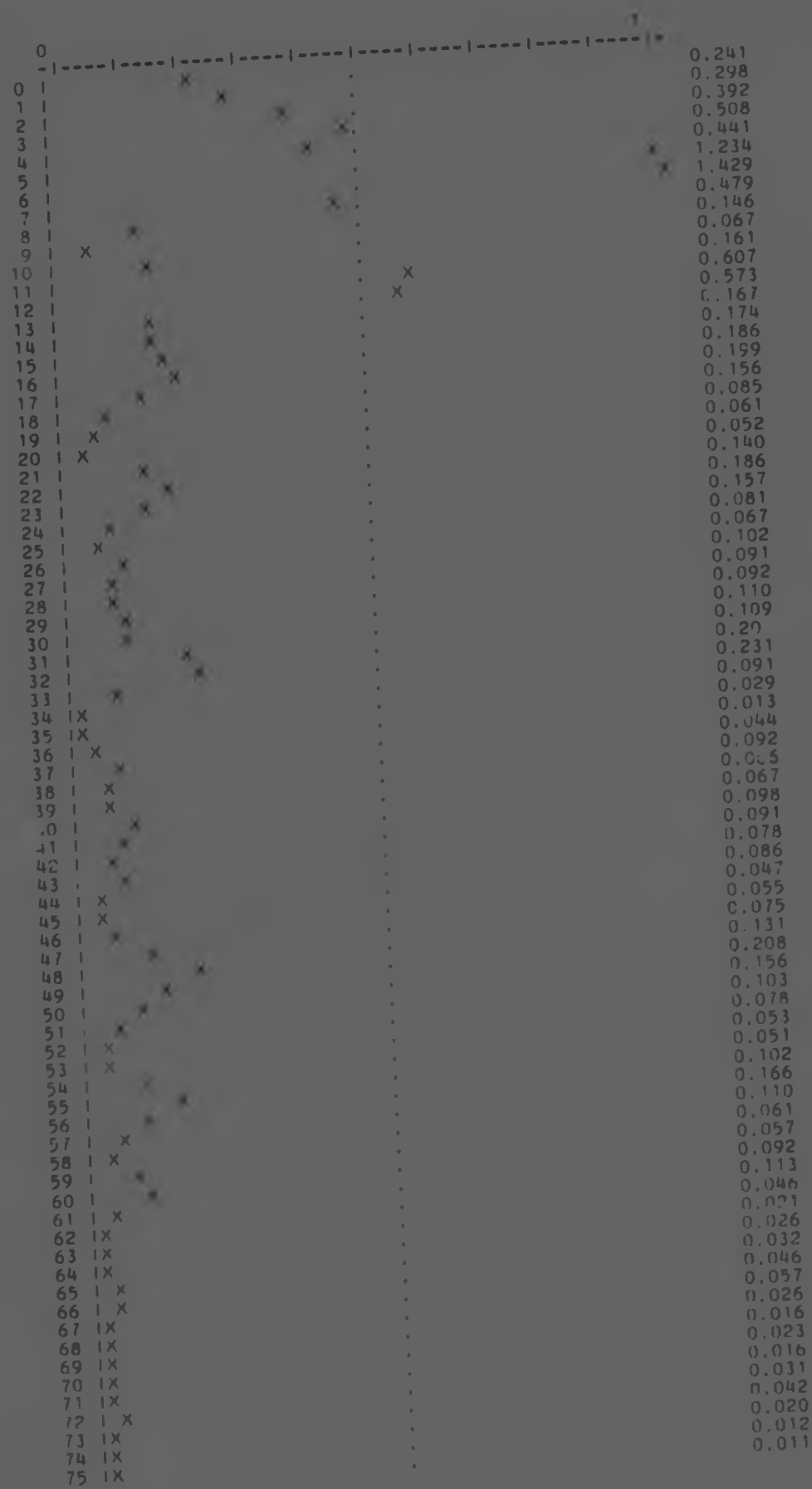


Fig. A19.4 Results from the program BUYBAL for the original data

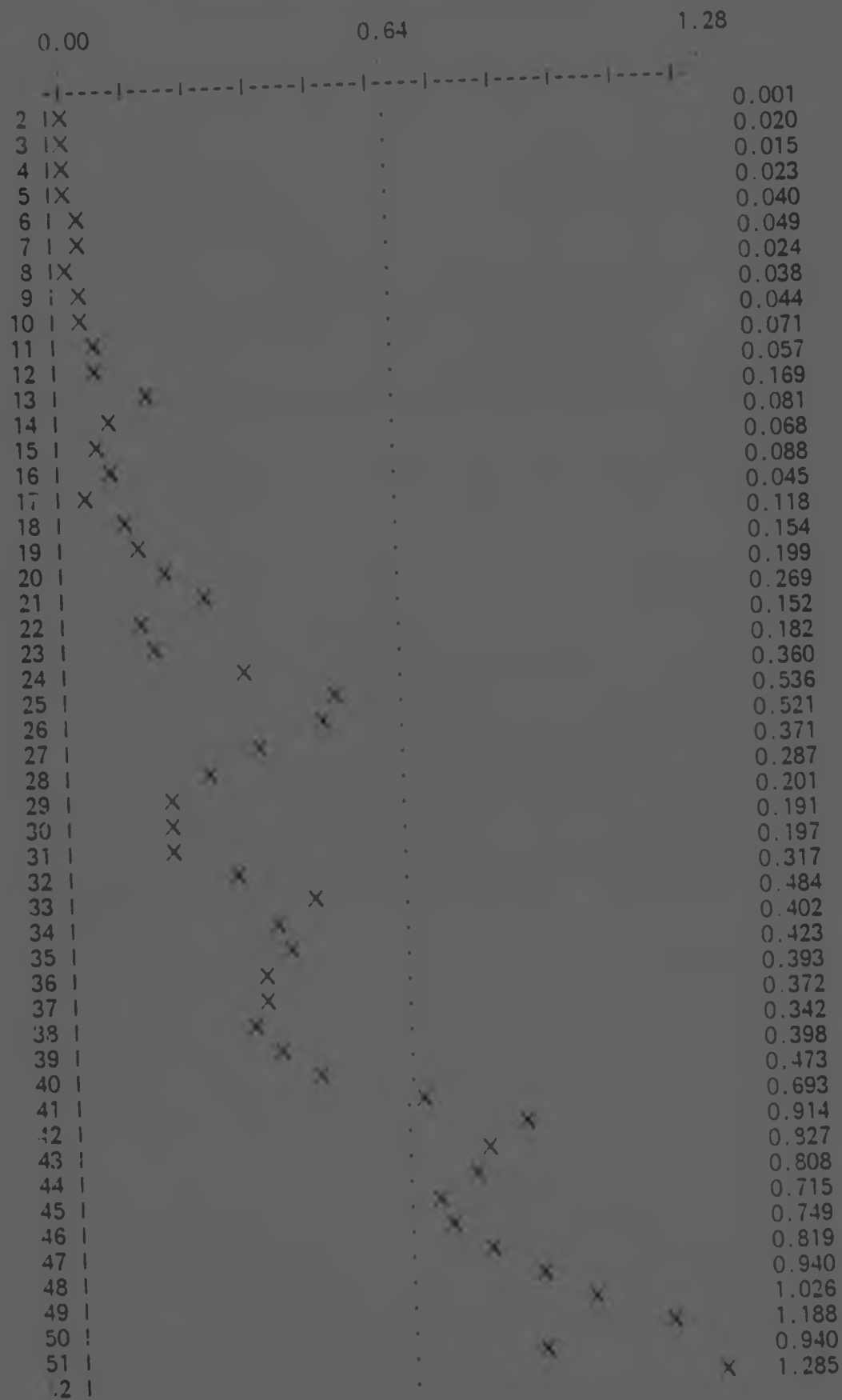




Fig. A19.5

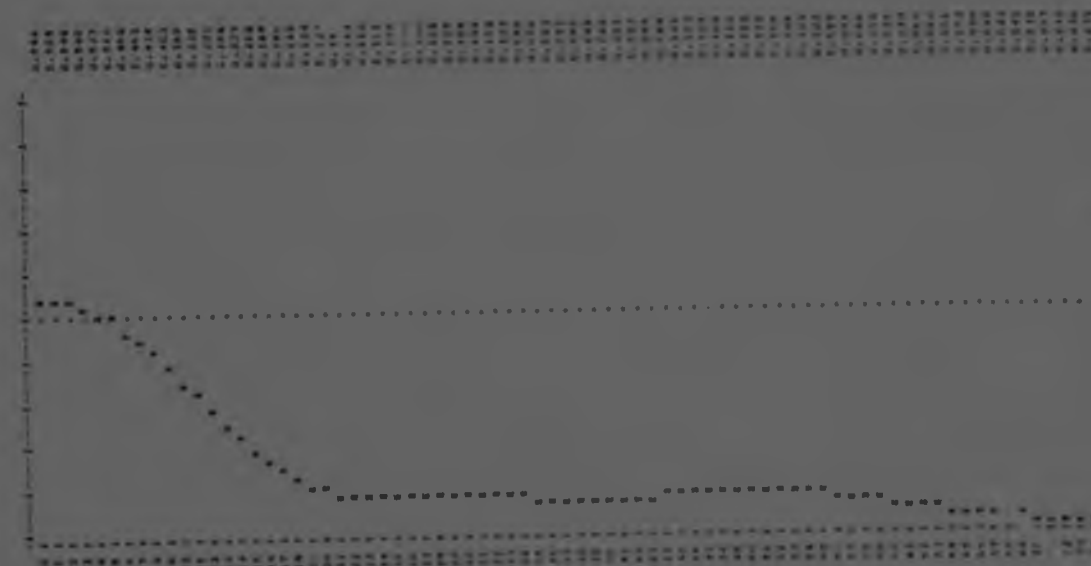


Fig. A19.6



Fig. A19.7

Spectrum estimates for transformed series (Parzen (b) window, truncation
points 7, 15 and 30 resp.)

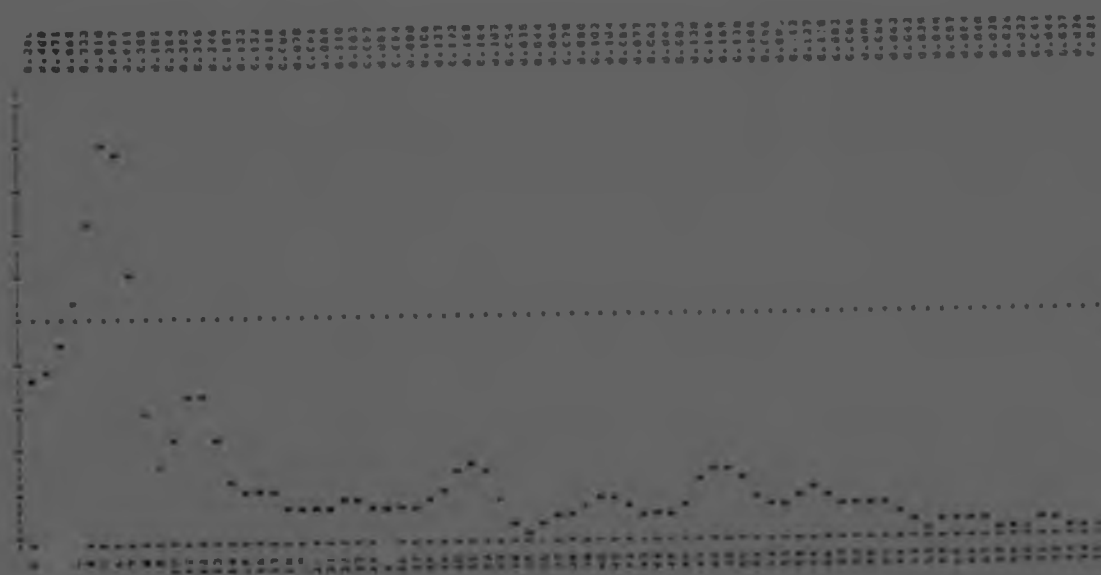


Fig. A19.8

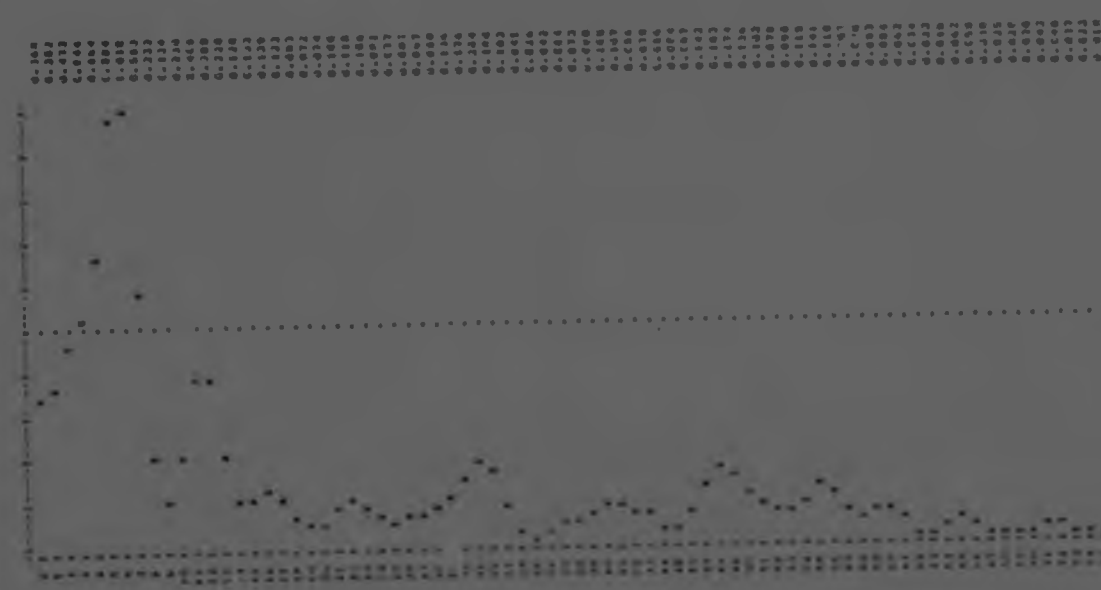


Fig. A19.9

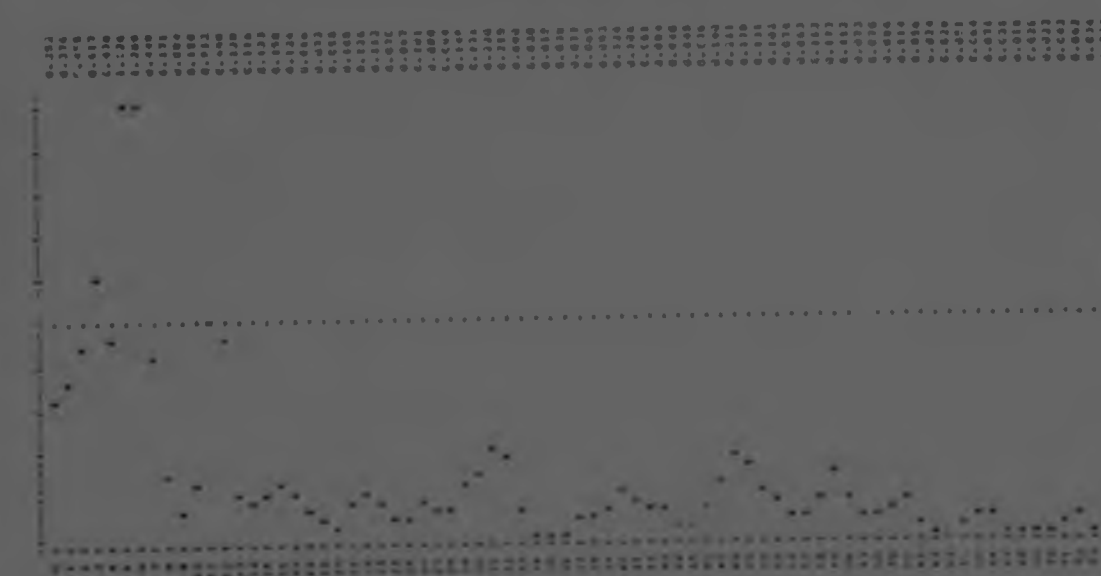


Fig. A19.10

Spectrum estimates for transformed series (Parzen (b) window, truncation points 60, 75 and 112 resp.)

A19.13



Fig. A19.8



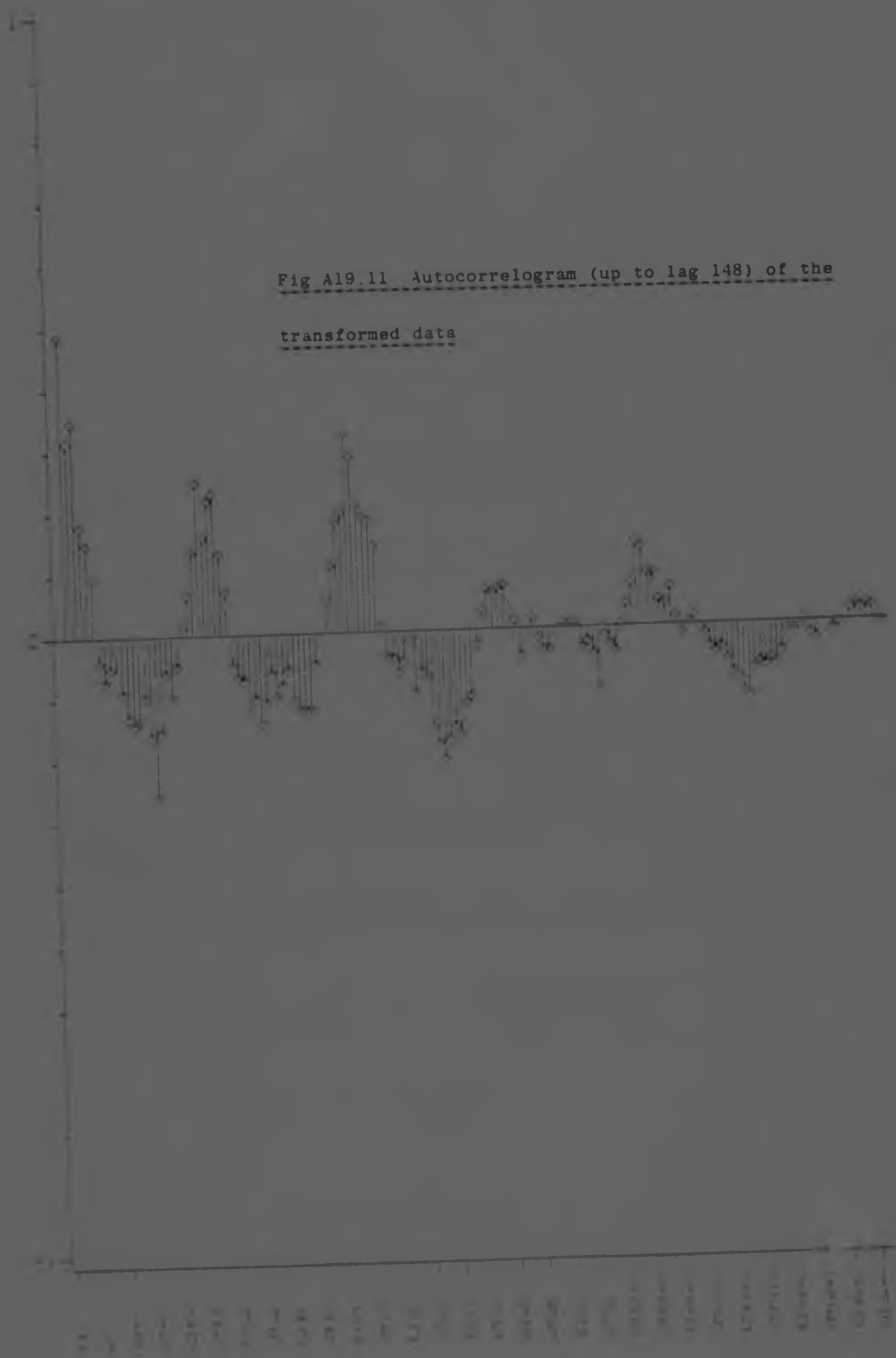
Fig. A19.9



Fig. A19.10

Spectrum estimates for transformed series (Parzen (b) window, truncation points 60, 75 and 112 resp.)

Fig A19.11 Autocorrelogram (up to lag 148) of the
transformed data



BIBLIOGRAPHY

- AKAIKE H (1970)
STATISTICAL PREDICTOR IDENTIFICATION
ANN. INST. MATH. STAT., V22, P203
- AKAIKE H (1979)
A BAYESIAN EXTENSION OF THE MINIMUM AIC PROCEDURE OF
AUTOREGRESSIVE MODEL FITTING
BIOA, V66, P237
- ANDERSON OD AND PERRYMAN MR (EDS.) (1981)
TIME SERIES ANALYSIS
NORTH HOLLAND PUBLISHING CO., NEW YORK
- ANDERSON RL (1942)
DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT
ANN. STAT., V13, P1
- ANDERSON TW (1971)
THE STATISTICAL ANALYSIS OF TIME SERIES
JOHN WILEY AND SONS INC., NEW YORK
- ANDERSON TW (1975)
MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS OF AR PROCESSES
WITH MA RESIDUALS AND OTHER COVARIANCE MATRICES WITH LINEAR
STRUCTURE
ANN. MATH. STAT., V3, P1283
- ANDERSON TW (1977)
ESTIMATION FOR ARIMA MODELS IN THE TIME AND FREQUENCY DOMAIN
ANN. MATH. STAT., V5, P842
- ANDERSON TW AND WALKER AM (1984)
ON THE ASYMPTOTIC DISTRIBUTION OF THE AUTOCORRELATIONS OF A
SAMPLE FROM A LINEAR STOCHASTIC PROCESS
ANN. STAT., V35, P1296
- ANSLEY CF (1979)
AN ALGORITHM FOR THE EXACT LIKELIHOOD OF A MIXED ARMA PROCESS
BIOA, V66, P59
- ANSLEY CF AND NEWBOLD P (1979)
ON THE FINITE SAMPLE DISTRIBUTION OF RESIDUAL AUTOCORRELATIONS
IN ARMA MODELS
BIOA, V66, P547
- ANSLEY CF, SPIVEY WA AND WROBLESKI WJ (1977)
A CLASS OF TRANSFORMATIONS FOR BOX-JENKINS SEASONAL MODELS
APP. STAT., V26, P173
- BARTLETT MS (1946)
ON THE THEORETICAL SPECIFICATION OF SAMPLING PROPERTIES OF
AUTOCORRELATED TIME SERIES
JRSS SUPPLEMENT, V8, P27
- BARTLETT MS (1950)
PERIODOGRAM ANALYSIS AND CONTINUOUS SPECTRA
BIOA, V37, P1
- BARTLETT MS (1978)
STOCHASTIC PROCESSES
CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE
- BHANSALI RJ (1979)
A MIXED SPECTRUM ANALYSIS OF THE LYNX DATA
JRSSA, V142, P199
- BHANSALI RJ (1980)

AUTOREGRESSIVE AND WINDOW ESTIMATES OF THE INVERSE CORRELATION
FUNCTION
BIOA,V67,P551

BINGHAM C, GODFREY MD AND TUKEY JW (1967)
MODERN TECHNIQUES OF POWER SPECTRUM ESTIMATION
IEEE TRANS. AUD. ELECT., V AU-15, P56

BLACKMAN RB AND TUKEY JW (1959)
THE MEASUREMENT OF POWER SPECTRUM FROM THE POINT OF VIEW OF
COMMUNICATIONS ENGINEERING
BELL SYSTEMS TECHNICAL J., V37, P185

BLOOMFIELD P (1970)
SPECTRAL ANALYSIS WITH RANDOMLY MISSING OBSERVATIONS
JRSSB, V32, P369

BLOOMFIELD P (1976)
FOURIER ANALYSIS OF TIME SERIES - AN INTRODUCTION
JOHN WILEY AND SONS INC., NEW YORK

BOCHNER S (1936)
LECTURES ON FOURIER ANALYSIS
PRINCETON UNIVERSITY PRESS, PRINCETON

BOHMAN H (1961)
APPROXIMATE FOURIER ANALYSIS OF DISTRIBUTION FUNCTIONS
ARKIV FUR MATHEMATIK, V4, P99

BOLCH BW AND HUANG CJ (1974)
MULTIVARIATE STATISTICAL METHODS FOR BUSINESS AND ECONOMICS
PRENTICE HALL, LONDON

BOX GEP AND COX DR (1964)
AN ANALYSIS OF TRANSFORMATIONS
JRSSB, V26, P211

BOX GEP AND JENKINS GM (1970, 1976)
TIME SERIES ANALYSIS : FORECASTING AND CONTROL
HOLDEN-DAY INC., SAN FRANCISCO

BOX GEP AND LJUNG GM (1978)
ON A MEASURE OF LACK OF FIT IN TIME SERIES MODELS
BIOA, V65, P297

BOX G AND LJUNG GM (1979)
THE LIKELIHOOD FUNCTION OF STATIONARY ARMA MODELS
BIOA, V66, P265

BOX GEP AND PIERCE DA (1970)
THE DISTRIBUTION OF RESIDUAL AUTOCORRELATIONS IN ARIMA TIME
SERIES MODELS
JASA, V65, P1509

BRENT RP (1973)
ALGORITHMS FOR MINIMIZATION WITHOUT DERIVATIVES
PRENTICE HALL INC., LONDON

BRILLINGER DR (1975)
TIME SERIES : DATA ANALYSIS AND THEORY
HOLT, RINEHART AND WINSTON, NEW YORK

BRILLINGER DR AND TIAO GC (EDS.) (1978)
DIRECTIONS IN TIME SERIES ANALYSIS
INSTITUTE OF MATHEMATICAL STATISTICS

BUYS-BALLOT JHD (1847)
LES CHANGEMENTS DE TEMPERATURE
UTRECHT

CAMPBELL MJ AND WALKER AM (1977)
A SURVEY OF STATISTICAL WORK ON THE MACKENZIE RIVER SERIES OF
ANNUAL CANADIAN LYNX TRAPPINGS FOR THE YEARS 1821-1934 AND A
NEW ANALYSIS
JRSSA, V140, P411

CHATFIELD C (1977)
SOME RECENT DEVELOPMENTS IN TIME SERIES ANALYSIS
JRSSA, V140, P492

CHATFIELD C (1979)
INVERSE AUTOCORRELATIONS
JRSSA, V142, P363

CHATFIELD C (1980) (2ND ED.)
THE ANALYSIS OF TIME SERIES: AN INTRODUCTION
CHAPMAN AND HALL, LONDON

CHATFIELD C AND PROTHERO DL (1973)
BOX-JENKINS SEASONAL FORECASTING: PROBLEMS IN A CASE STUDY
JRSSA, V136, P295

CLEVELAND WS (1972)
THE INVERSE AUTOCORRELATIONS OF A TIME SERIES AND THEIR
APPLICATIONS
TECHNO, V14, P277

COOLEY JW AND TUKEY JW (1965)
AN ALGORITHM FOR THE MACHINE CALCULATION OF COMPLEX FOURIER
SERIES
MATHEMATICS OF COMPUTATION, V19, P297

DANIELL PJ (1946)
DISCUSSION ON 'SYMPOSIUM ON AUTOCORRELATION IN TIME SERIES'
JRSS SUPPLEMENT, V8, P88

DAVIES N, SPEDDING T AND WATSON W (1980)
ARMA PROCESSES WITH NON-NORMAL RESIDUALS
JTSA, V1, P103

DAVIES N, TRIGGS CM AND NEWBOLD P (1977)
SIGNIFICANCE LEVELS OF THE BOX-PIERCE PORTMANTEAU STATISTIC IN
FINITE SAMPLES
BIOA, V64, P517

DRAPER NR AND SMITH H (1966)
APPLIED REGRESSION ANALYSIS
JOHN WILEY AND SONS INC., NEW YORK

DURBIN JR (1960)
THE FITTING OF TIME SERIES MODELS
REVIEW OF THE INTERNATIONAL STATISTICAL INSTITUTE, V28, P233

EPANCHNIKOV VA (1969)
NON-PARAMETRIC ESTIMATION OF A MULTIVARIATE PROBABILITY
DENSITY
THEORETICAL PROBABILITY APPLICATION, V14, P153

FEJER L (1910)
LEBESQUESCHE KONSTANTEN UND DIVERGENTE FOURIERREIHEN
J. FÜR DIE REINE UND ANGEWANDTE MATHEMATIK

FINDLEY DF (1981)
LARGE SAMPLE BEHAVIOUR OF THE S-ARRAY OF SEASONALLY
NONSTATIONARY ARMA SERIES
IN ANDERSON AND PERRYMAN (EDS.), TIME SERIES ANALYSIS,
NORTH HOLLAND PUBL. CO., P163

FISHER RA (1929)
TESTS OF SIGNIFICANCE IN HARMONIC ANALYSIS

PROC. RSSA, 25, 254

FLETCHER R AND POWELL MJD (1963)
A RAPIDLY CONVERGENT DESCENT MET FOR MINIMIZATION
COMP. J. V6, P163

GABR MM AND SUBBA RAO T (1981)
THE ESTIMATION OF SUBSET BILINEAR TIME SERIES MODELS WITH
APPLICATIONS
J TSA, V2, P155

HALEBAITH RF (1977)
DISCUSSION OF THE PAPERS BY MR CAMPBELL ET AL
JRSSA, V140, P456

GODFREY LG (1979)
TESTING THE ADEQUACY OF A TIME SERIES MODEL
BIOA, V66, P67

GOODMAN SR (1961)
SOME COMMENTS ON SPECTRAL ANALYSIS OF TIME SERIES
TECHNO, V3, P221

GOTTMAN JM (1981)
TIME SERIES ANALYSIS
CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE

GRANGER CWJ AND MORRIS MJ (1976)
TIME SERIES MODELLING AND INTERPRETATION
JRSSA, V139, P246

GRANGER CWJ AND NEWBOLD P (1976)
FORECASTING TRANSFORMED SERIES
JRSSB, V38, P189

GRANGER CWJ AND NEWBOLD P (1977)
FORECASTING ECONOMIC TIME SERIES
ACADEMIC PRESS INC., NEW YORK

GRAY HL, KELLEY GD AND MCINTIRE DD (1978)
A NEW APPROACH TO ARMA MODELLING
COMM. STAT., V B7(1), P1

HAMILTON DC AND WATTS DG (1981)
INTERPRETING PARTIAL AUTOCORRELATION COEFFICIENTS OF SEASONAL
TIME SERIES MODELS
BIOA, V65, P135

HANNAN EJ (1960)
TIME SERIES ANALYSIS
NORTH-HOLLAND MONOGRAPHS, LONDON

HANNAN EJ (1970)
MULTIPLE TIME SERIES
JOHN WILEY AND SONS INC., NEW YORK

HARRIS E (ED) (1977)
ADVANCED SEMINAR SPECTRAL ANALYSIS OF TIME SERIES
JOHN WILEY AND SONS INC., NEW YORK

HARTLEY HJ (1949)
TESTS OF SIGNIFICANCE IN HARMONIC ANALYSIS
IOA, V36, P194

HELI JW AND WOODWORTH D (1980)
AUTOMATIC BOX-JENKINS FORECASTING
JORS, V31, P413

HIPEL KW, MCLEOD AI AND LENNOX WC (1977)
ADVANCES IN BOX-JENKINS MODELLING 1

WRR,V13,P567

B5

HIPEL KW, MCLEOD AI AND MCKEAN EA (1977)
STATISTICAL MODELLING OF THE EFFECTS OF RESERVOIR OPERATION
JHY,V22,P97

HOSKING JRM (1980)
THE ASYMPTOTIC DISTRIBUTION OF THE SAMPLE INVERSE
AUTOCORRELATIONS OF AN ARMA PROCESS
BIOA,V67,P123

ISSERLIS L (1918)
ON THE FORMULA FOR THE PRODUCT MOMENT COEFFICIENT OF ANY
ORDER OF A NORMAL FREQUENCY DISTRIBUTION IN ANY NUMBER OF
VARIABLES
BIOA,V12,P134

JENKINS GM (1961)
GENERAL CONSIDERATIONS IN THE ANALYSIS OF SPECTRA
TECHNO,V3,P133

JENKINS GM AND WATTS DG (1968)
SPECTRAL ANALYSIS AND ITS APPLICATIONS
HOLDEN-DAY INC., SAN FRANCISCO

JONES RH (1962)
SPECTRAL ANALYSIS WITH REGULARLY MISSED OBSERVATIONS
JNN. STAT.,V31,P568

JONES RH (1965)
A REAPPRAISAL OF THE PERIODOGRAM IN SPECTRAL ANALYSIS
TECHNO,V7,P531

KANNAN D (1979)
AN INTRODUCTION TO STOCHASTIC PROCESSES
ELSEVIER NORTH HOLLAND INC., NEW YORK

KINGMAN JFC AND TAYLOR SJ (1966)
INTRODUCTION TO MEASURE AND PROBABILITY
CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE

KRISHNAIAH PR (ED.) (1969)
MULTIVARIATE ANALYSIS 2
ACADEMIC PRESS INC., NEW YORK

LAMPERTI J (1977)
STOCHASTIC PROCESSES
SPRINGER-VERLAG, NEW YORK INC.

LOMNICKI ZA (1961)
TESTS FOR DEPARTURE FROM NORMALITY IN THE CASE OF LINEAR
STOCHASTIC PROCESSES
METRIKA,V4,P37

MARQUARDT DW (1963)
AN ALGORITHM FOR LEAST SQUARES ESTIMATION OF NONLINEAR
PARAMETERS
J. OF THE SOCIETY FOR INDUSTRIAL AND APPLIED MATH.,V11,P431

MCCLAVE JT (1975)
SUBSET AUTOREGRESSION
TECHNO,V17,P213

MCCLAVE JT (1978)
ESTIMATING THE ORDER OF AR MODELS : MAX CHI-SQUARED METHOD
JASA,V73,P122

MCCLAVE JT (1978A)
ESTIMATING THE ORDER OF MA MODELS : MAX CHI-SQUARED METHOD
COMM. STAT.,V A7(3),P259

MCLEOD AI (1978)
ON THE DISTRIBUTION OF RESIDUAL AUTOCORRELATIONS IN BOX-
JENKINS MODELS
JRSSB, V40, P296

MCLEOD AI, HIPEL KW AND LENNOX WC (1977)
ADVANCES IN BOX-JENKINS MODELLING 2
WRR, V13, P577

MEHTA BM AND AHLERT RC (1981)
STOCHASTIC ANALYSIS AND TRANSFER FUNCTIONS FOR FLOWS OF THE
UPPER DELAWARE RIVER
ECO. MOD., V14, P59

NEAVE HR (1970)
AN IMPROVED FORMULA FOR THE ASYMPTOTIC VARIANCE OF SPECTRUM
ESTIMATES
ANN. STAT., V41, P70

NEAVE HR (1971)
THE EXACT ERROR IN SPECTRUM ESTIMATES
ANN. STAT., V42, P961

NEAVE HR (1972)
A COMPARISON OF LAG WINDOW GENERATORS
JASA, V67, P152

NERLOVE M, GREYER DM AND CARVALHO JL (1979)
ANALYSIS OF ECONOMIC TIME SERIES
ACADEMIC PRESS INC., NEW YORK

NEWBOLD P (1974)
THE EXACT LIKELIHOOD FOR A MIXED ARMA PROCESS
BIOA, V61, P423

NEWBOLD P (1981)
SOME RECENT DEVELOPMENTS IN TIME SERIES ANALYSIS
INT. STAT. REV., V49, P53

NEWTON HW (1958)
THE FACE OF THE SUN
PENGUIN

OZAKI T (1977)
ON THE ORDER DETERMINATION OF ARMA MODELS
APP. STAT., V26, P290

PACK DJ (1978)
WHAT WILL YOUR TIME SERIES ANALYSIS COMPUTER PACKAGE DO?
IN BRILLINGER AND TIAO (EDS.), DIRECTIONS IN TIME SERIES ANALYSIS,
INSTITUTE OF MATHEMATICAL STATISTICS, P60

PARZEN E (1957)
ON CHOOSING AN ESTIMATE OF THE SPECTRAL DENSITY FUNCTION OF A
STATIONARY TIME SERIES
ANN. STAT., V28, P921

PARZEN E (1957A)
ON CONSISTENT ESTIMATES OF THE SPECTRUM OF A STATIONARY TIME
SERIES
ANN. STAT., V23, P329

PARZEN E (1958)
ON ASYMPTOTICALLY EFFICIENT CONSISTENT ESTIMATES OF THE
SPECTRAL DENSITY FUNCTION OF A STATIONARY TIME SERIES
JRSSB, V20, P303

PARZEN E (1961)
AN APPROACH TO TIME SERIES ANALYSIS

ANN. STAT., V32, P951

B7.

PARZEN E (1961A)
MATHEMATICAL CONSIDERATIONS IN THE ESTIMATION OF SPECTRA
TECHNO, V3, P167

PARZEN E (1969)
MULTIPLE TIME SERIES MODELLING
IN KRISHNAIAH (1969), MULTIVARIATE ANALYSIS 2,
ACADEMIC PRESS INC., P389

PARZEN E (1974)
SOME RECENT ADVANCES IN TIME SERIES MODELLING
IEEE TRANS. AUTO. CON., V AC-19, P723

PARZEN E (1978)
TIME SERIES MODELLING, SPECTRAL ANALYSIS AND FORECASTING
IN BRILLINGER AND TIAO (EDS.), DIRECTIONS IN TIME SERIES ANALYSIS,
INSTITUTE OF MATHEMATICAL STATISTICS, P80

PARZEN E (1979)
NONPARAMETRIC STATISTICAL DATA MODELING
JASA, V74, P105

PARZEN E (1982)
ARARMA MODELS FOR TIME SERIES ANALYSIS AND FORECASTING
JFO, V1, P67

POWELL MJD (1964A)
AN EFFICIENT METHOD FOR MINIMIZING A SUM OF SQUARES OF
NONLINEAR FUNCTIONS WITHOUT CALCULATING DERIVATIVES
COMP. J., V7, P153

POWELL MJD (1964B)
A METHOD FOR MINIMIZING A SUM OF SQUARES OF NONLINEAR
FUNCTIONS WITHOUT CALCULATING DERIVATIVES
COMP. J., V7, P303

PRIESTLEY MB (1962)
BASIC CONSIDERATIONS IN THE ESTIMATION OF SPECTRA
TECHNO, V3, P167

PRIESTLEY MB (1962A)
THE ANALYSIS OF STATIONARY PROCESSES WITH MIXED SPECTRA - 1
JRSSB, V24, P215

PRIESTLEY MB (1962B)
THE ANALYSIS OF STATIONARY PROCESSES WITH MIXED SPECTRA - 2
JRSSB, V24, P511

PRIESTLEY MB (1963)
THE SPECTRUM OF A CONTINUOUS PROCESS DERIVED FROM A DISCRETE
PROCESS
BIOA, V50, P517

PRIESTLEY MB (1980)
STATE DEPENDENT MODELS : A GENERAL APPROACH TO NONLINEAR TIME
SERIES
JTSA, V1, P47

PRIESTLEY MB (1981)
SPECTRAL ANALYSIS AND TIME SERIES - VOL. 1
ACADEMIC PRESS INC., NEW YORK

PRIESTLEY MB (1981A)
SPECTRAL ANALYSIS AND TIME SERIES - VOL. 2
ACADEMIC PRESS INC., NEW YORK

QUENOUILLE MH (1949)
APPROXIMATE TESTS OF CORRELATION IN TIME SERIES

JRSSB,V11,P68

88

REILLY DP (1981)
RECENT EXPERIENCES WITH AN AUTOMATIC BOX-JENKINS MODELLING
ALGORITHM
IN ANDERSON AND PERRYMAN (EDS.), TIME SERIES ANALYSIS,
NORTH HOLLAND PUBL. CO., P493

ROBINSON EA (1980)
PHYSICAL APPLICATIONS OF STATIONARY TIME SERIES
CHARLES GRIFFIN AND CO. LTD., BUCKS(ENGLAND)

SHIBATA R (1976)
SELECTION OF THE ORDER OF AN AR MODEL BY AKAIKE'S INFORMATION
CRITERION
BIOA,V63,P117

SINGLETON RC AND POULTER TC (1967)
SPECTRAL ANALYSIS OF THE CALL OF A KILLER WHALE
IEEE TRANS. AUD. ELECT.,V AU-15,P104

SUBBA RAO T (1981)
ON THE THEORY OF BILINEAR TIME SERIES MODELS
JRSSB,V43,P244

SUBBA RAO T AND GABR MM (1980)
A TEST FOR LINEARITY OF STATIONARY TIME SERIES
JTSA,V1,P145

TITCHMARSH EC (1939)
THE THEORY OF FUNCTIONS
OXFORD UNIVERSITY PRESS, OXFORD

TONG H (1977)
SOME COMMENTS ON THE CANADIAN LYNX DATA
JRSSA,V140,P432

TONG H AND LIM KS (1980)
THRESHOLD AUTOREGRESSION, LIMIT CYCLES AND CYCLICAL DATA
JRSSB,V42,P245

TUKEY JW (1949)
THE SAMPLING THEORY OF POWER SPECTRUM ESTIMATES
PROCEEDINGS ON APPLICATIONS OF AUTOCORRELATION ANALYSIS OF
PHYSICAL PATTERNS

TUKEY JW (1967)
AN INTRODUCTION TO THE CALCULATIONS OF NUMERICAL SPECTRUM
ANALYSIS
IN HARRIS (ED.), ADVANCED SEMINAR ON SPECTRAL ANALYSIS
OF TIME SERIES, JOHN WILEY AND SONS, P25

VERA G AND GUERRERO VM (1981)
ESTIMATION OF SEASONAL FACTORS USING BOTH TRADITIONAL METHODS
AND BOX-JENKINS TECHNIQUES
IN ANDERSON AND PERRYMAN (EDS.), TIME SERIES ANALYSIS,
NORTH HOLLAND PUBL. CO., P607

WHITTAKER ET AND ROBINSON G (1924)
THE CALCULUS OF OBSERVATIONS
D. VAN NOSTRAND, NEW YORK

WHITTLE P (1952)
THE SIMULTANEOUS ESTIMATION OF A TIME SERIES HARMONIC
COMPONENT AND COVARIANCE STRUCTURE
TRABAJOS DE ESTADISTICA,V3,P43

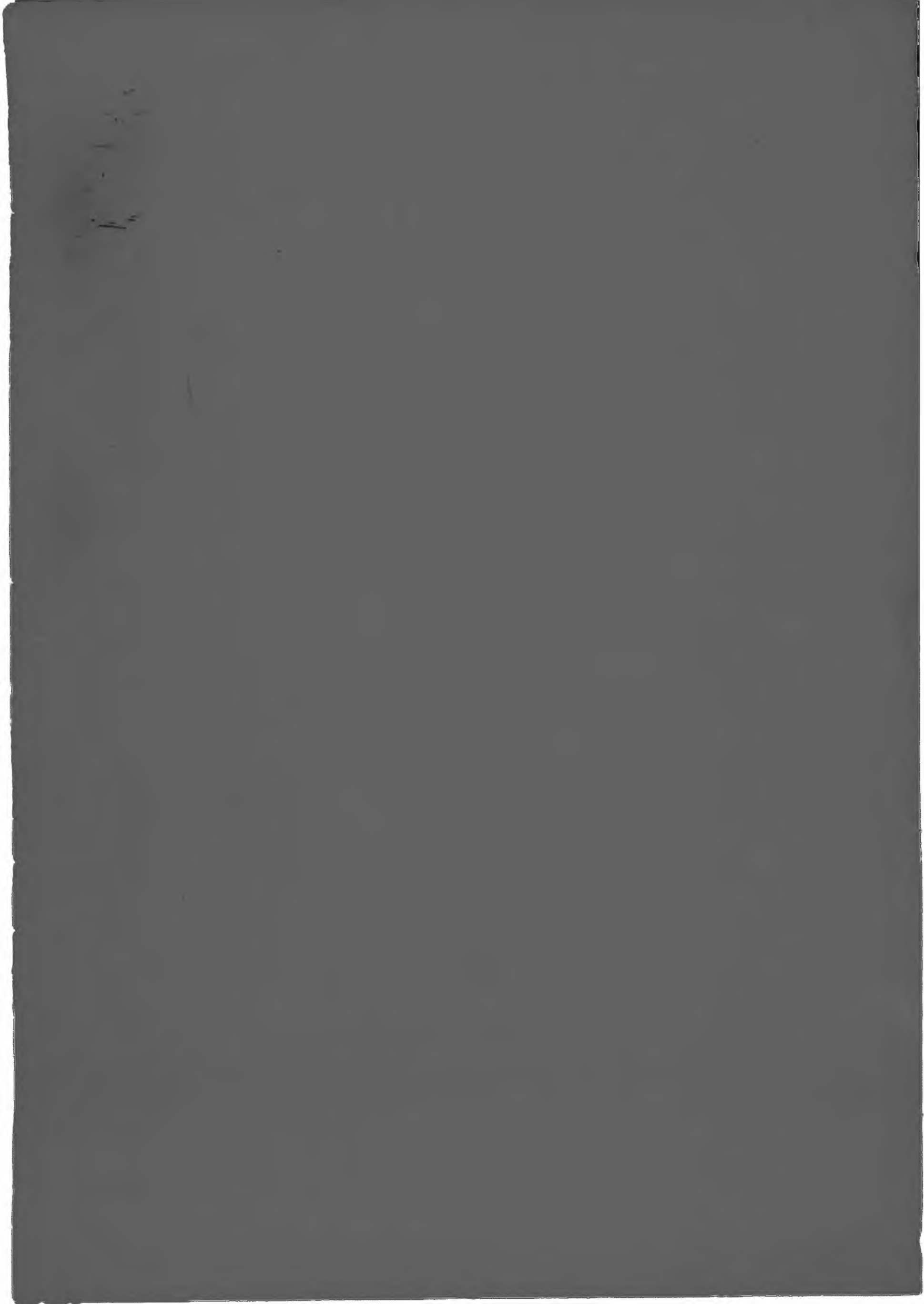
WIENER N (1930)
GENERALIZED HARMONIC ANALYSIS
ACT. MATH.,V55,P117

WONNACOTT TH (1961)
SPECTRAL ANALYSIS COMBINING A BARTLETT WINDOW WITH AN
ASSOCIATED INNER WINDOW
TECHNO,V3,P235

YULE GU (1927)
ON A METHOD OF INVESTIGATING PERIODICITIES IN DISTURBED SERIES
WITH SPECIAL REFERENCE TO WOLFEX'S SUNSPOT NUMBERS
PHIL. TRANS. A.V226,P267

JOURNAL ABBREVIATIONS

ACT. MATH.	-----	ACTA MATHEMATICA
ANN. INST. MATH. STAT.	-----	ANNALS OF THE INSTITUTE OF STATISTICAL MATHEMATICS
ANN. STAT.	-----	ANNALS OF STATISTICS
ANN. MATH. S.AT.	-----	ANNALS OF MATHEMATICAL STATISTICS
APP. STAT.	-----	APPLIED STATISTICS
BIOA	-----	BIOMETRIKA
COMM. STAT.	-----	COMMUNICATIONS IN STATISTICS
COMP. J.	-----	COMPUTER JOURNAL
ECO. MOD.	-----	ECOLOGICAL MODELLING
IEEE TRANS. AUD. ELECT.	-----	IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS
IEEE TRANS. AUTO. CON.	-----	IEEE TRANSACTIONS ON AUTOMATIC CONTROL
INT. STAT. REV.	-----	INTERNATIONAL STATISTICAL REVIEW
JASA	-----	JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
JFO	-----	JOURNAL OF FORECASTING
JHY	-----	JOURNAL OF HYDROLOGY
JORS	-----	JOURNAL OF THE OPERATIONS RESEARCH SOCIETY
JRSSA	-----	JOURNAL OF THE ROYAL STATISTICAL SOCIETY,SERIES A
JRSSB	-----	JOURNAL OF THE ROYAL STATISTICAL SOCIETY,SERIES B
JTSA	-----	JOURNAL OF TIME SERIES ANALYSIS
PHIL. TRANS.	-----	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON, SERIES A
PROC. RSSA	-----	PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON,SERIES A
TECHNO	-----	TECHNOMETRICS
WRR	-----	WATER RESOURCES RESEARCH



Author Kelly G S

Name of thesis Review and application of selected topics in discrete univariate time-series analysis 1984

PUBLISHER:

University of the Witwatersrand, Johannesburg

©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.