INFERENCE PROBLEMS FOR VECTOR LINEAR
TIME SERIES MODELS

by D. F. Nicholls

A thesis submitted for the degree of
Doctor of Philosophy in the Australian National University
The results presented in this thesis are my own except where otherwise stated.

[Signature]

1971.
ACKNOWLEDGEMENTS

I was introduced to the subject of this thesis by my supervisor, Professor E.J. Hannan, who suggested many of the problems considered here and who guided me throughout their investigation. I am greatly indebted to him.

I would also like to express my gratitude to the many other people whose help has contributed, in one way or another, to the work within this thesis, and apologize for the fact that their large number makes it difficult to mention each separately. However I would like to thank Mrs J. Radley in particular for her excellent typing of this manuscript. Finally I would like to thank the ANU Computer Centre for the use of their facilities.
# CONTENTS

Acknowledgements

Summary

## CHAPTER 1 INTRODUCTION

1.1 Preliminary

1.2 Linear Models

1.3 Estimation of the Parameters of Finite Linear Models

## CHAPTER 2 A GOODNESS OF FIT TEST FOR VECTOR AUTOREGRESSIVE MODELS

2.1 Introduction

2.2 Derivation of the Test

2.3 The Computational Procedure

2.4 A Comparison of Goodness of Fit Tests

2.5 Conclusion

## CHAPTER 3 THE ESTIMATION OF MOVING AVERAGE MODELS

3.1 Introduction

3.2 Convergence of the Iterative Procedure

3.3 An Extension of the Estimation Procedure

3.4 Alternative Estimation Procedures

3.5 Conclusion

Appendix
<table>
<thead>
<tr>
<th>CHAPTER 4</th>
<th>THE ESTIMATION OF MIXED AUTOREGRESSIVE MOVING AVERAGE MODELS WITH EXOGENOUS VARIABLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.2</td>
<td>The Estimation Procedure</td>
</tr>
<tr>
<td>4.3</td>
<td>The Distribution of the Estimates</td>
</tr>
<tr>
<td>4.4</td>
<td>The Computational Procedure</td>
</tr>
<tr>
<td>4.5</td>
<td>Numerical Results</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 5</th>
<th>DISTRIBUTED LAG MODELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>5.2</td>
<td>A Test for the Equality of the Parameters</td>
</tr>
<tr>
<td>5.3</td>
<td>Estimation and Properties of the Coefficients</td>
</tr>
<tr>
<td>5.4</td>
<td>Summary</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 6</th>
<th>EXTENSIONS TO THE VECTOR MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>The Vector Model</td>
</tr>
<tr>
<td>6.2</td>
<td>The Estimation Procedure and the Distribution of the Estimates</td>
</tr>
<tr>
<td>6.3</td>
<td>The Computational Procedure</td>
</tr>
<tr>
<td>6.4</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>

Bibliography 145
This thesis presents an investigation of a number of inference problems associated with linear time series models. Two linear models which are of fundamental importance in the analysis of time series are the moving average model and the autoregressive model. It is models such as these, and models constructed from them, with which we shall be concerned. The chapters may be summarized as follows: after a brief outline of the basic concepts of spectral theory chapter 1 is concerned with a discussion of the properties of a number of linear time series models. This introductory chapter concludes with a description of the estimation procedure, based on spectral methods, which is applied in the latter part of this thesis to obtain efficient estimates of the coefficients of mixed regression, autoregression, moving average and distributed lag models.

In chapter 2 a goodness of fit test is derived for the vector autoregressive model and an outline given of the computational procedure to be followed when applying the test. This test is compared with various other goodness of fit tests (for autoregressive models) occurring in the literature, the advantages and disadvantages of each being discussed.

Chapter 3 commences with a description of a procedure (due to Hannan [25]) for the efficient estimation of the coefficients of moving average models. After examining the rate of convergence of the iterative procedure (in the case of the first order moving average model) associated with this method, an extension of Hannan's procedure is described. This extension is introduced so as to prevent the estimation procedure from, on some occasions, giving estimates which fail to converge as the iterations proceed.
Finally in this chapter Hannan's estimation procedure is compared with two other methods due to Durbin [14].

Chapter 4 is concerned with the derivation of a method for the estimation of the coefficients of mixed autoregressive moving average models with exogenous variables. This procedure, which is based on the method of maximum likelihood, gives estimates which are shown to be asymptotically normally distributed and (asymptotically) efficient. After outlining the computational procedure for the application of this method, the results of a number of numerical experiments, using generated data, are presented in section 4.5. As we see from that section, these results help to illustrate the theoretical results obtained.

By placing constraints on the coefficients of the models considered in chapter 4 it is possible to efficiently estimate the coefficients of certain classes of rational distributed lag models. These models are discussed in chapter 5 and the associated estimation formulae derived. Finally, in chapter 6, the estimation procedure developed in chapter 4 is extended to consider the efficient estimation of the coefficients of vector mixed autoregressive moving average models with exogenous variables.
CHAPTER 1
INTRODUCTION

1.1 Preliminary

We shall in this section introduce the basic definitions and terminology which will be used throughout this thesis. In general no attempt will be made to present rigorous derivations of the theory to be outlined since the current literature contains an increasing number of excellent introductions to the spectral theory of time series (e.g. Whittle [55], Grenander and Rosenblatt [18], Hannan [22], [26], Jenkins and Watts [31], Parzen [44] and Anderson [3] to name a few). For the purpose of this thesis Hannan [26] has been taken as the basic reference and we shall, in what follows, make reference to many of the results derived in that book.

Let us consider a time series \( \{x(n); n = 0, 1, \ldots \} \), where \( x(n) \) is a vector of \( p \) components taken at time \( n \), i.e. \( x'(n) = (x_1(n) \ldots x_p(n)) \).

We shall (unless otherwise stated) assume that the mean is zero so that for \( j = 1, \ldots, p; \ n = 0, 1, \ldots \),

\[
E(x_j(n)) = 0.
\]

Furthermore we define the covariance matrix of this process to be \( \Gamma(m,n) \), where

\[
E(x(m)x'(n)) = \Gamma(m,n).
\]

The series \( \{x(n); n = 0, 1, \ldots \} \) is said to be second order (or wide sense) stationary if

\[
E(x(m)x'(n)) = \Gamma(0,n-m) = \Gamma(n-m) \quad \text{say},
\]  

(1.1.1)
i.e. the covariance matrix depends only on the distance $(n-m)$ between the two observations being considered. If $\gamma_{jk}(n)$ is the element in row $j$ column $k$ $(j,k = 1, \ldots, p)$ of $\Gamma(n)$ then $\gamma_{jk}(n) = \gamma_{kj}(-n)$, i.e. $\Gamma(n) = \Gamma'(-n)$ (where $\Gamma'(n)$ is the transpose of the matrix $\Gamma(n)$). It is often convenient to work in terms of the serial correlations $\rho_{jk}(n)$ (which are scale free), where

$$\rho_{jk}(n) = \gamma_{jk}(n) / (\gamma_{jj}(0)\gamma_{kk}(0))^{1/2}.$$ 

For $j = k$ we call these the autocorrelations and use the notation

$$\rho_k(n) = \gamma_k(n) / \gamma_k(0).$$

If $\Gamma(n)$ is the covariance matrix of a stationary vector process, then by Bochner's theorem

$$\Gamma(n) = \int_{-T}^{T} e^{i\lambda} dF(\lambda), \quad (1.1.2)$$

where $F(\lambda)$ is a matrix with Hermitian non-negative increments i.e. $F(\lambda_1) - F(\lambda_2) \leq 0$, $\lambda_1 \leq \lambda_2$. This function $F(\lambda)$ is uniquely defined if we require that (i) $\lim_{\lambda \to \infty} F(\lambda) = 0$, (ii) $F(\lambda)$ is continuous from the right. It is called the spectral distribution matrix and since it is Hermitian $F(\lambda) = F'(-\lambda)$. Rewriting $dF(\lambda)$ in the form

$$dF(\lambda) = \frac{1}{2} (dC(\lambda) - idQ(\lambda)) \quad \lambda > 0,$$

$$= dC(\lambda) \quad \lambda = 0,$$

(1.1.2) becomes

---

1 In what follows, when we speak of stationarity we shall, unless otherwise stated, mean second order stationarity.
\[ \Gamma(n) = \int_0^{\pi} \left( \cos n\lambda \, dC(\lambda) + \sin n\lambda \, dQ(\lambda) \right). \]

\(C(\lambda)\) and \(Q(\lambda)\) are referred to as the cospectral and quadrature spectral distribution matrices respectively.

Suppose \(F(\lambda)\) is assumed to be absolutely continuous. Then (1.1.2) may be expressed in the form

\[ \Gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda \]

\[ = \int_0^{\pi} (c(\lambda) \cos n\lambda + q(\lambda) \sin n\lambda) d\lambda, \]

(1.1.3)

where now \(f(\lambda), c(\lambda), q(\lambda)\) are respectively the spectral density, the cospectral density and the quadrature spectral density matrices. Furthermore from (1.1.3) it follows that

\[ f(\lambda) = (2\pi)^{-1} \sum_{n=-\infty}^{\infty} \Gamma(n) e^{-in\lambda}. \]

(1.1.4)

Just as (1.1.2) represents a transformation of the covariance matrix from the time to the frequency domain, there exists a similar transformation (or spectral representation) for \(x(n)\). Indeed we may let

\[ x(n) = \int_{-\pi}^{\pi} e^{-in\lambda} dz(\lambda) \]

\[ = \int_0^{\pi} \left( \cos n\lambda \, d\xi(\lambda) + \sin n\lambda \, d\eta(\lambda) \right), \]

(1.1.5)

where \(z(\lambda)\) is a complex vector process of orthogonal increments with

\[ dz(\lambda) = \frac{1}{\lambda} (d\xi(\lambda) + id\eta(\lambda)) \quad \lambda > 0, \]

\[ = d\xi(\lambda) \quad \lambda = 0, \]

and
\[ E(dz(\lambda)dz^*(\mu)) = \delta_{\lambda,\mu} dF(\lambda). \]

dz*(\mu) represents the transposed conjugate of dz(\mu), while \( \delta_{\lambda,\mu} \) is the usual Kronecker function i.e. \( \delta_{\lambda,\mu} \) is unity when \( \lambda = \mu \) and zero otherwise. Furthermore the only non zero covariances of \( d\xi(\lambda), d\eta(\lambda) \) are

\[ E(d\xi(\lambda)d\xi'(\mu)) = E(d\eta(\lambda)d\eta'(\mu)) = \delta_{\lambda,\mu} dC(\lambda), \]

\[ E(d\xi(\lambda)d\eta'(\mu)) = -E(d\eta(\lambda)d\xi'(\mu)) = \delta_{\lambda,\mu} dQ(\lambda). \]

In practice \( C(\lambda) \) and \( Q(\lambda) \) are replaced by two other expressions, which we shall discuss only for the absolutely continuous case (i.e. that where \( dF(\lambda) = f(\lambda)d\lambda \), so that \( dC(\lambda) = c(\lambda)d\lambda \) and \( dQ(\lambda) = q(\lambda)d\lambda \)). We introduce the quantities

\[ \sigma^2_{jk}(\lambda) = \frac{c_{jk}(\lambda) + q_{jk}(\lambda)}{c_{jj}(\lambda)c_{kk}(\lambda)} = \frac{|f_{jk}(\lambda)|^2}{f_{jj}(\lambda)f_{kk}(\lambda)}, \]

\[ \theta_{jk}(\lambda) = \tan^{-1}\left(\frac{q_{jk}(\lambda)}{c_{jk}(\lambda)}\right), \]

where \( c_{jk}(\lambda), q_{jk}(\lambda) \) and \( f_{jk}(\lambda) \) are the elements in row \( j \) column \( k \) \((j,k = 1,\ldots,p)\) of \( c(\lambda), q(\lambda) \) and \( f(\lambda) \) respectively. \( \sigma_{jk}(\lambda) \) and \( \theta_{jk}(\lambda) \) are called the coherence and the phase respectively and are two characteristics measuring the dependence of \( x_j(n) \) on \( x_k(n) \). (Indeed the coherence measures the strength of association in the sense of maximum correlation obtained by rephasing \( x_k(n) \) relative to \( x_j(n) \) by \( \theta_{jk}(\lambda) \)).

Suppose we now consider a sample of size \( N \) from the vector time series \( x(n) \) i.e. \( \{x(n); n = 1,\ldots,N\} \). We estimate the covariance matrix \( \Gamma(n) \) by
\[ C(n) = (N-n)^{-1} \sum_{m=1}^{N-n} x(m)x'(m+n) \quad n \neq 0, \quad (1.1.6) \]
\[ C(-n) = C'(n). \]

(The elements of \( C(n) \) are, at times, referred to as the sample serial covariances while the diagonal elements are called the sample autocovariances). \( C(n) \) is sometimes defined with the divisor \((N-n)\) replaced by \(N\). Although the division by \(N\) makes \( C(n) \) a biased estimator, as Parzen [45] points out this alternative estimate of \( \Gamma(n) \) has two desirable properties. Firstly it leads to a smaller mean square error for a given sample than does the unbiased one and secondly the biased estimator is positive definite, whereas the unbiased estimator is not. (For some matrix \( \tilde{C}(n) \), if \( a'\tilde{C}(n)a = c_a(n) \) (say) for all \( a \neq 0 \), then \( \tilde{C}(n) \) is positive definite if
\[ \sum_{\ell,m=1}^{N} c_a(\ell-m)z(\ell)z(m) > 0, \]
for some \( z(\ell) \) not all zero. If
\[ \tilde{C}(n) = N^{-1} \sum_{m=1}^{N-n} x(m)x'(m+n), \]
this condition is obviously satisfied). Furthermore, as Hannan ([26] p 210) shows, under fairly general conditions the sample covariance \((1.1.6)\) converges almost surely to the true covariance \( \Gamma(n) \). (We shall discuss this more fully later when dealing with particular linear models).

If we take the finite Fourier transform of the \( x(n) \) series so as to obtain
\[ w(\lambda_t) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} x(n)e^{in\lambda_t}, \]
with $\lambda_t = \lambda_{N-t} = 2\pi t/N; t = 1, \ldots, [N/2]$, then the periodogram ordinate at frequency $\lambda_t$ is defined as

$$I(\lambda_t) = w(\lambda_t)w^*(\lambda_t).$$

Furthermore, if the spectral density is continuous at $\lambda$, as $N \to \infty$ it is not hard to show that

$$E(I(\lambda)) = f(\lambda).$$

Although the periodogram is an inconsistent estimate of the spectral density (see Hannan [22] p 53) it still plays an important role in estimation problems in time series.

Finally in this first section we shall introduce some notation and results from matrix theory which will be constantly used throughout this thesis. If $A$ is an $(m \times n)$ matrix then by $A'$ we mean the transpose of $A$, $\overline{A}$ represents the complex conjugate of $A$ and $A^*$ the transposed conjugate of $A$ (i.e. $A^* = \overline{A}'$). $\|A\|$ is the norm of the matrix $A$, which we may take to be the positive square root of the greatest eigenvalue of $AA^*$, while $\text{tr}(A)$ is the trace of the matrix $A$. If a positive definite matrix $F$ is partitioned as

$$\begin{bmatrix}
A & \cdots & D \\
D' & \ddots & \vdots \\
& \cdots & E
\end{bmatrix},$$

where $A$ and $E$ are square, then $F^{-1}$ can be written in the form

---

2 By $[N/2]$ we mean the largest integer equal to or less than $N/2$. 

Furthermore some simplification shows that for the $(1,1)$ block of $F^{-1}$ we have

$$A^{-1}A^{-1}D(E-D'A^{-1}D)^{-1}D'A^{-1} = (A-DE^{-1}D')^{-1}. \quad (1.1.9)$$

We call vec$(A)$ the vector formed from the matrix $A$ by writing the columns of $A$ below each other from left to right. Consequently for the $(m \times n)$ matrix $A$, vec$(A)$ has $a_{jk}$ in row $(k-1)m+j$, $j = 1, \ldots, m; \ k = 1, \ldots, n$. If $B$ is a $(p \times q)$ matrix then by $A \otimes B$ we mean the tensor (or Kronecker) product of two matrices, which is the $(mp \times nq)$ partitioned matrix

$$\begin{bmatrix}
a_{11}B & a_{12}B & \cdots & a_{1n}B \\
a_{21}B & a_{22}B & \cdots & a_{2n}B \\
\cdots & \cdots & \cdots & \cdots \\
a_{m1}B & a_{m2}B & \cdots & a_{mn}B
\end{bmatrix} \quad (1.1.10)$$

Thus $A \otimes B$ has $a_{ij}b_{kl}$ in row $(i-1)p+k$, column $(j-1)q+l$, $i = 1, \ldots, m; \ j = 1, \ldots, n; \ k = 1, \ldots, p; \ l = 1, \ldots, q$. One result which proves useful in what follows (and the proof of which may be found in Neudecker [40]) is that if $A$, $B$ and $C$ are matrices such that the matrix product $ABC$ is defined, then

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B). \quad (1.1.11)$$

(When we come to consider vector linear models we shall be able to express many of our results much more succinctly by making use of this tensor notation).
1.2 Linear Models

In the analysis of time series (and more specifically prediction theory) uncorrelated processes form a basic building block in the representation of a wide range of stationary processes. The starting point for the consideration of such processes is the Wold decomposition theorem (see Wold [60]) which states that every second order stationary process $x(n)$ (we shall initially consider the scalar case for simplicity) may be represented as

$$x(n) = u(n) + v(n) = \sum_{j=0}^{\infty} \alpha(j)\epsilon(n-j) + v(n), \quad (1.2.1)$$

where

(i) $E(\epsilon(m)\epsilon(n)) = \delta_{m,n} \sigma^2$,

(ii) $\alpha(0) = 1$, $\sum_{j=0}^{\infty} \alpha^2(j) < \infty$,

(iii) $E(\epsilon(m)v(n)) = 0$,

(iv) the $v(n)$ process is deterministic i.e. $v(n)$ may be determined exactly from an infinite number of past values $v(m)$, $m \leq (n-1)$.

If $\sigma^2 > 0$ and $v(n)$ is absent then $x(n)$ is said to be purely nondeterministic and has the moving average representation

$$x(n) = \sum_{j=0}^{\infty} \alpha(j)\epsilon(n-j).$$

In this case the $x(n)$ process has an absolutely continuous density function $f_x(\lambda)$ given by

$$f_x(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \alpha(j)e^{ij\lambda} \right|^2.$$
The Wold decomposition theorem requires only that the \( \epsilon(n) \) are uncorrelated. For the purpose of this thesis however we shall require the \( \epsilon(n) \) to be identically and independently distributed. Indeed we shall be concerned with linear processes of the form

\[
x(n) = \sum_{j=0}^{\infty} \alpha(j) \epsilon(n-j), \quad \sum_{j=0}^{\infty} \alpha^2(j) < \infty, \quad \alpha(0) = 1,
\]

where the \( \epsilon(n) \) are identically and independently distributed with mean zero and variance \( \sigma^2 \), which we paraphrase as i.i.d. \((0, \sigma^2)\). (In a recent paper Hannan and Heyde [30] have shown that, subject to some reasonable additional conditions, the classical theory of inference goes through if this independence assumption is replaced by the weaker condition \( \mathbb{E}(\epsilon(n)|F_{n-1}) = 0 \), almost surely, for all \( n \), where \( F_n \) is the \( \sigma \)-field generated by the \( \epsilon(m), m \leq n \). In the stationary case for example their additional conditions are the regularity condition \( \sum_{j=1}^{\infty} j \alpha^2(j) < \infty \), together with the requirement \( \mathbb{E}(\epsilon^2(n)|F_{n-1}) = \sigma^2 \), almost surely).

Let us consider a moving average model of order \( p \), i.e.

\[
x(n) = \sum_{j=0}^{p} \alpha(j) \epsilon(n-j), \quad \sum_{j=0}^{p} \alpha^2(j) < \infty, \quad \alpha(0) = 1,
\]

where the \( \epsilon(n) \) are i.i.d. \((0, \sigma^2)\). For this model, in order that the \( \epsilon(n) \) be the prediction errors (in which case the model will be uniquely identified), we require all the zeros of the \( z \) transform

\[
g(z) = \sum_{j=0}^{p} \alpha(j) z^j
\]

to lie on or outside of the unit circle (see Hannan [26] chapter 3 for a discussion of prediction theory). When determining the asymptotically efficient estimates of the coefficients of the mixed autoregressive moving average model with exogenous variables (in chapter 4), for the asymptotic
theory to go through we shall in fact require all the zeros of \( g(z) \) to lie outside of the unit circle.

Suppose \( \hat{\alpha}(j) \) represents an estimate of \( \alpha(j) \), \( j = 1, \ldots, p \), for the model (1.2.3). If the spectral density estimate satisfies

\[
\hat{f}_x(\lambda) = \frac{1}{(2\pi)} \sum_{j=-p}^{p} c(j)e^{ij\lambda} \geq 0, \quad \text{for all } \lambda,
\]

(where the \( c(j) \) are the sample covariances of the \( x(n) \) process) and all the zeros of \( g(z) \) lie outside of the unit circle, then we may obtain the \( \hat{\alpha}(j) \) from

\[
\hat{f}_x(\lambda) = \left(\frac{\sigma^2}{2\pi}\right) \sum_{j=-p}^{p} \hat{\alpha}(j)e^{ij\lambda} \left|\frac{1}{2\pi} \sum_{j=-p}^{p} c(j)e^{ij\lambda}\right|^2, \quad \hat{\alpha}(0) = 1. \quad (1.2.4)
\]

By expressing the \( \hat{\alpha}(j) \) in terms of the autocorrelations of the \( x(n) \) process it can be shown (for a sample of size \( N \)) that the \( \sqrt{N}(\hat{\alpha}(j) - \alpha(j)) \), \( j = 1, \ldots, p \), are asymptotically normally distributed (see Anderson [3], section 5.7). Furthermore for \( x(n) \) defined by (1.2.3) it follows that the sample covariance (of the \( x(n) \) process) converges almost surely to the true covariance. Thus from (1.2.4) it follows that \( \hat{\alpha}(j) \) converges almost surely to \( \alpha(j) \).

Since it is often desirable to express \( x(n) \) in terms of its past values we are led to consider the autoregressive model (of order \( q \), say)

\[
\sum_{k=0}^{q} \beta(k)x(n-k) = \varepsilon(n) \quad , \quad \beta(0) = 1, \quad (1.2.5)
\]

where the \( \varepsilon(n) \) are i.i.d. \( (0, \sigma^2) \). For this model the spectral density of the \( x(n) \) process is given by

\[
f_x(\lambda) = \left(\frac{\sigma^2}{2\pi}\right) \left|\sum_{k=0}^{q} \beta(k)e^{ik\lambda}\right|^{-2}. \quad (1.2.6)
\]
Furthermore, in order that the $e(n)$ be the prediction errors we require that the corresponding $z$ transform

$$h(z) = \sum_{k=0}^{q} \beta(k)z^k$$

have all its zeros outside of the unit circle. Assuming this condition to be satisfied, then the autoregressive model may be expressed in terms of an infinite moving average of present and past values of the $e(n)$ with coefficients which converge exponentially to zero, and so the $x(n)$ process is stationary.

Suppose $\hat{\beta}(k)$ represents the estimate of the autoregressive coefficient $\beta(k), k = 1, \ldots, q,$ obtained from the Yule-Walker equations

$$\sum_{k=0}^{q} \hat{\beta}(k)e(k-j) = \delta_{0,j}\sigma^2, \quad \hat{\beta}(0) = 1, \quad j \geq 0.$$  

Then as Mann and Wald [38] prove, for a sample of size $N$, the $\sqrt{N}(\hat{\beta}(k)-\beta(k))$ are asymptotically normally distributed. (In their proof these authors assume that all moments of the $e(n)$ are finite. It is not hard to show however (see Hannan [26] pp 329-332 for example) that the finiteness of moments higher than the second is not needed). Furthermore, since the sample covariance of the $x(n)$ process (defined by (1.2.5)) converges almost surely to the true covariance, $\hat{\beta}(k)$ converges almost surely to $\beta(k), k = 1, \ldots, q.$

Combining the two models (1.2.3) and (1.2.5) we obtain the mixed autoregressive moving average model which may be written in the form

$$\sum_{k=0}^{q} \beta(k)x(n-k) = \sum_{j=0}^{P} \alpha(j)e(n-j), \quad \alpha(0) = \beta(0) = 1,$$  

(1.2.7)
where the $\epsilon(n)$ are i.i.d. $(0, \sigma^2)$. If all the zeros of the z transform

$$h(z) = \sum_{k=0}^{q} \beta(k)z^k$$

lie outside of the unit circle then the $x(n)$ process is stationary with spectral density

$$f_x(\lambda) = \frac{\sigma^2}{2\pi} \left( \sum_{j=0}^{p} |\alpha(j)e^{i\lambda j}|^2 / \sum_{k=0}^{q} |\beta(k)e^{ik\lambda}|^2 \right).$$

Furthermore in order that (1.2.7) be uniquely identified we require that all the zeros of

$$g(z) = \sum_{j=0}^{p} \alpha(j)z^j$$

lie on or outside of the unit circle and $g(z), h(z)$ must have no zeros in common.

In this section we have, to date, been concerned exclusively with scalar linear models, though all the results carry through to their vector counterparts. Indeed suppose we consider the vector moving average model

$$x(n) = \sum_{j=0}^{p} A(j)\epsilon(n-j) , \quad A(0) = I , \quad \sum_{j=0}^{p} \|A(j)\|^2 < \infty, \quad (1.2.8)$$

where now $x(n), \epsilon(n)$ are vectors of $v$ components while the $A(j), j = 1, \ldots, p$ are square matrices. The $\epsilon(n)$ are assumed to be i.i.d. $(0, \sigma)$. In order that this model be uniquely identified we assume that all the zeros of the determinant

$$\det(\sum_{j=0}^{p} A(j)z^j)$$

lie on or outside of the unit circle. The spectral density of the $x(n)$ process defined by (1.2.8) is given by
The vector autoregressive model may be expressed in the form

\[ q \sum_{k=0}^{q} B(k)x(n-k) = \epsilon(n) , \quad B(0) = I , \quad (1.2.9) \]

where \( x(n) \), \( \epsilon(n) \) are vectors of \( v \) components, the \( B(k) \) \( (k = 1, \ldots, q) \) are square matrices and the \( \epsilon(n) \) are assumed to be i.i.d. \( (0, G) \). In this case, in order that the \( x(n) \) process be stationary, we require that the determinant

\[ \det \left( \sum_{k=0}^{q} B(k)z^{-k} \right) \]

have all its roots outside of the unit circle. For this model

\[ f_x(\lambda) = (2\pi)^{-1} \left( \sum_{k=0}^{q} B(k)e^{ik\lambda} \right)^{-1} \left( \sum_{k=0}^{q} B(k)e^{ik\lambda} \right)^{*} . \]

Suppose \( \beta = \text{vec}(B(1); B(2); \ldots; B(q)) \) is the vector of \( vq \) components (defined below (1.1.9)), while \( \hat{\beta} \) is an estimate of \( \beta \) obtained from the Yule-Walker equations (for the vector model), i.e.

\[ \sum_{k=0}^{q} \hat{B}(k)C(k-j) = \delta_{0,j}\hat{G} , \quad j \geq 0 , \quad \hat{B}(0) = I . \]

Then for a sample of size \( N \) it is well known that \( \sqrt{N}(\hat{\beta} - \beta) \) has a distribution which converges, as \( N \to \infty \), to that of a normally distributed vector of random variables, and furthermore, \( \hat{\beta} \) converges almost surely to \( \beta \).

In practice when fitting an autoregressive model to a set of data we must decide on the order of the autoregression to be fitted (i.e. the size of \( q \) in (1.2.9) for example). A number of goodness of fit tests have been
developed for testing the order of such models. We shall, in chapter 2, derive such a test for the vector autoregressive model (1.2.9) and compare this test with other goodness of fit tests occurring in the literature.

The vector mixed autoregressive moving average model is of the form

$$\sum_{k=0}^{q} B(k)x(n-k) = \sum_{j=0}^{p} A(j)\epsilon(n-j), \quad A(o) = B(o) = I, \quad (1.2.10)$$

where \(x(n)\) and \(\epsilon(n)\) are vectors of \(v\) components, \(A(j), B(k)\) \((j = 1, \ldots, p; k = 1, \ldots, q)\) are square matrices and the \(\epsilon(n)\) are assumed to be i.i.d. \((0, G)\). In order that the \(x(n)\) process defined by (1.2.10) be stationary we require that all the zeros of the determinant

$$\det(\sum_{k=0}^{q} B(k)z^k) \quad (1.2.11)$$

lie outside of the unit circle. The spectral density of the \(x(n)\) process for this mixed model is given by

$$f_x(\lambda) = (2\pi)^{-1}(\sum_{k=0}^{q} B(k)e^{ik\lambda})^{-1}(\sum_{j=0}^{p} A(j)e^{ij\lambda}G(\sum_{j=0}^{p} A(j)e^{ij\lambda})^*(\sum_{k=0}^{q} B(k)e^{ik\lambda})^*^{-1}. \quad (1.2.12)$$

Furthermore, in order that the model (1.2.10) be uniquely identified, all the zeros of the determinant

$$\det(\sum_{j=0}^{p} A(j)z^j) \quad (1.2.12)$$

must lie on or outside of the unit circle. In addition we require that (1.2.11) and (1.2.12) have no zeros in common and the partitioned matrix

$$[A(p) : B(q)]$$

be of rank \(v\) (for a discussion of these conditions see Hannan [27]). In the latter part of this thesis, we shall be interested in efficiently
estimating the coefficients of such models as the mixed autoregressive moving average model with exogenous variables (i.e. variables which are independent of the \(e(n)\)), in both the scalar and vector case, and the rational distributed lag model.

1.3 Estimation of the Parameters of Finite Linear Models

The procedure which we shall use to efficiently estimate the coefficients of the models to be considered in this thesis is asymptotically equivalent to the method of maximum likelihood. If the data is Gaussian then we may calculate the maximum likelihood estimates and consider their limiting distribution. The estimates may be used however whether the data is Gaussian or not. Indeed let \(\{x(n); n = 0, 1, \ldots\}\) represent a scalar stationary nondeterministic time series. Then Whittle [57] and Walker [54] have proved (independently of each other) that if \(x(n)\) is a linear process with finite fourth moments i.e.

\[
x(n) = \sum_{j=0}^{\infty} a(j)e(n-j), \quad \sum_{j=0}^{\infty} a^2(j) < \infty,
\]

where the \(e(n)\) are i.i.d. \((0, \sigma^2)\) and \(E(e^4(n)) < \infty\), then the asymptotic distribution of the estimates obtained in this way will turn out to be independent of the Gaussian assumption. (Using the results of Hannan ([26] pp 329-332) however it can be shown that for this result to hold the finiteness of moments higher than the second is not required).

Suppose we are considering the estimation of the coefficients of a particular linear model. If it can be shown that the estimates obtained in the manner described above have (asymptotically) the same distribution as the maximum likelihood estimates, given only that the \(e(n)\) are
i.i.d. \((0, \sigma^2)\), then we say that our estimates are asymptotically efficient.  

Let us consider a sample of size \(N\) from the \(x(n)\) process. Then assuming that the \(e(n)\) are i.i.d. \((0, \sigma^2)\), we may set up the associated "likelihood function". It can be shown that the logarithm of this function is dominated by a function of the quadratic form \(x'\Gamma^{-1}_N x\) (see for example Walker [54]), where \(x\) is a vector of \(N\) components with \(x(j)\) in the \(j\)th place, and \(\Gamma_N\) is of the form

\[
\Gamma_N = \begin{bmatrix}
\gamma_x(0) & \ldots & \gamma_x(N-1) \\
\vdots & \ddots & \vdots \\
\gamma_x(N-1) & \ldots & \gamma_x(0)
\end{bmatrix},
\]

with \(\gamma_x(j) = \text{E}(x(k)x(k+j))\). Consequently maximizing this "likelihood" is asymptotically equivalent to minimizing \(x'\Gamma^{-1}_N x\).

The quadratic form \(x'\Gamma^{-1}_N x\) is not an easy expression to deal with however. In order to make it more manageable we Fourier transform the data and consider the estimation problem via spectral methods. Indeed proceeding heuristically, suppose \(U\) is the \((N \times N)\) unitary matrix with \(N^{-\frac{1}{2}}\exp(in\lambda_t)\) in row \(t\) column \(n\), where \(\lambda_t = 2\pi t/N\). Then since \(U^* = U^{-1}\), it follows that

\[
x'\Gamma^{-1}_N x = x'U^*(U\Gamma_N U^*)^{-1}Ux.
\]

---

3 Throughout this thesis when we speak of efficiency we shall, unless otherwise stated, always mean efficiency in the asymptotic sense.
The matrix $U_N^{-1}U^*$ is, to order $N^{-1}$, diagonal with $2\pi f_x(\lambda_u)$ in the $u$th place in the main diagonal. In order to verify this we consider the element in row $u$ column $v$ ($u, v = 1, \ldots, N$) of $U_N^{-1}U^*$, which is given by

$$
\frac{1}{N} \sum_{m,n=1}^{N} \gamma_x(n-m)e^{-i(m\lambda_u-n\lambda_v)} = \sum_{j=-N+1}^{N-1} \gamma_x(j)e^{-ij\lambda_u} \frac{1}{N} \sum_{j=-N+1}^{N-1} \Sigma(j)e^{i(\lambda_u-\lambda_v)} \quad (1.3.1)
$$

where, in the sum $\Sigma(j)$, $n$ lies between 1 and $N-j$ for $j$ positive and $-j+1$ and $N$ for $j$ negative or zero. For $\lambda_u \neq \lambda_v$ we have

$$\left| \sum_{j=-N+1}^{N-1} \gamma_x(j)e^{-ij\lambda_u} \frac{1}{N} \sum_{j=-N+1}^{N-1} \Sigma(j)e^{i(\lambda_u-\lambda_v)} \right| \leq \frac{1}{N} \sum_{j=-N+1}^{N-1} |j| |\gamma_x(j)| = O(N^{-1}),$$

since

$$\left| \frac{1}{N} \sum_{j=-N+1}^{N-1} \Sigma(j)e^{i(\lambda_u-\lambda_v)} \right| = \frac{\sin\left(\frac{1}{2}(N-|j|)(\lambda_u-\lambda_v)\right)}{N\sin\left(\frac{1}{2}(\lambda_u-\lambda_v)\right)} \leq \frac{1}{N} \quad (1.3.1),$$

and for the models we are to consider, $\gamma_x(j)$ converges exponentially to zero. If $\lambda_u = \lambda_v$ (1.3.1) becomes

$$\sum_{j=-N+1}^{N-1} (1 - \frac{|j|}{N}) \gamma_x(j)e^{-ij\lambda_u} = 2\pi f_x(\lambda_u) + O(N^{-1}).$$

It thus follows that $U_N^{-1}U^*$ is, to order $N^{-1}$, a diagonal matrix with $2\pi f_x(\lambda_u)$ in the $u$th place ($u = 1, \ldots, N$), as required.

Since $x'T_N^{-1}x$ is scalar we have

$$x'T_N^{-1}x = \text{tr}(x'T_N^{-1}x)$$

$$= \text{tr}((Ux)^*(U_N^{-1}U^*)^{-1}(Ux))$$

$$= \text{tr}((U_N^{-1}U^*)^{-1}(Ux)(Ux)^*).$$
But

\[(Ux)(Ux)^* = [N^{-1} \sum_{m,n=1}^{N} x(m)x(n) e^{i(m\lambda_s - n\lambda_t)}],\]

so that, to our order of approximation,

\[x'\Gamma_N^{-1}x = \text{tr} \left\{ (2\pi)^{-1} \begin{pmatrix} f^{-1}_x(\lambda_1) & 0 \\ 0 & f^{-1}_x(\lambda_N) \end{pmatrix} (Ux)(Ux)^* \right\} \]

\[= \sum_t f^{-1}_x(\lambda_t) I_x(\lambda_t), \tag{1.3.2} \]

where \(\lambda_t = \lambda_{N-t} = 2\pi t/N, \ t = 1, \ldots, \lfloor N/2 \rfloor\) and \(I_x(\lambda_t)\) is the periodogram ordinate of the \(x(n)\) process at frequency \(\lambda_t\), i.e.

\[I_x(\lambda_t) = (2\pi N)^{-1} \sum_{m,n=1}^{N} x(m)x(n)e^{-i(n-m)\lambda_t}.\]

Consequently minimizing \(x'\Gamma_N^{-1}x\) is asymptotically equivalent to minimizing (1.3.2).

We emphasize once more that the arguments we have presented here are only of a heuristic nature. For the models to be considered in this thesis (i.e. models of the form (1.2.7) for example) E.J. Hannan (in a yet to be published paper) has presented a rigorous proof of these (and other) results (which are more general than those proved by Walker [54]). Indeed under fairly weak conditions this author proves the following:

\[\text{For convenience of notation we shall, throughout this thesis, usually write } \sum_{t=1}^{N} \text{ for } \sum_{t=1}^{N}.\]
(i) The function

\[
(\sigma^2/2\pi N) \sum_t \{r^{-1}(\lambda_t)I(\lambda_t)\}
\]  

(1.3.3)

converges almost surely to a function \( Q(\theta) \) which has a single maximum at \( \theta = \theta_0 \), where \( \theta_0 \) is the true value of the parameter \( \theta \) being estimated.

(ii) Suppose \( \hat{\theta}_N, \hat{\sigma}_N^2 \) represent the estimates of the parameters obtained by minimizing (1.3.2), while \( \tilde{\theta}_N, \tilde{\sigma}_N^2 \) represent the estimates obtained from the maximization of the "true" likelihood function (i.e. the likelihood function formed under Gaussian assumptions). Then \( \hat{\sigma}_N, \tilde{\sigma}_N^2 \) converge almost surely to \( \hat{\sigma}_N, \tilde{\sigma}_N^2 \) respectively.

(iii) \( \sqrt{N}(\hat{\sigma}_N - \sigma^2_0), \sqrt{N}(\tilde{\sigma}_N^2 - \sigma^2_0) \) both converge in probability to zero. Thus \( \hat{\sigma}_N \) and \( \tilde{\sigma}_N \) have the same distributions in the sense that if \( \sqrt{N}(\hat{\sigma}_N - \sigma^2_0) \) is asymptotically normal so is \( \sqrt{N}(\tilde{\sigma}_N^2 - \sigma^2_0) \), and the limiting distributions are the same, and similarly for \( \hat{\sigma}_N, \tilde{\sigma}_N^2 \).

Finally we point out that if \( N \) is highly composite it is easier to compute in relation to (1.3.2) than it is to compute in relation to \( x' \Gamma^{-1}_N x \). This is because it is easier to compute the finite Fourier transform than all of the covariances and, in general, all of the covariances will be needed to obtain the statistics resulting from \( x' \Gamma^{-1}_N x \). An exception to this last statement is the autoregressive case for which \( \Gamma^{-1}_N \) takes on a simpler form. Even then however (1.3.2) has the advantage for it is only in relation to that formula that we are led by maximum likelihood to the customarily used Yule-Walker relations for the estimation of the autoregression.
Summarizing the results of this section it follows that minimizing (1.3.2) is asymptotically equivalent to maximizing the logarithm of the "true" likelihood function (formed under Gaussian assumptions), and furthermore (1.3.2) will have a single maximum at the true parameter value. For the various models to be considered it is the function (1.3.2) which we shall minimize and from which we shall be able to obtain efficient estimates (of the coefficients of the particular linear model under consideration).
CHAPTER 2

A GOODNESS OF FIT TEST FOR VECTOR AUTOREGRESSIVE MODELS

2.1 Introduction

The problem of determining a test of the goodness of fit of an autoregressive model has received considerable attention since Quenouille's [49] paper. We shall, in this chapter, be concerned with an extension of Quenouille's test to the vector case, the extension being due to Bartlett and Rajalakshman [5]. (As we show in section 2.4, the test to be derived in the next section is in fact an alternative derivation of the Bartlett-Rajalakshman test). Indeed we shall determine a goodness of fit test for a vector autoregressive model i.e. one which is generated by

\[ \sum_{j=0}^{q} B(j)x(n-j) = \epsilon(n), \quad B(0) = I, \]  

(2.1.1)

where the \( x(n) \), \( n = 1, \ldots, N \) are vectors of \( p \) observations, the \( B(j) \) are square matrices and the \( \epsilon(n) \) are identically and independently distributed with covariance matrix \( G_q \), which has typical element \( \sigma_{ij} \) in row \( i \), column \( j \); \( i, j = 1, \ldots, p \). Since we will be making mean corrections we may take

\[ E(x(n)) = E(\epsilon(n)) = 0. \]

Furthermore we assume that the z transform

\[ h(z) = \sum_{j=0}^{q} B(j)z^j \]

has determinant with all zeros outside of the unit circle.

More specifically we shall be concerned with forming a test of the null hypothesis that the autoregression is of order \( q \), against the alternative that it is of order \( (q+s) \). A natural statistic to use is
\[
\Lambda_q(s) = \frac{|\hat{G}_{q+s}|}{|\hat{G}_q|},
\]  

(2.1.2)

where \(|\hat{G}_r|\) is the determinant of the estimate of the covariance matrix obtained from the estimated residuals from a fitted autoregression of order \(r\). \(\Lambda_q(s)\) is an approximation to the likelihood ratio (see Anderson [2] Chapter 10) and it can be shown, under the null hypothesis, that \(-N \log(\Lambda_q(s))\) is asymptotically distributed as chi square with \(sp^2\) degrees of freedom. Furthermore

\[
\frac{|\hat{G}_{q+s}|}{|\hat{G}_q|} = \frac{|\hat{G}_{q+s}|}{|\hat{G}_{q+s-1}|} \cdots \frac{|\hat{G}_{q+1}|}{|\hat{G}_q|} = \prod_{i=1}^{s} \Lambda_{q+i-1}(1),
\]

where the \(\Lambda_{q+i-1}(1), i = 1, \ldots, s\) are, under the null hypothesis, asymptotically independent (see Hannan [26] pp 302-303). Thus, since

\[-N \log(\Lambda_q(s)) = -N \sum_{i=1}^{s} \log(\Lambda_{q+i-1}(1)),\]

we see that for each \(i = 1, \ldots, s\), \(-N \log(\Lambda_{q+i-1}(1))\) is asymptotically distributed as chi square with \(p^2\) degrees of freedom.

As well as the test due to Quenouille, Bartlett and Diananda [4] have derived a test for the case \(p = 1\). The usefulness of this test however is reduced by the fact that the effect of estimation, in the test statistic, is not easily allowed for. Bartlett and Rajalakshman, as already mentioned, have extended Quenouille's test to the vector case, while Walker [52] has derived an extension of this test to the case of multivariate autoregressive models with finitely dependent residuals.
The goodness of fit test we shall derive in the next section is based on a moving average of the sample covariance matrices so that, after appropriate normalization, there is no correlation between elements from different matrices in the sequence. As we have said, and will indeed show, this test is asymptotically equivalent to the test based on (2.1.2), and is in fact an alternative derivation of the test first used by Bartlett and Rajalakshman. Furthermore in the scalar case this test will be shown to reduce to Quenouille's test.

The defect in this test, and indeed in all the tests to which we have referred (with the exception of the test based on (2.1.2)), in both the scalar and vector cases, arises from the replacement of an estimate of $G_{q+s}$ (obtained from the residuals of a fitted autoregression of order $(q+s)$) by an estimate of $G_q$ (obtained from the residuals of a fitted autoregression of order $q$). This has the effect of reducing the calculations and is a valid procedure under the null hypothesis. However when the alternative hypothesis is true the magnitude of the test statistic will be reduced by this procedure and the test itself will have reduced power. We shall discuss this further below.

5 The research for this thesis was commenced while Professor Hannan was completing his book [26]. Consequently as the contents of section 2.2 were developed jointly with Professor Hannan, there is an obvious analogy between this section and section 6.7 of [26].
2.2 Derivation of the Test

Let us represent the possible factorizations of the spectral density matrix for the autoregressive model (2.1.1) by
\[ f(\lambda) = (2\pi)^{-1} h(j)^{-1} (e^{i\lambda})_q (j)^{(j)^*} (e^{i\lambda}), \]
where
\[ h(j)(z) = \sum_{k=0}^{q} B(j)(k) z^k \]
is the z transform of a filter which produces a sequence \( e^{(j)}(n) \), with
\[ E(e^{(j)}(m)e^{(j)}(n)) = \delta_{m,n} \cdot q, \]
from the input \( x(n) \). There is of course only one such factorization which corresponds to \( h(j)(z) \) having determinant with all of its zeros outside the unit circle, and \( h(j)(z) = I \). We shall call that factor \( h(z) \). The corresponding output is then \( e(n) \), the covariance matrix for this output process being \( q \).

Proceeding heuristically in order to obtain a test statistic, let
\[ D(j)(u) = (2\pi/N) \sum_{t=1}^{N} h(e^{i\lambda_t}) I(\lambda_t) h^{(j)}(e^{i\lambda_t}) e^{-iu\lambda_t}, u > 0, \]
where \( \lambda_t = 2\pi t/N, t = 1, 2, \ldots, N \), and \( I(\lambda_t) \) is the periodogram ordinate at frequency \( \lambda_t \). Replacing \( I(\lambda_t) \) by its expectation (which is near to \( f(\lambda_t) \)) and on approximating the sum by its integral we have
\[ E(D(j)(u)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{h} h^{(j)} \text{h}^{(j)*} e^{-iu\lambda} d\lambda \]
\[ = 0, \quad u > 0, \]
\[ (2.2.2) \]

6 In what follows we shall, for convenience of notation, where no confusion arises, omit argument variables.
since we may express $h^{-1}$ as an infinite series of zero and negative powers of $e^{i\lambda_t}$. (Indeed since all the zeros of the determinant of $h$ lie outside of the unit circle we may expand $h^{-1}$ in terms of a one sided Laurent series of zero and positive powers of $e^{i\lambda_t}$, with coefficients which decay exponentially, i.e.

$$h^{-1} = \sum_{k=0}^{\infty} \Delta(k)e^{i\lambda_t^k}, \quad \|\Delta(k)\| < k^{-|k|}, \quad \tau > 1.$$  \hspace{1cm} (2.2.3)

On taking transposed conjugates of both sides of this expression the result follows).

To investigate the covariance matrix of the statistic $D^{(j)}(u)$, let us introduce a tensor notation (using the usual summation notation for repeated suffixes). Now (see Hannan [26] p 249)

$$E(I_{ab}(\lambda_1)I_{cd}(\lambda_2)) = 0$$

$$= f_{ad}f_{cb}$$

$$= f_{ac}f_{db}$$

$$= f_{ad}f_{cb}f_{ac}f_{db}$$

$$= \lambda_1 \not\equiv \lambda_2,$$

$$\lambda_1 = \lambda_2 \not\equiv 0, \pm \pi,$$

$$\lambda_1 = -\lambda_2 \not\equiv 0, \pm \pi,$$

$$\lambda_1 = \lambda_2 = 0, \pm \pi.$$

Furthermore

$$d^{(j)}_{ab}(w) = (2\pi/N) \sum_{t} h_{ta} I_{ab} h^{(j)}_{t} e^{-iw\lambda_t}$$

$$= (2\pi/N) \sum_{t} h_{ta} I_{ab} h^{(j)}_{t} e^{-iw\lambda_t}, \quad w > 0,$$

and again approximating a sum by an integral, the covariance will ultimately be got from
\[ N\mathcal{E}(d_{pq}^{(j)}(u)d_{rs}^{(j)}(v)) \]

\[ \approx 2\pi \int_{-\pi}^{\pi} h_{pa} h_{qb} h_{rc} h_{sd} f_{ad} f_{cb} e^{-i(u+v)\lambda} d\lambda \]

\[ + 2\pi \int_{-\pi}^{\pi} h_{pa} h_{qb} h_{rc} (e^{-i\lambda}) h_{sd} (e^{-i\lambda}) f_{ac} f_{db} e^{-i(u-v)\lambda} d\lambda, \quad (2.2.4) \]

\[ \text{where} \]

\[ h_{rc}(e^{-i\lambda}) = \sum_{k=0}^{q} \beta_{rc}(k)e^{-ik\lambda}, \quad h_{sd}(e^{-i\lambda}) = \sum_{k=0}^{q} \beta_{sd}(k)e^{-ik\lambda}. \]

When \( h^{(j)} = h \), the right hand side of (2.2.4) becomes

\[ 2\pi \int_{-\pi}^{\pi} h_{pa} h_{*} h_{*} h_{rc} h_{*} h_{*} f_{ad} f_{cb} f_{*} b_{q} e^{-i(u+v)\lambda} d\lambda \]

\[ + 2\pi \int_{-\pi}^{\pi} h_{pa} h_{*} h_{*} h_{sd} h_{*} h_{*} f_{ac} f_{db} f_{*} b_{q} e^{-i(u-v)\lambda} d\lambda; \quad u, v > 0 \]

\[ = 0, \quad u \neq v \]

\[ = \sigma_{pr} \sigma_{sq}, \quad u = v. \]

This latter case (i.e., \( h^{(j)} = h \) and \( u = v \)), which generalizes the test due to Bartlett and Diananda, is not of interest however since, as we have said earlier, the effects of estimating the function \( h \) are not easily allowed for.

There is another case where (2.2.4) is null for \( u \neq v \) and this is where \( h^{(j)}(z) \) corresponds to that unique factorization in which \( h^{(j)}(z) \) has all its zeros inside the unit circle and \( B^{(j)}(q) = I \). We call this \( h^{(1)}(z) \). Now \( h^{(1)}(z) \) has real coefficient matrices so that
Then the right hand side of (2.2.4) becomes

\[ 2\pi \int_{-\pi}^{\pi} h_{pa} h_{qc} h_{rc} h_{sd} f_{ad} f_{cb} e^{i(u+v)\lambda} d\lambda \]

\[ + 2\pi \int_{-\pi}^{\pi} h_{pa} h_{qc} h_{rc} h_{sd} f_{ac} f_{db} e^{-i(u-v)\lambda} d\lambda. \]

Now if $h_{ed}$ represents a general element of $h^{-1}$ then we have

\[ h_{pa} f_{ad} = \frac{1}{2\pi} \sigma_{pe} h_{ed}, \]

and

\[ h_{rc} f_{cb} = \frac{1}{2\pi} \sigma_{rf} h_{fd} \]

so that the first term becomes

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} (\sigma_{pe} h_{ed} h_{qb}) (\sigma_{rf} h_{fd}) e^{-i(u-v)\lambda} d\lambda, \]

while the second term becomes

\[ 2\pi \int_{-\pi}^{\pi} (h_{pa} h_{qc} h_{rc} h_{sd}) (h_{ad} h_{cb} h_{ bq}) e^{-i(u-v)\lambda} d\lambda \]

\[ = \sigma_{pr} \sigma_{sq} (1) \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-i(u-v)\lambda} d\lambda. \]

Thus the right hand side of (2.2.4) becomes

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ (\sigma_{pe} h_{ed} h_{qb}) (\sigma_{rf} h_{fd}) e^{-i(u-v)\lambda} + \sigma_{pr} \sigma_{sq} (1) e^{-i(u-v)\lambda} \right\} d\lambda \]

\[ = 0 \quad \text{if} \quad u \neq v, \]

\[ = \sigma_{pr} \sigma_{sq} (1) \quad \text{if} \quad u = v, \]

where $\sigma_{sq}$ is the element in row $s$ column $q$ of $G_{s}^{(1)}$, with
\[ f(\lambda) = (2\pi)^{-1}h(1)^{-1}g(1)h(1)^{-1}. \]

It is this latter case which we shall consider further since, as we shall later show, it is this case which provides a generalization to Quenouille's test.

For \( j = 1 \), after some manipulation, we may write (2.2.1) in the form

\[
D^{(1)}(u) = \sum_{j,k=0}^{q} (1 - \frac{1}{N})B(j)C(j-k-u)B^{(1)'}(k), \quad u > 0,
\]

\[
= \sum_{j,k=0}^{q} (1 - \frac{1}{N})B(j)C'(u+k-j)B^{(1)'}(k), \quad u > 0, \quad (2.2.5)
\]

where the \( B^{(1)}(k) \), \( k = 0, ..., q \) are the coefficient matrices of \( h^{(1)} \) and \( C(j) = (N-j)^{-1} \sum_{n=1}^{N-j} x(n)x'(n+j) = C'(j) \), \( j = 0, 1, ... \), are the sample covariance matrices. We shall discuss the computational procedure in section 2.3. However from (2.2.5) we see that once we have obtained the estimates \( \hat{B}(j), \hat{B}^{(1)}(k) \) of \( B(j), B^{(1)}(k) \) respectively it is then a straightforward matter to compute \( \hat{D}^{(1)}(u) \), the estimate of \( D^{(1)}(u) \).

We now prove the following:

**Theorem 2.2.1**

For any fixed \( s > 0 \), the elements of all of the matrices \( \sqrt{N} \hat{D}^{(1)}(u) \), \( 0 < u \leq s \), are asymptotically jointly normal with zero mean and covariances which are zero between two elements from differing \( u \) and otherwise are

\[ \sigma_{pr}^{(1)} \]

between two elements in row \( p \) column \( t \) and row \( q \) column \( r \), where \( \sigma_{qt}^{(1)} \) is the typical element of \( G^{(1)}_{q} \), while \( \sigma_{pr} \) is the typical element of \( G_{q} \). Then

\[
N \sum_{u=1}^{s} \text{tr}[\hat{G}^{-1}(1)(u)\hat{G}(1)^{-1}D^{(1)'}(u)]
\]
is asymptotically normally distributed as chi square with \( sp^2 \) degrees of freedom.

Proof:  For simplicity we shall give the proof for the case where no mean corrections are made, though this of course has no effect asymptotically. From (2.2.5) it follows that

\[
\sqrt{N} D^{(1)}(u) = \frac{1}{\sqrt{N}} \sum_{j,k=0}^{q} B(j) \sum_{n=1}^{N} x(n-j)x'(n-k-u)B^{(1)'}(k) + O(N^{-\frac{3}{2}}),
\]

where we have added, to the first term, a fixed number of terms to bring that expression to simple form. If we let

\[
\epsilon^{(1)}(n) = \sum_{k=0}^{q} B^{(1)}(k)x(n-k), \quad B^{(1)}(q) = I,
\]

where the \( \epsilon^{(1)}(n) \) are identically and independently distributed with zero mean and covariance matrix \( G^{(1)}_q \) (which has typical element \( c_{ij}^{(1)} \) in row \( i \) column \( j; i,j = 1,\ldots,p \)), then from (2.2.6) it follows that

\[
\sqrt{N} D^{(1)}(u) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \epsilon(n)\epsilon^{(1)'}(n-u) + O(N^{-\frac{3}{2}}).
\]  

(2.2.7)

Since \( x(n-k-u), k \geq 0, u > 0 \) is independent of \( \epsilon(n) \) it is evident that \( \epsilon(n) \) is independent of \( \epsilon^{(1)}(n-u), u > 0 \) \( \epsilon^{(1)}(n) \) may of course be expressed as a linear combination of \( \epsilon(n-j), j \geq 0 \) with coefficient matrices that decrease exponentially. If we let

\[
\sqrt{N} d^{(1)}_{ij}(u) = N^{-\frac{1}{2}} \sum_{n=1}^{N} \epsilon_{i}(n)\epsilon_{j}'(n-u), \quad i,j = 1,\ldots,p; \quad u = 1,\ldots,s,
\]

the proof that the \( \sqrt{N} d^{(1)}_{ij}(u) \) are asymptotically jointly normal is not essentially different from the main step in the proof of the joint asymptotic normality of the elements of the matrices \( \delta(j) \), for which see Mann and Wald [38]. To be sure these authors assume that all moments of
the \( \epsilon(n) \) are finite. However as Hannan ([26] pp 329-332) points out, a closer examination of this proof shows that the finiteness of moments higher than the second is not needed. This is basically because
\[
\sum_{n=1}^{N} \epsilon(n) \epsilon(n)'(n-u),
\]
when expressed as a function of elements of the vectors \( \epsilon(n) \), involves no squared terms (i.e. no terms of the form \( \epsilon_i(n)\epsilon_j(n) \)), so that the variance is finite even if the fourth moment of the \( \epsilon(n) \) is infinite. Referring to Hannan's proof, the \( \sqrt{N} \bar{d}_{ij}^{(1)}(u) \) are of the same form as what he calls \( e_{ij}(v) \) and which, by the results of his chapter 4, he shows to be asymptotically normally distributed. Consequently using an identical argument to that for the \( e_{ij}(v) \) it follows that \( \sqrt{N} \bar{d}_{ij}^{(1)}(u) \), \( i,j = 1, \ldots, p; u = 1, \ldots, s \), are also asymptotically normally distributed.

The general element of the covariance matrix of (2.2.7) may be written as
\[
B(N^{-1} \sum_{m=1}^{N} \epsilon_p(m)\epsilon_t^{(1)}(m-u) + \sum_{n=1}^{N} \epsilon_r(n)\epsilon_q^{(1)}(n-v)); \quad u,v > 0.
\]

Since \( \epsilon_p(m) \) is, for \( m \neq n \), independent of \( \epsilon_t^{(1)}(m-u), \epsilon_q^{(1)}(n-v) \), and orthogonal to \( \epsilon_r(n) \) for \( m > n \), we see that the only contribution to the expectation can occur when \( m = n \) and \( u = v \), when we obtain
\[
N^{-1} \sum_{m=1}^{N} \sigma_{pr}^{(1)} = \sigma_{pr}^{(1)},
\]
so that
\[
N \mathbb{E}(d_{pq}^{(1)}(u)d_{rt}^{(1)}(v)) = 0, \quad u \neq v
\]
\[
= \sigma_{pr}^{(1)}, \quad u = v.
\]

(2.2.8)

To consider the effect of inserting estimates of \( B(j) \) and \( B^{(1)}(k) \) in (2.2.5), we form
\[
\sqrt{N} (\hat{S}^{(1)}(u) - D^{(1)}(u)) = \sqrt{N} \sum_{j,k=0}^{q} \left(1 - \frac{|u+q-k-j|}{N}\right) (\hat{B}(j)-B(j)) C'(u+q-k-j) \hat{B}^{(1)'}(q-k) \\
+ \sqrt{N} \sum_{j,k=0}^{q} \left(1 - \frac{|u+q-k-j|}{N}\right) B(j) C'(u+q-k-j) (\hat{B}^{(1)'}(q-k)-B^{(1)'}(q-k))
\]

(2.2.9)

Now \((\hat{B}(j)-B(j))\) converges to zero with probability one and so therefore must \((\hat{S}^{(1)'}(k)-B^{(1)'}(k))\). On the other hand, if \(v = u+q, v > q\), considering the second term on the right hand side of (2.2.9) we have

\[
N^{-1/2} \sum_{j=0}^{q} B(j) (1 - \frac{|v-k-j|}{N}) C'(v-k-j) \\
= N^{-1/2} \sum_{j=0}^{q} B(j) \frac{N}{n} x(n-j)x'(n+k-v), \quad v-k > 0,
\]

\[
= N^{-3/2} \sum_{n=1}^{N} \epsilon(n)x'(n+k-v),
\]

(where we have added a finite number of terms to the inner summation).

Referring once more to the proof of Theorem 1 in chapter 6 of Hannan [26] we see this to be asymptotically normally distributed with mean zero.

Thus the second term on the right hand side converges in probability to zero, and using a similar argument the same may be shown to be true for the first term. Thus it follows that the elements of the \(\sqrt{N} \hat{B}^{(1)}(u), u > 0\) have the asymptotic properties stated in the theorem.

Reverting to a tensor notation and writing

\[
d^{(1)}(u) = \text{vec}(D^{(1)'}(u)), \quad \hat{d}^{(1)}(u) = \text{vec}(\hat{D}^{(1)'}(u)),
\]

we see from (2.2.8) that the covariance matrix of \(\sqrt{N} \hat{d}^{(1)}(u)\) is \(G_q \otimes G^{(1)}_q\). From what we have proved therefore it follows that

\[
N \hat{d}^{(1)'}(u) (\hat{G}_q \otimes \hat{G}^{(1)}_q)^{-1} \hat{d}^{(1)}(u)
\]

(2.2.10)
is asymptotically distributed as chi square with $p^2$ degrees of freedom
and for each $u = 1, \ldots, s$, where $s$ is fixed a priori, these quantities
are asymptotically independently distributed. Furthermore from (1.1.11)
it follows that

$$\left( \hat{G}_q \otimes \hat{G}_q^{(1)} \right)^{-1} \hat{d}(1)(u) = \text{vec}(\hat{G}_q^{(1)})^{-1} \hat{D}(1)'(u) \hat{G}_q^{-1},$$

and since

$$(\text{vec}(A))' \text{vec}(B) = \text{tr}(AB') = \text{tr}(B'A),$$

we have

$$\hat{d}(1)'(u) (\hat{G}_q \otimes \hat{G}_q^{(1)})^{-1} \hat{d}(1)(u) = \text{tr}(\hat{G}_q^{-1} \hat{D}(1)(u) \hat{G}_q^{(1)})^{-1} \hat{D}(1)'(u)).$$

Thus from (2.2.10) it follows that

$$N \sum_{u=1}^{s} \text{tr}(\hat{G}_q^{-1} \hat{D}(1)(u) \hat{G}_q^{(1)})^{-1} \hat{d}(1)'(u))$$

is distributed as chi square with $sp^2$ degrees of freedom. This proves
the theorem.

2.3 The Computational Procedure

In order to test the null hypothesis that the vector autoregressive
model (2.1.1) is of order $q$ against the alternative that it is of order
$(q+s)$, $s > 0$, the first step in the computational procedure is to compute
$\hat{B}(j)$, $j = 1, \ldots, q$; $\hat{B}_q^{(1)}(k)$, $k = 0, \ldots, (q-1)$; $\hat{G}_q$ and $\hat{G}_q^{(1)}$. To do this
we proceed as follows: From (2.1.1) it is not hard to see that

$$\sum_{k=0}^{q} \hat{B}(k)C(k-j) = \delta_{0,j} \hat{G}_q , \quad \hat{B}(0) = I , \quad j = 0, \ldots, q. \quad (2.3.1)$$

An equivalent expression for $B^{(1)}(k)$, $k = 0, \ldots, q$ may be obtained by
considering the expression
\[ \sum_{k=0}^{q} \hat{B}(1)(k)C(k+j) = \int_{-\pi}^{\pi} \sum_{k=0}^{q} \hat{B}(1)(k)e^{ik\lambda} f(\lambda)e^{ij\lambda} d\lambda, \quad (2.3.2) \]

where \( \hat{B}(1)(q) = I \), and

\[ (\sum_{k=0}^{q} \hat{B}(1)(k)e^{ik\lambda} f(\lambda)(\sum_{k=0}^{q} \hat{B}(1)(k)e^{ik\lambda})^*) = (2\pi)^{-1} \hat{G}(1). \]

Now

\[ \sum_{k=0}^{q} \hat{B}(1)(k)e^{ik\lambda} f(\lambda) = (2\pi)^{-1} \hat{G}(1)(\sum_{k=0}^{q} \hat{B}(1)'(k)e^{i(q-k)\lambda})^{-1}e^{iq\lambda}, \]

and since the determinant of \( \sum_{k=0}^{q} \hat{B}(1)'(k)q^{-k} \) has no zeros within the unit circle, following a similar argument to that for (2.2.3) we may express

\[ (\sum_{k=0}^{q} \hat{B}(1)'(k)e^{i(q-k)\lambda})^{-1} \]

in terms of an infinite series of zero and positive powers of \( e^{i\lambda} \). Thus (2.3.2) becomes

\[ \hat{G}(1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \sum_{k=0}^{q} \hat{B}(1)'(k)e^{i(q-k)\lambda} \right)^{-1} e^{iq+1}\lambda \lambda d\lambda \]

\[ = 0, \quad j = -q+1, \ldots, 0 \]

\[ = \hat{G}(1)_{q}, \quad j = -q. \]

Hence we obtain

\[ \sum_{k=0}^{q} \hat{B}(1)(k)C(k+j) = 0, \quad j = -q+1, \ldots, 0, \quad \hat{B}(1)(q) = I \]

\[ = \hat{G}(1)_{q}, \quad j = -q. \quad (2.3.3) \]

These equations could alternatively be written as

\[ \sum_{k=0}^{q} \hat{B}(1)(q-k)C(j-k) = \delta_{q,j} \hat{G}(1)_{q}, \quad j = 0, \ldots, q, \quad \hat{B}(1)(q) = I. \quad (2.3.4) \]

\[ \text{Strictly speaking this will only be true if the C(j) are defined using the divisor N}^{-1} \text{ and not (N-|j|)\}^{-1} as Parzen recommends (see Hannan [26] p 332 for a discussion of this).} \]
In order to solve (2.3.1) and (2.3.4) to obtain the $\hat{B}(j)$, $\hat{G}(1)(k)$, $\hat{G}_q$ and $\hat{G}_q(1)$ we could, as Hannan ([26] p 403) suggests, express these equations in tensor notation and solve them directly. This method involves the evaluation and inversion of a $(p^2 q \times p^2 q)$ matrix however, and when $p$ and $q$ are large this becomes computationally arduous. The solutions of (2.3.1) and (2.3.4) can best be obtained by applying an iterative procedure (for the solution of such equations) developed by Whittle [58]. This procedure is in fact a generalization of a method first developed by Durbin [13] (for calculating the coefficients of scalar autoregressive models).

Whittle's procedure is most convenient in the present situation since its application requires us to solve two autoregressions simultaneously, one in terms of the $\hat{B}(j)$ and the other in terms of the $\hat{B}(1)(k)$ (so that we obtain both sets of estimates at once). To apply this procedure, for simplicity of notation we shall let

\[ \hat{B}(1)(q-k) = A_{q,k}, \quad \hat{B}(k) = \tilde{A}_{q,k} \]

(in order to emphasize the dependence upon the order $q$ of the fitted scheme) so that $A_{q,0} = \tilde{A}_{q,0} = I$, and (2.3.1), (2.3.4) become respectively

\[ \sum_{k=o}^{q} \tilde{A}_{q, k} C(k-l) = 0, \quad l = 1, \ldots, q, \quad (2.3.5) \]

\[ \sum_{k=o}^{q} A_{q, k} C(l-k) = 0, \quad l = 1, \ldots, q. \quad (2.3.6) \]

Initial estimates based on a first order autoregression are made and then the following iterative procedure is carried out for $j = 1, \ldots, (q-1)$. Set

\[ V_j = \sum_{k=0}^{j} A_{j, k} C(-k), \quad \tilde{V}_j = \sum_{k=0}^{j} \tilde{A}_{j, k} C(k), \]
and
\[ \Delta_j = \sum_{k=0}^{j} A_{j,k} c(j-k+1), \quad \tilde{\Delta}_j = \sum_{k=0}^{j} A_{j,k} c(-j+k-1), \]
and then form
\[ A_{j+1,j} = -\Delta_j \bar{y}_j^{-1}, \quad \tilde{A}_{j+1,j} = -\tilde{\Delta}_j \bar{y}_j^{-1}. \]

Then for \( k = 1, \ldots, j \) we apply the following recurrence relations

\[ A_{j+1,k} = A_{j,k} + A_{j+1,j}, \quad \tilde{A}_{j+1,k} = \tilde{A}_{j,k} + \tilde{A}_{j+1,j}. \]  

From equations (2.3.7) and (2.3.8) we obtain \( \hat{\beta}(q-k), \hat{\beta}(k) \) respectively (i.e. \( \hat{B}(l)^{(q-k)}, \hat{B}(k) \)), \( k = 1, \ldots, q \). Having obtained these estimates we see from (2.3.1) and (2.3.4) that it is a straightforward matter to determine \( \hat{G}_q, \hat{G}_q^{(1)} \). Furthermore it is now possible to compute the quantities

\[ \hat{D}_q^{(1)}(u) = \sum_{j,k=0}^{q} (1 - \frac{1 - (j-k-u)}{N}) \hat{B}(j) c(j-k-u) \hat{B}^{(1)'}(k), \quad u = 1, \ldots, s, \]  

and hence

\[ \frac{s}{N} \sum_{u=1}^{s} \text{tr}(\hat{G}_q^{(1)}(u) \hat{G}_q^{(1)'(u)}), \]

which is distributed as chi square with \( sp^2 \) degrees of freedom.

2.4 A Comparison of Goodness of Fit Tests

That the test devised in section 2.2 reduces to Quenouille's test in the scalar case can be seen by letting \( p = 1 \) in (2.3.4), which reduces to

\[ \sum_{k=0}^{q} \hat{\beta}(q-k) c(k-u) = 0, \quad u = 1, \ldots, q; \quad \hat{\beta}(q) = 1 \]

\[ = \hat{c}_q^{(1)^2}, \quad u = 0, \]
and \( \hat{\beta}^{(1)}(q-k) = \hat{\beta}(k) \), \( k = 0, \ldots, q \),

so that \( \hat{\sigma}^{(1)}^2 = \sum_{k=0}^{q} \hat{\beta}(k)c(k) = \hat{\sigma}^2 \).

Thus the expression (2.3.9) leads to

\[
\sqrt{N} \hat{a}(u) = (N-u+q)^{\frac{1}{2}} q \sum_{j,k=0}^{q} \hat{\beta}(j)\hat{\beta}(k)c(j-k-u)
\]

(where we have in (2.3.9) replaced the factor \( (1 - \frac{|j-k-u|}{N}) \) by unity and modified the factor \( \sqrt{N} \) accordingly), and the corresponding test statistic is

\[
\sum_{u=1}^{s} \frac{d^2(u)}{(\sum_{k=0}^{q} \hat{\beta}(k)c(k))^2},
\]

which is just Quenouille's test statistic.

As we have already mentioned the test we have derived in section 2.3 is an alternative derivation of the test due to Bartlett and Rajalakshman. Indeed the statistic on which the Bartlett-Rajalakshman test is based may be written in the form (see for example equation 5 in Walker [52])

\[
Q(u) = G^+C(u)H^-
\]  

(2.4.1)

where

\[
G^+ = \sum_{k=0}^{q} B^{(1)}(k)F^k, \quad H^- = \sum_{j=0}^{q} B(j)F^{-j}
\]

(with \( F^k \) the usual forward shift operator which changes \( t \) to \( (t+k) \)), and \( C(u) \) is the sample covariance matrix. If we transpose (2.4.1) we see this to be, save for minor notational changes, just what we have called \( D^{(1)}(u) \) (equation (2.2.5)).

This test that Bartlett and Rajalakshman have devised is developed explicitly for the case \( q = 1 \), i.e. for first order autoregressive models.
As the authors point out however this represents no real restriction
for in cases where \( q > 1 \) the model may be reduced to one for which \( q = 1 \)
by redefinition of the model so that \( x(n) \) becomes a vector of higher
dimension. For example the second order model
\[
x(n) + \beta(1)x(n-1) + \beta(2)x(n-2) = \epsilon(n)
\]
may be represented as the first order model
\[
\begin{pmatrix}
    x(n) \\
    y(n)
\end{pmatrix}
+ \begin{pmatrix}
    \beta(1) \\
    -1
\end{pmatrix}
\begin{pmatrix}
    x(n-1) \\
    y(n-1)
\end{pmatrix}
= \begin{pmatrix}
    \epsilon(n) \\
    0
\end{pmatrix},
\]
i.e. \( X(n) + B(1)X(n-1) = U(n) \) say,
and in this form the Bartlett-Rajalakshman test is applicable.

In order to compare the test we have developed in this chapter with
the test based on \((2.1.2)\) we see that in order to test the null hypothesis
of an autoregression of order \((q+r-1)\) against the alternative of an
autoregression of order \((q+r)\), considering \((2.1.2)\) we form
\[
\Lambda_{q+r-1}(1) = \frac{|\hat{G}_{q+r}|}{|\hat{G}_{q+r-1}|},
\]
and \(-N \log(\Lambda_{q+r-1}(1))\) is asymptotically distributed as chi square with \(p^2\)
degrees of freedom. Now
\[
-N \log(|\hat{G}_{q+r}|/|\hat{G}_{q+r-1}|) = -N \log(|\hat{g}_{q+r}|/|\hat{g}_{q+r-1}^\hat{G}_{q+r-1}G_{q+r}|)
\]
\[
= N \log(|I+\hat{g}_{q+r}^{-\frac{1}{2}}(\hat{G}_{q+r-1}G_{q+r})^\hat{G}_{q+r-1}^\frac{1}{2}|)
\]
and under the null hypothesis it is not hard to show that this becomes
\[
N \log(|I+\hat{g}_{q+r-1}^{-\frac{1}{2}}(\hat{G}_{q+r-1}G_{q+r})^\hat{G}_{q+r-1}^\frac{1}{2}|).
\]
If \( \lambda_i, i = 1, \ldots, p \) represent the eigenvalues of the matrix 

\[
\hat{G}^{-\frac{1}{2}}_{q+r-1} \left( \hat{G}_{q+r-1} - \hat{G}_{q+r} \right) \hat{G}^{-\frac{1}{2}}_{q+r-1}
\]

then

\[
-N \log(\Lambda_{q+r-1}(l)) = N \log(\prod_{i=1}^{p} (1+\lambda_i)) = N \sum_{i=1}^{p} \log(1+\lambda_i) = N \sum_{i=1}^{p} \lambda_i = N \text{tr}\left[\hat{G}^{-\frac{1}{2}}_{q+r-1} \left( \hat{G}_{q+r-1} - \hat{G}_{q+r} \right) \hat{G}^{-\frac{1}{2}}_{q+r-1}\right].
\]

Indeed since \( \lambda_i \geq 0, i = 1, \ldots, p \) then

\[
N \sum_{i=1}^{p} \lambda_i - N \sum_{i=1}^{p} \lambda_i^2 \leq N \sum_{i=1}^{p} \log(1+\lambda_i) = N \sum_{i=1}^{p} \lambda_i,
\]

But \( N \sum_{i=1}^{p} \lambda_i^2 \leq (N \sum_{i=1}^{p} \lambda_i) (\sum_{i=1}^{p} \lambda_i), \) which converges in probability to zero (since \( N \sum_{i=1}^{p} \lambda_i \) converges in distribution, while \( \sum_{i=1}^{p} \lambda_i \) converges almost surely to zero), so that

\[
N \sum_{i=1}^{p} \log(1+\lambda_i) = N \sum_{i=1}^{p} \lambda_i,
\]

as required. Hence under the null hypothesis

\[
-N \log(\Lambda_{q+r-1}(l)) \approx N \text{tr}\left[\hat{G}^{-\frac{1}{2}}_{q+r-1} \left( \hat{G}_{q+r-1} - \hat{G}_{q+r} \right) \hat{G}^{-\frac{1}{2}}_{q+r-1}\right]
\]  

is asymptotically distributed as chi square with \( p^2 \) degrees of freedom.

If we let

\[
\hat{e}' = (\hat{e}(1) \ldots \hat{e}(N)), \quad \hat{e}'(1)' = (\hat{e}(1)(1-r) \ldots \hat{e}(1)(N-r))
\]

then from (2.2.7) we have
\[
\hat{G}^{(1)}(r) = \frac{1}{N} \sum_{n=1}^{N} \hat{\epsilon}^{(1)}(n)\hat{\epsilon}'^{(1)}(n-r) = \frac{1}{N} \hat{\epsilon}'\hat{\epsilon}^{(1)}.
\]

Furthermore
\[
\hat{G}_{q+r-1} = \frac{1}{N} \sum_{n=1}^{N} \hat{\epsilon}(n)\hat{\epsilon}'(n) = \frac{1}{N} \hat{\epsilon}'\hat{\epsilon},
\]

and similarly
\[
\hat{G}(1)_{q+r-1} = \frac{1}{N} \hat{\epsilon}(1)\hat{\epsilon}'(1).
\]

For the goodness of fit test based on the vector moving average of the sample covariances, the test statistic is
\[
N \text{tr}(\hat{G}_{q+r-1}^{-1}\hat{D}(1)(r)\hat{G}_{q+r-1}^{-1}\hat{D}'(1)'(r)), \tag{2.4.3}
\]

which is asymptotically distributed as chi square with \( p^2 \) degrees of freedom. To show that the two tests based on (2.4.2) and (2.4.3) are, under the null hypothesis, asymptotically equivalent, we must show that
\[
N \text{tr}(\hat{G}_{q+r-1}^{-1}(\hat{G}_{q+r-1} - \hat{G}_{q+r}^{-1})\hat{G}_{q+r-1}^{-1}) = N \text{tr}(\hat{G}_{q+r-1}^{-1}(\hat{G}(1)(r)\hat{G}(1)^{-1}\hat{D}(1)'(r)). \tag{2.4.4}
\]

Now \((\hat{G}_{q+r-1} - \hat{G}_{q+r}^{-1})\) is the reduction in the estimated covariance matrix of the residuals due to the \((q+r)\)th term in the autoregression. This is of the form \(HSS'\), where \(H\) is the regression matrix for \(x(n)\) on \(x(n-q-r)\) after eliminating \(x(n-1), \ldots, x(n-q-r+1)\), while \(NS\) is the residual sum of squares from the regression of \(x(n-q-r)\) on \(x(n-q-r+1), \ldots, x(n-1)\). Thus we have
\[
x(n) + \sum_{j=1}^{q+r-1} \hat{B}(j)x(n-j) = H[x(n-q-r) + \sum_{k=1}^{q+r-1} \hat{B}(1)(q-k)x(n-r-(q-k))] + \hat{\eta}(n)
\]
say, i.e.
\[ \hat{e}(n) = H \hat{e}^{(1)}(n-r) + \hat{r}(n). \]

For a sample of size \( N \) it follows that
\[ \hat{e} = \hat{e}^{(1)} + \hat{r}, \]
where \( \hat{r}' = (\hat{r}(1) \ldots \hat{r}(N)), \) and so
\[ \hat{r}' = \frac{1}{N} \hat{e}^{(1)' \hat{e}^{(1)}}^{-1} \frac{1}{N} \hat{e}^{(1)' \hat{e}} = \hat{G}^{(1)-1}_{q+r-1} \hat{D}^{(1)'}(r), \]
\[ \hat{S} = \frac{1}{N} \hat{e}^{(1)' \hat{e}^{(1)}} = \hat{G}^{(1)}_{q+r-1}. \]

Consequently \( \hat{G}_{q+r-1}^{(1)} - \hat{G}_{q+r}^{(1)} = \frac{H \hat{S} H'}{\hat{D}^{(1)(1)}} = \hat{G}^{(1)}_{q+r-1} \hat{D}^{(1)'}(r), \) and from this we see that (2.4.4) is true, thus verifying that the two tests are, under the null hypothesis, asymptotically equivalent.

2.5 Conclusion

The test we have derived in section 2.2 is an extension of Quenouille's test to the vector case and is, as we have seen, an alternative derivation of the Bartlett-Rajalakshman test. Furthermore, when the null hypothesis is true, this test is asymptotically equivalent to the test based on (2.1.2), its advantage being computational simplicity. Indeed when testing the null hypothesis that the autoregression is of order \( q \) against the alternative that it is of order \( (q+s) \), for the test based on (2.2.5) we have only to compute \( \hat{G}^{(1)}_{q} \). For the test based on (2.1.2) however we have also to compute \( \hat{G}^{(1)}_{q+s} \) and, particularly if \( s \) is large, the extra computation required may not be insignificant.

We must again emphasize a disadvantage of the test (and indeed of all the tests we have discussed, with the exception of the test based on (2.2.2)) which we mentioned in section 2.1. In the test statistic the inverses of the covariance matrices obtained from the residuals (i.e. \( \hat{G}^{-1}_{q} \)
and $\hat{g}_q^{(1)}$ have been estimated assuming the null hypothesis to be true. Consequently if the alternative hypothesis is true, since $\hat{g}_q^{-1} > \hat{g}_q^{(1)}$ and $\hat{g}_q^{(1)-1} > \hat{g}_q$, the estimates we are using will be too small and so the test statistic will be reduced, as will the power of the test.
CHAPTER 3
THE ESTIMATION OF MOVING AVERAGE MODELS

3.1 Introduction

A number of procedures have been derived for the estimation of the coefficients of moving average models, one of the more recent being that due to Hannan [25]. This particular procedure enables us to determine estimates of the coefficients which are asymptotically efficient. The method does however have certain disadvantages, which are usually apparent when we are attempting to estimate coefficients such that the corresponding z transform has zeros occurring close to the boundary of the region in which they are restricted to lie. More specifically, suppose we represent a scalar moving average model by

\[ x(n) = \sum_{j=0}^{p} a(j)\epsilon(n-j) \quad , \quad a(0) = 1, \quad \sum_{j=0}^{p} a^2(j) < \infty, \]  

(3.1.1)

where the \( \epsilon(n) \) are i.i.d. \( (0, \sigma^2) \), and furthermore all the zeros of the z transform

\[ g(z) = \sum_{j=0}^{p} a(j)z^j \]  

(3.1.2)

are required to lie outside of the unit circle. This latter condition, as we have mentioned in chapter 1, is required to identify the model.

When Hannan's method is applied to estimate coefficients which are such that the zeros of \( g(z) \) are close to the boundary (of the acceptable region) the procedure, at times, does not work. As a simple example consider the first order moving average model

\[ x(n) = \epsilon(n) + \alpha \epsilon(n-1), \]  

(3.1.3)

where the \( \epsilon(n) \) are i.i.d. \( (0, \sigma^2) \). For this model the condition on the
corresponding z transform implies that we must have $|\alpha| < 1$. As we have stated, when attempting to estimate values of $\alpha$ near to unity the estimation procedure, on some occasions, does not work. This is due to the fact that when iterating the procedure, if we obtain an estimate of $|\alpha| > 1$, further iterations give estimates which "explode" in the sense that as we continue to iterate the estimates become larger and larger. Indeed when attempting to estimate a value of $\alpha = 0.96$, for the model \((3.1.3)\), with $N = 100$ observations, in forty replications of the experiment (using generated data) the iterative procedure exploded in five cases.

In its present form there is no necessity for the estimation procedure to give efficient estimates which are restricted to lie in the acceptable region. In what follows, when we talk of the acceptable region we mean the region in which the coefficients are confined such that the corresponding z transform has all of its roots outside of the unit circle. Such a region, for the model \((3.1.1)\), in terms of constraints on the $\alpha(j)$, $j = 1, \ldots, p$ may be determined in the following manner: considering the z transform \((3.1.2)\) suppose we compute the triangular array (sometimes called the Routh scheme)

\[
\begin{array}{cccccc}
1 & \alpha(1) & \cdots & \cdots & \cdots & \cdots \\
& a(0) & a(1) & \cdots & a(p-1) & \cdots \\
& b(0) & b(1) & \cdots & b(p-2) & \cdots \\
& & & \cdots & \cdots & \cdots \\
& & l(0) & l(1) & \cdots & \cdots \\
& & & m(0) & \cdots & \cdots \\
\end{array}
\]

where the first row consists of the coefficients of $g(z)$, the second row is obtained from the first according to the formula
a(r) = a(o)a(r)-a(p)a(p-r),
and succeeding rows are obtained in an analogous way. Thus
b(r) = a(o)a(r)-a(p-1)a(p-r-1), etc.

Then the necessary and sufficient condition for all the zeros of \( g(z) \) to
lie outside of the unit circle is that the elements of the first column
of the Routh scheme be all positive. (This is just Routh's criterion, a
proof of which may be found in Gantmacher \[16\]).

We shall in this chapter be concerned with an extension of Hannan's
procedure, the purpose of which will be to constrain the estimates to lie
in the acceptable region even when we are estimating coefficients for which
the corresponding z transform has zeros near the boundary. A further
problem we shall examine is the rate of convergence of the iterative
procedure used in this estimation procedure. Finally, in section 3.4
this method of estimation will be compared with two similar methods due
to Durbin \[14\].

To conclude this introduction we shall give a brief outline of
Hannan's estimation procedure (which is based on the method discussed in
section 1.3) as applied to the model \((3.1.1))\). In order to efficiently
estimate the \( a(j), j = 1, \ldots, p \) from the \( N \) observations \((x(n); \, n = 1, \ldots, N)\)
an initial estimate is required. Suppose we let \( I(\lambda_{t}) \) represent the
periodogram ordinate at \( \lambda_{t}; \, \lambda_{t} = \lambda_{N-t} = 2\pi t/N, \, t = 1, \ldots, [N/2]. \)\(^8\)

Furthermore if

\(^8\) Unless otherwise stated, \( \lambda_{t} \) will always be of this form.
\[ \bar{x} = N^{-1} \sum_{m=1}^{N} x(m), \]

\[ c(n) = (N-n)^{-1} \sum_{m=1}^{N-n} (x(m) - \bar{x})(x(m+n) - \bar{x}), \quad n = 0, \ldots, p, \]

then the spectral density estimate which we require is defined by

\[ \hat{r}(\lambda_t) = (2\pi)^{-1} (2\pi + 2 \sum_{n=1}^{p} c(n) \cos n\lambda_t). \]

Having determined these quantities, an initial estimate of \( \alpha \) (where \( \alpha = (\alpha(1) \ldots \alpha(p)) \)), is given by

\[ \hat{\alpha} = -A^{-1} \hat{a}, \]

where \( A \) has \( \hat{a}(k-\ell) \) in row \( k \) column \( \ell \), while \( a \) has \( \hat{a}(k) \) in row \( k \);

\[ \hat{a}(k) = N^{-1} \sum_{t} \hat{r}^{-2}(\lambda_t) I(\lambda_t) \cos k\lambda_t. \]

We may now use \( \hat{\alpha} \) to form

\[ \hat{g}(\lambda_t) = \sum_{j=0}^{p} \hat{\alpha}(j) e^{ij\lambda_t}, \quad \hat{g}(0) = 1, \]

and then

\[ \hat{a}(1)(k) = N^{-1} \sum_{t} |\hat{g}(\lambda_t)|^{-1} I(\lambda_t) \cos k\lambda_t, \quad k = 0, \ldots, p, \]

so that

\[ \hat{\alpha}(1) = -A(1)^{-1} \hat{a}(1), \quad (3.1.4) \]

where \( A(1), \hat{a}(1) \) are formed from the \( \hat{a}(1)(k) \) in an obvious way. Letting

\[ \tilde{\alpha}(1) = 2\hat{\alpha} - \hat{\alpha}(1), \]
Hannan proves that $\sqrt{N}(\tilde{\alpha}^{(1)} - \alpha)$ is asymptotically normally distributed and efficient. Furthermore if we express $\tilde{\alpha}^{(1)}$ in the form

$$\tilde{\alpha}^{(1)} = \hat{\alpha}^{(1)} + 2(\hat{\alpha} - \tilde{\alpha}^{(1)})$$

we see that the estimate $\tilde{\alpha}^{(1)}$ requires the addition of the "correction factor" $2(\hat{\alpha} - \tilde{\alpha}^{(1)})$ in order to make it efficient. Thus the steps in the estimation procedure are as follows: form $\hat{\alpha}$ and then use $\hat{\alpha}$ to form $\tilde{\alpha}^{(1)}$ and hence $\tilde{\alpha}^{(1)}$, as described above. The procedure is now iterated, the second iteration commencing by using $\tilde{\alpha}^{(1)}$ in place of $\hat{\alpha}$ to form $\tilde{\alpha}^{(2)}$ and hence $\tilde{\alpha}^{(2)}$, where

$$\tilde{\alpha}^{(2)} = 2\tilde{\alpha}^{(1)} - \tilde{\alpha}^{(2)}$$

and so on. In general we have, from the $k$th iteration,

$$\tilde{\alpha}^{(k)} = 2\tilde{\alpha}^{(k-1)} - \tilde{\alpha}^{(k)}, \quad k > 1. \quad (3.1.5)$$

An efficient estimate of the covariance matrix of $\sqrt{N}(\tilde{\alpha}^{(k)} - \alpha)$ is given by

$$\tilde{\Sigma}(k)^{-1},$$

where $\tilde{\Sigma}(k)$ has $\tilde{\sigma}(k)_{m-n}$ in row $m$ column $n$; $m, n = 1, \ldots, p$, and

$$\tilde{\sigma}(k)_{m-n} = \frac{1}{N} \sum_{t} |\tilde{g}(k)(\lambda_t)|^{-2} e^{i(m-n)\lambda_t},$$

with

$$\tilde{g}(k)(\lambda_t) = \sum_{j=0}^{p} \tilde{\alpha}^{(k)}(j)e^{ij\lambda_t}, \quad \tilde{\alpha}^{(k)}(0) = 1.$$

We emphasise that $\tilde{\alpha}^{(1)}$ is asymptotically efficient so that, in that sense, iteration is not necessary. However in practice one should iterate. Basically we are trying to solve a non-linear equation and the determination of an iterative scheme is not a disadvantage since any technique for solving the equation will have to be iterative.
3.2 Convergence of the Iterative Procedure

For simplicity we shall, in this section, consider the model (3.1.3).

In order to examine the rate of convergence of the iterative procedure used to obtain the efficient estimates defined in the previous section, we consider

\[ h(\alpha) = -\left( \sum_{t=1}^{N} |1+\alpha e^{-i\lambda_t}|^{-1} I(\lambda_t) \cos \lambda_t \right) / \left( \sum_{t=1}^{N} |1+\alpha e^{-i\lambda_t}|^{-1} I(\lambda_t) \right). \]  

(3.2.1)

From (3.1.4) and (3.1.5) it follows that, for \( k > 1 \),

\[ \tilde{\alpha}(k) = \tilde{\alpha}(k-1) - h(\tilde{\alpha}(k-1)). \]

Furthermore, letting

\[ m(\alpha) = 2\alpha - h(\alpha), \]

we have, for \( k > 1 \),

\[ \tilde{\alpha}(k) = m(\tilde{\alpha}(k-1)). \]

If we are considering a value \( \alpha \) near to some true value \( \alpha_0 \) say, it follows that if we start near enough to this true value and \( |m'(\alpha)| < 1 \), i.e. the absolute value of the first derivative of \( m(\alpha) \) is less than unity, then the iterative procedure will converge well (see for example Noble [41]).

From (3.2.1), if we let

\[ g(\lambda) = \frac{i\lambda_t}{1+\alpha e^{-i\lambda}}, \]

then it is not hard to see that if \(-1+\epsilon < \alpha < 1-\epsilon, \epsilon > 0\), then

\[ \lim_{N \to \infty} [h(\alpha)] = - \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(\lambda)|^{-1} |1+\alpha_0 e^{-i\lambda}|^2 \cos \lambda \, d\lambda \right] \]

\[ / \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(\lambda)|^{-1} |1+\alpha_0 e^{-i\lambda}|^2 \, d\lambda \right] \]

\[ = k(\alpha), \text{ say.} \]  

(3.2.3)
Indeed, considering the denominator of (3.2.1), for some \( \varepsilon > 0 \),

\(-1+\varepsilon < \alpha < 1-\varepsilon\), \( |1+\alpha e^{t}|^{-\frac{1}{4}} \) may be expanded in an infinite series of powers of \( e^{-t} \), the coefficients of which decay exponentially, i.e.

\[
|1+\alpha e^{t}|^{-\frac{1}{4}} = \sum_{j=-\infty}^{\infty} g_{j} e^{ij\lambda t}, \quad |g_{j}| < k\tau^{-|j|}, \quad \tau > 1.
\]

Now for some \( M, \nu \) it follows that

\[
\sum_{j=-\infty}^{\infty} g_{j} e^{ij\lambda t} = \sum_{j=-M}^{M} g_{j} e^{ij\lambda t} + \nu
\]

(where \( |\nu| \) may be made, uniformly in \( t \), arbitrarily small by choosing \( M \) sufficiently large). Thus

\[
\frac{1}{N} \sum_{t} |1+\alpha e^{t}|^{-\frac{1}{4}} I(\lambda t) = \frac{1}{N} \sum_{j=-M}^{M} g_{j} e^{ij\lambda t} I(\lambda t) + \frac{\nu}{N} \sum_{t} I(\lambda t)
\]

\[
= \frac{1}{2\pi} \sum_{j=-M}^{M} g_{j} \gamma(j) + \eta \quad \text{say},
\]

where \( \gamma(j) = E(x(n)x(n+j)) = \gamma(-j), \ j = 0, 1, \ldots, \ \eta = (2\pi)^{-1} \nu \gamma(0) \), and we have neglected terms which converge almost surely to zero. But

\[
\gamma(j) = \int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda) d\lambda,
\]

and we may choose \( \eta \) arbitrarily small so that we obtain, for \(-1+\varepsilon < \alpha < 1-\varepsilon\),

\[
\lim_{N \to \infty} \left[ \frac{1}{N} \sum_{t} |1+\alpha e^{t}|^{-\frac{1}{4}} I(\lambda t) \right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\sum_{j=-\infty}^{\infty} g_{j} e^{ij\lambda}) f(\lambda) d\lambda
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(\lambda)|^{-\frac{1}{4}} f(\lambda) d\lambda = (\sigma^{2}/2\pi) \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(\lambda)|^{-\frac{1}{4}} |1+\alpha e^{i\lambda}|^{2} d\lambda.
\]

Similarly
\[
\lim_{N \to \infty} \left[ \frac{1}{N} \sum_{t=0}^{N} e^{i\lambda t} \right] = (\sigma^2/2\pi) \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \, d\lambda.
\]

(3.2.5)

From (3.2.4) and (3.2.5) we obtain (3.2.3), as required.

Now
\[
h'(\alpha) = -\frac{\alpha}{\alpha - 1} \left[ \sum_t g(\lambda_t) \right] \left( \sum_t g(\lambda_t) \right) - 2 \left( \alpha \right) \left( \alpha + 1 \right)
\]

and, as we show in the appendix at the end of this chapter,
\[
\lim_{N \to \infty} \left[ \sum_t g(\lambda_t) \right] = 2, \quad \alpha \neq \pm 1.
\]

(3.2.6)

Furthermore,
\[
\lim_{\alpha \to \pm 1} \lim_{N \to \infty} h'(\alpha) = 0.
\]

(3.2.7)

As we have pointed out, a sufficient condition for the iterative procedure to converge is that \( |m'(\alpha)| < 1 \), for some \( \alpha \) near to the true value \( \alpha_0 \). From (3.2.2) we see this to be equivalent to \( 3 > h'(\alpha) > 1 \). (3.2.6) shows this condition to be (asymptotically) satisfied so that for \( \alpha \) in some suitable neighbourhood of \( \alpha_0 \), the iterative procedure will converge. However there is one region where this condition is not satisfied. Indeed as (3.2.7) indicates, for values of \( \alpha \) near the boundary of the acceptable region, \( h'(\alpha) \) is (asymptotically) near zero, so that in this region the iterative procedure may not converge.
To illustrate these theoretical results a number of sampling experiments were carried out for \( N = 100 \) observations (using generated data) and for various values of \( \alpha_0 \), namely 0.2, 0.5 and 0.9. The functions \( k(\alpha) \), \( h(\alpha) \) and \( h'(\alpha) \) were computed for various values of \( \alpha \) between -0.9 and 0.9, these results being recorded in Table 3.1 at the end of this section.

From these results we see that in all cases the sample values \( h(\alpha) \) are very close to their limiting values \( k(\alpha) \). For values of \( \alpha \) near to \( \alpha_0 \), \( h'(\alpha) \) is about 2, hence satisfying (3.2.6). Finally for values of \( \alpha \) near to the upper and lower limits, the function \( h'(\alpha) \) tends to zero in accordance with (3.2.7).
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\alpha_0 = 0.2$</th>
<th>$\alpha_0 = 0.5$</th>
<th>$\alpha_0 = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k(\alpha)$</td>
<td>$h(\alpha)$</td>
<td>$h'(\alpha)$</td>
</tr>
<tr>
<td>-0.9</td>
<td>-0.9959</td>
<td>-0.9957</td>
<td>0.0633</td>
</tr>
<tr>
<td>-0.8</td>
<td>-0.9840</td>
<td>-0.9780</td>
<td>0.2785</td>
</tr>
<tr>
<td>-0.7</td>
<td>-0.9592</td>
<td>-0.9466</td>
<td>0.3289</td>
</tr>
<tr>
<td>-0.6</td>
<td>-0.9200</td>
<td>-0.9124</td>
<td>0.3665</td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.8646</td>
<td>-0.8706</td>
<td>0.4860</td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.7904</td>
<td>-0.8122</td>
<td>0.6939</td>
</tr>
<tr>
<td>-0.3</td>
<td>-0.6953</td>
<td>-0.7295</td>
<td>0.9693</td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.5773</td>
<td>-0.6175</td>
<td>1.2709</td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.4355</td>
<td>-0.4760</td>
<td>1.5513</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.2708</td>
<td>-0.3090</td>
<td>1.7797</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.0870</td>
<td>-0.1222</td>
<td>1.9448</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1088</td>
<td>-0.0775</td>
<td>2.0351</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3062</td>
<td>0.2813</td>
<td>2.0227</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4929</td>
<td>0.4774</td>
<td>1.8740</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6568</td>
<td>0.6913</td>
<td>1.5821</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7892</td>
<td>0.7905</td>
<td>1.1925</td>
</tr>
<tr>
<td>0.7</td>
<td>0.8664</td>
<td>0.8896</td>
<td>0.7956</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9498</td>
<td>0.9520</td>
<td>0.4665</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9848</td>
<td>0.9858</td>
<td>0.2243</td>
</tr>
</tbody>
</table>
3.3 An Extension of the Estimation Procedure

The efficient estimates obtained using the method outlined in section 3.1 are, as we have seen (after \( \tilde{\alpha}^{(1)} \)), of the form

\[
\tilde{\alpha}^{(k)} = \tilde{\alpha}^{(k-1)} - \tilde{\alpha}^{(k)}.
\]

In general this estimation procedure works well. Defined in this way however there is no necessity for the \( \tilde{\alpha}^{(k)} \) to lie in the acceptable region. Consequently for the case when we are attempting to estimate coefficients for which the corresponding z transform has zeros close to the boundary of the unit circle it is possible that \( \tilde{\alpha}^{(k)} \), say (the efficient estimate of \( \alpha \) after \( k \) iterations) may fall outside of the acceptable region. If such a situation arises, as the iterations proceed the ensuing estimates will explode, and so the estimation procedure will not work.

It would be convenient for us to be able to introduce, in some way, a method by which the efficient estimates we obtain at the end of each iteration are confined to lie within the acceptable region. In order to do this we proceed as follows: when iterating the estimation procedure in the manner outlined in section 3.1 suppose \( \tilde{\alpha}^{(k)} \) is the first efficient estimate we obtain which is such that it lies outside of the acceptable region. In this case we introduce into this iteration a further step (such that the last estimate in this iteration will lie in the acceptable region). Indeed we use \( \tilde{\alpha}^{(k-1)} \) to form

\[
\tilde{g}^{(k-1)}(\lambda_t) = \sum_{j=0}^{p} \tilde{\alpha}^{(k-1)}(j)e^{ij\lambda_t}, \quad \tilde{\alpha}^{(k-1)}(0) = 1,
\]

and then form

\[
\hat{\alpha}^{(k)} = \frac{\tilde{g}^{(k)}(\lambda_t)}{\tilde{\alpha}^{(k-1)}}(\lambda_t), \quad (3.3.1)
\]
where $\tilde{\alpha}(k)$, $\tilde{\alpha}(k)$ (with typical elements $\tilde{a}(k)(m-n)$, $\tilde{a}(k)(m)$ respectively, $m,n = 1, \ldots, p$) are formed from the $g(k-1)(\lambda_t)$ in the same way as $\hat{\alpha}(1)$, $\hat{\alpha}(1)$ are formed from the $g(\lambda_t)$. Our reasons for doing this are based primarily on the fact that estimates of the form (3.3.1) have been shown to be such that the corresponding $z$ transform has all its zeros outside of the unit circle (and so these estimates will lie in the acceptable region, as required). For a discussion of this see Hannan [26] p 383. In fact applying Hannan's argument to the estimates defined by (3.3.1), a necessary and sufficient condition that the corresponding $z$ transform has all its zeros outside of the unit circle is that the matrix with entries $\tilde{a}(k)(m-n)$ in row $m$, column $n$; $m,n = 0, \ldots, p$ be positive definite. If this is not so it means that

$$\sum_t \left\{ |\tilde{g}(k-1)(\lambda_t)|^{-1} I(\lambda_t) \sum_{m,n=0}^p u(m)u(n) = 0 \right\} = 0$$

for some $u(m)$ not all zero, i.e.

$$\sum_t |\tilde{g}(k-1)(\lambda_t)|^{-1} w(\lambda_t) \sum_{m=0}^p u(m)e^{im\lambda_t} = 0,$$

where $I(\lambda_t) = w(\lambda_t)w^*(\lambda_t)$. Since $\sum_{m=0}^p u(m)e^{im\lambda_t}$ can be zero for at most $p$ values of $\lambda_t$, this implies that $(N-p)$ values of $w(\lambda_t)$ (and hence $I(\lambda_t)$) are null, which of course will not be so. Thus the result follows.

The computation of $\tilde{\alpha}(k)$ is now taken as the last step in the $k$th iteration. The $(k+1)$st iteration then proceeds as before except that now $\tilde{\alpha}(k)$ replaces $\tilde{\alpha}(k)$ when forming $\tilde{\alpha}(k+1)$. Indeed for this iteration we use $\tilde{\alpha}(k)$ to form $\tilde{\alpha}(k+1)$ and then use $\tilde{\alpha}(k+1)$ to form $\tilde{\alpha}(k+1) = 2\tilde{\alpha}(k) - \tilde{\alpha}(k+1)$. 
If \( \hat{\alpha}^{(k+1)} \) lies in the acceptable region then this completes the \((k+1)\)st iteration. (If \( \hat{\alpha}^{(k+1)} \) lies outside of this region however we introduce the extension of the procedure once more and form \( \hat{\alpha}^{(k+1)} \) (which we know will lie in the acceptable region), and this will then be the last step in the \((k+1)\)st iteration, and so on).\(^9\)

Following a similar proof to that occurring in the appendix of chapter 6 of Hannan [26] for the \( \sqrt{N}(\hat{\alpha}^{(k)}-\alpha) \), \( k \geq 1 \), it can be shown that the \( \sqrt{N}(\hat{\alpha}^{(k)}-\alpha) \) are asymptotically normally distributed and efficient.

An efficient estimate of the covariance matrix of \( \sqrt{N}(\hat{\alpha}^{(k)}-\alpha) \) is given by \( \hat{\Sigma}^{(k)} \), where \( \hat{\Sigma}^{(k)} \) has \( \hat{\Sigma}^{(k)}(m,n) \) in row \( m \), column \( n \); \( m,n = 1, \ldots, p \), and

\[
\hat{\Sigma}^{(k)}(m,n) = N^{-1} \sum_t \hat{g}^{(k)}(\lambda_t)^{-1} e^{i(m-n)\lambda_t},
\]

with \( \hat{g}^{(k)}(\lambda_t) \) defined in an obvious way.

A number of sampling experiments based on this extension (of the estimation procedure of section 3.1) were carried out for the model (3.1.3) with \( \alpha = 0.96 \). Twenty replications of the procedure were computed, new data being generated for each replication. When \( N = 100 \) observations were used for each replication the average value of the estimates of \( \alpha \) for these twenty replications was found to be 0.8751. A similar experiment with \( N = 200 \) gave an average \( \alpha \) value of 0.9017.

\(^9\) We emphasize that this extra step need only be included in say the \( r \)th iteration when \( \hat{\alpha}^{(r)} \) lies outside of the acceptable region.
At first sight these estimates for 0.96 appear to be surprisingly low. A closer examination of the situation however shows them to be not altogether unexpected. Indeed from the asymptotic theory we would expect the estimates for each replication to be asymptotically normally distributed with mean 0.96 and standard deviation

\[ [(1-\alpha^2)/N]^{1/2} = [(1-0.96^2)/100]^{1/2} = 0.028. \]

(For the model (3.1.3) it follows that

\[ \Phi = \frac{1}{2\pi} \int_{-\pi}^{\pi} |1+\alpha e^{i\lambda}|^{-2} d\lambda = (1-\alpha^2)^{-1}, \]

so that the theoretical variance of \( \hat{\alpha}(k) \) is given by \((1-\alpha^2)/N\). Thus 95% of the estimates would be expected to lie between 0.9051 and 1.0149. However we know that in this case we must have \(|\alpha| < 1\), so that the distribution of the estimates around 0.96 is truncated at unity. Consequently it would be expected and indeed is found that the distribution of the estimates is skew to the left of the true value (or, of course, to the right if we are attempting to estimate values near -1), thus helping to explain the mean values we have obtained.

The variance of the estimates will probably be affected as well by this distortion. Table 3.2 compares the theoretical (defined above) and the observed variance for each N, the observed variance being computed from the twenty estimates from which the average value of \( \alpha \) was calculated, in each case.
**Table 3.2**

<table>
<thead>
<tr>
<th></th>
<th>N = 100</th>
<th>N = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical variance</td>
<td>0.00078</td>
<td>0.00039</td>
</tr>
<tr>
<td>Observed variance</td>
<td>0.00619</td>
<td>0.00262</td>
</tr>
</tbody>
</table>

From this table we see the observed variance to be, in each case, much larger than the theoretical variance.

For N = 100 in the sixth replication the iterative procedure gave an estimate of 0.6178. As this was so low (as an estimate of 0.96) an examination of the likelihood function was made to verify that 0.6178 was the value of \( \alpha \) which, for this particular set of data, did in fact minimize

\[
\sum_t f^{-1}(\lambda_t) I(\lambda_t).
\]  

(3.3.3)

\( I(\lambda_t) \) and \( f(\lambda_t) \) are of course the periodogram ordinate and the spectral density respectively, each at frequency \( \lambda_t \). (As we have seen in section 1.3, maximizing the likelihood function is, asymptotically, equivalent to minimizing the function (3.3.3)). For the model (3.1.3) it is not hard to see that minimizing (3.3.3) is equivalent to minimizing

\[
\sum_t \{ I(\lambda_t)/(1+\alpha \cos \lambda_t+\alpha^2) \}.
\]  

(3.3.4)

Using the data of this replication and plotting the function (3.3.4) for values of \( \alpha \) from -0.99 to 0.99 in steps of 0.01, it was found that this function had a minimum which occurred between 0.61 and 0.62. This of course verifies that for this particular set of data the value 0.6178 was
indeed the estimate of $\alpha$ which minimized (3.3.3). Thus in this case the
poor agreement with the asymptotic theory appears to be due to the sample
size being too small in relation to the value of $\alpha$ being estimated, and
not to the procedure having failed to reproduce the properties of the
maximum likelihood procedure.

3.4 Alternative Estimation Procedures

As has already been mentioned, a number of estimation procedures have
appeared in the literature over the last ten years. Hannan [25] has
summarized a number of these, outlining the disadvantages associated with
each method. One procedure that warrants further investigation in the
present context however is that due to Durbin [14].

Although rather heuristic, Durbin's approach to the estimation problem
is similar to that used by Hannan. After a Fourier transformation of the
data, he minimizes the exponent of the resulting likelihood function which he
shows may be expressed (approximately) in the form (on neglecting constant
factors)

$$\sum_{k=1}^{2M} f^{-1}(\lambda_k)\hat{f}(\lambda_k),$$

where $\hat{f}(\lambda_k)$ represents the smoothed estimate of the spectral density $f(\lambda_k)$,
with $\lambda_k$ the midpoint of the $k$th band of frequencies, $k = 1, \ldots, 2M$.

Suppose $f(\lambda_k)$ is a rational function, say

$$f(\lambda_k) = K\left|\sum_{j=0}^{p} \alpha(j)e^{i\lambda_k j} / \sum_{s=0}^{q} \beta(s)e^{is\lambda_k s}\right|^2,$$

$\alpha(0) = \beta(0) = 1$.

---

10 I would like to thank Professor J. Durbin for his comments and
suggestions during the writing of this section.
(as is the case for a mixed autoregressive moving average model for example in which case the $\alpha(j)$, $j = 0, \ldots, p$ represent the moving average coefficients, while the $\beta(s)$, $s = 0, \ldots, q$ are the autoregressive coefficients). Substituting this in (3.4.1), it follows that the minimization of (3.4.1) with respect to the $\beta(s)$, $s = 1, \ldots, q$, for given values of $\alpha(j)$, $j = 1, \ldots, p$ is quite straightforward since we obtain a set of equations which are linear in the estimates required. Minimization with respect to $\alpha(j)$, $j = 1, \ldots, p$ for given values of $\beta(s)$, $s = 1, \ldots, q$ is not quite so simple however since these parameters occur in the denominator of (3.4.1). To simplify this problem Durbin showed that

$$\hat{I}(\lambda_k)\hat{f}^{-2}(\lambda_k) \frac{d(f(\lambda_k))}{d\theta} \equiv \hat{I}^{-1}(\lambda_k) \frac{d(f(\lambda_k))}{d\theta},$$

which has the effect that instead of minimizing (3.4.1) we may minimize

$$\sum_{k=1}^{2M} \hat{I}^{-1}(\lambda_k) f(\lambda_k),$$

whenever this is more convenient. (Kendall and Stuart ([34] p 485) refer to Durbin's procedure when considering the estimation of the coefficients of the mixed autoregressive moving average model from the observed spectrum).

Durbin's main estimation procedure, using his equations (29) and (30), is based on the minimization of (3.4.2). In addition he suggests an alternative procedure, using his equations (29) and (31), which is based on the minimization of (3.4.1). The estimates given by this alternative procedure correspond to what Hannan refers to as $\hat{\alpha}^{(1)}$, $\hat{\beta}^{(1)}$ (where $\alpha' = (\alpha(1) \ldots \alpha(p))$; $\beta' = (\beta(1) \ldots \beta(q))$) when estimating the parameters of the rational density function defined above. These estimates however, as Hannan proves, are not efficient.
In the case of the moving average model (3.1.1) the estimates given by Durbin's alternative procedure correspond to what we have called \( \hat{\alpha}^{(1)} \) in section 3.1, and which require the addition of the correction factor \( 2(\hat{\alpha} - \hat{\alpha}^{(1)}) \) to make them efficient. As a result the (asymptotic) variances of these estimates diverge as we iterate Durbin's alternative procedure. Indeed considering the model (3.1.3) for simplicity, suppose we let

\[
\hat{\alpha} = \left( \sum_{k=1}^{2M} f^{-2}(\lambda_k)I(\lambda_k) e^k \right) \left( \sum_{k=1}^{2M} f^{-2}(\lambda_k)I(\lambda_k) \right)^{-1}
\]

and \( \hat{\alpha}^{(j)} \), \( j = 1, 2, \ldots \) be the estimate of \( \alpha \) after \( j \) iterations of the estimation procedure (using 3.4.1)). Then, following a similar argument to that in the appendix of chapter 6 of Hannan [26], for \( j > 1 \) it can be shown (on neglecting terms which converge in probability to zero) that

\[
\sqrt{N}(\hat{\alpha}^{(j)} - \alpha) = \sqrt{N}(\hat{\alpha} - \alpha) + 2 \sqrt{N}(\hat{\alpha}^{(j-1)} - \alpha), \tag{3.4.3}
\]

where \( \sqrt{N}(\hat{\alpha} - \alpha) \) and \( \sqrt{N}(\hat{\alpha}^{(j)} - \alpha) \) are asymptotically normally distributed.

Suppose \( v_o \) and \( v_j \) represent the limiting variances of \( \sqrt{N}(\hat{\alpha} - \alpha) \) and \( \sqrt{N}(\hat{\alpha}^{(j)} - \alpha) \) respectively while \( c_j \) represents the limiting covariance between \( \sqrt{N}(\hat{\alpha} - \alpha) \) and \( \sqrt{N}(\hat{\alpha}^{(j)} - \alpha) \). Then from (3.4.3) we have

\[
v_j = v_{j-1} + 2v_{j-1}^3 + 4c_{j-1}
\]

\[
= (v_o v_{j-1})^{\frac{1}{2}}(e_{j-1} + 3e_{j-1}^{-1} + 4p_{j-1})
\]

say, where \( e_{j-1} = (v_o v_{j-1})^{\frac{1}{2}} \), \( p_{j-1} = c_{j-1} (v_o v_{j-1})^{-\frac{1}{2}} \). If \( e_{j-1} < 1 \), then for some \( \epsilon > 0 \), \( e_{j-1} = 1 - \epsilon \), and it follows that

\[
e_{j-1} + 3e_{j-1}^{-1} = 1 - \epsilon + 3(1 - \epsilon)^{-1} > 4 + \eta
\]
say, where \( \eta > 0 \). Furthermore the correlation coefficient \( \rho_{j-1} > -1 \), so that

\[
v_j > v_{j-1} + (v_0 v_{j-1} \frac{1}{2}(4+\eta-\eta)) = v_{j-1} + \eta(v_0 v_{j-1})^{\frac{1}{2}},
\]

i.e. \( e_j < e_{j-1} \), and since \( e_1 < 1 \) the induction is complete. From this we see that even in the case of the simple moving average model (3.1.3), as we iterate Durbin's estimation procedure the (asymptotic) variance of the estimate diverges.

For the model (3.1.1) minimizing (3.4.2) gives

\[
\hat{\alpha} = -\hat{\alpha}^{-1} \hat{\alpha},
\]

(3.4.4)

where for this particular case we have \( \hat{\alpha}' = (\hat{\alpha}(1) \ldots \hat{\alpha}(p)) \) and \( \hat{\alpha} \) has \( \hat{\alpha}(m-n) \) in row \( m \) column \( n \), while \( \hat{\alpha} \) has \( \hat{\alpha}(m) \) in row \( m \); \( m,n = 1,\ldots,p \) with

\[
\hat{\alpha}(m-n) = \sum_{k=1}^{2M} \hat{\alpha}^{-1}(\lambda_k) e^{i(m-n)\lambda_k}.
\]

This estimate corresponds to the estimate obtained from Durbin's main procedure, i.e. his equation (30). From the form of (3.4.4) we see that in this case we are unable to iterate the estimation procedure and furthermore the estimates are dependent on just how we choose our smoothed spectral estimates \( \hat{f}(\lambda_k), k = 1,\ldots,2M \).

To investigate this dependence further a number of numerical experiments were carried out using Durbin's method with data generated by the first order moving average model (3.1.3) with \( \alpha \) taking, in turn, the values \(-0.7, -0.2, 0.2, 0.5 \) and \( 0.96 \). For this model (3.4.4) becomes
\[ \hat{\alpha} = \frac{\sum_{k=1}^{2M} \hat{I}(\lambda_k) \cos \lambda_k}{\sum_{k=1}^{2M} \hat{I}(\lambda_k)}. \]  

(3.4.5)

In these experiments the \( \hat{I}(\lambda_k) \), \( k = 1, \ldots, 2M \) were estimated by averaging the periodogram ordinates over each of \( 2M \) bands. The number of bands was allowed to vary. Indeed \( M \) was taken in turn as \( 3, 5, 6, 10 \) and \( 15 \). With \( N = 108 \) observations twenty replications of each experiment (i.e. for each \( \alpha = -0.7, -0.2, 0.2, 0.5 \) and \( 0.96 \), and \( M = 3, 5, 6, 10 \) and \( 15 \)) were carried out using the generated data. (We chose \( N = 108 \) merely for computational convenience as this value may be factored a number of different ways, a fact which proves convenient when varying the bandwidth over which the spectral density is to be estimated). The results are recorded in Table 3.3 where, for each \( \alpha \) and \( M \), row (a) gives the mean of the estimated values of \( \alpha \) for the 20 runs, row (b) gives the observed variance of the 20 numbers and row (c) gives the mean square error (m.s.e.) for each estimate, where

\[ \text{m.s.e.} = (\text{bias})^2 + \text{variance}. \]
<table>
<thead>
<tr>
<th>α</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.7</td>
<td>-0.5683</td>
<td>-0.6230</td>
<td>-0.6441</td>
<td>-0.6929</td>
<td>-0.7253</td>
</tr>
<tr>
<td></td>
<td>0.0042</td>
<td>0.0040</td>
<td>0.0037</td>
<td>0.0058</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>0.0215</td>
<td>0.0099</td>
<td>0.0068</td>
<td>0.0059</td>
<td>0.0051</td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.2423</td>
<td>-0.1876</td>
<td>-0.1953</td>
<td>-0.2514</td>
<td>-0.3388</td>
</tr>
<tr>
<td></td>
<td>0.0049</td>
<td>0.0063</td>
<td>0.0049</td>
<td>0.0106</td>
<td>0.0076</td>
</tr>
<tr>
<td></td>
<td>0.0067</td>
<td>0.0065</td>
<td>0.0049</td>
<td>0.0132</td>
<td>0.0269</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1059</td>
<td>0.1965</td>
<td>0.1846</td>
<td>0.1067</td>
<td>-0.0336</td>
</tr>
<tr>
<td></td>
<td>0.0058</td>
<td>0.0082</td>
<td>0.0075</td>
<td>0.0123</td>
<td>0.0908</td>
</tr>
<tr>
<td></td>
<td>0.0147</td>
<td>0.0082</td>
<td>0.0077</td>
<td>0.0210</td>
<td>0.1454</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3543</td>
<td>0.4780</td>
<td>0.4662</td>
<td>0.3692</td>
<td>0.2034</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.0077</td>
<td>0.0086</td>
<td>0.0118</td>
<td>0.0112</td>
</tr>
<tr>
<td></td>
<td>0.0262</td>
<td>0.0082</td>
<td>0.0097</td>
<td>0.0289</td>
<td>0.0992</td>
</tr>
<tr>
<td>0.96</td>
<td>0.5256</td>
<td>0.7750</td>
<td>0.7931</td>
<td>0.7029</td>
<td>0.5899</td>
</tr>
<tr>
<td></td>
<td>0.0036</td>
<td>0.0045</td>
<td>0.0066</td>
<td>0.0107</td>
<td>0.0147</td>
</tr>
<tr>
<td></td>
<td>0.1923</td>
<td>0.0387</td>
<td>0.0345</td>
<td>0.0768</td>
<td>0.1517</td>
</tr>
</tbody>
</table>

From Table 3.3, considering the case $\alpha = -0.7$ for example, we see that the best estimate (in the sense of minimum mean square error), for the few values of $M$ we have chosen, occurs when $M = 15$. For $\alpha = 0.5$ however the best estimate occurs when $M = 5$. Choosing the best estimate in each case (i.e. for each value of $\alpha$ we have chosen) it is not hard to see that the variance associated with each of these estimates is (with the exception of the case $\alpha = 0.96$) not significantly different from the theoretical variance of the corresponding efficient estimate, defined below (3.3.2). Indeed for $\alpha = 0.2$ and $M = 6$ for example the observed variance is 0.0075, as compared to a theoretical variance of $$(1-(0.2)^2)/108 = 0.0089,$$
and the ratio of these two variances (namely 1.19) is not significant on an F distribution with \( \infty \) and 19 degrees of freedom (for which \( F_{\infty,19} = 1.88 \) at the 5\% level of significance). However for this same value of \( \alpha \) when \( M = 15 \) the observed variance is 0.0908, and in this case the ratio of the two variances (i.e. 10.20) is obviously significant on an F distribution with 19 and \( \infty \) degrees of freedom (indeed \( F_{19,\infty} = 1.91 \) at the 1\% level of significance).

Furthermore, with regard to the estimates of the mean, we have

\[
P(\bar{\alpha} - 1.96\sigma/\sqrt{n} < \bar{\alpha} < \bar{\alpha} + 1.96\sigma/\sqrt{n}) = 0.95,
\]

where of course \( \bar{\alpha} \) is estimated over 20 replications, so that \( n = 20 \). For the case we have just considered, namely \( \alpha = 0.2 \) (for which \( \sigma^2 = 0.0089 \)), on evaluation we find

\[
P(0.1586 < \bar{\alpha} < 0.2414) = 0.95.
\]

When \( M = 6 \) the estimated mean lies well within this region. However when \( M = 15 \) we see the estimate obtained, namely -0.0336, lies outside of the acceptable region.

With the help of this numerical example therefore we see that both the mean and variance (and hence the efficiency) of the estimates obtained are dependent on just how we choose the bandwidth over which to estimate the spectral density. Indeed for the few cases we have considered above we see that one particular choice of \( M \) may lead to estimates which are not significantly different (in both mean and variance) from the efficient estimates defined in section 3.1. However another choice of \( M \) may lead to estimates which are significantly different from the corresponding efficient estimates.
3.5 Conclusion

We have in this chapter extended the estimation procedure first proposed by Hannan [25] in such a way as to prevent the iterative procedure from giving estimates which explode as the iterations proceed (i.e. we have constrained the estimates in such a way that the corresponding z transform has all its roots outside of the unit circle). Of course if the number of observations N was large enough the possibility of such an occurrence would be small. In practice however it is often the case that such large amounts of data are not available. Furthermore in section 3.2, for the case of a first order moving average model, we have examined the rate of convergence of the iterative procedure used in this estimation method.

An examination of two methods of estimation due to Durbin shows one of his methods to give estimates which in fact correspond, in the case of the model (3.1.1), to what we have called $\hat{\alpha}^{(1)}$ in section 3.1, and hence are not fully efficient. The estimates obtained by Durbin's second estimation procedure are based on smoothed estimates of the spectral density function. The variance and hence the efficiency of the estimates obtained by this method however are dependent on just how the bandwidth (over which we estimate the spectral density) is chosen.
APPENDIX

Proof of equation (3.2.6)

In order to verify (3.2.6) suppose we let

\[ g = 1 + \alpha e^{i\lambda_t} \]

\( (g \) is of course a function of \( \lambda_t \). For convenience however we have omitted this argument variable). Then from (3.2.1) we have

\[ h(\alpha) = -N \sum_{t=1}^{N} |g|^{-\frac{1}{2}} I(\lambda_t) \cos \lambda_t / (\sum_{t=1}^{N} |g|^{-\frac{1}{2}} I(\lambda_t)) \]

On taking derivatives of both sides of this expression and following a similar type of argument as for (3.2.5), it can be shown that

\[ \lim_{N \to \infty} h'(\alpha) = -4 \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-\frac{1}{2}} f(\lambda) \cos \lambda d\lambda \right] \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-\frac{1}{2}} f(\lambda) \cos (\alpha \cos \lambda) d\lambda \right] \]

\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-\frac{1}{2}} f(\lambda) d\lambda \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-\frac{1}{2}} f(\lambda) d\lambda \right]^{-2} \]  

(A.3.1)

Furthermore, for \( \alpha = \pm 1 \), on simplification we find

\[ \lim_{N \to \infty} h'(\alpha_0) = -4 \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-\frac{1}{2}} e^{i\lambda \lambda_0} d\lambda \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-\frac{1}{2}} (\alpha_0 e^{i\lambda}) d\lambda \right) \right] \]

\[ = -\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-2} \lambda_0(d\lambda) \left( \frac{\alpha_0}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-\frac{1}{2}} e^{i\lambda \lambda_0} d\lambda + \frac{1}{4\pi} \int_{-\pi}^{\pi} |g_0|^{-\frac{1}{2}} (e^{2i\lambda_0} + 1) d\lambda \right) \]

\[ = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-2} d\lambda \right]^{-2} \]  

(A.3.2)

where \( g_0 = 1 + \alpha e^{i\lambda} \).
By residues it follows that

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-2} e^{-i\lambda} d\lambda = (1-\alpha_0^2)^{-1},
\]

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-2} e^{i\lambda} d\lambda = -\alpha_0 (1-\alpha_0^2)^{-1},
\]

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-1} e^{i\lambda} d\lambda = (1+\alpha_0^2)(1-\alpha_0^2)^{-1},
\]

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-1} e^{i\lambda} d\lambda = -2\alpha_0 (1-\alpha_0^2)^{-1},
\]

and

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_0|^{-1} e^{2i\lambda} d\lambda = \alpha_0^2 (3-\alpha_0^2)(1-\alpha_0^2)^{-3}.
\]

Substituting these values in (A.3.2) above we find, on simplification,
that for \( \alpha_0 \neq \pm 1 \),

\[
\lim_{N \to \infty} [h'(\alpha_0)] = [-\alpha_0 (1-\alpha_0^2)^{-1} \alpha_0 (1-\alpha_0^2)^{-1} + (1-\alpha_0^2)^{-1} 2(1+\alpha_0^2)(1-\alpha_0^2)^{-2}] / (1-\alpha_0^2)^{-2}
\]

\[
= 2,
\]

as required.
CHAPTER 4

THE ESTIMATION OF MIXED AUTOREGRESSIVE MOVING AVERAGE MODELS
WITH EXOGENOUS VARIABLES

4.1 Introduction

A number of procedures have been developed for the estimation of mixed
autoregressive moving average models. As we have already shown (see (1.2.7)),
these models may be expressed (in the scalar case) in the form

\[ \sum_{k=0}^{q} \beta(k) y(n-k) = \sum_{j=0}^{p} \alpha(j) e(n-j), \quad \alpha(0) = \beta(0) = 1. \] (4.1.1)

One reason for considering such models lies in the fact that they are more
general than (say) the scalar autoregressive model. Consequently when the
number of parameters to be estimated may be restricted by small amounts of
data being available their use may be necessary. Indeed suppose we wish to
fit a linear model to a particular set of data. Proceeding in a somewhat
ad hoc fashion, one method of determining which model best fits this data is
to begin by fitting, in turn, first and second order autoregressive models to
the data. Having done this we now apply a suitable goodness of fit test,
say the likelihood ratio test (see section 2.1 for a discussion of this test
in the vector case), to determine whether the extra term in the autoregression
has significantly changed the residual variance. If there is a significant
change we next fit an autoregression of order three and carry out the
likelihood ratio test once more (comparing the goodness of fit of this model
with the autoregression of order two), and so on, until the residual
variance becomes (nearly) constant. For example let us assume that the
likelihood ratio test shows there to be no significant difference between
the residual variances of autoregressions of order \( r \) and \( (r+1) \) say. The
fitting procedure would then be terminated and our model of best fit would, in this case, be an autoregression of order \( r \).

Suppose we only had 80 data points available however and it turned out that an autoregression of quite high order (e.g. ten or more) had to be fitted before the residual variance became (nearly) constant. Such a situation (i.e. the estimation of a large number of parameters from a relatively small amount of data) is unsatisfactory of course and so we are led to consider more general models such as (4.1.1). By fitting models of the form (4.1.1) the number of parameters to be estimated (in order that the variance of the residuals shall become constant) may be less than for the case when a high order autoregression is fitted. The estimation procedure for obtaining efficient estimates of these parameters is considerably more complicated to apply however (as compared to the procedure for estimating the parameters of the autoregressive model).

Considering the model (4.1.1), prior to developing a satisfactory method for the estimation of the coefficients of such models via spectral methods, Hannan [25] summarizes a number of estimation procedures already appearing in the literature. Indeed he discusses procedures due to Box and Jenkins [6], Durbin [12] and Walker [53] and points out the disadvantages associated with each method. Letting

\[
\mathbf{u}(n) = \sum_{k=0}^{q} \beta(k)\mathbf{y}(n-k) = \sum_{j=0}^{p} \alpha(j)\varepsilon(n-j) , \quad \alpha(0) = \beta(0) = 1,
\]

Parzen [46] has, in a recent paper, obtained efficient estimates of the \( \beta(k) \) from efficient estimates of the covariances of the \( \mathbf{u}(n) \) process. Furthermore using an algorithm due to Wilson [59] he then obtains efficient estimates
of the $\alpha(j)$, $j = 1, \ldots, p$ from these efficient covariance estimates.

Phillips [47] has considered the estimation problem by applying the Gauss-Newton procedure (an iterative technique for the solution of sets of non-linear equations) to the non-linear normal equations obtained from the likelihood function. (This approach is similar to that used by Box and Jenkins). More recently Pagan [42] has considered the estimation problem from a computational viewpoint by considering a number of algorithms for the solution of the non-linear normal equations. Indeed this author compares such iterative procedures as the Gauss-Newton and the Newton-Raphson techniques, and applies these procedures to a large number of Monte Carlo studies.

We shall be concerned with an extension of the estimation procedure used by Hannan to the more general mixed autoregressive moving average model in which exogenous variables are present. Such models are of importance in Econometrics for, as we shall see in the next chapter, they are a generalization of the rational distributed lag model. Furthermore, by considering this estimation procedure (under certain general conditions to be defined below), we shall be able to establish central limit theorems for the resulting estimates.

The mixed autoregressive moving average model with exogenous variables may be represented mathematically in the form

$$\sum_{k=0}^{q} \beta(k)y(n-k) + \sum_{k=1}^{r} \delta(k)x_k(n) = \sum_{j=0}^{p} \alpha(j)e(n-j)$$, \quad \alpha(0) = \beta(0) = 1. \quad (4.1.2)$$

In what follows we shall assume the following general conditions to hold:
(i) The $e(n)$ are i.i.d. $(0, \sigma^2)$.

(ii) All the zeros of the $z$ transform:
\[
g(z) = \sum_{j=0}^{p} \alpha(j)z^j, \quad h(z) = \sum_{k=0}^{q} \beta(k)z^k
\]
lie outside of the unit circle.

(iii) The $x_j(n)$ come from infinite sequences which satisfy, almost surely:
\[
\lim_{N \to \infty} N^{-1} \sum_{m=1}^{N} x_j(m)x_k(m+n) = \gamma_{jk}(n) = \gamma_{kj}(-n), \quad j, k = 1, \ldots, r;
\]
and
\[
n = 0, 1, \ldots,
\]
indeed of the $e(n)$ sequence.

(iv) $\delta' = (\delta(1) \ldots \delta(r)) \neq 0$, so that the model (4.1.2) is identified.
If $\delta = 0$, then as Hannan ([26] pp 388-389) shows, we must assume that $g(z)$, $h(z)$ have no roots in common and furthermore either $\alpha(p)$ or $\beta(q)$ is non-null.

The condition (ii) is reasonable, as we can see from the discussion of linear models in section 1.2. Indeed the restriction on the zeros of $g(z)$ is applied to uniquely identify the model, while that on the zeros of $h(z)$ is necessary if there is to exist a stationary solution $y(n)$ to (4.1.2), which is expressible in terms of the $e(n-k)$, $x_j(n-k)$, $k \geq 0$.
(Of course the possibility of stationarity will depend on the $x_j(n)$ sequence also).

As a consequence of (iii) we have
\[
\gamma_{jk}(n) = \int_{-\pi}^{\pi} e^{in\lambda} dF_{jk}(\lambda),
\]
where the matrix $F(\lambda)$, with $F_{jk}(\lambda)$ as entry in row $j$, column $k$ ($j, k = 1, \ldots, r$), is the spectral distribution matrix of the $x(n)$ sequence. It seems preferable
to treat the $x_j(n)$ as fixed sequences as in (iii), rather than to prescribe them stochastically, as this makes the treatment less restrictive.

Finally in this introduction we shall describe the sense in which we speak of (asymptotically) efficient estimates in the present context. If the $e(n)$ and $x_j(n)$ are Gaussian we may obtain the maximum likelihood estimates and consider their limiting distribution. Their asymptotic distribution will turn out to be independent of the Gaussian assumption given only that the $e(n)$ are i.i.d. $(0, \sigma^2)$ and the $x_j(n)$ are suitably restricted. As we pointed out in section 1.3 this is discussed in Walker [54]. (He deals basically with the univariate case without exogenous variables but it is clear that for the $x_j(n)$ suitably restricted the results will extend to our case). In what follows we shall (as we have already mentioned) establish a central limit theorem for our estimates (under the conditions imposed on the model (4.1.2)) and furthermore we shall speak of them as being (asymptotically) efficient if their limiting distribution is the same as that of the maximum likelihood estimates.

The likelihood equations are non-linear and are difficult to solve. The procedure which we shall follow is based on that outlined in section 1.3. As we have seen in that section, maximizing the likelihood function is asymptotically equivalent to minimizing the function (1.3.2). It is this function which we consider further and from which we shall derive asymptotically efficient (in the sense defined above) estimates of the parameters of the model (4.1.2).
4.2 The Estimation Procedure

Suppose we let

\[ \sum_{k=0}^{q} \beta(k)y(n-k) + \sum_{k=1}^{r} \delta(k)x(n) = \sum_{j=0}^{p} \alpha(j)\epsilon(n-j) = z(n), \quad (4.2.1) \]

say, where \( \alpha(o) = \beta(o) = 1 \), and consider a sample of size \( N \)
(i.e. \( z(1), \ldots, z(N) \)). Then from section 1.3 it follows that maximizing

the likelihood function is, asymptotically, equivalent to minimizing

\[ \frac{1}{N} \sum_{t=1}^{N} z_{t}^{-1}(\lambda_{t})I_{z}(\lambda_{t}). \quad (4.2.2) \]

\( I_{z}(\lambda_{t}) \) and \( f_{z}(\lambda_{t}) \) are respectively the periodogram ordinate and the spectral density at frequency \( \lambda_{t} = \lambda_{N-t} = 2\pi t/N, \ t = 1,2, \ldots, [N/2]. \)

In what follows we shall (for convenience) omit argument variables so that we will, for example, write \( I_{z} \) for \( I_{z}(\lambda_{t}) \), \( f_{z} \) for \( f_{z}(\lambda_{t}) \) etc.

We now introduce the generating (or response) functions

\[ g = \sum_{j=0}^{p} \alpha(j)\lambda_{t}^{j}, \quad h = \sum_{k=0}^{q} \beta(k)\lambda_{t}^{k}, \]

and the vectors \( \alpha, \beta, \delta \) with \( \alpha(j), \beta(k), \delta(\ell) \) in the \( j \)th, \( k \)th and \( \ell \)th place respectively, \( j = 1, \ldots, p; \ k = 1, \ldots, q; \ \ell = 1, \ldots, r \). Furthermore let

\( I_{x}, \ I_{y} \) represent the periodograms of the \( x(n) \) and \( y(n) \) sequences respectively, and \( I_{xy}, \ I_{yx} \) represent the cross periodograms between these two sequences. Then from (4.2.1) it is not hard to show that

\[ f_{z} = (\sigma^{2}/2\pi)|g|^{2} \]

\[ ^{11} \text{In what follows, unless otherwise stated, } \lambda_{t} \text{ will always be of this form.} \]
and (on neglecting end terms which, asymptotically, will not affect our results)

\[ I_z = |h|^2 I_y + 5' I_x 5 + 5h 5' I_{xy} + hI_{yx} 5. \]  

(4.2.3)

Substituting these values in (4.2.2), since the Jacobian of the transformation is unity (and on neglecting the constant multiplier \((2\pi/\sigma^2)\)), it follows that maximizing the likelihood function is asymptotically equivalent to minimizing the Hermitian form \( H \), where

\[ H = \frac{1}{N} \sum_{t} |g|^{-2} \left( |h|^2 I_y + 5' I_x 5 + 2h 5' I_{xy} \right). \]  

(4.2.4)

Let \( \hat{\alpha}(1) \), \( \hat{\beta}(1) \) and \( \hat{\delta}(1) \) represent the estimates obtained from the minimization of \( H \) with respect to \( \alpha \), \( \beta \) and \( \delta \) respectively. As we shall see the resulting equations are highly non-linear. We linearize these equations by inserting (where applicable) initial consistent estimates \( \hat{\alpha}, \hat{\beta} \) and \( \hat{\delta} \) of \( \alpha \), \( \beta \) and \( \delta \) respectively. (We shall discuss the computation of these latter estimates below). Indeed the partial derivatives of \( H \) with respect to \( \beta(\ell), \ell = 1, \ldots, q \) and \( \delta(k), k = 1, \ldots, r \) respectively lead to

\[ \frac{q}{\sum_{k=0}^{q} \hat{\beta}(1)(k) \left( \frac{1}{N} \sum_{t} |\hat{g}|^{-2} I_y e^{i(k-\ell)\lambda_t} \right)} = - \frac{1}{N} \sum_{t} |\hat{g}|^{-2} \hat{\delta}(1)' I_{xy} e^{-i\ell\lambda_t} \]  

(4.2.5)

\[ \ell = 1, \ldots, q, \quad \text{and} \]

\[ \frac{q}{\sum_{\ell=0}^{q} \hat{\beta}(1)(\ell) \left( \frac{1}{N} \sum_{t} |\hat{g}|^{-2} I_{xy} e^{-i\ell\lambda_t} \right)} = - \frac{1}{N} \sum_{t} |\hat{g}|^{-2} I_x \hat{\delta}(1), \]  

(4.2.6)

where

\[ \hat{g} = \sum_{j=0}^{q} \hat{\alpha}(j)e^{ij\lambda_t}, \quad \hat{\alpha}(0) = 1. \]
With some manipulation we may express (4.2.5) and (4.2.6) in a more concise form. Indeed from (4.2.5), since \( \hat{\beta}^{(1)}(\omega) = 1 \), we have, on taking transposed conjugates of both sides (for convenience)

\[
\sum_{\ell=1}^{q} \hat{\beta}^{(1)}(\ell) \left( \frac{1}{N} \sum_t \hat{g}^{-2I_y} e^{i(k-\ell)\lambda_t} \right) + \left( \frac{1}{N} \sum_t \hat{g}^{-2I_{yx}} e^{ik\lambda_t} \right) \delta^{(1)}(1) = - \frac{1}{N} \sum_t \hat{g}^{-2I_y} e^{ik\lambda_t},
\]

(4.2.7)

for \( k = 1, \ldots, q \), while from (4.2.6) we have, since \( \hat{\beta}^{(1)}(\omega) = 1 \),

\[
\sum_{\ell=1}^{q} \hat{\beta}^{(1)}(\ell) \left( \frac{1}{N} \sum_t \hat{g}^{-2I_{xy}} e^{-i\ell\lambda_t} \right) + \left( \frac{1}{N} \sum_t \hat{g}^{-2I_x} \right) \delta^{(1)}(1) = - \frac{1}{N} \sum_t \hat{g}^{-2I_{xy}}.
\]

(4.2.8)

From (4.2.7) and (4.2.8) it follows that

\[
\left( \begin{array}{c}
\hat{\beta}^{(1)} \\
\vdots \\
\delta^{(1)}
\end{array} \right) = - \hat{D}^{-1} \hat{A},
\]

(4.2.9)

where

\[
\hat{D} = \begin{bmatrix}
\frac{1}{N} \sum_t \hat{g}^{-2I_y} e^{i(k-\ell)\lambda_t} & \frac{1}{N} \sum_t \hat{g}^{-2I_{yx}} e^{ik\lambda_t} \\
\vdots & \vdots \\
\frac{1}{N} \sum_t \hat{g}^{-2I_{xy}} e^{-i\ell\lambda_t} & \frac{1}{N} \sum_t \hat{g}^{-2I_x}
\end{bmatrix},
\]

and
Here $k,l$ vary from 1 to $q$ so that the $(1,1)$ block of $\hat{D}$ is a $(q \times q)$ matrix, the $(1,2)$ block is a $(q \times r)$ matrix and the $(2,2)$ block is an $(r \times r)$ matrix. Also the $(1,1)$ block of $\hat{d}$ is a $(q \times 1)$ vector while the $(2,1)$ block is a $(r \times 1)$ vector.

If we now let

$$w_y = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} y(n)e^{in\lambda_t}, \quad w_x = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} x(n)e^{in\lambda_t},$$

(where $x'(n) = (x(1)...x(r))$ so that $w_x$ is a vector of $r$ components), then the partial derivative of $H$ with respect to $\alpha(k), k = 1,...,p$ leads to

$$\frac{\partial}{\partial \alpha(k)} \left( \sum_{\ell=0}^{p} \alpha(1)(\ell) \left( \frac{1}{N} \sum_{t} \hat{e}|^{-1} h(1)|w_y + h(1)'|w_x|^2 e^{i(k-\ell)\lambda_t} \right) = 0, \right.$$  

i.e.

$$\frac{\partial}{\partial \alpha(k)} \left( \sum_{\ell=0}^{p} \alpha(1)(\ell) \left( \frac{1}{N} \sum_{t} \hat{e}|^{-1} \hat{z}(1)e^{i(k-\ell)\lambda_t} \right) = 0, \right.$$

where

$$\hat{h}(1) = \sum_{j=0}^{q} \hat{\beta}(1)(k)e^{ik\lambda_t}, \quad \hat{\beta}(1)(0) = 1,$$

and

$$\hat{z}(1) = \left| \hat{h}(1)|w_y + h(1)'|w_x|^2 \right|^2.$$

Furthermore if $\hat{\alpha}(1)$ is the vector with $\hat{\alpha}(1)(\ell)$ in the $\ell$th row, $\ell = 1,...,p$, then since $\hat{\alpha}(1)(0) = 1$, (4.2.10) may be written more concisely in the form
\( \hat{\alpha}(1) = -\hat{A}(1)^{-1}\hat{a}(1) \),

(4.2.11)

where \( \hat{A}(1) \) has \( \hat{a}(1)(k-l) \) in row \( k \) column \( l \), while \( \hat{a}(1) \) has \( \hat{a}(1)(k) \) in row \( k \),

\( k, l = 1, \ldots, p \) and

\[
\hat{a}(1)(k) = \frac{1}{N} \sum_{t} \hat{g}_{t}^{-1} \hat{z}(1)^{k\lambda_{t}}.
\]

(4.2.12)

Now (4.2.9) and (4.2.11) give us the estimates which we obtain by minimizing \( H \). To solve these equations however, as we mentioned earlier, we require initial estimates \( \hat{\beta} \) and \( \hat{\delta} \) of \( \beta \) and \( \delta \) respectively in order to obtain an initial estimate \( \hat{\alpha} \) of \( \alpha \), which in turn is required to form the function \( \hat{g} \) occurring on the right hand side of both (4.2.9) and (4.2.11).

Initial consistent estimates of \( \beta \) and \( \delta \) may be obtained using instrumental variables (see Liviatan [36]). Indeed suppose we let

\[
c_{y}(n) = (N-n)^{-1} \sum_{m=1}^{N-n} (y(m) - \bar{y})(y(m+n) - \bar{y}), \quad n = 0, \ldots, p+q,
\]

\[
c_{jk}(0) = N^{-1} \sum_{m=1}^{N} (x_{j}(m) - \bar{x}_{j})(x_{k}(m) - \bar{x}_{k}), \quad j, k = 1, \ldots, r,
\]

and

\[
c_{yk}(n) = (N-n)^{-1} \sum_{m=1}^{N-n} (y(m) - \bar{y})(x_{k}(m+n) - \bar{x}_{k}), \quad k = 1, \ldots, r; \quad n = 0, \ldots, p+q,
\]

where

\[
\bar{y} = N^{-1} \sum_{m=1}^{N} y(m) \quad , \quad \bar{x}_{k} = N^{-1} \sum_{m=1}^{N} x_{k}(m) \quad , \quad k = 1, \ldots, r.
\]

Then from (4.2.1) it is not hard to see that

\[
\sum_{k=1}^{q} c_{y}(p+j-k)\hat{\delta}(k) + \sum_{k=1}^{r} c_{yk}(p+j)\hat{\delta}(k) = -c_{y}(p+j), \quad j = 1, \ldots, q; \quad \hat{\delta}(0) = 1,
\]

(4.2.13)
and

\[
\sum_{k=1}^{q} c_{y}(k)\hat{\beta}(k) + \sum_{k=1}^{r} c_{j}(o)\hat{\delta}(k) = -c_{y}(o), \quad j = 1, \ldots, r; \quad \hat{\beta}(o) = 1. \tag{4.2.14}
\]

If we now introduce the partitioned row vectors

\[
\rho' = (\beta' ; \delta'), \quad \rho'' = (\hat{\beta}' ; \hat{\delta}'),
\]

then from (4.2.13) and (4.2.14) we may write the equations for the initial estimates \( \hat{\beta}, \hat{\delta} \) as

\[
\hat{\rho} = \hat{\rho}^{-1} \hat{C}, \tag{4.2.15}
\]

where for \( j, k = 1, \ldots, q \), (and \( c_{y}(n), c_{x}(o) \) defined in an obvious way),

\[
\hat{C} = \begin{bmatrix}
    c_{y}(p+j-k) & c_{y}(p+j) \\
    \cdots & \cdots \\
    c_{y}(k) & c_{x}(o)
\end{bmatrix},
\]

\[
\hat{C} = \begin{bmatrix}
    c_{y}(p+j) \\
    \cdots \\
    c_{y}(o)
\end{bmatrix}.
\]

From the results of chapter 4 of Hannan [26] we see that under fairly general conditions the sample covariance converges almost surely to the true covariance. Thus (using a similar argument as for (3.2.3)) from (4.2.1) it follows that \( \hat{C} \) converges almost surely to
\[ C = \begin{bmatrix}
\int_{-\pi}^{\pi} |h|^{-2} e^{i(p+1-k)\lambda} \left( e^{2i \theta} \right) |g|^2 d\lambda d\theta dF(\theta) \delta - \int_{-\pi}^{\pi} |h|^{-1} e^{i(p+1)\lambda} dF(\lambda) \\
\int_{-\pi}^{\pi} e^{-i k \lambda} dF(\lambda) \\
\int_{-\pi}^{\pi} dF(\lambda)
\end{bmatrix}
\]

(4.2.16)

Here \( j, k \) vary from 1 to \( q \) so that the \((1,1)\) block of \( C \) is a \((q \times q)\) matrix.

Since \( F(\lambda) \) is \((r \times r)\) the \((1,2)\) block is a \((q \times r)\) matrix and the \((2,2)\) block an \((r \times r)\) matrix.

In order that \( \hat{p} \) exists, our final assumption for the model (4.1.2) is

(v) \( C \) is non singular.

Having obtained initial estimates \( \hat{\beta}, \hat{\delta} \) of \( \beta \) and \( \delta \) respectively we may now compute an initial estimate \( \hat{\alpha} \) of \( \alpha \). Indeed letting

\[ \hat{z}(n) = \sum_{j=0}^{q} \hat{\beta}(j)y(n-j) + \sum_{k=1}^{r} \hat{\delta}(k)x_k(n), \quad n = 1, \ldots, N; \quad \hat{\beta}(0) = 1, \]

and then forming

\[ \hat{c}_z(n) = (N-n)^{-1} \sum_{m=1}^{N-n} (\hat{z}(m) - \hat{z})(\hat{z}(m+n) - \hat{z}), \quad n = 0, \ldots, p, \]

where

\[ \hat{\frac{z}{z}} = N^{-1} \sum_{m=1}^{N} \hat{z}(m), \]

we can determine

\[ \hat{f}_z = (2\pi)^{-1}(\hat{c}_z(0) + 2 \sum_{n=1}^{P} \hat{c}_z(n) \cos n\lambda_t) \]  

(4.2.17)
and
\[ I^*_z = \hat{w}_z \hat{w}_z^* , \] \hspace{1cm} (4.2.18)

where
\[ \hat{w}_z = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} \hat{z}(n) e^{i\lambda_n t}. \]

From (4.2.11) and (4.2.12) we see that if we multiply both \( \hat{A}^{(1)} \) and \( \hat{a}^{(1)} \) by \( (2\pi /\sigma^2)^2 \), and replace \( (\sigma^2 /2\pi)|\hat{g}|^2 \) by \( \hat{f}_z \), an initial estimate of \( \alpha \) is given by
\[ \hat{\alpha} = -\hat{A}^{-1} \hat{a} \] \hspace{1cm} (4.2.19)

where \( \hat{A} \) has \( \hat{a}(k-l) \) in row \( k \) column \( l \), while \( \hat{a} \) has \( \hat{a}(k) \) in row \( k \); \( k, l = 1, \ldots, p \), and
\[ \hat{a}(k) = \frac{1}{N} \sum_{t} \hat{f}_z \hat{z} e^{ik\lambda_n t}, \]

with \( \hat{f}_z \) and \( I^*_z \) defined by (4.2.17) and (4.2.18) respectively.

The initial estimate \( \hat{\alpha} \) given by (4.2.19) may now be used to form
\[ \hat{g} = \sum_{j=0}^{p} \hat{a}(j) e^{ij\lambda_n t}, \quad \hat{a}(0) = 1, \]

and hence
\[ (\sigma^2 /2\pi) = \frac{1}{N} \sum_{t} |\hat{g}|^{-2} \hat{f}_z. \]

Having obtained \( \hat{g} \), from (4.2.9) we may compute \( \hat{\beta}^{(1)}, \hat{\delta}^{(1)} \) and hence \( \hat{h}^{(1)} \).

Using (4.2.11) it is now possible to compute \( \hat{\alpha}^{(1)} \).

Suppose we now compute the matrices \( \hat{F}, \hat{\Omega} \), which have typical elements \( \hat{F}(k-l), \hat{\Omega}(k-l) \) respectively, where
\[ \hat{F}(k-l) = \frac{1}{N} \sum_{t} |\hat{g}|^{-2} e^{i(k-l)\lambda_n t}, \quad k, l = 1, \ldots, p, \]
\[ \hat{\omega}(k-\ell) = \frac{1}{N} \sum_{t} (\hat{g} \hat{h})^{-1} e^{-i(k-\ell)\lambda t}, \quad k = 1, \ldots, q; \quad \ell = 1, \ldots, p, \]

with
\[ \hat{h} = \sum_{j=0}^{q} \hat{\beta}(j)e^{ij\lambda t}, \quad \hat{\beta}(0) = 1. \]

As we shall prove in the next section, our first efficient estimate of \( \alpha \) may now be computed as
\[ \tilde{\alpha}(1) = \left[ I - (\hat{\alpha}(1)/2\pi) \right]^{-1} \hat{\beta}(1) \hat{\alpha}(1) \hat{\beta}(1). \] (4.2.20)

Having determined \( \tilde{\alpha}(1) \) we may use this estimate to form \( \tilde{\beta}(1), \tilde{\delta}(1) \) which are the same as \( \hat{\beta}, \hat{\delta} \) respectively except that \( \hat{g} \) is replaced by \( \tilde{g}(1) \), where
\[ \tilde{g}(1) = \sum_{j=0}^{P} \tilde{\alpha}(1)(j)e^{ij\lambda t}, \quad \tilde{\alpha}(1)(0) = 1, \]

and hence compute our first efficient estimates of \( \beta, \delta \) as
\[ \tilde{\beta}(1) = \tilde{\beta}(1), \quad \tilde{\alpha}(1) = \tilde{\alpha}(1). \] (4.2.21)

Usually it will be worthwhile carrying out a number of iterations of this procedure, i.e. use \( \tilde{\alpha}(1), \tilde{\beta}(1), \tilde{\delta}(1) \) in place of \( \hat{\alpha}, \hat{\beta}, \hat{\delta} \) respectively to form \( \hat{\alpha}(2) \) and hence \( \tilde{\alpha}(2), \tilde{\beta}(2), \tilde{\delta}(2) \) etc. (We shall examine this more fully when we discuss the computational procedure in section 4.4).

Finally we note that the estimates \( \hat{\alpha}, \hat{\beta}, \hat{\delta} \) of \( \alpha, \beta, \delta \) respectively will not be efficient since they are derived from \( \hat{\alpha}, \hat{\beta} \) which are not efficient. As we shall later show however, on iterating, efficient estimates may be obtained by using \( \tilde{\alpha}(1), \tilde{\beta}(1) \) to form \( \tilde{\beta}(1), \tilde{\alpha}(1) \) (which may then be used to compute \( \tilde{\alpha}(2) \) and hence \( \tilde{\beta}(2), \tilde{\delta}(2) \) etc).
4.3 The Distribution of the Estimates

In order to determine the distribution of the estimates we have obtained in the previous section, suppose we let

\[ \hat{\alpha} = -A^{-1}a, \quad \hat{\rho} = -D^{-1}d, \]

where \( A, a \) are defined in the same way as \( \hat{A}^{(1)}, \hat{a}^{(1)} \) in (4.2.11) but with \( \hat{s}, \hat{h}, \hat{\delta} \) replaced by \( s, h, \delta \) respectively. Furthermore \( D, d \) are defined in the same way as \( \hat{D}, \hat{d} \) in (4.2.9) but with \( \hat{s} \) replaced by \( s \).

\( \hat{\alpha} \) and \( \hat{\rho} \) are of course not computable. We shall in fact derive the estimates \( \tilde{\alpha}^{(1)}, \tilde{\rho}^{(1)} \) defined by (4.2.20) and (4.2.21) respectively and then express these estimates in terms of \( \hat{\alpha}, \hat{\rho} \), whose distribution we are able to obtain.

As we have mentioned earlier, under fairly general conditions the sample covariance converges almost surely to the true covariance. Hannan ([26] p 210) proves that this is indeed true for the sequence \( z(n) \), \( n = 1, \ldots, N \) defined by (4.2.1). Consequently \( \hat{f}_z \) converges almost surely to \( f_z \) and so, since

\[ \hat{f}_z = \left( \sigma^2 / 2\pi \right) |\hat{s}|^2, \]

it follows that \( \hat{s} \) converges almost surely to \( \alpha \). Thus from (4.2.9) it follows that \( \hat{\rho}^{(1)}, \hat{\delta} \) converge almost surely to \( \rho, \Delta \) respectively where, for \( k, \ell = 1, \ldots, q, \)

\[ \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \vdots & \vdots \\ \Delta_{21} & \Delta_{22} \end{bmatrix}, \]

(4.3.1)

with
\[ \Delta_{11} = \left[ \frac{\sigma^2}{2\pi} \right] \frac{1}{2\pi} \int_{-\pi}^{\pi} |h|^{-2} e^{i(k-\ell)\lambda} d\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} h^{-1} e^{i(k-\ell)\lambda} dF(\lambda) \beta, \]

\[ \Delta_{12} = \Delta_{21}^* = -\left[ \frac{1}{2\pi} \right] \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} h^{-1} e^{i(k-\ell)\lambda} dF(\lambda), \]

and

\[ \Delta_{22} = \left[ \frac{1}{2\pi} \right] \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} dF(\lambda). \]

Now

\[ \sqrt{N}(\hat{\rho}(1)-\rho) = -\sqrt{N}[\hat{D}^{-1}\hat{a}-\hat{D}^{-1}\hat{d}] \]
\[ = -\hat{D}^{-1}\sqrt{N}[-(\hat{D}-D)\hat{D}^{-1}\hat{a}+(\hat{a}-\hat{d})] \]
\[ = -\hat{D}^{-1} [\sqrt{N}(\hat{D}-D)\hat{\rho}(1)+\sqrt{N}(\hat{a}-\hat{d})]. \] (4.3.2)

But \( D^{-1} \) converges almost surely to \( \Delta^{-1} \) and, as we shall show below, the term in the square brackets is of the form \( J_N \sqrt{N}(\hat{a}-\alpha) \) say, where \( J_N \) converges almost surely to \( -(\sigma^2/2\pi)(\Omega, \Omega) \) (with \( \Omega \) defined by (4.3.7)), while \( \sqrt{N}(\hat{\omega}-\omega) \) converges in distribution (in fact \( \sqrt{N}(\hat{\omega}-\omega) \) is asymptotically normal (see section 1.2)). Thus on neglecting terms which converge in probability to zero it follows that

\[ \sqrt{N}(\hat{\rho}(1)-\rho) = (\sigma^2/2\pi)\Delta^{-1}(\Omega, \Omega) \sqrt{N}(\hat{\omega}-\omega), \]

where \((\hat{\omega}-\omega)\) is the vector with \((\hat{\omega}(j)-\omega(j))\) in the \( j \)th row, \( j = 1, \ldots, p \).

Indeed if \( \omega' = (1 \cdots p') \), then (since \( \hat{\rho}(1) \) converges almost surely to \( \rho \))
\[ \sqrt{N} (\delta(1) - \bar{\rho}) \]
\[ = -\Delta^{-1} \sqrt{N} [ (\hat{\rho} - \delta) ; (\hat{\delta} - \rho) ] \theta \]
\[ = -\Delta^{-1} \left[ \begin{array}{c}
\frac{1}{N} \sum_t \hat{v}_t \beta(I_y e^{i(k-t)\lambda_t}) \\
\frac{1}{N} \sum_t \hat{v}_t \delta e^{i(k-t)\lambda_t}
\end{array} \right] \theta, \quad k = 1, \ldots, q; \quad \ell = 0, \ldots, q,
\]

where
\[ \hat{v} = \sqrt{N} (|\hat{g}|^{-2} - |g|^{-2}) \]
\[ = \sqrt{N} |\hat{g}|^{-2} |g|^{-2} (|g|^{2} - |\hat{g}|^{2}) \]

and which, on neglecting terms which converge in probability to zero, becomes
\[ -|g|^{-1} \sqrt{N} ((\hat{g} - g) \delta + (\hat{g} - g) e). \quad (4.3.3) \]

Thus after some simplification we obtain
\[ \sqrt{N} (\delta(1) - \bar{\rho}) = \sqrt{N} \Delta^{-1} \]
\[ = \Delta^{-1} \left[ \begin{array}{c}
\frac{1}{N} \sum_t |g|^{-2} g^{1-1} (\hat{g} - g) (I_y \sum_{\ell=0}^{q} \beta(\ell) e^{i(k-t)\lambda_t} + I_{yx} \delta e^{ik \lambda_t}) \\
\frac{1}{N} \sum_t |g|^{-2} g^{1-1} (\hat{g} - g) (I_{xy} \sum_{\ell=0}^{q} \beta(\ell) e^{-i\ell \lambda_t} + I_{x} \delta)
\end{array} \right] \]

\[ + \sqrt{N} \Delta^{-1} \left[ \begin{array}{c}
\frac{1}{N} \sum_t |g|^{-2} g^{1-1} (\hat{g} - g) (I_y \sum_{\ell=0}^{q} \beta(\ell) e^{i(k-t)\lambda_t} + I_{yx} \delta e^{ik \lambda_t}) \\
\frac{1}{N} \sum_t |g|^{-2} g^{1-1} (\hat{g} - g) (I_{xy} \sum_{\ell=0}^{q} \beta(\ell) e^{-i\ell \lambda_t} + I_{x} \delta)
\end{array} \right], \]
where \( k = 1, \ldots, q \). But

\[
I_y \sum_{\ell=0}^q \beta(\ell) e^{i(k-\ell)\lambda_t} + I_y \delta e^{i k \lambda_t} = (hI_y + I_y \delta) e^{i k \lambda_t},
\]

and from (4.2.1) it follows that \((hI_y + I_y \delta)\) converges almost surely to

\[(\sigma^2/2\pi) |g|^2 h^{-1}.
\]

Furthermore

\[
I_{xy} \sum_{\ell=0}^q \beta(\ell) e^{-i \ell \lambda_t} + I_x \delta = hI_{xy} + I_x \delta,
\]

and on neglecting terms which converge almost surely to zero, from (4.2.1) this is seen to be zero. Thus on neglecting terms which converge in probability to zero, it follows that

\[
\sqrt{N}(\hat{\sigma}(1) - \rho) = \sqrt{N} \Delta^{-1} \left[ \begin{array}{c} (\sigma^2/2\pi) \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} g^{-1}(\hat{g} - g) |g|^{2h-1} e^{i k \lambda_t} d\lambda \\ 0 \end{array} \right]
\]

\[+ \sqrt{N} \Delta^{-1} \left[ \begin{array}{c} (\sigma^2/2\pi) \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} g^{-1}(\hat{g} - g) |g|^{2h-1} e^{i k \lambda_t} d\lambda \\ 0 \end{array} \right] \quad (4.3.4)
\]

\[= (\sigma^2/2\pi) \Delta^{-1} \left[ \begin{array}{c} \frac{1}{2\pi} \int_{-\pi}^{\pi} (g \hat{h})^{-1} e^{-i(k+\ell)\lambda} d\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} (g \hat{h})^{-1} e^{-i(k-\ell)\lambda} d\lambda \\ 0 \end{array} \right] \sqrt{N}(\hat{\sigma} - \alpha), \quad (4.3.5)\]
k = 1, \ldots, q; \ l = 1, \ldots, p \ (\text{where for convenience we have taken complex}
conjugates of both sides of (4.3.4)), with \( (\hat{\alpha} - \alpha)' = ((\hat{\alpha}(1) - \alpha(1)) \ldots (\hat{\alpha}(p) - \alpha(p))) \)
and \( \hat{\alpha}(o) = \alpha(o) = 1. \)

Since all the zeros of \( g \) lie outside of the unit circle, we may express \( g^{-1} \) as a series of zero and positive powers of \( e^{i\lambda t} \), with coefficients
which decay exponentially, i.e.
\[
g^{-1} = \sum_{u=0}^{\infty} g_u e^{iu\lambda t}, \quad |g_u| < k^{-u}, \quad \tau > 1,
\]
and on taking complex conjugates of both sides of (4.3.6) we see that \( g^{-1} \)
may be expressed in terms of zero and negative powers of \( e^{i\lambda t} \). A similar
argument holds for both \( h^{-1} \) and \( h^{-1} \). Thus it is not hard to see that
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} (g^{-1}h^{-1}) - i(k\lambda) \lambda d\lambda = 0, \quad k = 1, \ldots, q; \ \ell = 1, \ldots, p.
\]
Furthermore we define \( \Omega \) to be the \((q \times p)\) matrix with \( \omega(k-\ell) \) in row \( k \),
column \( \ell \), \( k = 1, \ldots, q; \ \ell = 1, \ldots, p \), where
\[
\omega(k-\ell) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (g^{-1}h^{-1}) - i(k\lambda) \lambda d\lambda. \tag{4.3.7}
\]
Then, on neglecting terms which converge in probability to zero, from (4.3.5)
it follows that
\[
\sqrt{N}(\hat{\alpha}^{(1)} - \alpha) = (\sigma^2 / 2\pi) \Delta^{-1}(\Omega) \sqrt{N}(\hat{\alpha} - \alpha),
\]
as required. Thus
\[
\sqrt{N}(\hat{\alpha}^{(1)} - \alpha) = \sqrt{N}(\hat{\alpha} - \alpha) + (\sigma^2 / 2\pi) \Delta^{-1}(\Omega) \sqrt{N}(\hat{\alpha} - \alpha). \tag{4.3.8}
\]
Before proceeding to find a similar type of expression for \( \sqrt{N}(\hat{\alpha}^{(1)} - \alpha) \)
in terms of \( \hat{\alpha} \) and \( \hat{\alpha} \) we note that \( \hat{\alpha}^{(1)} \) (defined below (4.2.11)) converges
almost surely to \((\sigma^2 / 2\pi) \Phi\), where \( \Phi \) is the \((p \times p)\) matrix with \( \phi(k-\ell) \) in
row $k$, column $\ell$; $k, \ell = 1, \ldots, p$, and

$$\phi(k-\ell) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} e^{i(k-\ell)\lambda} d\lambda.$$ 

Indeed since $\hat{\rho}^{(1)}$ converges almost surely to $\rho$, $|\hat{\mathbf{v}}^{(1)} Y + \hat{\mathbf{v}}^{(1)} X|^2$ converges almost surely to $(\sigma^2/2\pi)|g|^2$. Thus on neglecting terms which converge almost surely to zero

$$\hat{\mathbf{A}}^{(1)} = \frac{1}{N} \sum_t |g|^{-2} |\hat{\mathbf{V}}^{(1)} Y + \hat{\mathbf{V}}^{(1)} X|^2 e^{i(k-\ell)\lambda}$$

$$= (\frac{\sigma^2}{2\pi}) \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} e^{i(k-\ell)\lambda} d\lambda \right), \quad k, \ell = 1, \ldots, p,$$

$$= (\frac{\sigma^2}{2\pi}) \vec{\phi}. \quad (4.3.9)$$

Similarly $\hat{\mathbf{a}}^{(1)}$ converges almost surely to $(\sigma^2/2\pi)\vec{\phi}$, where $\vec{\phi}$ is the $(p \times 1)$ vector defined by

$$\phi = \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} e^{i k \lambda} d\lambda \right), \quad k = 1, \ldots, p.$$ 

Now

$$\sqrt{N}(\hat{\alpha}^{(1)} - \alpha) = \sqrt{N}[-(\hat{\mathbf{A}}^{(1)} - \mathbf{A})^{-1} \hat{\mathbf{a}}^{(1)} - \mathbf{a}]$$

$$= -\mathbf{A}^{-1} \sqrt{N}[-(\hat{\mathbf{A}}^{(1)} - \mathbf{A}) \hat{\mathbf{A}}^{(1)} \hat{\mathbf{a}}^{(1)} (\hat{\mathbf{a}}^{(1)} - \mathbf{a})]$$

$$= -\mathbf{A}^{-1} \sqrt{N}[-(\hat{\mathbf{A}}^{(1)} - \mathbf{A}) \hat{\mathbf{A}}^{(1)} (\hat{\mathbf{a}}^{(1)} - \mathbf{a})].$$

But $\hat{\alpha}^{(1)}$ converges almost surely to $\alpha$ (since $\hat{\mathbf{A}}^{(1)}$, $\hat{\mathbf{a}}^{(1)}$ converge almost surely to $\frac{\sigma^2}{2\pi} \vec{\phi}$, $(\sigma^2/2\pi)\vec{\phi}$ respectively), and using a similar argument as for $\hat{\mathbf{A}}^{(1)}$ above we can show that $\mathbf{A}$ converges almost surely to $(\sigma^2/2\pi) \vec{\phi}$, so that on neglecting terms which converge in probability to zero we have
\[ \sqrt{N} \left( \hat{\gamma}^{(1)} - \gamma^* \right) = -\frac{2\pi}{\sigma^2} \mathbf{I} \sqrt{N} \left[ (\hat{\mu}^{(1)} - \mu^*) \Delta + (\hat{\sigma}^{(1)} - \sigma^*) \right]. \] (4.3.10)

In order to reduce the right hand side of (4.3.10) into a more suitable form we proceed as follows: if we let

\[ \hat{\gamma} = \sqrt{N} \left| \gamma \right|^{-1} \hat{g}^{(1)} w_y + \hat{g}^{(1)'} w_x \]

By following a similar type of argument to that for (4.3.3), on neglecting terms which converge in probability to zero it can be shown that

\[ \sqrt{N} \left( \left| \gamma \right|^{-1} - \left| \gamma^* \right|^{-1} \right) = -2 \left| \gamma \right|^{-1} \sqrt{N} \left[ \left( \hat{\gamma} - \gamma \right) + \left( \hat{\gamma}^* - \gamma^* \right) \right]. \] (4.3.12)

Furthermore

\[ \sqrt{N} \left( \left| \hat{\gamma}^* \right|^{-1} - \left| \gamma^* \right|^{-1} \right) = -2 \left| \gamma \right|^{-1} \sqrt{N} \left[ \hat{\gamma}^* \left( \hat{\gamma} - \gamma \right) + \left( \hat{\gamma} - \gamma \right) \right]. \] (4.3.13)

But \( \hat{\gamma} \) converges almost surely to \( \gamma \), so that the second term on the right hand side of (4.3.13) becomes

\[ \sqrt{N} \left( \hat{\gamma} - \gamma \right) \begin{pmatrix} \gamma_x & \gamma_y \\ \gamma_y & \gamma_x \end{pmatrix} \begin{pmatrix} \hat{\gamma}^{(1)} - \gamma^* \\ \hat{\gamma}^{(1)'} - \gamma^* \end{pmatrix} \]
\[
= \sqrt{N} (h_{xy} + \delta_{xy} \gamma_{xy} : h_{yx} + \delta_{yx} \gamma_{yx}) \left( \begin{array}{c} \hat{\alpha}(1) \varepsilon \\ \delta(1) \end{array} \right),
\]
and on neglecting terms which converge in probability to zero, this becomes
\[
((\sigma^2/2\pi) |g|^{1/2} : 0) \sqrt{N} \left( \begin{array}{c} \hat{\alpha}(1) \varepsilon \\ \delta(1) \end{array} \right). \tag{4.3.14}
\]

We now have
\[
\sqrt{N} [\alpha^{(1)} - \alpha] \alpha^{(1)} - \alpha]
\]
\[
= \left[ \frac{1}{N} \sum_{k=1}^{p} \alpha(k) \Sigma e^{i(t-k)\lambda_t}(\hat{K}-K) + \frac{1}{N} \sum_{t} \lambda_t(\hat{K}-K) \right]
\]
\[
= \left[ \frac{1}{N} \sum_{t} \Re \lambda_t(\hat{K}-K) \right], \quad t = 1, \ldots, p.
\]
From (4.3.11), (4.3.12) and (4.3.13) the right hand side becomes, on neglecting terms which converge in probability to zero,\(^\text{12}\)
\[
\left[ \frac{1}{N} \sum_{t} \Re \lambda_t \sqrt{N} [\alpha^{(1)} - \alpha] \alpha^{(1)} - \alpha] b'w_{y} \right]^2
\]
\[
+ \delta'w_{y} \delta'w_{y}(\hat{K} - K), \quad t = 1, \ldots, p. \tag{4.3.15}
\]

\(^{12}\) We shall not continue to emphasize the fact that we are neglecting terms which converge in probability to zero. When in fact such terms are being neglected this will be obvious from the discussion.
Now

\[ |b'_{x,v}|^2 = |h|^2 I_y + h^2 I_{xy} + h'I_y + hI_{yx}, \]

and on neglecting terms which converge almost surely to zero, this becomes

\[(\sigma^2/2\pi)|g|^2. \] Thus from (4.3.15) we have

\[
\left[ \frac{1}{N} \sum_t |g|^{-\frac{1}{2}} e^{-\frac{1}{2} \lambda t} \sqrt{N} \left(-2 (g^{-1}(\hat{g}-\hat{g}) + g^{-1}(\frac{1}{g\hat{g}})) |b'_{x,v}|^2 \right) \right] 
\]

\[ = -2(\sigma^2/2\pi) \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} [g^{2/2\pi} \sqrt{N}(g^2-g) + |g|^{1/2\pi} \sqrt{N}(g^2-g)] e^{i\ell \lambda} d\lambda \right], \quad \ell = 1, \ldots, p.
\]

(4.3.16)

But

\[
\left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} g^{2/2\pi} \sqrt{N}(g^2-g)e^{i\ell \lambda} d\lambda \right] 
\]

\[ = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} g^{2/2\pi} e^{i(\ell+k)\lambda} d\lambda \right] \sqrt{N}(\hat{\alpha}-\alpha), \quad k, \ell = 1, \ldots, p.
\]

From (4.3.6) however we see that

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} g^{2/2\pi} e^{i(\ell+k)\lambda} d\lambda = 0 \quad , \quad k, \ell = 1, \ldots, p,
\]

since the exponent under the integral sign will always be greater than zero.

Furthermore,

\[
\left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{2/2\pi} \sqrt{N}(g^2-g)e^{i\ell \lambda} d\lambda \right] 
\]

\[ = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{2/2\pi} e^{i(\ell-k)\lambda} d\lambda \right] \sqrt{N}(\hat{\alpha}-\alpha), \quad k, \ell = 1, \ldots, p,
\]

\[ = \Phi \sqrt{N}(\hat{\alpha}-\alpha).\]
Thus (4.3.16) becomes

\[ \frac{1}{N} \sum_{\gamma} |e|^{-\frac{i}{2} \lambda t} \sqrt{N} \left[ -2 \left( g^{-1} (\gamma g) + \gamma^{-1} (\gamma^{-1} g) \right) b' \nu \nu^* \right]^2 \]

\[ = -2 \left( \frac{\partial^2}{\partial \theta^2} \right) \sqrt{N} (\alpha - \alpha). \]  

(4.3.17)

Referring once more to (4.3.15) it follows from (4.3.14) that

\[ \frac{1}{N} \sum_{\gamma} |e|^{-\frac{i}{2} \lambda t} \sqrt{N} \ b' \nu \nu^* \hat{\gamma} \quad (4.3.17) \]

\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| e \right|^{-\frac{i}{2} \left( \frac{\partial^2}{\partial \theta^2} \right) |g|^{2n-1} : 0} \sqrt{N} \left( \hat{\beta} \quad \ldots \quad \hat{\beta} \right) \quad (4.3.18) \]

where \( \alpha \) is defined by (4.3.7). Also on taking complex conjugates of (4.3.14) it is not hard to see that

\[ \frac{1}{N} \sum_{\gamma} |e|^{-\frac{i}{2} \lambda t} \sqrt{N} \ b' \nu \nu^* \hat{\gamma} \quad (4.3.19) \]

However, from (4.3.6) we see that the exponent under the integral sign will always be positive, and so this integral is zero. Thus

\[ \frac{1}{N} \sum_{\gamma} |e|^{-\frac{i}{2} \lambda t} \sqrt{N} \ b' \nu \nu^* \hat{\gamma} = 0. \]  

(4.3.19)
From equations (4.3.10)-(4.3.19) therefore it follows that
\[
\sqrt{N}(\hat{\alpha}^{(1)} - \alpha) = -(2\pi/\sigma^2) \Phi^{-1}\sqrt{N}[(\hat{\alpha}^{(1)} - \alpha) + \sigma^2/\pi]\]
\[
= -(2\pi/\sigma^2) \Phi^{-1}[\sqrt{N}(\hat{\alpha} - \alpha) + (\sigma^2/2\pi)(\alpha' : 0) \sqrt{N}(\hat{\beta}^{(1)} - \rho)]
\]
\[
= 2 \sqrt{N}(\hat{\alpha} - \alpha) - \Phi^{-1}(\alpha' : 0) \sqrt{N}(\hat{\beta}^{(1)} - \rho),
\]
so that, on neglecting terms which converge in probability to zero, we have
\[
\sqrt{N}(\hat{\alpha}^{(1)} - \alpha) = \sqrt{N}(\hat{\alpha} - \alpha) + 2\sqrt{N}(\hat{\beta}^{(1)} - \rho).
\] (4.3.20)

From (4.3.8) and (4.3.20) it follows that
\[
\sqrt{N}(\hat{\alpha}^{(1)} - \alpha)
\]
\[
= -[I-(\sigma^2/2\pi) \Phi^{-1} \Phi^{-1} \Delta^{-1}(\alpha' : 0)] \sqrt{N}(\hat{\alpha} - \alpha) - \Phi^{-1}(\alpha' : 0) \sqrt{N}(\hat{\beta} - \rho),
\] (4.3.21)
where \(\hat{\alpha}^{(1)}\) is a weighted sum of \(\hat{\alpha}\) and \(\hat{\alpha}^{(1)}\). Indeed from (4.3.8) and (4.3.20) we have
\[
\sqrt{N}(\hat{\alpha}^{(1)} - \alpha) + 2\sqrt{N}(\hat{\beta}^{(1)} - \rho).
\]

We may rewrite the left hand side as
\[
\sqrt{N}(\hat{\alpha}^{(1)} - \alpha) + \sqrt{N}(\hat{\alpha} - \alpha) = (I+H) \sqrt{N}[(-H)^{-1}(\hat{\alpha}^{(1)} + (I+H)^{-1}H \hat{\beta} - \alpha)]
\]
\[
= (I+H) \sqrt{N}(\hat{\alpha}^{(1)} - \alpha), \quad \text{say.}
\]
Thus \(\sqrt{N}(\hat{\alpha}^{(1)} - \alpha)\)
\[
= -[I-(\sigma^2/2\pi) \Phi^{-1} \Phi^{-1} \Delta^{-1}(\alpha' : 0)] \sqrt{N}(\hat{\alpha} - \alpha) - \Phi^{-1}(\alpha' : 0) \sqrt{N}(\hat{\beta} - \rho),
\]
as required. Furthermore
\[ \tilde{\alpha}^{(1)} = (I+H)^{-1}\hat{\alpha}^{(1)} + (I+H)^{-1}H\hat{\omega}, \] (4.3.22)

so that if \( C = (\sigma^2/2\pi) \Phi^{-1}(\Omega' : 0) \Delta^{-1}(\Omega) \), then \( H = -2I+C \), and

\[ \tilde{\alpha}^{(1)} = -[I-C]^{-1}\hat{\alpha}^{(1)} + [I-C]^{-1}[2I-C]\hat{\omega}. \]

But \( [I-C]^{-1}[2I-C] = [I-C]^{-1}[I+(I-C)] \)

\[ = [I-C]^{-1}I. \]

Hence \( \tilde{\alpha}^{(1)} = [I-C]^{-1}(\hat{\alpha}^{(1)} + \hat{\omega}) \),

i.e. \( \tilde{\alpha}^{(1)} = [I-(\sigma^2/2\pi) \Phi^{-1}(\Omega' : 0) \Delta^{-1}(\Omega)]^{-1}(\hat{\alpha}^{(1)} + \hat{\omega}) \),

which we see to be of the same form as (4.2.20). From (4.3.22) it follows that since \( \hat{\alpha} \) and \( \hat{\alpha}^{(1)} \) both converge almost surely to \( \alpha \), then \( \tilde{\alpha}^{(1)} \) converges almost surely to \( \alpha \).

Referring to equations (4.3.8) and (4.3.21), it is not hard to see that

\[ \sqrt{N}(\tilde{\rho}^{(1)} - \rho) = \sqrt{N}(\hat{\rho} - \rho) + (\sigma^2/2\pi) \Delta^{-1}(\Omega) \sqrt{N}(\tilde{\alpha}^{(1)} - \alpha) \]

\[ = \sqrt{N}(\hat{\rho} - \rho) - (\sigma^2/2\pi) \Delta^{-1}(\Omega)[I-(\sigma^2/2\pi) \Phi^{-1}(\Omega' : 0) \Delta^{-1}(\Omega)]^{-1} \]

\[ (\sqrt{N}(\hat{\alpha}-\alpha) - \Phi^{-1}(\Omega' : 0) \sqrt{N}(\hat{\rho} - \rho)). \] (4.3.23)

If we now let

\[ E = (\sigma^2/2\pi) \Delta^{-1}(\Omega) \quad \text{and} \quad F = \Phi^{-1}(\Omega' : 0) \]

then

\[ \sqrt{N}(\tilde{\rho}^{(1)} - \rho) = \sqrt{N}(\hat{\rho} - \rho) - E(I-FE)^{-1}(\sqrt{N}(\hat{\alpha}-\alpha) - F \sqrt{N}(\hat{\rho} - \rho)) \]

\[ = (I+E(I-FE)^{-1}F) \sqrt{N}(\hat{\rho} - \rho) - E(I-FE)^{-1} \sqrt{N}(\hat{\alpha}-\alpha). \] (4.3.24)
But
\[ E(I_{-F})^{-1} = (E^{-1} - F)^{-1} = (I_{-E})^{-1}, \]
and so
\[ I + E(I_{-E})^{-1}F = I + (I_{-E})^{-1}EF \]
\[ = (I_{-EF})^{-1} = (I_{-EF})^{-1}EF = (I_{-EF})^{-1}. \]
Consequently (4.3.24) becomes
\[ \sqrt{N}(\rho)_{(1)} - \rho = (I_{-EF})^{-1} \sqrt{N}(\rho - \rho) - (I_{-EF})^{-1}E \sqrt{N}(\alpha - \alpha) \]
\[ = (I_{-EF})^{-1} \left( \sqrt{N}(\rho - \rho) - E \sqrt{N}(\alpha - \alpha) \right), \]
i.e.
\[ \sqrt{N}(\rho)_{(1)} - \rho \]
\[ = \left[ I - (\sigma^2 / 2\pi) A^{-1} \right]^{-1} \left( \tau(\rho - \rho) - (\sigma^2 / 2\pi) A^{-1} \right) \]
\[ \sqrt{N}(\rho - \rho), \sqrt{N}(\alpha - \alpha), \sqrt{N}(\alpha - \alpha), \sqrt{N}(\rho - \rho) \]
\[ \text{so that if we now determine the} \]
\[ \text{asymptotic distribution of} \sqrt{N}(\alpha - \alpha), \sqrt{N}(\rho - \rho) \text{we will effectively have the} \]
\[ \text{asymptotic distribution of} \sqrt{N}(\alpha(1) - \alpha), \sqrt{N}(\rho(1) - \rho). \]

Since \( A \) converges almost surely to \((\sigma^2 / 2\pi) I\), we may replace
\[ \sqrt{N}(\alpha - \alpha) = - \sqrt{N} A^{-1}(\alpha - \alpha) \]
by \(- I^{-1} u\). \( u \) has typical element given by
\[ u(k) = (2\pi / \sigma^2) N^{-\frac{1}{2}} \sum_{t, l=0}^{P} \alpha(l) |g|^{-l} \sum_{x, y} \sum_{\omega} \sum_{\omega}^{N} e^{i(k - l)t} \]
\[ \equiv (2\pi / \sigma^2) N^{-\frac{1}{2}} \sum_{t} g^{-1} \sum_{e}^{i(kt)} \]
since \( h_{y, \chi} \psi_{x, \chi} \cong gw_{e} \), and \( |gw_{e}|^{2} = |g|^{2} r_{e} \),

with 
\[
I_{e} = (2\pi k)^{-1} \sum_{m,n=1}^{N} \epsilon(m)\epsilon(n) e^{-i(n-m)\lambda_{t}}.
\]

From (4.3.6) we see that we may let 
\[
g^{-1} = \sum_{j=0}^{\infty} g_{j} e^{ij\lambda_{t}}, \quad |g_{j}| < K \tau^{-|j|}, \quad \tau > 1,
\]
so that

\[
u(k) \cong (N^{-1/2}/\sigma^{2}) \sum_{m,n=1}^{N} \epsilon(m)\epsilon(n) g_{j} \frac{1}{N} \sum_{t=1}^{N} \epsilon_{n-m-k+\ell N}.
\]

But

\[
\frac{1}{N} \sum_{t=1}^{N} \epsilon_{n-m-k+\ell N} = 0, \quad j \not\equiv (n-m-k) \mod N
\]

\[
= 1, \quad j \equiv (n-m-k) \mod N,
\]

(i.e. \( j = n-m-k+\ell N; \ell = 0, 1, 2, \ldots \)). Thus

\[
u(k) \cong (N^{-1/2}/\sigma^{2}) \sum_{m,n=1}^{N} \epsilon(m)\epsilon(n) \sum_{\ell} g_{n-m-k+\ell N},
\]

where the sum is over all \( \ell \), subject to \( g_{\ell} = 0, a < 0 \). Since \( g_{\ell} \) converges exponentially to zero, for each fixed \( n,m,k \) and \( \ell \not\equiv 0, N^{-1/2} \sum_{\ell} g_{n-m-k+\ell N} \) obviously converges to zero. Thus we are left to consider

\[
u(k) \cong (N^{-1/2}/\sigma^{2}) \sum_{m,n=1}^{N} \epsilon(m)\epsilon(n) g_{n-m-k}
\]

\[
= (N^{-1/2}/\sigma^{2}) \sum_{n=1}^{N} \epsilon(n) \sum_{m=1}^{N} g_{n-m-k} \epsilon(m), \quad \text{(4.3.27)}
\]

since \( g_{a} = 0, a < 0 \), so that for \( k = 1, \ldots, p \),

\[
(N^{-1/2}/\sigma^{2}) \sum_{n=1}^{N} \epsilon(n) \sum_{m=N+1}^{N} g_{n-m-k} \epsilon(m) = 0.
\]
Thus we may write \((4.3.27)\) in the form

\[
\begin{align*}
\mathbf{u}(k) &= (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} \varepsilon(n) \sum_{j=0}^{\infty} g_j \varepsilon(n-j-k), \\
&= (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} \mathbf{e}(n) \mathbf{g}(n-j-k),
\end{align*}
\]

where the added terms will not (asymptotically) affect our results (see Hannan [26] p 412 for a related discussion). Hence, to our order of approximation, it follows that

\[
\mathbf{u}(k) = (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} \varepsilon(n) \mathbf{v}(n-k),
\]

where

\[
\mathbf{v}(n-k) = \sum_{j=0}^{\infty} g_j \varepsilon(n-j-k).
\]

This implies that \(\mathbf{v}(n)\) is generated by the autoregressive relation

\[
\sum_{j=0}^{\infty} \alpha(j) \mathbf{v}(n-j) = \varepsilon(n).
\]

Thus we may replace \(\sqrt{N}(\hat{\alpha} - \alpha)\) by \(\mathbf{P}^{-1}\mathbf{u}\), where \(\mathbf{u}\) has (approximately) typical element

\[
\mathbf{u}(k) = (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} \varepsilon(n) \mathbf{v}(n-k), \quad k = 1, \ldots, p, \quad (4.3.28)
\]

and

\[
\sum_{j=0}^{\infty} \alpha(j) \mathbf{v}(n-j) = \varepsilon(n).
\]

For \(\mathbf{u}(k)\) of the form \((4.3.28)\) we see (from the proof of Theorem 1 Chapter 6 of Hannan [26] p 330) that \(\mathbf{u}\) is, asymptotically, normally distributed and hence it follows that \(\sqrt{N}(\hat{\alpha} - \alpha)\) is asymptotically normally distributed.
In a similar manner the distribution of $\sqrt{N}(\hat{\rho} - \rho)$ is obtained from

$$\sqrt{N}(\hat{\rho} - \rho) = -\sqrt{N}(D^{-1}d_{+\rho})$$

$$= -\sqrt{N}D^{-1}(d_{+\rho}),$$

and on neglecting terms which converge in probability to zero we have

$$\sqrt{N}(\hat{\rho} - \rho) = -\sqrt{N} \Delta^{-1}[d : D](\hat{\rho})$$

$$= -\sqrt{N} \Delta^{-1} \left[ \begin{array}{ccc}
\frac{1}{N} \sum_{t} |g|^{-2} I_{y} e^{-i(k-\ell)\lambda_{t}} & \cdots & \frac{1}{N} \sum_{t} |g|^{-2} I_{y} e^{i\lambda_{t}} \\
\cdots & \cdots & \cdots \\
\frac{1}{N} \sum_{t} |g|^{-2} I_{x} e^{-i\lambda_{t}} & \cdots & \frac{1}{N} \sum_{t} |g|^{-2} I_{x}
\end{array} \right] \left( \begin{array}{c}
1 \\
\vdots \\
\beta \\
\vdots \\
\delta
\end{array} \right)$$

$$k = 1, \ldots, q; \; \ell = 0, \ldots, q$$

$$= -\sqrt{N} \Delta^{-1} \left[ \begin{array}{c}
\frac{1}{N} \sum_{t} |g|^{-2} e^{i\lambda_{t}} (H_{y} + I_{y} \delta) \\
\cdots \cdots \\
\frac{1}{N} \sum_{t} |g|^{-2} (I_{x} e^{i\lambda_{t}} + I_{x})
\end{array} \right]$$

$$\approx -\sqrt{N} \Delta^{-1} \left[ \begin{array}{c}
\frac{1}{N} \sum_{t} |g|^{-1} I_{y} e^{i\lambda_{t}} \\
\cdots \cdots \\
\frac{1}{N} \sum_{t} |g|^{-1} I_{x} e^{i\lambda_{t}}
\end{array} \right], \; k = 1, \ldots, q,

(4.3.29)$$
since from (4.2.1) it is not hard to show that
\[ h_{y+x} \delta \equiv i_{y+g} \quad \text{and} \quad i_{y+x} h + \delta \equiv i_{x+g}, \]
where \( i_{y+g} \) is the cross periodogram between the \( y(n) \) and the \( e(n) \) series, with \( i_{x+g} \) defined in an obvious way. Thus from (4.3.29) we see that
\[ \sqrt{N}(\delta - \rho) = -[\frac{1}{2\pi N^2}]^{-1} \text{h}\ldots \]
\[ \text{(4.3.30)} \]
say, where \( \nu \) has typical element
\[ \nu(k) \equiv (2\pi N^2)^{-\frac{1}{2}} \sum_{t} \xi_{t} e^{ik\lambda_{t}}, \]
\[ = (N^{-\frac{1}{2}}\sigma^2)^{2\pi} \sum_{t} \xi_{t} e^{ik\lambda_{t}}, \quad k = 1, \ldots, q. \]
But from (4.3.6) (and the discussion below that equation) it follows that
we may let
\[ h_{g-1} = \sum_{u=0}^{\infty} f_{u} e^{iu\lambda_{t}}, \quad h_{-1} = \sum_{s=0}^{\infty} h_{s} e^{is\lambda_{t}}, \]
where \( f_{u} \) and \( h_{s} \) decrease exponentially, and
\[ i_{y+g} = (2\pi N)^{-1} \sum_{m,n=1}^{N} \nu(m) e^{i(m-n)\lambda_{t}} \]
so that
\[ \nu(k) \equiv (N^{-\frac{1}{2}}\sigma^2)^{N} \sum_{m,n=1}^{N} \nu(m) e^{i(m-n)\lambda_{t}}. \]
Now
\[ \sum_{t=1}^{N} \frac{1}{N} e^{i(k+u+s+m-n)\lambda_{t}} = 0, \quad s \neq (n-m-u-k) \mod N \]
\[ = 1, \quad s = (n-m-u-k) \mod N, \]
(i.e. \( s = n-m-u-k+tN, \quad t = 0, 1, \ldots \)), and so
\[ v(k) = (N^{-1/2}/\sigma^2) \sum_{m,n=1}^{N} \sum_{u=0}^{\infty} \Sigma f_u y(m)e(n)h_{n-m-u-k}\ell N, \]

where the sum is over all \( \ell \) for which the summand is defined. Following an identical argument to that for the \( u(k) \), we may (asymptotically) neglect all values of \( \ell \) except \( \ell = 0 \), so that we are left with

\[ v(k) = (N^{-1/2}/\sigma^2) \sum_{m,n=1}^{N} \sum_{u=0}^{\infty} f_u y(m)e(n)h_{n-m-u-k} \]

\[ = (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} e(n) \sum_{u=0}^{\infty} \sum_{m=1}^{\infty} f_u y(m)h_{n-m-u-k} \]

where the added terms do not change the value of \( v(k) \). If we now let \( j = (n-m-u-k) \), then (repeating the same argument as for \( u(k) \)) to our order of approximation, it follows that

\[ v(k) = (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} e(n) \sum_{j=0}^{\infty} \sum_{u=0}^{\infty} f_u y(n-u-j-k)h_j. \]

But

\[ \sum_{u=0}^{\infty} f_u y(n-u-j-k) = e(n-j-k), \]

so that

\[ v(k) = (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} e(n) \sum_{j=0}^{\infty} h_j e(n-j-k) \]

\[ = (N^{-1/2}/\sigma^2) \sum_{n=1}^{N} e(n) \tau(n-k), \quad k = 1, \ldots, q, \quad \text{(4.3.31)} \]

where

\[ \tau(n-k) = \sum_{j=0}^{\infty} h_j e(n-j-k). \]
This means that $\tau(m)$ is generated by the autoregressive relation

$$
\sum_{j=0}^{q} \beta(j)\tau(m-j) = \epsilon(m). \tag{4.3.32}
$$

For $v(k)$, $k = 1, \ldots, q$ in the form (4.3.31) it can be seen (from the proof of Theorem 1 Chapter 6 of Hannan [26] p 330) that $v$ is, asymptotically, normally distributed.

Considering (4.3.30) once more we have

$$
w = \left(\frac{2\pi/\sigma^2}{\sigma^2}\right) N^{-\frac{1}{2}} \Sigma g^{-1} t x e. \tag{4.3.33}
$$

Now from the lemma in section 3 of Hannan [28] it immediately follows that in this form $w$ is asymptotically normally distributed. Thus from (4.3.29) it can be seen that $\sqrt{N}(\hat{p}-p)$ is asymptotically normally distributed. As a result, since (as we have already seen) $\sqrt{N}(\hat{\alpha}-\alpha)$ is asymptotically normally distributed, from (4.3.21) and (4.3.25) it follows that $\sqrt{N}(\hat{\alpha}^{(1)}-\alpha)$ and $\sqrt{N}(\hat{\rho}^{(1)}-\rho)$ are asymptotically normally distributed.

All that now remains to be done is to determine the asymptotic covariance matrix of $\sqrt{N}(\hat{\alpha}^{(1)}-\alpha)$, $\sqrt{N}(\hat{\rho}^{(1)}-\rho)$. The method we employ is just an extension of that used by Hannan ([26] p 413) to obtain the asymptotic covariance of the efficient estimates of the coefficients of the mixed autoregressive moving average model. Indeed from (4.3.28), (4.3.31) and (4.3.33) we may determine the asymptotic covariance of $u$, $v$ and $w$ and hence that of $\sqrt{N}(\hat{\alpha}-\alpha)$ and $\sqrt{N}(\hat{\rho}-\rho)$. Having obtained the asymptotic covariance matrix of $\sqrt{N}(\hat{\alpha}-\alpha)$ and $\sqrt{N}(\hat{\rho}-\rho)$, it is then a straightforward matter to determine the asymptotic covariance matrix of $\sqrt{N}(\hat{\alpha}^{(1)}-\alpha)$ and $\sqrt{N}(\hat{\rho}^{(1)}-\rho)$. After some tedious manipulation it can be shown that, asymptotically,
where \( \Phi \) is defined above (4.3.9), while \( \Omega \) and \( \Delta \) are defined by (4.3.7) and (4.3.1) respectively.

We may verify that these estimates are efficient by showing that the covariance matrix (4.3.34) is the same as the asymptotic covariance matrix obtained from the matrix of second derivatives of the logarithm of the likelihood function. Indeed if \( \tilde{\alpha}^{(1)}, \tilde{\beta}^{(1)} \) and \( \tilde{\delta}^{(1)} \) are asymptotically normal and efficient then (following Walker [54] p 365),

\[
\text{Cov}[\sqrt{N}(\tilde{\alpha}^{(1)}-\alpha), \sqrt{N}(\tilde{\beta}^{(1)}-\beta), \sqrt{N}(\tilde{\delta}^{(1)}-\delta)] = \lim_{N \to \infty} N \left[ \begin{array}{ccc} \Phi & -\Omega' & 0 \\ -\Omega & (2\pi/\sigma^2)\Delta & 0 \\ 0 & 0 & 0 \end{array} \right],
\]

where \( \Phi \) takes the values \( \alpha, \beta \) and \( \delta \) respectively, and for the model (4.2.1) we are considering

\[
\log L = \frac{1}{2} \sum_{t}^{T} z_{t}^{-1}(\lambda_{t})I_{x}(\lambda_{t}).
\]

On evaluation (4.3.35) is found to be the same as (4.3.34), as required.

We now summarize the results of this chapter in the following:

**Theorem 4.3.1**

Under conditions (i)-(v) the vector \( (\tilde{\alpha}^{(1)'} : \tilde{\beta}^{(1)'} : \tilde{\delta}^{(1)'} ) \) converges almost surely to \( (\alpha' : \beta' : \delta' \) and \( \sqrt{N}[\tilde{\alpha}^{(1)}-\alpha)' : (\tilde{\beta}^{(1)}-\beta)' \) has a distribution converging to the multivariate normal distribution with zero mean and covariance matrix defined by (4.3.34). The estimate is asymptotically
efficient. The estimates \( \hat{\Phi}^{(1)} \), \( \hat{\Omega}^{(1)} \) and \( \hat{\Delta}^{(1)} \) converge almost surely to \( \Phi, \Omega \) and \( \Delta \) respectively.

Finally in this section we make the following comments with regard to this theorem:

(i) As we have mentioned earlier, the procedure will usually be iterated, say \((j+1)\) times, so that \( \hat{\Phi}^{(j)} \), \( \hat{\Omega}^{(j)} \) and \( \hat{\Delta}^{(j)} \) will be available and these will be asymptotically efficient.

(ii) It seems plausible that we could have generalized the condition (iii) regarding the \( x_j(n) \) so as to include trending terms which must increase more slowly than exponentially, as well as those already specified. In fact all the results of this chapter will probably go through under the more general specifications, usually referred to as Grenander's conditions (for a full discussion of these conditions and their implications we refer to either Hannan [26] p 77 or Terrell [51] pp 19-21):

\[
\lim_{N \to \infty} d_j^2(N) = \lim_{N \to \infty} N \sum_{n=1}^N x_j(n)^2 = \infty, \quad j = 1, \ldots, r,
\]

\[
\lim_{N \to \infty} \left[ \frac{x_j^2(N)}{d_j^2(N)} \right] = 0, \quad j = 1, \ldots, r,
\]

\[
\lim_{N \to \infty} \left[ \frac{\sum_{m=1}^N x_j(m)x_k(m+n)}{d_j(N)d_k(N)} \right] = \rho_{jk}(n), \quad j, k = 1, \ldots, r,
\]

where by this last condition we mean that this limit exists with probability 1. As these authors point out, the second condition is included to exclude cases of exponential rates of increase (which need special consideration) and
furthermore it ensures that \( p_{jk}(n) \) is free from end effects, so that, for example

\[
\lim_{N \to \infty} \frac{\Sigma x_j(m)x_k(m+n)}{d_j(N)d_k(N)} = p_{jk}(n).
\]

We have not emphasized these more general conditions here, for by doing so the complexity of the proof of the asymptotic results would be increased.

(iii) This final comment relates to the size of \( N \). For \( N < 500 \) the calculations are well within the capacity of modern computing equipment. However as \( N \) increases, the labour involved in computing the Fourier transforms \( w(\lambda_t) \) becomes dominant. If \( N \) is highly composite then much larger values may easily be handled by adopting the Cooley-Tukey technique for the calculation of finite Fourier transforms (see Hannan [26] p 263). If really large values of \( N \) are to be used it may be worthwhile replacing the periodograms by smoothed estimates of the corresponding spectra. Provided the smoothing is not too radical there is no doubt that Theorem 4.3.1 will continue to hold.

4.4 The Computational Procedure

We shall, in this section, outline the computational procedure (along similar lines to that described in Hannan and Nicholls [29]) associated with the method of estimation developed in this chapter. Furthermore the application of this procedure to generated data shall be discussed and the numerical results presented in the next section. For the model

\[
\Sigma \beta(k)y(n-k) + \Sigma \delta(k)x_k(n) = \Sigma \alpha(j)e(n-j) = z(n),
\]

\( k=0 \)

\( k=1 \)

\( j=0 \)
where \( \alpha(0) = \beta(0) = 1 \), we wish to compute efficient estimates of the \( \alpha(j), \beta(k), \delta(m), j = 1, \ldots, p; k = 1, \ldots, q; m = 1, \ldots, r \).

For samples of size \( N \) we first form

\[
\bar{x}_k = N^{-1} \sum_{n=1}^{N} x_k(n), \quad k = 1, \ldots, r,
\]

\[
\bar{y} = N^{-1} \sum_{n=1}^{N} y(n),
\]

and use these to compute the sample covariances

\[
c_{yk}(m) = (N-m)^{-1} \sum_{n=1}^{N-m} (y(n) - \bar{y})(y(m+n) - \bar{y}), \quad m = 0, \ldots, p+q,
\]

\[
c_{jk}(o) = N^{-1} \sum_{n=1}^{N} (x_j(n) - \bar{x}_j)(x_k(n) - \bar{x}_k), \quad j, k = 1, \ldots, r,
\]

\[
c_{jy}(m) = (N-m)^{-1} \sum_{n=1}^{N-m} (y(n) - \bar{y})(x_j(m+n) - \bar{x}_j), \quad m = 0, \ldots, p+q; j = 1, \ldots, r.
\]

We also form

\[
w_y(\lambda_t) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} y(n) e^{i n \lambda_t}, \quad \lambda_t = \lambda_{N-t} = 2\pi t/N, \quad t = 1, \ldots, [N/2],
\]

\[
w_j(\lambda_t) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} x_j(n) e^{i n \lambda_t}, \quad j = 1, \ldots, r.
\]

---

13 As we pointed out in footnote 11, unless otherwise stated, \( \lambda_t \) is always of this form so that we shall not continually refer to it as such.
It is now possible to form the periodograms

\[ I_{y}(\lambda_t) = w_y(\lambda_t)w_y^*(\lambda_t), \quad I_{jk}(\lambda_t) = w_j(\lambda_t)w_k^*(\lambda_t), \]

and the cross periodograms

\[ I_{yj}(\lambda_t) = w_y(\lambda_t)w_j^*(\lambda_t), \quad I_{yj}(\lambda_t) = w_j(\lambda_t)w_y^*(\lambda_t), \]

where \( j,k = 1,\ldots,r \). Having computed these statistics we are now in a position to carry out step 1, which is not repeated when we iterate the estimation procedure.

**Step 1**

(i) Initial estimates of \( \beta \) and \( \delta \).

Solve

\[
\sum_{k=1}^{q} c_{y}(p+j-k)\hat{\beta}(k) + \sum_{k=1}^{r} c_{yk}(p+j)\hat{\delta}(k) = -c_{y}(p+j), \quad j = 1,\ldots,q,
\]

\[
\sum_{k=1}^{q} c_{yj}(k)\hat{\beta}(k) + \sum_{k=1}^{r} c_{jk}(o)\hat{\delta}(k) = -c_{yj}(o), \quad j = 1,\ldots,r.
\]

The solutions of these equations give \( \hat{\beta}(k), \hat{\delta}(j), k = 1,\ldots,q; j = 1,\ldots,r \).

(ii) Initial estimate of \( \alpha \).

Having estimated \( \hat{\beta}' = (\hat{\beta}(1)\ldots\hat{\beta}(q)), \hat{\delta}' = (\hat{\delta}(1)\ldots\hat{\delta}(r)) \), we may now use these to form

\[
\hat{z}(n) = \sum_{j=0}^{q} \hat{\beta}(j)y(n-j) + \sum_{k=1}^{r} \hat{\delta}(k)x_k(n), \quad \hat{\beta}(o) = 1, \quad n = 1,\ldots,N,
\]

and hence

\[
\hat{c}_z(m) = (N-m)^{-1} \sum_{n=1}^{N-m} (\hat{z}(n)-\hat{z}) (\hat{z}(m+n)-\hat{z}), \quad m = 0,\ldots,p,
\]

where
\[ \hat{z} = N^{-1} \sum_{n=1}^{N} \hat{z}(n). \]

(In practice when \( N \) is fixed we will of course only be able to compute \( \hat{z}(n) \) for \( n = q+1, \ldots, N \), so that obvious minor changes will have to be made to the formulae presented. For the Monte Carlo experiments we consider in the next section however we are able to obtain \( \hat{z}(n) \), \( n = 1, \ldots, N \), by generating \( (N+q) \) random variables, from which the \( \hat{z}(n) \) are constructed).

We may now compute
\[ \hat{r}_z(\lambda_t) = (2\pi)^{-1} (\hat{c}_z(o) + 2 \sum_{n=1}^{p} \hat{c}_z(n) \cos n\lambda_t) \]
and
\[ \hat{I}_z(\lambda_t) = \hat{w}_z(\lambda_t) \hat{w}_z^*(\lambda_t), \]
where
\[ \hat{w}_z(\lambda_t) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} \hat{z}(n) e^{in\lambda_t}. \]

Forming
\[ \hat{a}(k) = N^{-1} \sum_{t} \hat{s}_{z}^{-2}(\lambda_t) \hat{I}_z(\lambda_t) e^{ik\lambda_t}, \quad k = 0, \ldots, p, \]
we may then set up the \((p \times p)\) matrix
\[ \hat{A} = [\hat{a}(k-\ell)] \quad , \quad k, \ell = 1, \ldots, p, \]
and the \((p \times 1)\) vector
\[ \hat{a} = [\hat{a}(k)] \quad , \quad k = 1, \ldots, p. \]

If \( \hat{\alpha} = (\hat{\alpha}(1) \ldots \hat{\alpha}(p)) \), our initial estimate of \( \alpha \) (where \( \alpha' = (\alpha(1) \ldots \alpha(p)) \)) is then defined as
\[ \hat{\alpha} = -\hat{A}^{-1} \hat{a}. \]
This completes step 1.
Step 2

Form

$$\hat{g}(\lambda_t) = \sum_{k=0}^{p} \hat{\gamma}(k)e^{ik\lambda_t}, \quad \hat{h}(\lambda_t) = \sum_{k=0}^{q} \hat{\beta}(k)e^{ik\lambda_t},$$

$$\hat{\alpha}(0) = \hat{\beta}(0) = 1, \text{ and}$$

$$\left( \hat{g}^2/2\pi \right) = N^{-1} \sum_{t} \hat{g}(\lambda_t)|^{-2} \hat{f}_{\hat{z}}(\lambda_t).$$

We now compute

$$\begin{pmatrix} \hat{\beta}(1) \\ \vdots \\ \hat{\beta}(1) \end{pmatrix} = -\hat{D}^{-1} \hat{a},$$

where

$$\hat{D} = \begin{bmatrix} \hat{E} & \hat{B} \\ \hat{C} & \hat{D} \\ \hat{B}^* & \hat{C}^* \end{bmatrix}, \quad \hat{a} = \begin{bmatrix} \hat{G} \\ \hat{R} \end{bmatrix},$$

and, for $k, l = 1, \ldots, q,$

$$\hat{E} = N^{-1} \sum_{t} \hat{g}(\lambda_t)|^{-2} I_{xy}(\lambda_t)e^{ik-l\lambda_t}, \quad \hat{B} = N^{-1} \sum_{t} \hat{g}(\lambda_t)|^{-2} I_{yx}e^{ik\lambda_t},$$

$$\hat{C} = N^{-1} \sum_{t} \hat{g}(\lambda_t)|^{-2} I_{x}(\lambda_t), \quad \hat{G} = N^{-1} \sum_{t} \hat{g}(\lambda_t)|^{-2} I_{y}e^{ik\lambda_t},$$

$$\hat{H} = N^{-1} \sum_{t} \hat{g}(\lambda_t)|^{-2} I_{xy}(\lambda_t).$$

$I_{x}(\lambda_t), I_{yx}(\lambda_t)$ and $I_{xy}(\lambda_t)$ have typical elements $I_{jk}(\lambda_t), I_{yj}(\lambda_t)$ and $I_{jy}(\lambda_t)$ respectively, $j, k = 1, \ldots, r$. We now use $\hat{\beta}(1), \hat{\beta}(1)$ to compute

$$\hat{z}(1)(n) = \sum_{j=0}^{q} \hat{\beta}(1)(j)y(n-j) + \sum_{k=1}^{r} \hat{\delta}(1)(k)x_k(n), \quad \hat{\beta}(1)(0) = 1,$$

$$n = 1, \ldots, N,$$
and hence $1 \hat{Z}(1)(\lambda_t)$ in an obvious way. By doing this we are now in a position to form

$$\hat{a}^{(1)}(k) = N^{-1} \sum_t (\sigma^2/2\pi)|\hat{g}(\lambda_t)|^2 \hat{Z}^{(1)}(\lambda_t)e^{ik\lambda_t}, \quad k = 0, \ldots, p.$$ 

(It is convenient, from a computational viewpoint, to introduce the factor $(\sigma^2/2\pi)^2$ into the denominator of $\hat{a}^{(1)}(k)$. This factor will be accounted for when we form the efficient estimate $\tilde{\alpha}^{(1)}$ defined by (4.4.2)). Forming the $(p \times p)$ matrix

$$A^{(1)} = [\hat{a}^{(1)}(k-\ell)]$$

and the $(p \times 1)$ vector

$$\hat{a}^{(1)} = [\hat{a}^{(1)}(k)],$$

$k, \ell = 1, \ldots, p$, we then compute

$$\hat{\alpha}^{(1)} = -A^{(1)-1}A^{(1)}.$$

Next we compute $\hat{\omega}$ which has $\hat{\omega}(k-\ell)$ in row $k$ column $\ell$; $k = 1, \ldots, p$;

$$\ell = 1, \ldots, p,$$

where

$$\hat{\omega}(k-\ell) = N^{-1} \sum_t (\hat{g}(\lambda_t)\hat{\lambda}_t)^{-1}e^{-i(k-\ell)\lambda_t}.$$

It is now possible to compute

$$\tilde{\alpha}^{(1)} = [\hat{\alpha}^{(1)}]^{(1)}D^{-1}\hat{\alpha}^{(1)} + \hat{\alpha}^{(1)};$$

where $D^{(1)}$ is the $(1,1)$ block of $D^{-1}$. Now use $\tilde{\alpha}^{(1)}$ to form

$$\tilde{g}^{(1)}(\lambda_t) = \sum_{k=0}^p \tilde{\alpha}^{(1)}(k)e^{ik\lambda_t}, \quad \tilde{\alpha}^{(1)}(0) = 1,$$

and hence $\tilde{D}^{(1)}$, $\tilde{d}^{(1)}$ by replacing $\hat{g}(\lambda_t)$ by $\tilde{g}^{(1)}(\lambda_t)$ in $\hat{D}$, $\hat{d}$ respectively.

Then compute
\[
\begin{bmatrix}
\tilde{\beta}^{(1)} \\
\vdots \\
\tilde{\delta}^{(1)}
\end{bmatrix} = -\tilde{\delta}^{(1)-1}d^{(1)}. 
\quad (4.4.3)
\]

Finally we require an efficient estimate of \((\sigma^2/2\pi)\) for the next iteration. To obtain this estimate we use \(\tilde{\beta}^{(1)}, \tilde{\delta}^{(1)}\) to form
\[
\tilde{z}^{(1)}(n) = \sum_{j=0}^{q} \tilde{\beta}^{(1)}(j)y(n-j) + \sum_{k=1}^{r} \tilde{\delta}^{(1)}(k)x_k(n), \quad \tilde{\beta}^{(1)}(0) = 1, \quad n = 1, \ldots, N.
\]

Then compute \(\hat{c}_z^{(1)}(m), \ m = 0, \ldots, p\) and hence \(\hat{f}_z^{(1)}(\lambda_t)\) (in the same way as we formed \(\hat{c}_z(m)\) and hence \(\hat{f}_z(\lambda_t)\) from the \(\hat{z}(n), n = 1, \ldots, N\) and use these to form
\[
(\sigma^{(1)}^2/2\pi) = N^{-1} \sum_t |\tilde{\delta}^{(1)}(\lambda_t)|^{-2}\hat{f}_z^{(1)}(\lambda_t),
\]

which replaces the estimate (4.4.1) at the start of step 2.

Equations (4.4.2) and (4.4.3) give our first efficient estimates of \(\alpha, \beta\) and \(\delta\). We now iterate the procedure by using \(\tilde{\alpha}^{(1)}, \tilde{\beta}^{(1)}\) and \(\tilde{\delta}^{(1)}\) in place of \(\hat{\alpha}, \hat{\beta}^{(1)}\) and \(\hat{\delta}^{(1)}\) respectively, and repeating step 2 to obtain \(\tilde{\alpha}^{(2)}, \tilde{\beta}^{(2)}\) and \(\tilde{\delta}^{(2)}\). This procedure is usually iterated a number of times so that in general we have, for \(k \geq 1\),
\[
\tilde{\alpha}^{(k+1)} = [I - \tilde{\alpha}^{(k)}]^{-1} \tilde{\alpha}^{(k)} D^{(k)} \tilde{\alpha}^{(k)}]^{-1} (\tilde{\alpha}^{(k)} \alpha^{(k+1)} + \tilde{\alpha}^{(k)}).
\]

4.5 Numerical Results

The computational procedure we have outlined was applied to data generated by the relation
\[
y(n) + \beta y(n-1) + 8x(n) = \epsilon(n) + \alpha \epsilon(n-1), \quad (4.5.1)
\]
where $\alpha = 0.5$, $\beta = -0.8$, $\delta = 0.3$. The $x(n)$ sequence was generated by

$$x(n) - 0.6x(n-1) = \eta(n). \quad (4.5.2)$$

Both the $y(n)$ and the $x(n)$ sequences were (very nearly) stationary, while the $\epsilon(n)$ and $\eta(n)$ sequences were taken to be independent of each other and each consisted of independent random variables with zero mean and uniform distribution. The ratio $\theta = \sigma^2_\eta / \sigma^2_\epsilon$ (where $\sigma^2_\eta$ and $\sigma^2_\epsilon$ represent the variances of the $\eta(n)$ and the $\epsilon(n)$ sequences respectively), which determines the relative importance of the $\epsilon(n)$ and $\eta(n)$ sequences in the generation of $y(n)$, was taken as either 1 or 9.

Using the above values of $\alpha$, $\beta$ and $\delta$ we are able to evaluate the theoretical (asymptotic) covariance matrix directly, in each case, from (4.3.34). Indeed for the model (4.5.1) above (for which $k, \ell = 1$) we have (on dropping the argument $\lambda$ for convenience),

$$\Phi = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| g \right|^2 d\lambda = (1-\alpha^2)^{-1}$$

and

$$\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\overline{\eta} \overline{\epsilon})^{-1} d\lambda = (1-\alpha\beta)^{-1}.$$ 

But $\alpha = 0.5$ and $\beta = -0.8$, so that in this case

$$\Phi = 1.3333 \quad , \quad \Omega = \Omega' = 0.7143.$$ 

A direct evaluation of $(2\pi / \sigma^2_\epsilon) \Delta$ is somewhat more laborious than for $\Phi$ or $\Omega$. Indeed
\[
\Delta = \begin{bmatrix}
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} f_y \, d\lambda \\
\cdots \\
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} f_{xy} e^{i\lambda} \, d\lambda \\
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} f_x \, d\lambda
\end{bmatrix}
\]

But \( f_y = (\sigma^2_e / 2\pi) |g|^2 |h|^{-2} |\delta|^2 f_x \delta \), so that
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} f_y \, d\lambda = (\sigma^2_e / 2\pi) \frac{1}{2\pi} \int_{-\pi}^{\pi} |h|^{-2} d\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} |h|^{-2} \delta f_x \delta \, d\lambda,
\]
and on evaluation the right hand side becomes
\[
2.7778(\sigma^2_e / 2\pi) + 0.5164(\sigma^2_\eta / 2\pi).
\]
Furthermore it is not hard to show that
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} f_x \, d\lambda = 1.1218(\sigma^2_\eta / 2\pi),
\]
while
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} f_{xy} e^{i\lambda} \, d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} |g|^{-2} f_{xy} e^{-i\lambda} \, d\lambda
\]
\[
= -0.1770(\sigma^2_\eta / 2\pi).
\]
Thus, for the model (4.5.1), the theoretical (asymptotic) covariance matrix of the efficient estimates of \( \alpha \), \( \beta \) and \( \delta \) is given by
\[
\text{Cov}(\sqrt{N}(\bar{\alpha} - \alpha), \sqrt{N}(\bar{\beta} - \beta), \sqrt{N}(\bar{\delta} - \delta)) \approx \begin{bmatrix}
1.3333 & -0.7143 & 0 \\
-0.7143 & 2.7778 + 0.5164\theta & -0.1770 \\
0 & -0.1770 & 1.1218\theta
\end{bmatrix}^{-1}
\]
(4.5.3)
where of course $\vartheta = (\sigma^2 / \sigma^2_\varepsilon)$. For particular values of $\vartheta$ it is a straightforward matter to evaluate this covariance matrix.

A number of sampling experiments were carried out for various values of $\vartheta$ and $N$, and for each $\vartheta$, $N$ a total of 20 replications of the experiment was made. In Table 4.1 we summarize the results of these sampling experiments. Indeed in this Table for each $\vartheta$, $N$ row (a) gives the mean of the observed values of the relevant parameters for the 20 runs, row (b) shows the theoretical variance obtained from (4.5.3) and row (c) shows the observed variance of the 20 numbers.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\vartheta$</th>
<th>$\tilde{\alpha}$</th>
<th>$\tilde{\beta}$</th>
<th>$\tilde{\varsigma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>0.5217</td>
<td>-0.7537</td>
<td>0.3117</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>0.0085</td>
<td>0.0035</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.0109</td>
<td>0.0030</td>
<td>0.0097</td>
</tr>
<tr>
<td>100</td>
<td>9</td>
<td>0.5105</td>
<td>-0.7706</td>
<td>0.3072</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>0.0079</td>
<td>0.0015</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.0108</td>
<td>0.0014</td>
<td>0.0010</td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>0.4979</td>
<td>-0.7650</td>
<td>0.2970</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>0.0042</td>
<td>0.0017</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.0032</td>
<td>0.0021</td>
<td>0.0051</td>
</tr>
<tr>
<td>200</td>
<td>9</td>
<td>0.4920</td>
<td>-0.7750</td>
<td>0.3014</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>0.0040</td>
<td>0.0007</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.0038</td>
<td>0.0012</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

The iterations were continued for each run until successive estimates of all parameters differed by not more than 0.005. For $\vartheta = 1$, $N = 100$, in 10 cases the estimates were stable after the third iteration and in 18 after the sixth. The two exceptional cases had both stabilized by the ninth iteration. For $\vartheta = 1$, $N = 200$, in 15 cases the estimates were stable after the third iteration, in 19 after the fifth and all were stable after 6 iterations. (The total running time for the program of 20 series for the
case $N = 100$, including generation of the data was 14.66 minutes on an
IBM 360/50).

From the results we have obtained there is no doubt that $\tilde{\beta}$ is biased
towards zero. Walker [53], for the case where $\delta = 0$ and $N = 100$, from an
asymptotic expansion of the mean of $\tilde{\beta}$ evaluates this bias as 0.042, which
is close to the bias observed above.

The same model was also estimated using $N = 40$ for $\theta = 9$. The results,
for 20 replications, are summarized in Table 4.2.

Table 4.2

<table>
<thead>
<tr>
<th></th>
<th>$\tilde{\alpha}$</th>
<th>$\tilde{\beta}$</th>
<th>$\tilde{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>0.5249</td>
<td>-0.6707</td>
<td>0.3249</td>
</tr>
<tr>
<td>(b)</td>
<td>0.0198</td>
<td>0.0037</td>
<td>0.0026</td>
</tr>
<tr>
<td>(c)</td>
<td>0.0411</td>
<td>0.0411</td>
<td>0.0061</td>
</tr>
</tbody>
</table>

In this experiment the ratio of the theoretical to the observed variance is
significant in all cases at the 1% level of significance ($F_{19,\infty} = 1.91$ at
the 1% level), so that the agreement with the asymptotic theory is not good
for such a small value of $N$.

Summarizing the results of this section we see that the numerical
results we have obtained help to illustrate, for $N$ sufficiently large, the
asymptotic theory we have developed in this chapter.

4.6 Conclusion

By extending the estimation procedure developed by Hannan [25] we have,
in this chapter, been able to estimate the coefficients of mixed autoregressive
moving average models with exogenous variables. The estimates we have
obtained have been shown to be asymptotically normally distributed and efficient. In section 4.5 these theoretical results have been applied (along the lines of the computational procedure described in section 4.4) to estimate the parameters of a particular model (of the form (4.2.1)) using generated data. As we have seen from the numerical results of that section, provided $N$ is sufficiently large, the estimation procedure works well.

In the next chapter, as a particular example of the model (4.2.1), we shall consider the estimation of certain classes of distributed lag models. Furthermore proceeding along rather heuristic lines we shall, in chapter 6, show that the asymptotic results of the present chapter may be extended to the vector mixed autoregressive moving average model with exogenous variables.
Chapter 5
Distributed Lag Models

5.1 Introduction

A large literature has been developed since the 1950's on distributed lag models in the field of Econometrics. Attention has, until recent years, been centred primarily on the geometric distributed lag model, first used systematically by Koyck [35]. Hannan [23] has obtained estimates of, and examined the properties of, the parameters of this model. He obtains these estimates by means of a transformation of the data from the time to the frequency domain, and then applying maximum likelihood methods to this transformed data. Jorgenson [32] has suggested that any arbitrary distributed lag model may be approximated to any desired degree of accuracy by a rational distributed lag model, while Dhrymes [10] has applied Hannan's method to the more general rational distributed lag model. As an application of the theory developed in the previous chapter, we shall here relate both the geometric and the rational distributed lag models to the mixed autoregressive moving average model with exogenous variables (i.e. the model (4.2.1)). It will be found that by placing restrictions on (4.2.1) we are able to obtain efficient estimates of the parameters of these models.

The geometric distributed lag model is usually expressed in the form

\[ y(n) = \sum_{j=0}^{\infty} \lambda^j x(n-j) + \epsilon(n), \quad |\lambda| < 1, \]

(5.1.1)

where the \( x(n-j) \) are the exogenous variables. It is not hard to see that

\[ \lambda y(n-1) = \sum_{j=0}^{\infty} \lambda^{j+1} x(n-j-1) + \lambda \epsilon(n-1), \]
i.e. \( \lambda y(n-1) = \delta \sum_{k=1}^{\infty} \lambda^k x(n-k) + \lambda e(n-1), \)

so that on subtracting this last expression from (5.1.1) we have

\[ y(n) - \lambda y(n-1) = 8x(n) + e(n) - \lambda e(n-1). \]  (5.1.2)

In the form (5.1.2) we see the geometric distributed lag model to be just a special case of (4.2.1). Indeed if \( p = q = r = 1, \alpha(1) = \beta(1) = -\lambda, \)
\( \delta(1) = -\delta \) and the exogenous variable is lagged (i.e. \( x_k(n) = x(n-k+1), \)
\( k = 1 \)), then (4.2.1) reduces to (5.1.2).

The rational distributed lag model may be expressed in the form

\[ y(n) + \frac{C(L)}{B(L)} x(n) = e(n), \]  (5.1.3)

where

\[ B(L) = \sum_{j=0}^{q} \beta(j)L^j, \quad C(L) = \sum_{k=1}^{r} \delta(k)L^k \]

(with \( L \) the lag operator, i.e. \( L^m z(n) = z(n-m) \)), and the \( x(n) \) are the exogenous variables. Referring once more to (4.2.1), if \( p = q, \alpha = \beta \) and the exogenous variables are lagged, then (4.2.1) reduces to (5.1.3). Comparing (5.1.2) and (5.1.3) it is obvious that the geometric distributed lag model is a special case of the rational distributed lag model. Consequently in what follows we shall concentrate on the more general rational distributed lag model. Before proceeding to the efficient estimation of the parameters of these distributed lag models however we shall derive a test of the equality of the coefficients of the autoregressive and moving average parts of (4.2.1).
5.2 A Test for the Equality of the Parameters

If $\alpha(j)$, $\beta(k)$ are the jth and kth elements of the vectors $\alpha$, $\beta$ respectively, $j = 1, \ldots, p$; $k = 1, \ldots, q$, and $p = q$ in the model (4.2.1), we want to set up a test of the null hypothesis that $\alpha = \beta$ against the alternative that $\alpha \neq \beta$. From the previous chapter we know that, after $j$ iterations of the estimation procedure, $\sqrt{n}(\hat{\alpha}(j) - \alpha)$, $\sqrt{n}(\hat{\beta}(j) - \beta)$ and $\sqrt{n}(\hat{\delta}(j) - \delta)$ are asymptotically normally distributed with covariance matrix

$$
\begin{bmatrix}
\Phi & -\Omega' \\
\Omega & (2\pi/\sigma^2)\Delta
\end{bmatrix}^{-1}
$$

(5.2.1)

$\Delta$ and $\Omega$ are defined by (4.3.1) and (4.3.7) respectively while $\Phi$ is defined above (4.3.9).

When the null hypothesis is true we have $g = h$ so that $\Omega' = \Phi = \Omega$, while

$$(2\pi/\sigma^2)\Delta = \begin{bmatrix}
\Phi + \theta & \psi \\
\psi^* & \xi
\end{bmatrix}$$

say, where now for $k, l = 1, \ldots, q$,

$$\Phi = \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} |h|^{-2} e^{i(k-l)\lambda} d\lambda\right],$$

$$\theta = (2\pi/\sigma^2) \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} |h|^{-3} e^{i(k-l)\lambda} \hat{\delta} dF(\lambda) \hat{\delta}\right],$$
\[
\psi = -(2\pi/\sigma^2)[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left|h \right|^{-2} e^{i k \lambda} d\lambda],
\]
and
\[
\zeta = (2\pi/\sigma^2)[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left|h \right|^{-2} d\lambda].
\]

Thus when \( \alpha = \beta \), (5.2.1) becomes

\[
\begin{bmatrix}
\bar{\Phi} & -\bar{\Phi} & \cdots & 0 \\
-\bar{\Phi} & \bar{\Phi}+\theta & \cdots & \psi \\
0 & \cdots & -\bar{\Phi} & \bar{\Phi}
\end{bmatrix}^{-1}
\]

Furthermore, from (1.1.9) it follows that the (1,1) block of (5.2.2) is

\[
\begin{bmatrix}
\bar{\Phi} & -\bar{\Phi} \\
-\bar{\Phi} & \bar{\Phi}+\theta
\end{bmatrix} - \begin{bmatrix} 0 & \psi \\ \psi & 0 \end{bmatrix}^{-1}
\]

say, and from (1.1.8) it is not hard to see that this becomes

\[
\begin{bmatrix}
-\bar{\Phi}^{-1+\Lambda} & -\Lambda^{-1} \\
\cdots & \cdots \\
\Lambda^{-1} & -\Lambda^{-1}
\end{bmatrix}
\]

where \( \Lambda = (\theta-\psi_\zeta^{-1}\psi^*) \). (5.2.3) is just the asymptotic covariance matrix of \( \sqrt{N}(\alpha(j)-\alpha) \) and \( \sqrt{N}(\beta(j)-\beta) \) when \( \alpha = \beta \).

If we let \( y = \sqrt{N}(\alpha(j)-\beta(j)) \) then, under the null hypothesis, it is not hard to see that \( \text{Cov}(y) \equiv -\bar{\Phi}^{-1} \), and so
will be asymptotically distributed as chi square with q degrees of freedom.

We have thus proved the following:

**Theorem 5.2.1**

For the model (4.2.1) assume that \( p = q \). To test the null hypothesis that \( \alpha = \beta \) against the alternative that \( \alpha \neq \beta \), let \( \tilde{\alpha}(j) \) and \( \tilde{\beta}(j) \) be the efficient estimates of \( \alpha \) and \( \beta \) respectively after \( j \) iterations. Then the test statistic is, under the null hypothesis, given by (5.2.4) and is asymptotically distributed as chi square with \( q \) degrees of freedom.

### 5.3 Estimation and Properties of the Coefficients

Suppose we let \( p = q \) and \( \alpha = \beta \) in (4.2.1) so that this model becomes

\[
\sum_{j=0}^{q} \beta(j)y(n-j) + \sum_{k=1}^{r} \delta(k)x_k(n) = \sum_{j=0}^{q} \beta(j)e(n-j), \quad \beta(0) = 1. \tag{5.3.1}
\]

To obtain efficient estimates of the coefficients of (5.3.1) we require that \( \delta \neq 0 \) (where \( \delta \) is the vector of \( r \) components with \( \delta(j) \) in the \( j \)th place) for this model to be identified (i.e. so that condition (iv) on the model (1.2.1) is satisfied).

Let \( \tilde{\alpha}(j) \) and \( \tilde{\beta}(j) \) represent the efficient estimates of \( \alpha \), \( \beta \) respectively (after \( j \) iterations) for the model (4.2.1). Then considering the model (5.3.1) the best estimate of \( \beta \) is obtained as a linear combination of \( \tilde{\alpha}(j) \) and \( \tilde{\beta}(j) \), say \( A_1\tilde{\alpha}(j) + A_2\tilde{\beta}(j) \), \( A_1 + A_2 = 1 \). If \( U = \sqrt{n}(A_1\tilde{\alpha}(j) + A_2\tilde{\beta}(j)) \) then, under the null hypothesis that \( \alpha = \beta \), from (5.2.3) it is not hard to show that
\begin{align*}
\text{Cov}(u) & \equiv A_1 (\Phi^{-1} + \Lambda^{-1}) \Lambda_1^\prime + A_1^\prime (I - A_1')(I - A_1) \Lambda^{-1} \Lambda_1^\prime + (I - A_1') A_1 \Lambda^{-1} (I - A_1') \\
& = A_1 (\Phi^{-1} + \Lambda^{-1}) = \Gamma,
\end{align*}

say. Consequently the best linear combination of \( \hat{\alpha}(j) \) and \( \hat{\beta}(j) \) (in the sense that for all \((q \times 1)\) vectors \(a \neq 0\), \(a' \Gamma a\) is a minimum) is that for which \(A_1 = 0\), i.e. \(\hat{\beta}(j)\) is the best estimate of \(\beta\). Thus for the model \((5.3.1)\) it follows, from section 4.2, that the coefficients \(\beta, \delta\) are (asymptotically) efficiently estimated (after \((j+1)\) iterations of the estimation procedure, \(j \geq 0\)) by

\begin{equation}
\begin{aligned}
\hat{\beta}(j + 1) &= \tilde{w}(j)^{-1} \tilde{w}(j), \\
\tilde{w}(j) &= \begin{bmatrix}
\frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{i(k-\ell)\lambda_t} & \cdots & \cdots & \frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{ik\lambda_t} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\ell\lambda_t} & \cdots & \cdots & \frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\lambda_t} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\lambda_t} & \cdots & \cdots & \frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\lambda_t}
\end{bmatrix},
\end{aligned}
\end{equation}

where, for \(k, \ell = 1, \ldots, q\),

\begin{align*}
\tilde{w}(j) &= \begin{bmatrix}
\frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{i(k-\ell)\lambda_t} & \cdots & \cdots & \frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{ik\lambda_t} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\ell\lambda_t} & \cdots & \cdots & \frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\lambda_t} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\lambda_t} & \cdots & \cdots & \frac{1}{N} \sum_{t} [\tilde{h}(j)]^{-1} \delta(j)' \tilde{\delta}(j) e^{-i\lambda_t}
\end{bmatrix},
\end{align*}

Using an identical argument to that for \(\hat{\beta}, \hat{\delta}\) in section 4.3 it can be shown that \(\tilde{w}(j), \tilde{w}(j)\) converge almost surely to
\[ W = (\sigma^2/2\pi) \begin{bmatrix} \theta & \psi \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\
\psi^* & \zeta \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} \tau \\ \cdot \end{bmatrix}, \quad (5.3.3) \]

respectively, where \( \theta, \psi, \zeta \) are defined above (5.2.2),

\[ \tau = \frac{1}{2\pi} \int_{-\pi}^{\pi} |h|^{-\frac{1}{2}} e^{ik\lambda} \delta \, dF(\lambda) \delta, \quad k = 1, \ldots, q, \]

and

\[ \nu = -\frac{1}{2\pi} \int_{-\pi}^{\pi} |h|^{-2} \, dF(\lambda). \]

Comparing (5.3.3) and (4.3.1) it follows (when \( \alpha = \beta \)) that

\[ \bar{W} = \begin{bmatrix} \Delta^{-1} & \bar{\Phi} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\
\bar{\Phi} & \cdot \end{bmatrix}, \quad (5.3.4) \]

where now of course \( g = h \) in the definitions of \( \Delta \) and \( \bar{\Phi} \). Thus from a computational point of view it is a straightforward matter to compute the efficient estimate \( \tilde{\rho}(j+1) \) of \( \rho \) (after \( j+1 \) iterations of the estimation procedure, \( j \geq 0 \)) as

\[ \tilde{\rho}(j+1) = \begin{bmatrix} \tilde{\beta}(j) - (\tilde{\sigma}(j)^2/2\pi)^2 & \tilde{\alpha}(j) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\alpha}(j) \\ \cdot \end{bmatrix}, \quad (5.3.5) \]

where \( \tilde{\beta}(j), \tilde{\alpha}(j), \tilde{\alpha}(j) \) and \( (\tilde{\sigma}(j)^2/2\pi) \) are defined in section 4.4 except that now \( g = h \) in these definitions.

In order to determine the asymptotic covariance matrix of \( \tilde{\rho}(j+1) \) we consider (5.2.2), which may be partitioned as
From (1.1.8) it follows that the (2,2) block of this matrix is given by

\[
\begin{bmatrix}
\Phi & \Phi+\theta & 0 \\
0 & \Phi & \psi \\
-\Phi & \psi^* & \xi
\end{bmatrix}^{-1}
\]

and referring to (5.3.3) it follows that

\[
\text{Cov}\left( \frac{n}{n} (p^+)^2 - p \right) = \frac{\sigma^2}{2\pi} W_1.
\]

Hence for computational purposes the covariance matrix of \( \sqrt{N}(\tilde{\beta}^{(j+1)} - \rho) \) is estimated by

\[
(\sigma^{(j+1)} / 2\pi) \left[ \tilde{D}^{(j+1)} - (\sigma^{(j+1)} / 2\pi)^2 \right]^{-1},
\]

with \( \tilde{D}^{(j+1)} \), \( \tilde{A}^{(j+1)} \) and \( (\sigma^{(j+1)} / 2\pi) \) defined in an obvious way. We summarize these results in the following:

**Theorem 5.3.1**

For the model (5.3.1) the vector \( \beta^{(j+1)}' = (\beta^{(j+1)} : \tilde{\beta}^{(j+1)}') \) converges almost surely to \( \rho' = (\beta' : \beta') \), and \( \sqrt{N}(\tilde{\beta}^{(j+1)} - \rho) \) is asymptotically normally distributed with zero mean and covariance matrix \( (\sigma^2 / 2\pi) W^{-1} \) (where \( W \) is defined by (5.3.3)). The \( \tilde{\beta}^{(j+1)} \), \( j \geq 0 \), are (asymptotically) efficient and the covariance matrix of \( \sqrt{N}(\tilde{\beta}^{(j+1)} - \rho) \) is efficiently estimated by (5.3.7).
As a particular case of (5.3.1) suppose we now lag the exogenous variables i.e. let \( x_k(n) = x(n-k+l) \), \( k = 1, \ldots, r \). This model then becomes

\[
\sum_{j=0}^{q} \beta(j)y(n-j) + \sum_{k=1}^{r} \delta(k)x(n-k+l) = \sum_{j=0}^{q} \beta(j)e(n-j), \quad \beta(0) = 1, \tag{5.3.8}
\]

which, from (5.1.3), we see to be the rational distributed lag model. As well as the restriction on the zeros of \( h(z) \), in order that this model be identified, we require that all the zeros of the \( z \) transform

\[
m(z) = \sum_{k=1}^{r} \delta(k)z^{k-1}
\]

lie outside of the unit circle and \( h(z), m(z) \) have no zeros in common. Furthermore either \( \beta(q) \) or \( \delta(r) \) must be non-null. (For a related discussion see Hannan [26] pp 388-389). In this case Theorem 5.3.1 continues to hold (with minor alterations brought about by the lagging of the exogenous variables). Indeed for this model efficient estimates of the coefficients are, after \((j+1)\) iterations, given by

\[
p(j+1) = \tilde{g}(j)^{-1}s(j),
\]

where if

\[
m(\lambda_1) = \sum_{k=1}^{r} \delta(k)e^{(k-1)\lambda_1},
\]

then \( \tilde{g}(j), \tilde{s}(j) \) can be shown to converge almost surely to \( S, s \) respectively where for \( k, l = 1, \ldots, q; \ u, v = 1, \ldots, r, \)
Thus by considering (5.3.8) as a special case of (4.2.1) we are able to obtain efficient estimates of the coefficients of the rational distributed lag model (and hence the geometric lag model which is, of course, a particular case of (5.3.8)).

5.4 Summary

Considering the various distributed lag models in turn we begin with the geometric distributed lag model

\[ y(n) = 8 \sum_{j=0}^{\infty} \lambda^j x(n-j) + u(n), \quad |\lambda| < 1, \]

(5.4.1)

where \( u(n) \) is a disturbance independent of \( x(m) \), for all \( m, n \). As we have already mentioned Hannan [23] has estimated the coefficients (and considered the properties of the estimates) of this model by applying maximum likelihood procedures to the Fourier transformed data. In obtaining these estimates he assumes that \( u(n) \) is a linear process satisfying
where the \( \epsilon(n) \) are identically and independently distributed. In (5.4.1) if the \( u(n) \) are assumed to be identically and independently distributed then the geometric distributed lag model is a special case of the rational distributed lag model

\[
y(n) + \frac{C(L)}{B(L)} x(n) = \epsilon(n),
\]

where \( B(L) = \sum \beta(j)L^j \), \( C(L) = \sum \delta(j)L^j \), and the \( \epsilon(n) \) are identically and independently distributed. By placing restrictions on the coefficients of the mixed autoregressive moving average model with exogenous variables (i.e. the model (4.2.1)), as we have seen in the previous section, we are able to efficiently estimate the coefficients of models of the form (5.4.2).

A class of rational distributed lag models which is of a more general form than those obtained by placing constraints on (4.2.1) (and for which (5.4.1) is a special case) may be expressed in the form

\[
y(n) + \frac{C(L)}{B(L)} x(n) = u(n),
\]

where \( B(L) \), \( C(L) \) are as for (5.4.2) while the \( u(n) \) are now assumed to be correlated. Dhrymes [10] has considered the estimation of such models assuming that the \( x(n) \) and \( u(n) \) processes are mutually independent covariance stationary stochastic processes each possessing a spectral density. Indeed this author obtains estimates of the parameters of such models using the spectral techniques first suggested by Hannan [23] in the case of the geometric distributed lag model (5.4.1).
Pierce [48] has also considered models of the form (5.4.3) for the case where \( u(n) \) may be represented by a stationary linear process, i.e.

\[
 u(n) = \sum_{j=0}^{\infty} \alpha(j)e(n-j), \quad \sum_{j=0}^{\infty} |\alpha(j)| < \infty, \tag{5.4.4}
\]

and the \( e(n) \) are identically and independently distributed. Using Jorgenson's result that any linear lag function may be approximated by a rational one, Pierce approximates (5.4.4) by a rational lag function i.e.

\[
 u(n) = \frac{A(L)}{D(L)} e(n),
\]

with \( A(L) = \sum a(j)L^j \) and \( D(L) = \sum d(j)L^j \). In this case (5.4.3) may be expressed in the form

\[
 y(n) + \frac{C(L)}{B(L)} x(n) = \frac{A(L)}{D(L)} e(n), \tag{5.4.5}
\]

where the \( x(n) \) are the exogenous variables and the \( e(n) \) are identically and independently distributed. For the model (5.4.5) Pierce applies least squares procedures to estimate the coefficients of the lag function and the residuals.

In practice of course the more general model may not necessarily be the best model to fit to a particular set of data. Indeed as we pointed out in section 4.1, because of the increased computations required to obtain efficient estimates of the more general models, their use may not be justified.
CHAPTER 6
EXTENSIONS TO THE VECTOR MODEL

6.1 The Vector Model

In this final chapter we shall outline, somewhat heuristically, the extension of the estimation procedure of chapter 4 to the vector autoregressive moving average model with exogenous variables. This vector model may be represented in the form:

$$\sum_{k=0}^{q} \left( \begin{array}{c}
\beta_{11}(k) \cdots \beta_{1v}(k) \\
\cdots \cdots \\
\beta_{v1}(k) \cdots \beta_{vv}(k)
\end{array} \right) \left( \begin{array}{c}
y_1(n-k) \\
\cdots \\
y_v(n-k)
\end{array} \right) + \left( \begin{array}{c}
\delta_{11} \cdots \delta_{1u} \\
\cdots \\
\delta_{v1} \cdots \delta_{vu}
\end{array} \right) \left( \begin{array}{c}
x_1(n) \\
\cdots \\
x_u(n)
\end{array} \right)$$

$$= \sum_{j=0}^{p} \left( \begin{array}{c}
\alpha_{11}(j) \cdots \alpha_{1v}(j) \\
\cdots \cdots \\
\alpha_{v1}(j) \cdots \alpha_{vv}(j)
\end{array} \right) \left( \begin{array}{c}
\varepsilon_1(n-j) \\
\cdots \\
\varepsilon_v(n-j)
\end{array} \right) = \left( \begin{array}{c}
z_1(n) \\
\cdots \\
z_v(n)
\end{array} \right)$$

i.e.

$$\sum_{k=0}^{q} B(k)y(n-k) = \sum_{j=0}^{p} A(j)e(n-j) = z(n), \quad (6.1.1)$$

$$A(o) = B(o) = I_v,$$ where $I_v$ is the $(v \times v)$ unit matrix. $y(n)$, $e(n)$ and $z(n)$ are vectors of $v$ components while $x(n)$ is a vector of $u$ components.

The conditions which we require to be satisfied in order that we may efficiently estimate the coefficients of the model (6.1.1) are simply vector extensions of those which we required for the model (4.2.1). Indeed for the model (6.1.1) we assume:
(i) The $\epsilon(n)$ are i.i.d. $(0, G)$ and have finite fourth moments.

(ii) All zeros of the determinants of the $z$ transforms

$$g(z) = \sum_{j=0}^{p} A(j)z^j, \quad h(z) = \sum_{k=0}^{q} B(k)z^k,$$

lie outside of the unit circle.

(iii) The $x_j(n)$, the exogenous variables, come from infinite sequences which satisfy, almost surely,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} x_j(m)x_k(m+n) = \gamma_{jk}(n) = \gamma_{kj}(-n), \quad j, k = 1, \ldots, r; \quad n = 0, 1, \ldots,$$

independently of the $\epsilon(n)$ sequence.

(iv) $\Delta \not= 0$, so that the model is identified.\(^{14}\)

The requirement that the $\epsilon(n)$ have finite fourth moments is not needed in the scalar case. However in the vector case for the sample covariance to converge almost surely to the true covariance this assumption seems to be required (see Hannan [26] p 210).\(^{15}\) The condition on the zeros of the determinant of $g(z)$ is required so as to uniquely identify the model while that on the determinant of $h(z)$ is required if there is to exist a stationary solution to (6.1.1) (stationarity will of course depend on the form of the $x_j(n)$ sequence as well). The condition (iii) implies that

$$\gamma_{jk}(n) = \int_{-\pi}^{\pi} e^{in\lambda} dF_{jk}(\lambda),$$

\(^{14}\) For a discussion of the identification problem for systems of equations (which include those of the form (6.1.1)) see Hannan [27].

\(^{15}\) In a yet to be published paper E.J. Hannan has been able to show that this condition is in fact not required for the vector case.
where $F_{jk}(\lambda)$ is the element in row $j$, column $k$; $j,k = 1, \ldots, u$, of the spectral distribution matrix $F(\lambda)$ of the $x(n)$ sequence.

If the data is Gaussian we may calculate the maximum likelihood estimates and consider their limiting distribution. As we have already mentioned Whittle [57] and Walker [54] have independently investigated the asymptotic properties of the estimates obtained by this procedure given only that the $\varepsilon(n)$ are identically and independently distributed. Indeed these authors show that the asymptotic distribution of the estimates obtained in this way is independent of the Gaussian assumption. It seems intuitively reasonable to assume that these results hold in the vector case and we shall in fact make such an assumption in this chapter. The asymptotic results we obtain in the next section verify (at least for the model (6.1.1)) that these arguments do extend to the vector case.

6.2 The Estimation Procedure and the Distribution of the Estimates

In order to efficiently estimate the coefficients of the model (6.1.1) the procedure we use is identical to that for the scalar model (4.2.1), though it is mathematically more complicated. Indeed suppose we take a sample of size $N$, and let

$$z = \begin{pmatrix} z_1(1) & \cdots & z_1(N) \\ \vdots & \ddots & \vdots \\ z_v(1) & \cdots & z_v(N) \end{pmatrix}.$$  

Reverting to tensor notation we let $Z = \text{vec}(z')$ so that the covariance matrix of $Z$ may be expressed as
\[ \Gamma_N = \begin{pmatrix} \Gamma_{11} & \cdots & \Gamma_{1v} \\ \vdots & \ddots & \vdots \\ \Gamma_{v1} & \cdots & \Gamma_{vv} \end{pmatrix}, \]

where for \( i, j = 1, \ldots, v; m,n = 1, \ldots, N, \)

\[ \Gamma_{ij} = [\Gamma_{ij}(m-n)] = \begin{pmatrix} \gamma_{ij}(0) & \cdots & \gamma_{ij}(N-1) \\ \vdots & \ddots & \vdots \\ \gamma_{ij}(N-1) & \cdots & \gamma_{ij}(0) \end{pmatrix}, \]

with \( \gamma_{ij}(n) = E(z_i(m)z_j(m+n)). \)

Letting \( U \) be the \((N \times N)\) unitary matrix with element \( \frac{1}{N} \text{in} \lambda_t^2 \) in row \( t \) column \( n \) once more then, as for the scalar model, we want to minimize \( \sum \Gamma_N^{-1}Z \), where now

\[ \sum \Gamma_N^{-1}Z = \text{tr}\{[(I_v \otimes U_N)Z](I_v \otimes U_N)^{-1}[(I_v \otimes U_N)^*][(I_v \otimes U_N)Z]\}. \quad (6.2.1) \]

But

\[ (I_v \otimes U_N)^{-1}[(I_v \otimes U_N)^*] = [(I_v \otimes U_N)\Gamma_N(I_v \otimes U_N)^*]^{-1}, \]

so that on rearrangement and simplification (6.2.1) becomes (approximately)

\[ \sum \Gamma_N^{-1}Z = \text{tr}\{[(I_v \otimes U_N)^{-1}[(I_v \otimes U_N)^*]^{-1}[(I_v \otimes U_N)Z][I_v \otimes U_N)^*]\}

\[ = \text{tr} \left\{ (2\pi)^{-1} \begin{pmatrix} f^{-1}_{z_{\lambda_1}} & 0 & \cdots & 0 \\ \cdots & \ddots & \cdots & \cdots \\ 0 & \cdots & f^{-1}_{z_{\lambda_N}} & 0 \end{pmatrix} \begin{pmatrix} 2\pi I_{z_2}(\lambda_1) & \cdots & W_z(\lambda_N,1) \\ \cdots & \cdots & \cdots \\ W_z(\lambda_1,N) & \cdots & 2\pi I_{z_N}(\lambda_N) \end{pmatrix} \right\} \]

\[ = \sum_{t=1}^{N} \text{tr}(f^{-1}_{z_{\lambda_t}}I_{z_2}(\lambda_t)), \]
where \( \lambda_t = \lambda_{N-t} = 2\pi t / N; \ t = 1, \ldots, [N/2] \), while \( f_z(\lambda_t) \) and \( W_z(\lambda_s, t) \) have \( f_{jk}(\lambda_t) \) and \( w_{jk}(\lambda_s, t) \) respectively in row \( j \) column \( k; \ j, k = 1, \ldots, r \) with

\[
 f_{jk}(\lambda_t) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \gamma_{jk}(m)e^{-im\lambda_t},
\]

\[
 w_{jk}(\lambda_s, t) = N^{-1} \sum_{m, n=1}^{N} z_j(m)z_k(n)e^{i(m\lambda_s - n\lambda_t)}.
\]

When \( s = t \), \( w_{jk}(\lambda_s, t) = 2\pi I_{jk}(\lambda_t) \), and so \( W(\lambda_s, t) = 2\pi I(\lambda_t) \). Thus maximizing the likelihood function is, in the vector case, asymptotically equivalent to minimizing

\[
 \frac{1}{N} \sum_{i=1}^{N} \text{tr}\{f_z^{-1}(\lambda_t)I_z(\lambda_t)\}. \tag{6.2.2}
\]

On dropping the argument variable \( \lambda_t \) once more, suppose we now let

\[
 g = \sum_{j=0}^{p} A(j)e^{ij\lambda_t}, \quad h = \sum_{k=0}^{q} B(k)e^{ik\lambda_t},
\]

where \( A(0) = B(0) = I_v \), and define both the periodograms \( I_z, I_y, I_x \) and the cross periodograms \( I_{xy}, I_{yx} \) in an obvious manner. Then from (6.1.1) it follows that

\[
 f_z = (2\pi)^{-1}(gg^*) \text{,}
\]

and (on neglecting end terms which, asymptotically, will not affect our results)

\[
 I_z = hI_y h^* + \Delta I_x \Delta^* + \Delta I_{xy} h^* + h I_{yx} \Delta^*.
\]

Thus on neglecting the constant factor \( 2\pi \), from (6.2.2) it follows that we want to minimize the Hermitian form \( J \), where
\[ J = \frac{1}{N} \sum_{t} \text{tr}\{(gGg^*)^{-1}I_{z}\} \]
\[ = \frac{1}{N} \sum_{t} \text{tr}\{(gGg^*)^{-1}(h_{y}h^{*}+2\Delta I_{xy}h^{*}+\Delta I_{x^{'}})\}. \quad (6.2.3) \]

The estimates of \( A(j) \), \( B(k) \) and \( \Delta \) obtained by minimizing (6.2.3) shall be denoted by \( \hat{A}^{(1)}(j) \), \( \hat{B}^{(1)}(k) \) and \( \hat{\Delta}^{(1)} \) respectively, \( j = 1, \ldots, p; \ k = 1, \ldots, q. \)

As we shall see these estimates depend upon initial consistent estimates of the coefficients which we shall in turn denote by \( \hat{A}(j) \), \( \hat{B}(k) \) and \( \hat{\Delta} \), \( j = 1, \ldots, p; \ k = 1, \ldots, q. \)

The partial derivatives of \( J \) with respect to \( \beta_{mn}(k) \) and \( \delta_{sw} \), \( m,n,s = 1, \ldots, v; \ w = 1, \ldots, u; \ k = 1, \ldots, q, \) respectively lead to

\[ \frac{1}{N} \sum_{t} (gGg^*)^{-1} \left[ \hat{h}(1)_{y} + \hat{\Delta}(1)_{xy} \right] e^{-ik\lambda_{t}} = 0 \quad (6.2.4) \]
\[ \frac{1}{N} \sum_{t} (gGg^*)^{-1} \left[ \hat{h}(1)_{xy} + \hat{\Delta}(1)_{x^{'}} \right] = 0 \quad (6.2.5) \]

where, of course,

\[ \hat{g} = \sum_{j=0}^{p} \hat{A}(j)e^{ij\lambda_{t}}, \quad \hat{h}(1) = \sum_{\ell=0}^{q} \hat{B}(1)(\ell)e^{i\ell\lambda_{t}}, \]

and \( \hat{A}(0) = \hat{B}(1)(0) = I_{y}. \) In order to express the estimates \( \hat{B}(1)(\ell), \ \ell = 1, \ldots, q, \hat{\Delta}(1) \) in a more suitable form we proceed as follows: Equations (6.2.4) and (6.2.5) may be rewritten as

\[ \sum_{\ell=1}^{q} \frac{1}{N} \sum_{t} (gGg^*)^{-1} \hat{B}(1)(\ell)_{y} e^{-i(\ell-\ell')\lambda_{t}} + \frac{1}{N} \sum_{t} (gGg^*)^{-1} \hat{\Delta}(1)_{xy} e^{-ik\lambda_{t}} \]
\[ = -\frac{1}{N} \sum_{t} (gGg^*)^{-1} I_{y} e^{-ik\lambda_{t}}, \]
respectively. Now from (1.1.11) we have vec(ABC) = (C' ⊗ A)vec(B), so that these two equations may be expressed in the form

\[
\sum_{\ell=1}^{q} \frac{1}{N} \Sigma_{t} (g^{g_{*}})^{-1} B_{t}^{\ell} (1)_{y} \Psi^{\ell} \lambda_{t} e^{i \ell \lambda_{t}} + \frac{1}{N} \Sigma_{t} (g^{g_{*}})^{-1} \Delta_{t}^{(1)}_{x}
\]

\[
= - \frac{1}{N} \Sigma_{t} (g^{g_{*}})^{-1} I_{x y}^{t}
\]

where \( k, \ell = 1, \ldots, q \), and if

\[
\hat{\Psi}(k) = \frac{1}{N} \Sigma_{t} (g^{g_{*}})^{-1} I_{y x}^{t} e^{-i k \lambda_{t}},
\]

then by vec(\( \hat{\Psi}(k) \)) we mean vec(\( \hat{\Psi}(1):\ldots:\hat{\Psi}(q) \)). Thus if we let

\[
\hat{\rho}^{(1)} = vec(B^{(1)}(1):\ldots:B^{(1)}(q):\Delta^{(1)}),
\]

from (6.2.6) and (6.2.7) it follows that

\[
\hat{\rho}^{(1)} = -B^{-1} \hat{\Delta},
\]

where now for \( k, \ell = 1, \ldots, q \),
Considering the partitioned matrix $\hat{D}$, the (1,1) block is a $(qv^2 \times qv^2)$ matrix, the (1,2) block is a $(qv^2 \times vu)$ matrix and the (2,2) block is a $(vu \times vu)$ matrix. For the vector $\hat{d}$ the (1,1) block is a $(qv^2 \times 1)$ vector while the (2,1) block is a $(vu \times 1)$ vector.

Having obtained these estimates of $B(\ell)$, $\ell = 1, \ldots, q$, $A$, we may now use them to obtain the estimates $\hat{A}(1)(\ell)$ of $A(\ell)$, $\ell = 1, \ldots, p$. Indeed the partial derivative of $J$ with respect to $\alpha_{mn}(k)$, $m,n = 1, \ldots, v$; $k = 1, \ldots, p$, after some manipulation leads to

$$\sum_{\ell=0}^{p} \left( \frac{1}{N} \Sigma_{t} (g\hat{G}g^*)^{-1}e^{-i(k-\ell)\lambda_{t}} \right) I_{y} \Sigma_{t} (g\hat{G}g^*)^{-1}e^{-i\kappa\lambda_{t}} \hat{A}(1)(\ell) = 0,$$

where

$$I_{y} = \hat{h}(1)_{y} + \hat{h}(1)^{*} \hat{h}(1)_{x} + \hat{h}(1)^{*} \hat{h}(1)_{xy} + \hat{h}(1)^{*} \hat{h}(1)_{yx} + \hat{h}(1)_{y} \hat{h}(1)^{*} (1).$$
Suppose we let \( \hat{\alpha}^{(l)} = \text{vec}(\hat{\alpha}^{(l)}(1) : \ldots : \hat{\alpha}^{(l)}(p)) \). Then using (1.1.11) once more we may express (6.2.9) in the form
\[
\hat{\alpha}^{(l)} = \hat{\alpha}^{(l)} - \hat{\alpha}^{(l)},
\]
where \( \hat{\alpha}^{(l)} \) has
\[
\frac{1}{N} \sum_{t=1}^{N} E \left[ \left( (\hat{g} g^*)^{-1} \hat{z}(1) (\hat{g} g^*)^{-1} \right) e^{-(k-\ell)\lambda_t} \right]
\]
as the block in the \( k \)th row, \( \ell \)th column of blocks, \( k, \ell = 1, \ldots, p \), and
\[
\hat{\alpha}^{(l)} = \text{vec} \left[ \frac{1}{N} \sum_{t=1}^{N} (\hat{g} g^*)^{-1} \hat{z}(1) (\hat{g} g^*)^{-1} e^{-ik\lambda_t} \right],
\]
the term in the square brackets being the \( k \)th row of blocks, \( k = 1, \ldots, p \).

From (6.2.8) and (6.2.10) it follows that we require initial consistent estimates \( \hat{A}(\ell) \) of \( A(\ell) \), \( \ell = 1, \ldots, p \), as well as an initial estimate \( \hat{G} \) of \( G \). As in the scalar case we determine initial consistent estimates \( \hat{B}(k) \) of \( B(k) \), \( k = 1, \ldots, q \), and \( \hat{\Delta} \) of \( \Delta \) by instrumental variables, and we use these to form the \( \hat{A}(\ell) \). Now suppose \( \Gamma_y(s) \), \( \Gamma_{yx}(s) \) represent the covariances of the \( y(n) \) and \( x(n) \) sequences respectively while \( \Gamma_{xy}(s) \) and \( \Gamma_{yx}(s) \) represent the cross covariances between the two sequences (where \( \Gamma_{yx}(s) = E(y(m)x'(m+s)) = \Gamma_{xy}'(-s) \)). Then, as we shall see in the next section, for the particular estimation equations (obtained by instrumental variables) we have chosen, in order that initial estimates \( \hat{B}(k) \) and \( \hat{\Delta} \) of \( B(k) \) and \( \Delta \) exist, our final assumption for the model (6.1.1) is:

(v) For \( j, k = 1, \ldots, q \), the matrix
\[
C = \begin{bmatrix}
\Gamma_y(k-p-j) & \Gamma_{yx}(k) \\
\cdots & \cdots & \cdots & \cdots \\
\Gamma_{xy}(-p-j) & \Gamma_x(c)
\end{bmatrix}
\]
is non singular.
Having obtained the initial estimates $\hat{A}(k)$, $\hat{B}(k)$ and $\hat{\Lambda}$, if we let

$$\hat{z}(n) = \sum_{k=0}^{q} \hat{B}(k) y(n-k) + \hat{A} x(n), \quad n = 1, \ldots, N,$$

$$\hat{z} = N^{-1} \sum_{n=1}^{N} \hat{z}(n),$$

and

$$\hat{C}_z(m) = (N-m)^{-1} \sum_{n=1}^{N-m} (\hat{z}(n)-\hat{z}) (\hat{z}(m+n)-\hat{z})', \quad m = 0, \ldots, p,$$

with

$$\hat{C}_z(-m) = \hat{C}_z(m),$$

then

$$\hat{f}_z(\lambda_t) = (2\pi)^{-1} \sum_{m=-p}^{p} \hat{C}_z(m) e^{im\lambda_t},$$

and $G$ may be estimated as

$$\hat{G} = (2\pi/N) \sum_{t} \hat{f}_z(\lambda_t, \lambda_t) \hat{G}^{-1}(\lambda_t).$$

We now introduce the matrices

$$\Delta = \begin{bmatrix}
\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-i(k-\ell)\lambda} (d\tilde{F}_y'(\lambda) \otimes (gGg^*)^{-1}) & \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-i\ell\lambda} (d\tilde{F}_y'(\lambda) \otimes (gGg^*)^{-1}) \\
\vdots & \vdots \\
\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\ell\lambda} (d\tilde{F}_y'(\lambda) \otimes (gGg^*)^{-1}) & \frac{1}{2\pi} \int_{-\pi}^{\pi} (d\tilde{F}_x'(\lambda) \otimes (gGg^*)^{-1})
\end{bmatrix}$$

(6.2.11)

where $k, \ell = 1, \ldots, q$, and the spectral distribution matrices $F_x(\lambda)$, $F_y(\lambda)$, $F_{yx}(\lambda)$ and $F_{xy}(\lambda)$ are defined in an obvious way,
\[ \mathbf{\Phi} = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} (g^{-1}(gG^*)_{G}^{-1} \lambda^{1} \otimes (gG^*)^{-1} e^{-1}(k-\ell)\lambda d\lambda), \right) \] 
(6.2.12)

\[ k, \ell = 1, \ldots, p, \text{ and} \]

\[ \Omega = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} (g^{-1}(gG^*)_{G}^{-1} \lambda^{1} \otimes (gG^*)^{-1} e^{-1}(k-\ell)\lambda d\lambda), \right) \] 
(6.2.13)

\[ k = 1, \ldots, q; \ \ell = 1, \ldots, p. \] Then our first efficient estimate of the \((pv^2 \times 1)\) vector \(\alpha\) is (asymptotically) given by

\[ \tilde{\alpha}^{(1)} = \left[ \frac{1}{pv^2} \int_{-\pi}^{\pi} \mathbf{\Phi}^{-1}(\mathbf{w}' : 0)\Delta^{-1}(\mathbf{\Omega})^{-1}(\hat{\alpha}^{(1)})_+\hat{\alpha}, \right) \] 
(6.2.14)

where \(\mathbf{\Phi}^{(1)}\) is defined in an obvious way and \(\hat{\alpha}\) is formed from the \(\hat{A}(k)\) in a similar way to that in which \(\hat{A}^{(1)}\) is formed from the \(\hat{A}^{(1)}(k), k = 1, \ldots, p.\)

(We discuss the computation of \(\hat{\alpha}\) in the next section).

For computational purposes we have

\[ \tilde{\alpha}^{(1)} = \left[ \frac{1}{pv^2} \hat{A}^{-1}(\mathbf{I}_p \otimes \mathbf{G}^{-1} \otimes I_\nu)(\mathbf{w}' : 0)\hat{\Omega}^{-1}(\mathbf{0})^{-1}(\hat{\alpha}^{(1)})_+\hat{\alpha}, \right) \] 
(6.2.15)

where, if \(\hat{I}_z\) is the periodogram ordinate formed from the \(\hat{z}(n)\) process,

\[ \hat{A} = \left[ \frac{1}{N} \sum_{t} I_\nu \otimes \left( \hat{f}_z^{-1} \hat{I}_z^{-1} \otimes e^{-1}(k-\ell)\lambda_t \right) \right], k, \ell = 1, \ldots, p, \]

and furthermore \(\hat{A}\) converges almost surely to \(2\pi(I_p \otimes \mathbf{G}^{-1} \otimes I_\nu)\). Indeed since \(\hat{I}_z\) converges almost surely to \((2\pi)^{-1}(gG^*)\), \(\hat{A}\) converges almost surely to

\[ 2\pi\left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} I_\nu \otimes (gG^*)^{-1} e^{-1}(k-\ell)\lambda d\lambda, \right]. \]

However
\[ \Phi = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} (g^{-1}(gGg^*)g^{*^{-1}})' \otimes (gGg^*)^{-1} \mathrm{e}^{-\mathrm{i}(k-\ell)\lambda} \mathrm{d}\lambda \right] \]

\[ = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \otimes (gGg^*)^{-1} \mathrm{e}^{-\mathrm{i}(k-\ell)\lambda} \mathrm{d}\lambda \right] , \quad k, \ell = 1, \ldots, p, \]

and since

\[ (G \otimes I_v)(I_v \otimes (gGg^*)^{-1}) = G \otimes (gGg^*)^{-1}, \]

it follows that \( \hat{A} \) converges almost surely to \( \frac{1}{2\pi} (I_v \otimes G^{-1} \otimes I_v) \Phi \), as required.

Suppose we now let \( \beta = \text{vec}(B(1): \ldots : B(q)) \) and \( \delta = \text{vec}(\Delta) \), with \( \hat{\beta}^{(1)} \) and \( \hat{\delta}^{(1)} \) defined in an obvious way. Then, having obtained efficient estimates \( \tilde{\alpha}^{(1)} \) of \( \alpha \), efficient estimates of \( \beta \) and \( \delta \) are given by

\[ \hat{\beta}^{(1)} = \left( \begin{array}{c} \hat{\beta}^{(1)} \\ \vdots \\ \hat{\delta}^{(1)} \end{array} \right) = \tilde{D}^{(1)} \tilde{d}^{(1)}, \tag{6.2.16} \]

where \( \tilde{D}^{(1)} \) and \( \tilde{d}^{(1)} \) are defined in the same way as \( \hat{D}, \hat{d} \) (see (6.2.8)) except that \( \hat{g} \) and \( \hat{G} \) are replaced by \( \tilde{g}^{(1)} \) and \( \tilde{G}^{(1)} \) respectively, where

\[ \tilde{g}^{(1)} = \sum_{j=0}^{P} \tilde{A}^{(1)}(j) e^{\mathrm{i}j\lambda} \mathrm{t} \quad , \quad \tilde{A}^{(1)}(0) = I_v \], and \( \tilde{g}^{(1)} \) is defined by (6.3.3).

In order to determine the asymptotic distribution of the estimates (6.2.14) and (6.2.16) we use a similar derivation to that which was used to obtain the asymptotic distribution of the corresponding estimates in the scalar case. Indeed proceeding heuristically suppose in section 4.3 we replace \( \alpha^2 \) by \( G \) and assume \( \Delta, \Phi \) and \( \Omega \) to be now defined by (6.2.11), (6.2.12) and (6.2.13) respectively. If we now reproduce a similar type of argument to that in section 4.3 it will be found that \( \sqrt{N}(\hat{\alpha}^{(1)} - \alpha), \sqrt{N}(\hat{\beta}^{(1)} - \beta) \) are asymptotically normally distributed and efficient (where \( \tilde{\alpha}^{(1)}, \tilde{\beta}^{(1)} \) are the
vectors of $p_1v^2$ and $(q_1v^2 + vu)$ components defined by (6.2.14) and (6.2.16) respectively). Furthermore the asymptotic covariance matrix of these estimates is given by

$$\text{Cov}[\sqrt{N}(\hat{\alpha}^{(1)} - \alpha), \sqrt{N}(\hat{\beta}^{(1)} - \beta), \sqrt{N}(\hat{\delta}^{(1)} - \delta)] = \begin{bmatrix} \Phi & -\Omega' & 0 \\ -\Omega & 2\pi\Delta & 0 \\ 0 & 0 & 0 \end{bmatrix}^{-1}, \quad (6.2.17)$$

(where, in this covariance matrix, $\Delta, \Phi$ and $\Omega$ are, of course, defined by (6.2.11), (6.2.12) and (6.2.13) respectively). We thus have the following: Theorem 6.2.1

For the vector model (6.1.1) let $\alpha = \text{vec}(A(1):...:A(p))$ and $\rho = \text{vec}(B(l):...:B(q):\Delta)$. Under conditions (i)-(v), for $\hat{\alpha}^{(1)}, \hat{\rho}^{(1)}$ defined by (6.2.14) and (6.2.16) respectively, $(\hat{\alpha}^{(1)}', \hat{\rho}^{(1)}')$ converges almost surely to $(\alpha', \rho')$. Furthermore $\sqrt{N}[(\hat{\alpha}^{(1)} - \alpha') : (\hat{\rho}^{(1)} - \rho')]$ has a distribution converging to the multivariate normal distribution with zero mean and covariance matrix defined by (6.2.17). The estimate is asymptotically efficient.

The estimation procedure will usually be iterated a number of times so that after $j$ iterations we have $\hat{\alpha}^{(j)}, \hat{\rho}^{(j)}$. Thus we may efficiently estimate $\Phi, \Omega$ and $\Delta$ by $\hat{\Phi}^{(j)}, \hat{\Omega}^{(j)}$ and $\hat{\Delta}^{(j)}$ respectively. Furthermore, for $j > 1$, $\hat{g}^{(j)} = (2\pi/N) \sum_{t}^{\hat{g}^{(j-1)}} (\lambda_t)^{-1}(\lambda_t)^{-1}(\lambda_t)^{-1}(\lambda_t)^{-1}, \quad (6.2.18)$
where \( \tilde{g}^{(j)}(\lambda_t) \) is formed in an obvious way, while \( \tilde{f}_z^{(j-1)}(\lambda_t) \) is obtained from \( \tilde{B}^{(j-1)}(k) \), \( k = 1, \ldots, q \) and \( \tilde{z}^{(j-1)} \) in the same way as \( \tilde{f}_z(\lambda_t) \) is obtained from the \( \tilde{B}(k) \), \( k = 1, \ldots, q \) and \( \tilde{z} \). \( \tilde{g}(1) \) is defined (in the next section) by equation (6.3.3).

6.3 The Computational Procedure

The computations involved in applying the estimation procedure for the vector model are completely analogous to those for the scalar model. For the model (6.1.1), i.e.

\[
\sum_{k=0}^{q} B(k)y(n-k)+Ax(n) = \sum_{j=0}^{p} A(j)e(n-j) = z(n),
\]

\( B(0) = A(0) = I_v \), where \( y(n-k) \), \( e(n-j) \) are vectors of \( v \) components while \( x(n) \) is a vector of \( u \) components, we will outline the computational procedure to be applied in order to obtain efficient estimates of the \( A(j), B(k), D; j = 1, \ldots, p; k = 1, \ldots, q \).

Choosing a sample of size \( N \) we first form

\[
\bar{y} = N^{-1} \sum_{m=1}^{N} y(m), \quad \bar{x} = N^{-1} \sum_{m=1}^{N} x(m),
\]

and hence

\[
\begin{align*}
C_{yy}(n) &= (N-n)^{-1} \sum_{m=1}^{N-n} (y(m)-\bar{y})(y(m+n)-\bar{y})', \quad n = 0, \ldots, p+q, \\
C_{xx}(o) &= N^{-1} \sum_{m=1}^{N} (x(m)-\bar{x})(x(m)-\bar{x})', \\
C_{yx}(n) &= (N-n)^{-1} \sum_{m=1}^{N-n} (y(m)-\bar{y})(x(m+n)-\bar{x})', \quad n = 0, \ldots, p+q,
\end{align*}
\]

with \( C_{y}(-n) = C_{y}^t(n) \) and \( C_{xy}(-n) = C_{yx}^t(n) \). We also compute
\[ I_y(\lambda_t) = w_y(\lambda_t)w_y^*(\lambda_t), \quad I_x(\lambda_t) = w_x(\lambda_t)w_x^*(\lambda_t), \]

and
\[ I_{yx}(\lambda_t) = w_y(\lambda_t)w_x^*(\lambda_t), \quad I_{xy}(\lambda_t) = w_x(\lambda_t)w_y^*(\lambda_t), \]

where
\[ w_y(\lambda_t) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} y(n)e^{i\lambda_t}, \quad w_x(\lambda_t) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} x(n)e^{i\lambda_t}. \]

\[ \lambda_t = \lambda_{N-t} = \frac{2\pi t}{N}, \quad t = 1, \ldots, [N/2], \]

and will be assumed to be of this form throughout this section. We are now in a position to carry out Step 1 of the computational procedure.

**Step 1**

In this step we obtain initial consistent estimates of the coefficients. These estimates will not be required after we obtain \( \hat{\alpha}^{(1)}, \hat{\rho}^{(1)} \) (the first efficient estimates of \( \alpha, \rho \) respectively) so that this step is not repeated when we iterate the estimation procedure. For \( j = 1, \ldots, q \) we obtain the following relations by instrumental variables

\[ \sum_{k=1}^{q} \hat{B}(k)C_y(k-p-j) + \hat{\Delta}C_{xy}(-p-j) = -C_y(-p-j), \]

\[ \sum_{k=1}^{q} \hat{B}(k)C_{yx}(k) + \hat{\Delta}C_x(o) = -C_{yx}(o). \]

(In order that these equations may be solved, as we mentioned earlier, for \( j, k = 1, \ldots, q \) the matrix

\[
\begin{bmatrix}
\Gamma_y(k-p-j) & \cdots & \Gamma_y(k) \\
\cdots & \cdots & \cdots \\
\Gamma_{xy}(-p-j) & \cdots & \Gamma_x(o)
\end{bmatrix}
\]

must be non singular). The solution of these equations gives the initial
estimates \( \hat{B}(k) \), \( k = 1, \ldots, q \), \( \hat{z} \), which we may now use to form

\[
\hat{z}(n) = \sum_{k=0}^{q} \hat{B}(k)y(n-k) + \hat{\Delta}x(n), \quad n = 1, \ldots, N,
\]

and hence

\[
\hat{C}_z(n) = (N-n)^{-1} \sum_{m=1}^{N-n} (\hat{z}(n)-\hat{z})(\hat{z}(m+n)-\hat{z})', \quad n = 0, \ldots, p,
\]

with \( \hat{C}_z(-n) = \hat{C}_z(n) \). We may now compute

\[
\hat{f}_z(\lambda_t) = (2\pi)^{-\frac{1}{2}} \sum_{n=-p}^{p} \hat{C}_z(n) e^{in\lambda_t}
\]

and

\[
\hat{I}_z(\lambda_t) = \hat{w}_z(\lambda_t) \hat{w}_z(\lambda_t)',
\]

where

\[
\hat{w}_z(\lambda_t) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} \hat{z}(n) e^{in\lambda_t}.
\]

An initial estimate of the \( (pq^2 \times 1) \) vector \( \alpha \) is then computed as

\[
\hat{\alpha} = -\hat{A}^{-1} \hat{\Delta},
\]

where for \( k, \ell = 1, \ldots, p \), the block in row \( k \) column \( \ell \) of blocks of \( \hat{A} \) is given by

\[
\left[ \frac{1}{N} \sum_{t} \hat{f}_z^{-1}(\lambda_t) \hat{I}_z(\lambda_t) \hat{f}_z^{-1}(\lambda_t)e^{-i(k-\ell)\lambda_t} \right], \quad (6.3.2)
\]

while

\[
\hat{\alpha} = \text{vec}\left[ \frac{1}{N} \sum_{t} \hat{f}_z^{-1}(\lambda_t) \hat{I}_z(\lambda_t) \hat{f}_z^{-1}(\lambda_t)e^{-i\lambda_t} \right],
\]

the term in the square brackets here being the \( k \)th row of blocks, \( k = 1, \ldots, q \).
Step 2

Form

\[ \hat{g}(\lambda_t) = \sum_{j=0}^{p} \hat{A}(j)e^{ij\lambda_t}, \quad \hat{h}(\lambda_t) = \sum_{k=0}^{q} \hat{B}(k)e^{ik\lambda_t}, \]

where \( \hat{A}(o) = \hat{B}(o) = I \), and

\[ \hat{G} = (2\pi/N) \sum_{t} \hat{g}^{-1}(\lambda_t) \hat{f}(\lambda_t) \hat{g}^{*}(\lambda_t). \]

Then for \( k, \ell = 1, \ldots, q \), compute

\[ \hat{D} = \begin{bmatrix} \hat{K} & \cdots & \hat{L} \\ \hat{L}^* & \cdots & \hat{M} \end{bmatrix}, \quad \hat{d} = \begin{bmatrix} \hat{S} \\ \ldots \\ \hat{S} \end{bmatrix}, \]

where

\[ \hat{K} = \frac{1}{N} \sum_{t} \hat{I}_{x}^{'}(\lambda_t) \otimes (\hat{g}(\lambda_t)\hat{g}^{*}(\lambda_t))^{-1} e^{-1(k-\ell)\lambda_t}, \]

\[ \hat{L} = \frac{1}{N} \sum_{t} \hat{I}_{xy}(\lambda_t) \otimes (\hat{g}(\lambda_t)\hat{g}^{*}(\lambda_t))^{-1} e^{-1k\lambda_t}, \]

\[ \hat{M} = \frac{1}{N} \sum_{t} \hat{I}_{x}^{'}(\lambda_t) \otimes (\hat{g}(\lambda_t)\hat{g}^{*}(\lambda_t))^{-1}, \]

\[ \hat{R} = \text{vec}\left( \frac{1}{N} \sum_{t} (\hat{g}(\lambda_t)\hat{g}^{*}(\lambda_t))^{-1} \hat{I}_{y}(\lambda_t)e^{-1k\lambda_t} \right), \]

\[ \hat{S} = \text{vec}\left( \frac{1}{N} \sum_{t} (\hat{g}(\lambda_t)\hat{g}^{*}(\lambda_t))^{-1} \hat{I}_{yx}(\lambda_t) \right), \]

and hence

\[ \hat{p}(1) = \begin{bmatrix} \hat{p}(1) \\ \vdots \\ \hat{p}(1) \end{bmatrix} = -\hat{D}^{-1}\hat{d}. \]
It is now possible to use $\hat{\beta}(1)$, $\hat{\delta}(1)$ to form $\hat{z}(1)(n)$, $n = 1, \ldots, N$, and hence $I^2(\lambda_t)$ in the same way as we used $\hat{z}(n)$, $n = 1, \ldots, N$, to form $I^2(\lambda_t)$. We then compute

$$\hat{\alpha}(1) = -\hat{\alpha}(1)^{-1} \hat{a}(1),$$

with $\hat{\alpha}(1)$, $\hat{\delta}(1)$ defined in the same way as $\hat{\alpha}$, $\hat{\delta}$ except that $I^2(\lambda_t)$ is replaced by $I^2(\lambda_t)$, and $\hat{f}_z(\lambda_t)$ is now defined by

$$\hat{f}_z(\lambda_t) = (2\pi)^{-1} g(\lambda_t)G^*(\lambda_t).$$

Furthermore if we form the matrix

$$\hat{\Omega} = \left[ \frac{1}{N} \sum_t \left( \hat{g}^{-1}(\lambda_t)(\hat{g}(\lambda_t)\hat{G}^*(\lambda_t))\hat{\eta}^{-1}(\lambda_t) \right) \otimes \left( \hat{g}(\lambda_t)\hat{G}^*(\lambda_t) \right)^{-1} e^{-(k-l)\lambda_t} \right],$$

where $k = 1, \ldots, q; \ell = 1, \ldots, p$, then the first efficient estimate of $\alpha$ (for $\hat{\alpha}$ defined by (6.3.2)) may be computed as

$$\tilde{\alpha}(1) = \left[ I_{pv}^2 \hat{A}^{-1}(I_p \otimes \hat{G}^{-1} \otimes I_V) (\hat{\Omega} : 0) \hat{D}^{-1}(\hat{\Omega}) \right]^{-1} (\hat{\alpha} - \hat{\alpha}(1)) \hat{\alpha}.$$

We may now use $\tilde{\alpha}(1)$ to form $\tilde{D}(1)$, $\tilde{d}(1)$ and hence the efficient estimate

$$\tilde{\beta}(1) = \left( \begin{array}{c} \tilde{\beta}(1) \\ \vdots \\ \tilde{\delta}(1) \end{array} \right) = -\hat{D}(1)^{-1} \tilde{d}(1),$$

where $\tilde{D}(1)$, $\tilde{d}(1)$ are defined as were $\hat{D}$, $\hat{d}$ but using

$$\tilde{g}(1)(\lambda_t) = \sum_{j=0}^{p} \hat{A}(1)(j)e^{ij\lambda_t}, \quad \hat{A}(1)(0) = I_V,$$

(with the $\hat{A}(1)(j)$, $j = 1, \ldots, p$, formed in an obvious way from the $\tilde{\alpha}(1)$ in
place of \( g(\lambda_t) \) and \( \tilde{G}^{(1)} \) in place of \( \hat{G} \). The estimate \( \tilde{G}^{(1)} \) is given by

\[
\tilde{G}^{(1)} = (2\pi/N) \sum_{t} \tilde{G}^{(1)}(\lambda_t) \tilde{f}_z(\lambda_t) \tilde{G}^{(1)\ast}(-1)(\lambda_t),
\]

where \( \tilde{f}_z(\lambda_t) \) is computed in Step 1.

The procedure is now iterated a number of times repeating Step 2 using \( \tilde{\alpha}^{(1)} \), \( \tilde{\beta}^{(1)} \) in place of \( \hat{\alpha} \), \( \hat{\beta} \) to form \( \tilde{\alpha}^{(2)} \), \( \tilde{\beta}^{(2)} \) and so on, so that in general we have, for the \((k+1)\)st iteration, \( k \geq 1 \),

\[
\tilde{\alpha}^{(k+1)} = [I - A^{(k)} \otimes \bar{G}^{(k)}]^{-1} \bar{G}^{(k)} (\tilde{\alpha}^{(k)}')^{-1} \tilde{\alpha}^{(k+1)} + \tilde{\alpha}^{(k)}
\]

where, for \( k > 1 \), \( \tilde{G}^{(k)} \) is defined by (6.2.18).

6.4 Conclusion

By generalizing the estimation procedure developed in chapter 4 we have, in this chapter, obtained estimates of the coefficients of the vector mixed autoregressive moving average model with exogenous variables. The estimates we have derived, namely (6.2.14) and (6.2.16), are asymptotically normally distributed (with asymptotic covariance matrix given by (6.2.17)) and efficient. Finally, as section 6.3 indicates, the computational procedure to be applied in order to obtain these estimates is completely analogous to that for the scalar model.
BIBLIOGRAPHY


Whittle, P. "On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix". Biometrika, 50 (1963), 129-134.


Wold, H. "A Study in the Analysis of Stationary Time Series". Almqvist and Wiksell, Uppsala (1938).