ESTIMATION FOR VECTOR LINEAR TIME SERIES MODELS

by

W.T.M. Dunsmuir

A thesis submitted for the degree of

Doctor of Philosophy

in the Australian National University

Canberra

May 1976

DECLARATION

The results of this thesis are my own except where otherwise acknowledged.

ACKNOWLEDGEMENTS

My foremost thanks go to Professor E.J. Hannan who introduced me to the problems discussed in this thesis. For his guidance, encouragement and criticism I am very grateful. The work of Chapters 2 and 3 which is also contained in Dunsmuir and Hannan (1976) was developed jointly with Professor Hannan.

This research was supported by an Australian Government Postgraduate Research Award supplemented by the Australian National University.

I have greatly enjoyed my time spent in both Departments of Statistics at the Australian National University and would like to thank the members, students and visitors of those departments. I would especially like to thank Dr R.L. Tweedie of CSIRO (for his encouragement and assistance with proof reading this thesis) and D.K. Pickard (for various discussions).

My final thanks must go to Mrs B. Geary who, from a poorly prepared manuscript, has produced a first class typescript.

ABSTRACT

This thesis is concerned with asymptotic properties of estimation in various models for vector time series observed in discrete time. All the models considered will be taken to depend on a finite number of parameters. The estimation of these parameters will be achieved by maximising the Gaussian likelihood (or spectral equivalents to this likelihood) although Gaussianity is not required for any of the results to follow. Chapter 1 gives a brief introduction to the theory of multiple time series and each model to be covered in the thesis is introduced. The first main part of this thesis (Chapters 2, 3, 4) is concerned with stationary series. In Chapter 2 the strong law of large numbers and the central limit theorem for estimators of the parameters specifying a general class of stationary, ergodic non-deterministic time series models are established. In Chapter 3 the autoregressive moving average model is introduced. There, aspects of identification and suitable "topologies" for the parameter space are discussed. Following this the strong law of large numbers and central limit theorem are established, under quite general conditions. In Chapter 4 signal plus noise models are discussed. There the signal and noise are taken to be stationary vector sequences of the type discussed in Chapter 2. Also in Chapter 4 an extension of the central limit theorem of Chapter 2 is given. A brief section is also devoted to the special case of a scalar autoregressive signal observed with white noise. The second main part of this thesis, Chapter 5, is concerned with multiple linear regression models in which the residual vector is taken to be a stationary process of the type discussed in Chapter 2 or is taken to be an autoregressive moving average. There the strong law of large numbers and the central limit theorem for the parameters specifying the residual process and the regression coefficients will be established under general conditions.

CONTENTS

I. MATRIX NOTATION

The reference Neudecker (1969) contains many useful identities relating to Kronecker matrix products. In the following table A will be an $s \times s$ matrix, B a $t \times t$ matrix and x a column vector.

Symbol	Meaning
$\mathrm{tr}(A)$, $\mathrm{tr}\, A$	trace of A
$\det(A)$, $\det A$	determinant of A
$\mathrm{adj}(A)$, $\mathrm{adj}\, A$	adjoint of A
$\lambda_1(A)$	smallest eigenvalue of A where A is non-negative definite.
$\lambda_s(A)$	largest eigenvalue of A where A is non-negative definite.
$\lambda_j(A)$	jth smallest eigenvalue of A where A is non-negative definite.
$\mathrm{diag}\{a_1, \ldots, a_s\}$	the diagonal matrix with a_j in the jth position.
$A \otimes B$	the Kronecker matrix product of A and B. The resultant $(st \times st)$ matrix is $\begin{bmatrix} a_{11}B & \cdots & a_{1s}B \\ \vdots & & \vdots \\ a_{s1}B & \cdots & a_{ss}B \end{bmatrix}$
$\mathrm{Vec}(A)$, $\mathrm{Vec}\, A$	the vectorised form of A, the resultant $s^2 \times 1$ vector being $(\mathrm{Vec}\, A)$ where

$$(\mathrm{Vec}\, A)' = \{a_{11}, \ldots, a_{s1}; a_{12}, \ldots, a_{s2}; \ldots; a_{1s}, \ldots, a_{ss}\}$$

A^*	complex conjugate transpose of A
A', x'	transpose of A, of x. (The notation A^* and A' will be used interchangeably when A is real.)

$\|A\|$	the norm of A defined by $\|A\|^2 = \mathrm{tr}(AA^*)$
$\|x\|$	the norm of x defined by $\|x\|^2 = x^*x$.

II. INTERNAL REFERENCING

All theorems, lemmas, corollaries and equations are labelled with a 2 digit number X.Y where X refers to the section number and Y to the theorem, etc. number of that section. Conditions will be labelled CX.Y where X and Y are as just stated. Some examples of usage in referencing are:

equation (2.3.1)	- equation 1 of section 3 of Chapter 2 [or (2.3.1)].
(2.3)	- equation 3 in section 2 of the current chapter.
Theorem 2.3.1	- theorem 1 in section 3 of Chapter 2.
Lemma 2.3	- lemma 3 in section 2 of the current chapter.
C2.3	- condition 3 in section 2 of the current chapter.
2.C2	- all the conditions labelled in section 2 of Chapter 2.
2.C2.1	- condition 1 in section 2 of Chapter 2.
§2.3	- section 3 of Chapter 2.
§1	- section 1 of the current chapter.

Corrigenda for

ESTIMATION FOR VECTOR LINEAR TIME SERIES MODELS

by

W. T. M. Dunsmuir

Page	Line	Was	Should Be
1	10	Section 3	Section 4
	12	Section 4	Section 5
	12	Section 5	Section 6
	15	Section 6	Section 7
	18	Section 2	Section 3
2	eqn. (2.4)	$\int_{-\pi}^{\pi} f(\omega)\,d\omega$	$\int_{-\pi}^{\omega} f(\lambda)\,d\lambda$
	-6	$F^{(3)}$	$F^{(3)}(\omega)$
3	eqn. (2.6)	$\varepsilon(m-j)$	$\varepsilon(n-j)$
	7	(2.4)	(2.5)
	footnote	...nondeterministic should	...nondeterministic" should
5	10	...be written	...be written, apart from a constant, as
	15	are consistent	are strongly consistent
7	-1	...enable a	...enables a
11	10	...x-process θ_x	...x-process as θ_x
13	-1	...Sections 1.4	...Sections 4
14	1	and (1.5)	and 5
	eqn. (7.3) (both places)	$-I_N \otimes \Delta y_N$	$-(I_N \otimes \Delta) y_N$
15	1	$y_j(N)^2$	$y_j(n)^2$

Page	Line	Was	Should Be
17	eqn. (1.5)	$\int$	$\int_{-\pi}^{\pi}$
18	-6	IV.6	VI.6
19	10	...are strongly	...is strongly
24	2	..of θ_n to θ_0	...of a sequence θ_n to θ_0
26	-5	$\varepsilon.\lVert G(0)\rVert$.	$\varepsilon.\lVert G(0)\rVert/(2\pi)$.
27	1	= tr	$= \frac{1}{2\pi}$ tr
	2	= tr	$= \left(\frac{1}{2\pi}\right)^2$ tr
	2	$G(\ell)$	$G(-\ell)$
	-3	note (i)	note b(i)
28	6	$(2\pi)^{-1}\int$	$\int$
	10	$\lambda_1(K_0)$	$\lambda_1(K_0)/(2\pi)$
	12	$(\lambda_1(K_0)+\varepsilon_N)$	$(\lambda_1(K_0)+\varepsilon_N)/(2\pi)$
33	10	$\hat{Q}_N(p^{-1})$	$\hat{Q}_N(P^{-1})$
	-2	$\xi(n)\xi(n)'$	$\xi(m)\xi(n)'$
34	1	$y(0)$	$y(1)$

Page	Line	Was	Should Be
36	4	are required	is required
	-8	log det σ^2	log σ^2
37	-4	non-deterministic	nondeterministic
39	11	criteria	criterion
42	4	$E(\varepsilon(m)\varepsilon(m)'\|F_{n-1})$	$E(\varepsilon(n)\varepsilon(n)'\|F_{n-1})$
	10	$x(n)$,	$x(m)$,
	-2	$E(x(n)\|F_{\varepsilon}(n-1))$ $- E(x(n)\|F_{\varepsilon}(n-1))$	$E(x(n)\|F_{\varepsilon}(n-1))$ $- E(\dot{x}(n)\|F_{\varepsilon}(n-1))$
43	-3	$\underset{\rightarrow}{p}$	$\underset{\rightarrow}{P}$
44	-5	$\Big\|_{(\dot{\theta}_N, \bar{\mu}_N)}$	$\Big\|_{(\bar{\theta}_N, \bar{\mu}_N)}$
	-3	$\bar{\mu}_N - \mu_0$ and $\bar{\theta}_N - \theta_0$	$\|\|\bar{\mu}_N - \mu_0\|\|$ and $\|\|\bar{\theta}_N - \theta_0\|\|$
45	3	$\frac{\partial}{\partial\theta}$	$\frac{\partial}{\partial\theta_j}$
	4	$\frac{\partial}{\partial\theta}$	$\frac{\partial}{\partial\theta_j}$

age	Line	Was	Should Be
47	-7	$+ E_G$	$+ 2E_G$
48	6	$\gamma_{cd}(m'-m)\gamma_{cc}(n'-n) + \gamma_{dd}(n'-m)\gamma_{dc}(m'-n)$	$\gamma_{dd}(m'-m)\gamma_{cc}(n'-n) + \gamma_{dc}(n'-m)\gamma_{cd}(m'-n)$
52	-5	$k^*(\overline{\theta}_N)$	$k^*(\overline{\theta}_N)^{-1}$
	-1	as the r.h.s.	on the r.h.s.
58	-1	greatest	smallest
60	7	$(2\pi)^{-1}$	(2π)
64	2	$-N^{-1/2}$	$-(\sqrt{N}/2)$
66	11	belongs	belong
75	2	c(ii)	C(ii)
76	2	$\int$	$\frac{1}{2\pi}\int$
	3	$\int$	$\frac{1}{2\pi}\int$
77	-2	straight forward	straightforward
82	-6	2C.2	C2.3
97	6	subset	subset of

age	Line	Was	Should Be
101	-9	MacKenszie	MacKenzie
102	-7	Theorem 4.1	Theorem 5.1
103	-3	$s \times s$	$N \times N$
	-1	I_s	I_N
104	7	assumption is	assumption is, apart from a constant
105	-2	$\left[\begin{array}{c:c} & a \\ b & \cdots \\ & 0 \end{array}\right]$	$\left[\begin{array}{c:c} & -a \\ b & \cdots \\ & 0 \end{array}\right]$
106	eqn. (6.9)	append the column vector $\begin{pmatrix} cy \\ 0 \end{pmatrix}$ at end of equation	
	eqn. (6.11)	append the column vector $B_N x_N$ at end of equation	
107	eqn. (6.13)	insert $\sum_{j=1}^{ps}$ before b_3	

Page	Line	Was	Should Be
108	9-10	The matrix in these lines should be replaced by	

$$\begin{bmatrix} \dot{a}(I_p \otimes K)\dot{a}' & [\dot{a}(I_p \otimes K) \;\vdots\; 0]\alpha_N' \\ \dot{\alpha}_N \begin{bmatrix} (I_p \otimes K)a' \\ 0 \end{bmatrix} & a_N(I_{N+p} \otimes K)\alpha_N' \end{bmatrix}$$

where $\dot{a} = \begin{bmatrix} A(0) & & 0 \\ \vdots & \ddots & \\ A(p-1) & \cdots & A(0) \end{bmatrix}$

Page	Line	Was	Should Be
	-9	$\alpha_N(I_N \otimes K)\alpha_N'$	$\alpha_N(I_{N+p} \otimes K)\alpha_N'$
	-3	$(Z-Y'X^{-1}Y)$	$(Z-Y'X^{-1}Y)^{-1}$

Page	Line	Was	Should Be
116	7	κ_{abcd}	σ_{abcd}
	8	κ_{abcd}	σ_{abcd}
	12	Insert: "We call κ_{abcd} the fourth cumulant between $\varepsilon_a(m)$, $\varepsilon_b(m)$, $\varepsilon_c(m)$ and $\varepsilon_d(m)$ so that $\kappa_{abcd} = \sigma_{abcd} - K_{ab}(\theta_0)K_{cd}(\theta_0) - K_{ac}(\theta_0)K_{bd}(\theta_0) - K_{ad}(\theta_0)K_{bc}(\theta_0)$	
117	-6	As in the...	As in the proof...
118	5	Theorem 13.14	Theorem 3.15
	-3	$c_{ud}(t+p-m)$	$c_{ud}(\ell+p-m)$
119	1	$\gamma_{ru}(p-m)\gamma_{tv}(q-n)$	$\gamma_{rv}(q-m)\gamma_{tu}(p-n)$
	2	$f_{ru}(\lambda)$	$f_{rv}(\lambda)$
	3	$f_{tv}(\lambda)$	$f_{tu}(\lambda)$

Page	Line	Was	Should Be
119	9	$\cdot\ L_N(\omega)\int$	$L_N(\omega)\ \cdot\ \frac{1}{2\pi}\int$
	11	first term	second term
	13	second term	first term
120	2	$\frac{1}{2\pi}\int \ldots d\lambda . d\omega \Big\vert$	$\overline{\frac{1}{2\pi}\int \ldots d\lambda . d\omega} \Big\vert$
133	12	p.77):	p.77)):
134	-1 (both places)	$I_s \otimes \Delta$	$I_N \otimes \Delta$
140	2	$\ldots y(m+\ell)'] \times p(\ell)$	$y(m+\ell)'d(N)^{-1}] \times p(\ell)$
141		Interchange f and	f^{-1} throughout the page.

CHAPTER 1

INTRODUCTION

1. Introduction

In this first chapter the necessary theoretical background and the time series models to be considered will be introduced. Section 1 will include a brief account of theoretical aspects of stationary vector time series.

This thesis broadly divides into consideration of two basic situations. The first is where an observed vector time series is modelled as a stationary process. In Section 3 the general finite parameter model for such a time series is discussed. Autoregressive moving average models are then introduced in Section 4 and in Section 5 models for a stationary signal observed with noise are discussed. The second basic situation we will consider is that in which two vector series are observed at each time point. In this respect regression models are considered in Section 6.

Estimation of the parameters specifying these models will be based on a criterion developed from the Gaussian likelihood. This will be introduced and discussed in Section 2. A spectral approximation to this likelihood is also given there. The major part of this thesis is the development of an asymptotic theory of estimation based on such criteria. However, some practical aspects will be discussed.

It is the purpose of this introductory chapter to outline the scope and application of the theory to be presented in later chapters.

2. Some theoretical aspects of stationary vector processes

In this section we will discuss some of the theory concerning stationary vector processes in discrete time. The interval of time will be taken as $t = 1$ throughout this thesis. The following treatment will be brief since it is all available in, for example, Hannan (1970), and is only included in order to establish notation and to provide the correct setting for the results to follow.

If $x(n)$ is a stationary time series, of s components, we may form the covariances

$$\Gamma(n) = E\left(x(m)x(m+n)'\right) = \Gamma(-n)' . \tag{2.1}$$

Then (see p. 46, Hannan 1970) there exists a matrix function of $\omega \in [-\pi, \pi]$, $F(\omega)$, with Hermitian non-negative definite increments $F(\omega+\lambda) - F(\omega)$, $\lambda \geq 0$, and $F(-\pi) = 0$, for which

$$\Gamma(n) = \int_{-\pi}^{\pi} e^{in\omega} F(d\omega) . \tag{2.2}$$

Furthermore the spectrum (or spectral distribution matrix) F is uniquely defined by requiring it to be continuous from the right. In general F has Lebesgue decomposition

$$F(\omega) = F^{(1)}(\omega) + F^{(2)}(\omega) + F^{(3)}(\omega) \tag{2.3}$$

wherein $F^{(1)}(\omega)$ is absolutely continuous with respect to Lebesgue measure on $[-\pi, \pi]$ so that

$$F^{(1)}(\omega) = \int_{-\pi}^{\pi} f(\omega) d\omega . \tag{2.4}$$

The matrix function $f(\omega)$ will be referred to as the spectral density matrix corresponding to $x(n)$. $F^{(2)}(\omega)$ is the discrete part of F and $F^{(3)}$ the singular part.

If $x(n)$ is a zero mean stationary process of s components then the Wold decomposition (p. 158, Hannan (1970)) gives the representation

$$x(n) = u(n) + v(n) , \tag{2.5}$$

where $u(n)$ is the following one sided moving average of serially uncorrelated vectors $\varepsilon(m)$,

$$u(n) = \sum_{j=0}^{\infty} C(j)\varepsilon(m-j) \ , \quad \mathrm{tr} \sum_{0}^{\infty} C(j)KC(j)^* < \infty \tag{2.6}$$

$$E\varepsilon(m) = 0 \ , \quad E\varepsilon(m)\varepsilon(n)' = \delta_{mn}K$$

(δ_{mn} is the Kronecker delta) and $v(n)$, referred to as the deterministic part of $x(n)$, has zero mean and is uncorrelated with $u(n)$. The spectrum of u is absolutely continuous with spectral density matrix f_u (say). In order to associate the decomposition (2.3) of the spectrum with that of the process (2.4) note that

$$\det(K) = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\,\det\left(2\pi f(\omega)\right)d\omega\right\} \tag{2.7}$$

(see Hannan 1970, p. 162) where f is as in (2.4). If $\det K > 0$ then the process is said to be of full rank. In this case it is possible to write

$$f_u(\omega) = f(\omega) \ , \quad F_v(\omega) = F^{(2)}(\omega) + F^{(3)}(\omega)$$

where F_v is the spectrum of the deterministic process $v(n)$. Thus in this full rank case the condition that the spectrum of $x(n)$ is absolutely continuous is equivalent to requiring that $x(n)$ be purely nondeterministic[1] (i.e. $v(n)$ is null in (2.5)). Since the deterministic component of the observed process will usually be modelled as a signal or regression component it appears to be not unduly restrictive to require that $x(n)$ be purely nondeterministic. When $f(\omega) = (2\pi)^{-1}kKk^*$ but K is singular, $x(n)$ is still non-deterministic if K is not null. The case where (2.7) does not hold is very complex in general because of the possibility of a decomposition of f (see (2.9) below) in which the range space of k varies with ω. This cannot happen when f is rational for then if (2.7) does not hold it can only be because K is singular. This last aspect of the rational spectral density case is of no real interest for then there

---

[1] Strictly speaking the term "*linearly* purely nondeterministic should be used here. In Chapter 5 a different notion of purely nondeterministic will be introduced.

would exist linear forms in $x(n)$, $x(n-1)$, ..., $x(n-m)$, for m finite, that are identically zero, a.s., and discoverable from a long (but finite) stretch of data. The rational spectrum case underlies much of the work to follow and it will be assumed in these cases that K is nonsingular, i.e. we will consider only the full rank case. Moreover in cases, other than rational spectra, we will also assume that K is full rank to avoid the difficulties mentioned above.

It also follows that when $x(n)$ is purely nondeterministic and of full rank the spectral density matrix f uniquely factorises as

$$f(\omega) = (2\pi)^{-1} k\left(e^{i\omega}\right) K k\left(e^{i\omega}\right)^* \tag{2.9}$$

where k is the matrix function with complex argument z ,

$$k(z) = \sum_0^\infty C(j) z^j \tag{2.10}$$

having elements that are holomorphic within the unit circle (i.e. for $|z| < 1$) and $k(0) = I_s$. When $x(n)$ is as just described we will rewrite $x(n)$ for $u(n)$ in (2.6) to get

$$x(n) = \sum_0^\infty C(j)\varepsilon(n-j) \tag{2.11}$$

where the $C(j)$ in this expression are the same as in (2.10) and the $\varepsilon(n)$ are the one step prediction errors for predicting $x(n)$ by a linear combination of $x(m)$, $m \leq n-1$.

3. The Gaussian likelihood as a basis for estimation

If $x(n)$ is a zero mean Gaussian vector process then the likelihood may be constructed from a sample $x_N' = [x(1)', \ldots, x(N)']$ as

$$L_N = (2\pi)^{-Ns/2}\left(\det \Gamma_N\right)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2} x_N' \Gamma_N^{-1} x_N\right\} \tag{3.1}$$

where

$$\Gamma_N = Ex_N x_N' \tag{3.2}$$

so that Γ_N has $\Gamma(n-m)$ in the (m, n)th block of $s \times s$ elements. In this case knowledge of Γ_N is equivalent to knowledge of $\Gamma(0), \ldots, \Gamma(N-1)$ which may be obtained from f via (2.2). Hence as the sample size N increases we may, in principle, determine all the covariances and thus f. However in order to obtain a reasonable asymptotic theory of estimation for the $\Gamma(n)$ or equivalently f we must treat these as depending upon a finite number of parameters which we write as the vector θ. Then f will be written $f(\omega; \theta)$ and we will also write $\Gamma(l; \theta)$ and $\Gamma_N(\theta)$. In this new notation $-2N^{-1} \log L_N$ from (3.1) may be written

$$\hat{L}_N(\theta) = N^{-1} \log \det \Gamma_N(\theta) + N^{-1} x_N' \Gamma_N^{-1}(\theta) x_N \tag{3.3}$$

which when minimised with respect to θ yields the estimate $\hat{\theta}_N$. Even when $x(n)$ is not a Gaussian process (3.3) may still be used for estimation. *The main aim of this thesis is to demonstrate that under a wide range of circumstances such a procedure leads to estimators which are consistent* $\left(\hat{\theta}_N \xrightarrow{\text{a.s.}} \theta_0\right.$, *the true value*$\left.\right)$ *and are asymptotically efficient in the sense that* $N^{\frac{1}{2}}\left(\hat{\theta}_N - \theta_0\right)$ *converges to normality with asymptotic covariance matrix matching that obtained by maximum likelihood via* (3.3) *when the data are Gaussian. It is in this sense that we shall use the term "asymptotically efficient".*

The assumption that θ is a *finite* vector of parameters is needed in order that such an asymptotic theory may be established. Moreover many practically arising models (i.e. in which the parameters θ have physically identifiable meanings) will conform to such a specification. It is, however, possible to obtain asymptotic theories of estimation of the spectral density f when f is not finitely parameterised. A recent example of such a treatment is given by Berk (1974). We will not discuss such problems in this thesis.

The use of $\hat{L}_N(\theta)$ may be difficult in practice. For this reason two main alternatives will be considered which in some cases (notably ARMA models to be introduced in Section 5) do lead to easier computational procedures. The first is

$$\bar{L}_N(\theta) = \log\det K(\theta) + (2\pi)^{-1}\int_{-\pi}^{\pi} \operatorname{tr}\left(f^{-1}(\omega;\ \theta)I(\omega)\right)d\omega \tag{3.4}$$

where f and K are as before and $I(\omega)$ is the (matrix) periodogram defined in (3.6) below. This approximation to $\hat{L}_N(\theta)$ has its origins in the work of Whittle (1951) and will be very important in the proofs of asymptotic results in later chapters. However in practice one would often replace the integral in (3.4) by a summation over certain frequencies to arrive at

$$\tilde{L}_N(\theta) = \log\det K(\theta) + N^{-1}\sum_t \operatorname{tr}\left[f^{-1}(\omega_t;\ \theta)I(\omega_t)\right] \tag{3.5}$$

where $I(\omega_t)$ is defined as

$$I(\omega_t) = w(\omega_t)w(\omega_t)^*\ , \tag{3.6}$$

and $w(\omega_t)$ is the discrete Fourier transform (DFT) of the data

$$w(\omega_t) = \frac{1}{\sqrt{2\pi N}}\sum_1^N x(m)e^{im\omega_t}\ . \tag{3.7}$$

Another main aim of this thesis is to show that estimation based on $\tilde{L}_N$ *and* $\bar{L}_N$ *is asymptotically equivalent to estimation based on* $\hat{L}_N$ *under very general conditions.*

4. The general linear process

In Section 1.2 it was observed that the most general representation of a stationary purely nondeterministic process of full rank was given by the one sided moving average (2.11). We introduce into this the parametric representation

$$x(n) = \sum_{0}^{\infty} C(j;\ \theta)\varepsilon(n-j)\ ,\quad E\varepsilon(m)\varepsilon(n)' = \delta_{mn}K(\theta)\ , \tag{4.1}$$

so that the corresponding spectral density matrix is

$$f(\omega;\ \theta) = (2\pi)^{-1}k\left(e^{i\omega};\ \theta\right)K(\theta)k\left(e^{i\omega};\ \theta\right)^* \tag{4.2}$$

where A^* denotes the complex conjugate transpose of A. As it stands the parameterisation of (4.1) in this way is very general since we have not, for example, insisted that k and K be specified by separately varying parameters. Under quite general regularity conditions on k and K as functions of θ and on the space to which θ belongs we will show in Chapter 2 that the use of the Gaussian likelihood (3.3) provides consistent estimators even when the Gaussian assumptions are not satisfied. For this the main requirement is that $x(n)$ be ergodic (so that $x(n)$ is also strictly stationary). This assumption ensures that the probabilistic structure of $x(n)$ may be determined arbitrarily accurately from one sufficiently long record of the process. In other words any part of the probabilistic mechanism generating $x(n)$ which is constant over a single realisation will be treated as constant. Since (usually) only one realisation will be available this assumption is not restrictive (see p. 202, Mackey, 1974). The main use we make of ergodicity is to conclude that when $x(n)$ is strictly stationary and ergodic with finite mean square (and thus second order stationary) the sample covariances of lag n

$$G(n) = N^{-1}\sum_{m=1}^{N-n} x(m)x(m+n)' = G(-n)'\ ,\quad 0 \le n < \infty \tag{4.3}$$

converge a.s. to $\Gamma(n)$. (See p. 203, Hannan 1970.) This observation will be used repeatedly in this thesis.

We also give in Chapter 2 one form of the central limit theorem for $\hat{\theta}_N$ but for the treatment given there we have required k to be parametrised by θ and K by μ where θ and μ may vary independently. This, along with two assumptions on $\varepsilon(m)$ to be given below, enable a central limit theorem to be proved for the estimators $\hat{\theta}_N$ of θ specifying only k , without requiring the existence of moments higher than the second for $\varepsilon(m)$. (Of course the central limit theorem for $\hat{\mu}_N$ specifying K requires the existence of fourth moments for $\varepsilon(m)$.) This central limit theorem has wide application since its validity does not depend on the probability distribution of $\varepsilon(m)$ (or $x(n)$) in a very strong way. The conditions mentioned above for $\varepsilon(m)$ are

$$E\left(\varepsilon(m) \mid F_{m-1}\right) = 0 \quad \text{a.s.} \tag{4.4}$$

and

$$E\left(\varepsilon(m)\varepsilon(m)' \mid F_{m-1}\right) = K(\mu) \quad \text{a.s.} \tag{4.5}$$

where F_m is the σ-algebra generated by $\varepsilon(n)$, $n \leq m$. The first, (4.4), is just the condition that $\varepsilon(m)$ be a sequence of martingale differences (which is clearly satisfied if the $\varepsilon(m)$ are i.i.d.). As pointed out in Hannan and Heyde (1972) the condition (4.4) is a natural one when linear models are being constructed for $x(n)$. That is linear models are usually considered in order that prediction of $x(m)$ based in $x(n)$, $n \leq m-1$, will be linear prediction. The condition (4.4) is equivalent to the condition that the best linear predictor for $x(m)$ is also the best predictor for $x(m)$ (both based on $x(n)$, $n \leq m-1$ and both best in the least squares sense). Hannan and Heyde's proof of this statement is repeated in Chapter 2.

In Chapter 4 an alternative central limit theorem to that of Chapter 2 is given. The need for this arises because it may not always be possible to specify the separate parameterisation of k and K introduced in the last paragraph. Then it has been necessary to assume the existence of fourth moments for the $\varepsilon(m)$. The conditions required are extensions of (4.4) and (4.5) to higher moments and will be introduced in Chapter 4. This will be discussed further in Section 6 of this chapter where models for "signal plus noise" are introduced.

5. Autoregressive-moving average models for multiple time series

A special form for the spectral density matrix (2.9), but one of wide application, arises when the elements of f (and thus of k defined by (2.10)) are rational functions. Such cases correspond to the autoregressive moving average (ARMA) model

$$\sum_{j=0}^{q} B(j)x(n-j) = \sum_{j=0}^{p} A(j)\varepsilon(n-j) \;, \quad B(0) = A(0) = I_s \tag{5.1}$$

$$E\varepsilon(m) = 0 \;, \quad E\varepsilon(m)\varepsilon(n)' = \delta_{mn}K \;.$$

We assume that $x(n)$ is stationary and define

$$h(z) = \sum_{j=0}^{q} B(j)z^j \;, \quad g(z) = \sum_{j=0}^{p} A(j)z^j \;. \tag{5.2}$$

Then the corresponding spectral density matrix is

$$f(\omega) = \frac{1}{2\pi} k\left(e^{i\omega}\right)Kk\left(e^{i\omega}\right)^* \;, \quad k(z) = h^{-1}(z)g(z) \;. \tag{5.3}$$

It is well known (Hannan, 1969, 1971b) that many different structures of the form (5.1) will give rise to the same f. Two such structures will be said to be equivalent. That is, if we denote by θ the elements of $A(j)$, $B(j)$ and K in the following way

$$\theta' = \{\mathrm{vec}([B(1)|\ldots|B(q)|A(1)|\ldots|A(p)]') \quad K_{11}, \ldots, K_{1s}, K_{22}, \ldots, K_{2s}, \ldots, K_{ss}\} \tag{5.4}$$

then θ belongs to the Cartesian product $\mathrm{R}^{(p+q)s^2} \times K$ where R^u is u-dimensional Euclidean space and K is the $s(s+1)/2$-dimensional space of all symmetric positive definite matrices. Since the likelihood criterion $\hat{L}_N(\theta)$ (see (3.3)) depends only on f (or k and K in the factorisation (2.9)) then some method of choosing a unique representative of each equivalence class of θ-points is required. There are several ways of proceeding. In Chapter 3 the method of "simple identification" (Hannan, 1969), "triangular identification" (Hannan, 1971b) and "scalar identification" (Zellner and Palm, 1974) will be defined and discussed.

In Chapter 3 we will also discuss, for ARMA models, the strong law of large numbers, methods of identification and their relationship to the central limit theorem. The conditions required for the general theory of Chapter 2 are stronger than necessary for the corresponding theory for ARMA models. For example, the SLLN of Chapter 2 requires the parameter θ to belong to a set with compact closure. For the ARMA model this requirement may be relaxed because of the polynomial nature of h and g. Moreover, in the ARMA treatment the SLLN will be proved for k_N and K_N corresponding to the equivalence class of points in $R^{(p+q)s^2} \times K$ for which $\hat{L}_N(\theta)$ is minimised. This convergence translates into convergence of equivalence classes of θ-points, where θ is as in (5.4), but does not mean, necessarily, that $\hat{\theta}_N$ converges in the Euclidean topology to the true value. However for the CLT it is necessary to consider coordinates when p and q are fixed. In this regard it will be shown in Chapter 3 that "simple identification" enables the elements of θ in (5.4) to be used as coordinates. Similar discussion will be given for other modes of identification.

Constraints, other than constraints arising from identification, may naturally occur on the elements of $A(j)$, $B(j)$, K specifying (5.1). These will usually correspond to some physical meaning. For example if $x(n)$ is the discrete time sampled record of a continuous time process $x(t)$ where $x(t)$ is generated by a system of linear differential equations with white noise input (see Hannan, 1970, IV.8) then constraints in $A(j)$, $B(j)$ will be introduced. Another example of this will be discussed in the following section.

6. Models for a stochastic signal observed with noise

Up to this point we have considered a stationary vector process which is directly observed. Many examples arise where this is not possible, but instead the signal $y(n)$ is contaminated with added noise $x(n)$ so that $z(n)$ is observed where

$$z(n) = y(n) + x(n) . \tag{6.1}$$

In general $y(n)$ and $x(n)$ may be taken to be of the form (4.1) with spectral densities f_y and f_x (respectively) specified by (4.2). We will now take the parameters specifying the y-process as θ_y and those specifying the x-process θ_x. Then if $x(n)$ and $y(n)$ are independent the spectral density of $z(n)$ is

$$f_z(\omega;\ \theta_y,\ \theta_x) = f_y(\omega;\ \theta_y) + f_x(\omega;\ \theta_x) . \tag{6.2}$$

Before discussing this situation in generality we will specialise to the case where $y(n)$ is a scalar autoregression of degree q given by

$$\sum_0^q \beta(j)y(n-j) = \varepsilon_y(n) , \quad E\varepsilon_y(n)\varepsilon_y(m) = \delta_{nm}\sigma_y^2 , \quad \beta(0) = 1 \tag{6.3}$$

and $x(n)$ is a sequence of i.i.d. $(0,\ \sigma_x^2)$ random variables. (This situation has recently been considered by Pagano (1974) for the case where $y(n)$ and $x(n)$ are Gaussian.) The spectral density of $z(n)$ is then

$$f_z\left(\omega;\ \beta,\ \sigma_y^2,\ \sigma_x^2\right) = \frac{\sigma_y^2}{2\pi}\left|\sum_0^q \beta(j)e^{ij\omega}\right|^{-2} + \frac{\sigma_x^2}{2\pi} . \tag{6.4}$$

As is well known (see Pagano (1974), for example) this may be re-written as

$$f_z\left(\omega;\ \beta,\ \alpha,\ \sigma_z^2\right) = \frac{\sigma_z^2}{2\pi}\,\frac{\left|\sum_0^q \alpha(j)e^{ij\omega}\right|^2}{\left|\sum_0^q \beta(j)e^{ij\omega}\right|^2} , \quad \alpha(0) = \beta(0) = 1 , \tag{6.5}$$

where $\alpha(1), \ldots, \alpha(q)$ and σ_z^2 are functions of the original parameters $\beta(1), \ldots, \beta(q), \sigma_y^2, \sigma_x^2$. It also follows from the Wold decomposition that $z(n)$ has representation

$$z(n) = \sum_{0}^{\infty} C(j;\, \beta,\, \alpha)\varepsilon_z(n-j) \;, \quad E\varepsilon_z(m)\varepsilon_z(n) = \delta_{mn}\sigma_z^2 \;. \tag{6.6}$$

At first sight it would appear that even if it were assumed that $\varepsilon_z(m)$ in this representation satisfy (4.4) (the martingale requirement) and (4.5) then the central limit theorem to be presented in the next chapter would not apply because *both* σ_z^2 and $\{\alpha(1), \ldots, \alpha(q), \beta(1), \ldots, \beta(q)\}$ are functions of another set of parameters $\left\{\beta(1), \ldots, \beta(q), \sigma_y^2, \sigma_x^2\right\}$. However it is not always necessary to require that this latter set of original parameters be efficiently estimated directly. (For example the prediction of $z(n)$ or $y(n)$ on the past of $z(n)$ may only require estimates of the new parameters or a subset of them - we will discuss this in more detail in Chapter 4.) Indeed, in Chapter 4, it will be demonstrated that we can always take $\sigma_z^2, \beta(1), \ldots, \beta(q)$ and $\alpha(q)$ as freely varying parameters (except, of course, for boundary constraints arising, for example, from stationarity requirements) while $\alpha(1), \ldots, \alpha(q-1)$ $(q > 1)$ will depend only on $\beta(1), \ldots, \beta(q)$ and $\alpha(q)$ but not σ_z^2. Then the central limit theorem of the next chapter applies to $\{\beta(1), \ldots, \beta(q), \alpha(q)\}$ when (4.4) and (4.5) are assumed for $\varepsilon_z(n)$. (These are not unnatural assumptions to make in this special case as will be demonstrated in Chapter 4.)

In the general set-up, (6.1) and (6.2), the considerations of the last paragraph may no longer apply. To clarify this consider first of all the fact that the spectral density $f_z(\omega;\, \theta_y,\, \theta_x)$ in (6.2) may be factorised (see(2.9)) as

$$f_z(\omega;\ \theta_z) = (2\pi)^{-1} k\left(e^{i\omega};\ \theta_z\right) K(\theta_z) k\left(e^{i\omega};\ \theta_z\right)^* , \tag{6.7}$$

wherein it has been implicitly assumed that there exists a vector of parameters θ_z for which this holds. In this general case there is no obvious way of selecting a partition of θ_z into parameters specifying K and parameters specifying k . This is the first reason for presenting an extension of the CLT of Chapter 2, in Chapter 4. Another consideration arises as follows. $z(n)$ may be decomposed as

$$z(n) = \sum_{0}^{\infty} C\left(j;\ \theta_z\right) \varepsilon_z(n-j) \tag{6.8}$$

where $\varepsilon_z(n)$ are the one-step linear prediction errors for predicting $z(n)$ from its past. However it may not always be a natural assumption to require that these be martingale differences (for example, if the best linear predictor of the signal $y(n)$ is required). For this second reason a variation to the extended central limit theorem of Chapter 4 is given wherein assumptions of the type (4.4) and (4.5) are imposed on the innovations $\varepsilon_y(n)$ and $\varepsilon_x(n)$ and not necessarily on $\varepsilon_z(n)$.

7. Regression models

Until now we have been discussing the case where only one vector series has been observed at each time point and this has been stationary and non-deterministic. However a wide range of problems arise, for example, in which there is an input to a system which may be observed and an output which is also observed. The first case we consider is that of a vector $z(n)$ of dimension $s \times 1$ which is regressed on a vector $y(n)$ of dimension $t \times 1$ as follows

$$z(n) = g\left(y(n);\ \psi\right) + x(n) \tag{7.1}$$

where $x(n)$ is a stationary process of the type discussed in Sections 1.4

and (1.5) with parameters θ specifying its spectrum and $g\{y(n);\psi\}$ is a function of the regression (or exogenous) variables $y_1(n), \ldots, y_t(n)$ and a vector of parameters ψ. This specification does not exclude lagged values of $y(n)$ since we can always redefine them at a new time point. The case where $g\{y(n);\psi\} = \Delta y(n;\phi)$ (i.e. a non-linear regression model) has been discussed by Hannan (1971a) for $s = 1$ and Robinson (1972) for $s > 1$ when the spectral density of $x(n)$ is known. (Some discussion is given concerning the estimation of f_x.) Here we shall be content to study a simpler model than the above non-linear regression; that is we consider the linear regression model

$$z(n) = \Delta y(n) + x(n) \tag{7.2}$$

where $z(n)$, $x(n)$ are $s \times 1$ and $y(n)$ is $t \times 1$ so that Δ is $s \times t$.

Since both $z(n)$ and $y(n)$ are observed at $n = 1, 2, \ldots, N$ we will treat the $y(n)$ as if they are deterministic since this makes the treatment more general. The analogue of the likelihood (3.3) in this case is

$$\hat{L}_N(\theta, \Delta) = N^{-1}\left\{\log \det \Gamma_N(\theta) + \left(z_N - I_N \otimes \Delta y_N\right)'\Gamma_N^{-1}(\theta)\left(z_N - I_N \otimes \Delta y_N\right)\right\} \tag{7.3}$$

where $A \otimes B$ denotes the Kronecker product of the two matrices. Now $\hat{L}_N$ is minimised by choice of $\hat{\theta}_N$ and $\hat{\Delta}_N$. One may also define the analogues of $\tilde{L}_N$, $\bar{L}_N$ for this model - see Chapter 5. We will discuss the asymptotic theory for these estimators in Chapter 5 when the spectral density matrix f of $x(n)$ satisfies conditions similar to those of Chapter 2 and $y(n)$ satisfies Grenander's conditions (Hannan, 1970, p. 77):

$$\left.\begin{array}{ll}\text{(i)} & \lim_{N\to\infty} d_j(N)^2 = \infty\ ,\quad d_j(N)^2 = \sum_1^N y_j(N)^2\ ,\quad 1 \le j \le t\ ;\\[2ex] \text{(ii)} & \lim_{N\to\infty} \dfrac{y(N)^2}{d_j(N)^2} = 0\ ,\quad 1 \le j \le t\ ;\\[2ex] \text{(iii)} & \lim_{N\to\infty} \dfrac{\sum_{m=1}^{N} y_j(m)y_k(m+n)}{d_j(N)d_k(N)} = \Gamma_{jk}^{(Y)}(n),\ |\Gamma_{jk}^{(Y)}(n)| < 1,\ -\infty < n < \infty\ ;\end{array}\right\} \tag{7.4}$$

and for which $\Gamma^{(Y)}(0) = \left[\Gamma_{jk}^{(Y)}(0)\right]$ is positive definite. Since efficient estimation of Δ will in general require efficient estimation of θ $\{$specifying the spectrum of $x(n)$ $\}$ the joint estimation of θ, Δ needs to be considered. Hannan (1973a) established the general results for the case $s = 1$, $t \ge 1$, and we extend this to the vector case $s > 1$, $t \ge 1$, in Chapter 5.

CHAPTER 2

ASYMPTOTIC THEORY FOR THE GENERAL FINITE PARAMETER MODEL

1. Introduction

In this chapter the strong law of large numbers (SLLN) and the central limit theorem (CLT) will be discussed for various estimators of the parameters θ specifying the general finite parameter model for a stationary purely nondeterministic and ergodic vector time series. In Chapter 1 this model was given as

$$x(n) = \sum_{j=0}^{\infty} C(j;\ \theta)\varepsilon(n-j) \tag{1.1}$$

where

$$\sum_{j=0}^{\infty} \|C(j;\ \theta)\|^2 < \infty\ , \quad C(0;\ \theta) = I_s$$

(where $\|A\|^2$ is the matrix norm $\mathrm{tr}(AA^*)$) and

$$E\varepsilon(m) = 0\ , \quad E\varepsilon(m)\varepsilon(n)' = \delta_{mn}K(\theta)\ .$$

Here $x(n)$ and $\varepsilon(n)$ are $s \times 1$ vectors so that $C(j;\ \theta)$ and $K(\theta)$ are $s \times s$ matrices. The spectral density matrix corresponding to $x(n)$ given by (1.1) is

$$f(\omega;\ \theta) = (2\pi)^{-1}k\left(e^{i\omega};\ \theta\right)K(\theta)k\left(e^{i\omega};\ \theta\right)^* \tag{1.2}$$

where

$$k\left(e^{i\omega};\ \theta\right) = \sum_{j=0}^{\infty} C(j;\ \theta)e^{ij\omega}$$

(see Chapter 1).

When $x(n)$ is sampled at N consecutive time points the following minimisation criterion (also introduced in Chapter 1) may be used to obtain an estimator of the vector of parameters θ .

$$\hat{L}_N(\theta) = N^{-1} \log \det \Gamma_N(\theta) + \hat{Q}_N(\theta)\ , \tag{1.3}$$

$$\hat{Q}_N(\theta) = N^{-1}x_N'\Gamma_N^{-1}(\theta)x_N \ .$$

We recall that this is $(-2N^{-1})$ times the log-likelihood constructed on the assumption that $x_N' = (x(1)' \vdots \ldots \vdots x(N)')$ is a vector of observations from a multivariate Gaussian distribution with zero mean and covariance matrix

$$Ex_Nx_N' = \Gamma_N(\theta) \ . \tag{1.4}$$

Here the (m, n)th block of $s \times s$ elements in $\Gamma_N(\theta)$ is

$$\Gamma(n-m;\ \theta) = \int e^{i(n-m)\omega}f(\omega;\ \theta)d\omega \ . \tag{1.5}$$

We define $G(-n)' = G(n) = N^{-1}\sum_1^{N-n} x(m)x(m+n)'$ as the sample serial covariance. Although the estimation procedures to be considered are based on $\hat{L}_N$ (or approximations to $\hat{L}_N$ to be shortly introduced) the assumption of Gaussianity for x_N is *not* required for any of the results of this chapter.

The criterion (1.3) is usually a difficult function to minimise computationally since, in part, the inversion of the $Ns \times Ns$ matrix Γ_N will be required. (However, for the ARMA model there exist either approximations to Γ_N which are easier to deal with or efficient routines for the inversion of Γ_N. See §3.6 for an example of this.)

Whittle (1951) for the scalar case $(s = 1)$ first introduced an approximation to $\hat{L}_N(\theta)$ which is based on the spectral density $f(\omega;\ \theta)$ and the periodogram calculated from $x(1), \ldots, x(N)$. To obtain the (matrix) periodogram one forms the discrete Fourier transform (DFT) at frequency ω as

$$w(\omega) = \frac{1}{\sqrt{2\pi N}}\sum_{m=1}^{N} x(m)e^{im\omega} \tag{1.6}$$

and then the periodogram

$$I(\omega) = w(\omega)w(\omega)^* . \tag{1.7}$$

Later Whittle (1953) derived the corresponding approximation to $\hat{L}_N$, when $s > 1$, as

$$\overline{L}_N(\theta) = \log \det K(\theta) + \overline{Q}_N(\theta) , \tag{1.8}$$

$$\overline{Q}_N(\theta) = (2\pi)^{-1} \int_{-\pi}^{\pi} \operatorname{tr}\left(f^{-1}(\omega;\ \theta)I(\omega)\right) d\omega .$$

However, because of the integral appearing in $\overline{Q}_N$ this approximate form would not usually be used in practice. Instead one could consider

$$\tilde{L}_N(\theta) = \log \det K(\theta) + \tilde{Q}_N(\theta) , \tag{1.9}$$

$$\tilde{Q}_N(\theta) = (N')^{-1} \sum_t \operatorname{tr}\left[f^{-1}\left(\omega_t:\ \theta\right)I\left(\omega_t\right)\right] ,$$

in which the sum $\sum_t$ is over the frequencies

$$\omega_t = 2\pi t/N' , \quad -N'/2 < t \le [N'/2]$$

and N' is chosen to be some highly composite number not less than N. As is well known this facilitates the efficient computation of $w\left(\omega_t\right)$ via the fast Fourier transform (FFT). Frequently N' is chosen to be the least power of 2 not less than N. In this case $N'/N \le 2$. However it may benefit the estimation of θ via $\tilde{L}_N$ to choose N' at least $2N$ even when there is a power of 2 so that $N \le 2^M < 2N$. Nonetheless we will assume below that $N'/N \le b$ where b is finite.

A discussion of the reasonableness of the approximation (1.8) to (1.3) is given in Hannan (1970, Section IV.6) and will not be repeated here. The validity of this approximation (and the approximation $\tilde{L}_N$ in (1.9)) under general circumstances is the subject of this chapter. The estimators obtained by minimizing $\hat{L}_N$, $\overline{L}_N$, $\tilde{L}_N$ will be referred to as $\hat{\theta}_N$, $\overline{\theta}_N$, $\tilde{\theta}_N$ respectively. The space over which this minimisation is performed will be described in the next section. The true value of θ will be called θ_0.

That the approximation $\overline{L}_N$ leads to consistent and asymptotically efficient estimators was indicated by Whittle (1951, 1953). Whittle (1961), for the case $s = 1$, presents some general results concerning the asymptotic properties of $\overline{\theta}_N$ while Walker (1964), treating the same case, gives rigorous proofs of the weak law of large numbers and the central limit theorem. (The results of Walker and Whittle will be discussed more fully at the end of the next section.) Hannan (1973a) extended the treatment for $s = 1$ to cases previously not covered and for the first time demonstrated rigorously that all procedures $(\hat{L}_N, \overline{L}_N, \tilde{L}_N)$ yield asymptotically equivalent estimators in that each of $\hat{\theta}_N, \overline{\theta}_N, \tilde{\theta}_N$ are strongly consistent and that $N^{\frac{1}{2}}(\hat{\theta}_N-\theta_0)$, $N^{\frac{1}{2}}(\overline{\theta}_N-\theta_0)$, $N^{\frac{1}{2}}(\tilde{\theta}_N-\theta_0)$ possess the same asymptotic distribution. In this chapter corresponding results to Hannan's (1973a) are presented for the vector case $(s > 1)$. Section 2 is devoted to the SLLN and §3 to the CLT. Davies (1973) has also considered the vector case but his approach is very different from that presented here. (We will discuss this in §3.) Other treatments for special models (principally vector ARMA models) are available in the literature but a discussion of these treatments will be postponed until Chapter 3 where a complete discussion of the ARMA case is given.

In the formulation of the model (1.1) with spectral density matrix (1.2) k and K are parameterised by the same parameter θ. In the case $s = 1$ it is possible to normalise the covariance matrix $\Gamma_N(\theta)$ by division by $K(\theta)$ (here a scalar) and *if* K is taken as a freely varying parameter (i.e. not constrained by the parameters specifying k) then $\hat{L}_N(\theta, K)$ may be concentrated with respect to K for fixed θ yielding an expression which is in some ways easier to deal with (see Hannan 1973a). However when $s > 1$ this is usually no longer possible for $\hat{L}_N$ and for $\tilde{L}_N$, $\overline{L}_N$ does not lead to any significant simplification of the proofs.

Also there are cases where K and k must be parameterised by the same vector of parameters. An example of this is the problem of estimation in a model for a signal observed with noise which we discuss in Chapter 4. In the SLLN it is the parameterisation $k(e^{i\omega};\ \theta)$, $K(\theta)$ we discuss. However in the CLT given in this chapter it is required that K be parameterised by μ and k be parameterised by θ where θ and μ vary independently of each other. This is done in order that a CLT may be obtained without assumptions on the existence of moments for $\varepsilon(m)$ higher than the second. We return to the CLT for the case where K is not separately parameterised in Chapter 4.

The final remark of this section concerns the relationship of the treatment given in this chapter for the general finite parameter model (1.1) to the vector ARMA model. The conditions we introduce in order to prove the SLLN for the model (1.1) are stronger than those required to obtain the same results for the ARMA model. However the proofs for this latter case (which are given in Chapter 3) are in many respects very similar to those we present in the next section. Also there have been suggested models which do not have spectral density matrices with rational elements in $e^{i\omega}$ (as in the ARMA case). The following examples illustrate this.

EXAMPLE 1. Bloomfield (1973) presents the following exponential model for the spectral density of a single stationary time series

$$f(\omega;\ \theta) = \frac{\sigma^2}{2\pi}\left|k\left(e^{i\omega};\ \theta\right)\right|^2$$

where

$$k(z;\ \theta) = \exp\left\{\sum_{r=1}^{p} \theta_r z^r\right\}, \quad -\infty < \theta_j < \infty .$$

EXAMPLE 2. Dzhapharidze and Yaglom (1974) consider models of the type

$$f(\omega;\ \theta) = \sum_{k=0}^{p} \theta_k g_k\left(e^{i\omega}\right)$$

where the g_k are known real even functions of $\omega \in [-\pi, \pi]$. For example, if the θ_k are taken as the serial covariances from a moving average process of degree p then the $g_k(e^{i\omega})$ are just $2\cos(k\omega)$.

2. The Strong Consistency of the Estimators

In this section the almost sure convergence of $\hat{\theta}_N$, $\bar{\theta}_N$, $\tilde{\theta}_N$ to θ_0 is established under certain conditions which mainly relate to the behaviour of the spectrum $f(\omega; \theta)$ as a function of $(\omega; \theta)$. In the factorisation of $f(\omega; \theta)$ given in equation (1.2), $k(z; \theta)$ has elements which are analytic within a disc containing the region $|z| \le 1$ and has no zeros for $|z| < 1$. In the following treatment we have chosen not to bound the poles of $k(z; \theta)$ away from the unit circle. Further the definition of $\bar{L}_N$, $\tilde{L}_N$ require us to consider f^{-1} and thus k^{-1} . However we do not want to impose the requirement that det $k(z; \theta)$ have no zeros on the unit circle $|z| = 1$. To enable us to treat such cases we have further assumed that we may factorise $k(z; \theta)$ as

$$k(z; \theta) = h^{-1}(z; \theta)g(z; \theta) \tag{2.1}$$

in which each of h and g at least have elements which are analytic within the unit circle and their determinants have no zeros there. Then if we take our true parameters θ_0 to be such that also $\det h(z; \theta_0) \neq 0$ $\forall |z| = 1$ and $\det K(\theta_0) > 0$, $\|K(\theta_0)\| < \infty$ then the corresponding k has the properties mentioned above and the corresponding f is integrable and is the spectrum of a full rank process.

The above is formalised in the following conditions which shall be assumed throughout the remainder of this chapter.

C2.1: The true parameter θ_0 belongs to a set Θ in a topological space

where

$$\Theta = \{\theta : \det h(e^{i\omega}; \theta) \neq 0,\ \omega \in [-\pi, \pi],\ \det K(\theta) > 0,\ \|K(\theta)\| < \infty\} .$$

C2.2: The closure $\overline{\Theta}$ of Θ is compact and $g(e^{i\omega}; \theta)$, $h(e^{i\omega}; \theta)$ have elements which are continuous in $(\omega; \theta) \in [-\pi, \pi] \times \overline{\Theta}$. These elements are also analytic within the unit circle and $\det g(z)$, $\det h(z)$ have no zeros there. Also $g(0; \theta) = h(0; \theta) = I_s$ $\forall\theta \in \overline{\Theta}$.

C2.3: $k(e^{i\omega}; \theta) \not\equiv k(e^{i\omega}; \theta_0)$ for all $\theta \in \overline{\Theta}$ and $\theta \neq \theta_0$.

C2.4: $K(\theta)$ is a continuous matrix function of $\theta \in \Theta_K$ where

$$\Theta_K = \{\theta \in \overline{\Theta} : \|K(\theta)\| < \infty\} .$$

C2.5: The interior of Θ is Θ_0 where

$$\Theta_0 = \{\theta \in \Theta : \det g(e^{i\omega}; \theta) \neq 0,\ \omega \in [-\pi, \pi]\} .$$

C2.6: $\inf_{\theta \in \Theta_0} L(\theta) = L(\theta_0) = \log \det K(\theta_0) + s$ where

$$L(\theta) = \log \det K(\theta) + Q(\theta) ,$$

$$Q(\theta) = (2\pi)^{-1} \int_{-\pi}^{\pi} \operatorname{tr}\left(f^{-1}(\omega; \theta) f(\omega; \theta_0)\right) d\omega .$$

Before proceeding to the proof of the SLLN some comments on the above conditions are made.

a(i) Clearly $\Theta_0 \subseteq \Theta \subseteq \Theta_K \subseteq \overline{\Theta}$.

a(ii) No assumptions are made concerning the innovation vectors $\varepsilon(n)$ other than that they are ergodic, serially uncorrelated and have covariance matrix with finite norm and nonzero determinant.

a(iii) The specification that $\theta_0 \in \Theta$ is quite general in that it does not restrict the poles of $k(e^{i\omega}; \theta)$ to be bounded away from the unit circle and $\det k(e^{i\omega}; \theta)$ could be zero at certain values of ω. To illustrate consider the simple ARMA model with $s = p = q = 1$

given by

$$x(n) + \beta x(n-1) = \varepsilon(n) + \alpha\varepsilon(n-1) .$$

Here β does not have to be restricted *a priori* to be such that $|\beta| \le 1 - \delta$, $\delta > 0$, and we may allow any α such that $|\alpha| \le 1$.

a(iv) Since Θ is not necessarily a closed set the point at which $\hat{L}_N$ (say) is minimised may not lie in Θ but $\overline{\Theta}\backslash\Theta$ for N not large. In this regard the continuity conditions on h and g with respect to $\theta \in \overline{\Theta}$ are less restrictive than would be the requirement that k be continuous in $\theta \in \overline{\Theta}$.

a(v) A condition of the type given in C2.3 is required in order that there be a unique minimum in $L(\theta)$ which we will show to be the asymptotic limit of $\hat{L}_N$, $\overline{L}_N$, $\tilde{L}_N$. That is since $L(\theta)$ depends only on the spectral density matrices corresponding to θ and θ_0 this limit must possess a unique minimum with respect to $f(\omega;\, \theta)$ as θ ranges over $\overline{\Theta}$ in order that the asymptotic theory be meaningful. As we shall see in the proofs C2.3 guarantees this.

a(vi) Closely related to (v) is the choice of a suitable topology for Θ. We have not described any particular topology for Θ. Such a description relates to the parameterization of the functions h, g and K as well as to the condition C2.3. In general one might proceed by initially taking θ to be a u-vector of real numbers lying in some subset B of R^u. If there is not a one-to-one correspondence between the points in B and the different spectral density matrices f obtained at these points, the points in B that yield the same f would be grouped together into equivalence classes. This is a sensible way to proceed when parameter estimation is based only on second order quantities and thus only on f. A suitable topology for this derived space of equivalence classes would be the quotient (or identification) topology. In fact this is the way we have proceeded in the ARMA case discussed in the next chapter. Note however that convergence of θ_n to θ_0 in the quotient topology does not in general imply the convergence in the Euclidean topology of individual elements in the basic space $\mathcal{B}$. This relates to the ideas of identification again discussed fully in the next chapter for the ARMA case.

a(vii) Since θ_0 is not restricted to lie in Θ_0 (where $\det g(z;\theta)$ has no zeros on the unit circle) we require a condition such as C2.6. In the ARMA case this condition is fulfilled and the proof of this is given in Chapter 3.

We consider now the proof of the strong consistency of the estimators. The following theorem concerns $\overline{\theta}_N$, $\tilde{\theta}_N$ while the case $\hat{\theta}_N$ is dealt with later in this section.

THEOREM 2.1. *Let $x(n)$ be a stationary, ergodic purely non-deterministic vector process with representation* (1.1) *and spectrum $f(\omega;\theta_0)$ where $\theta_0 \in \Theta$. Assume also that conditions* C2 *hold. Then the estimators $\tilde{\theta}_N$ and $\overline{\theta}_N$ minimising $\tilde{L}_N$ and $\overline{L}_N$ (respectively) over Θ converge a.s. to θ_0.*

The proof of this result is lengthy so we proceed in steps with the use of some lemmas. Before considering these however some preliminary notation and often used results will be introduced.

b(i). Since $\overline{\theta}_N$, $\tilde{\theta}_N$, $\hat{\theta}_N$ or θ_0 are not necessarily restricted to Θ_0, we will require in our proofs, a device to avoid any difficulties associated with $\det g(z;\theta)$ having a zero on the unit circle. In this regard it is convenient to introduce, for any $\eta > 0$,

$$\Phi_\eta(\omega;\theta) = 2\pi\left(\frac{l(\omega;\theta)^* K^{-1}(\theta) l(\omega;\theta)}{d(\omega;\theta)+\eta}\right) \tag{2.2}$$

where $l = Gh$, G is the adjoint of g , and $d = |\det g|^2$. Both l and d are continuous in $(\omega;\ \theta)\varepsilon[-\pi,\ \pi] \times \overline{\Theta}$, so that in particular, $d(\omega;\ \theta)$ is bounded uniformly above by γ (say) which is finite. When f^{-1} has been replaced by Φ_η we refer to $\hat{Q}_N$ as $\hat{Q}_{N,\eta}$ and similarly to $\overline{Q}_{N,\eta}$, $\tilde{Q}_{N,\eta}$ for the other estimation procedures. When f^{-1} is replaced by Φ_η in $Q(\theta)$ (see C2.6) the result will be referred to as $Q_\eta(\theta)$. The definition of Φ_η as $\left(|k(e^{i\omega};\ \theta)|^2+\eta\right)^{-1}$, when $s = 1$, given in Hannan (1973a) (and subsequently corrected to the form (2.2)) will not have the properties we require since k is not necessarily bounded on $\overline{\Theta}$.

b(ii). It will be convenient to define, for any $0 < \delta_1 < \delta_2 < \infty$ the following subset of $\overline{\Theta}$;

$$\Theta_{\delta_1,\delta_2} = \left\{\theta \in \overline{\Theta} : \delta_1 \le \lambda_1\left(K(\theta)\right),\ \|K(\theta)\| \le \delta_2\right\}$$

where $\lambda_1(A)$ denotes the smallest eigenvalue of A . (We will sometimes use $\lambda_s(A)$ to denote the largest eigenvalue.) Then for any such δ_1, δ_2 , $\Phi_\eta(\omega;\ \theta)$ is a uniformly continuous matrix function of $(\omega;\theta) \in [-\pi,\pi] \times \Theta_{\delta_1,\delta_2}$. Our proof will show that our sequence of estimators will eventually enter such a set.

b(iii). A device frequently employed in the proof is to replace f (f^{-1} or Φ_η as the case may be) by the (matrix) Cesaro sum of its Fourier series to M terms. Let $g(\omega)$ be a continuous matrix function of $\omega \in [-\pi,\ \pi]$ with $g(-\pi) = g(\pi)$ and let $P_M(\omega)$ be the Cesaro sum to M terms of the Fourier series for $g(\omega)$. Then in almost the same way as given in Hannan (1970, pp. 506-507) it follows that $P_M(\omega)$ converges uniformly to $g(\omega)$ for $\omega \in [-\pi,\ \pi]$. That is given $\varepsilon > 0$ an M , finite, exists such that $\sup_{\omega\in[-\pi,\pi]} \|P_M(\omega)-g(\omega)\| \le \varepsilon$. When g is a uniformly continuous function of a parameter θ in a space $\overline{\Theta}$ in addition to ω, it follows that also for $\varepsilon > 0$,

$$\sup_{\theta\in\overline{\Theta}} \sup_{\omega\in[-\pi,\pi]} \|P_M(\omega;\ \theta)-g(\omega;\ \theta)\| \leq \varepsilon \ ;$$

that is M can be chosen independently of $\theta \in \overline{\Theta}$.

LEMMA 2.2. *(a) For all* $\theta \in \Theta_0$, $\overline{Q}_N(\theta) \xrightarrow{\text{a.s.}} Q(\theta)$ *and* $\tilde{Q}_N(\theta) \xrightarrow{\text{a.s.}} Q(\theta)$.

(b) For all $\eta > 0$, $0 < \delta_1 < \delta_2 < \infty$, $\overline{Q}_{N,\eta}(\theta) \xrightarrow{\text{a.s.}} Q_\eta(\theta)$ *and* $\tilde{Q}_{N,\eta}(\theta) \xrightarrow{\text{a.s.}} Q_\eta(\theta)$, *both uniformly in* $\theta \in \Theta_{\delta_1,\delta_2}$.

Proof. Consider the proof of *(a)* for $\tilde{Q}_N$. This proof is almost exactly the same as that given for the case $s = 1$ in Hannan (1973a). Since $f(\omega;\ \theta)$ is continuous in ω and $\det f(\omega;\ \theta) > 0$, $\omega \in [-\pi, \pi]$ whenever $\theta \in \Theta_0$, $f^{-1}(\omega;\ \theta)$ exists and is also continuous in $\omega \in [-\pi, \pi]$. Then if P_M is the Cesaro sum to M terms of the Fourier series $P(l)$ for f^{-1} it follows that for any $\varepsilon > 0$, there is an $M < \infty$ such that

$$\sup_{\omega\in[-\pi,\pi]} \left\|f^{-1}-P_M\right\| \leq \varepsilon \ .$$

Writing $\tilde{Q}_N(P_M)$ for $\tilde{Q}_N(\theta)$ with f^{-1} replaced by P_M it follows that

$$|\tilde{Q}_N(P_M)-\tilde{Q}_N(\theta)| \leq (N')^{-1} \sum_t \left\|P_M-f^{-1}\right\|\left(\operatorname{tr} I(\omega_t)\right) \leq \varepsilon.\|G(0)\| \ .$$

But by ergodicity $G(0) \xrightarrow{\text{a.s.}} \Gamma(0)$ which has bounded norm so that for N sufficiently large

$$|\tilde{Q}_N(P_M)-\tilde{Q}_N(\theta)| \leq \varepsilon.b \ , \quad b < \infty \ . \tag{2.3}$$

Thus consideration of $\tilde{Q}_N(\theta)$ reduces to consideration of

$$\tilde{Q}_N(P_M) = \text{tr} \sum_{l=-M}^{M} \left(1 - \frac{|l|}{M}\right) P(l) \left\{ (N')^{-1} \sum_t I(\omega_t) e^{-il\omega_t} \right\}$$

$$= \text{tr} \sum_{l=-M}^{M} \left(1 - \frac{|l|}{M}\right) P(l) \{G(l) + E(l)\} \tag{2.4}$$

where $E(-l)' = E(l) = N^{-1} \sum_{1}^{N-N'-l} x(N'+n+l)x(n)'$, if $-M \le l \le 0$, and is zero otherwise. But $E(l)$ consists of at most M terms of the form $N^{-1}x(N'+n+l)x(n)'$ which, by ergodicity, converge to zero a.s. Hence

$$\tilde{Q}_N(P_M) \xrightarrow{\text{a.s.}} \text{tr}\left[(2\pi)^{-1} \int P_M(\omega) f(\omega;\ \theta_0) d\omega\right] \quad \text{as} \quad N \to \infty .$$

However this limit can be taken to be arbitrarily close to $Q(\theta)$ by a similar argument to that leading to (2.3) above. In the proof for $\overline{Q}_N$ the terms $E(l)$ do not appear in (2.4); it is thus slightly easier. The proof of *(b)* differs only in the replacement of Φ_η by P_M where now for any given approximation error $\varepsilon > 0$, $M < \infty$ can be chosen independently of $\theta \in \Theta_{\delta_1, \delta_2}$. □

LEMMA 2.3. *There exists a* $\delta > 0$ *and an* $N_0 < \infty$ *, both independent of* $\theta \in \overline{\Theta}$ *such that for all* $N \ge N_0$ *and all* $\theta \in \overline{\Theta}$,

$$\inf_{\alpha'\alpha=1} \alpha' \left\{ (2\pi)^{-1} \int_{-\pi}^{\pi} k^{-1}(e^{i\omega};\ \theta) I(\omega) k^{-1}(e^{i\omega}; \theta)^* d\omega \right\} \alpha \ge \delta .$$

The same result holds for

$$(N')^{-1} \sum_t k^{-1}(e^{i\omega_t};\ \theta) I(\omega_t) k^{-1}(e^{i\omega_t};\ \theta)^* .$$

Proof. For any $\alpha'\alpha = 1$ it follows that

$$\alpha' \left(\frac{1}{2\pi} \int k^{-1}(\theta) I(\omega) k^{-1}(\theta)^* d\omega \right) \alpha \ge \frac{1}{\gamma} \text{tr} \left\{ \frac{1}{2\pi} \int [l(\omega;\ \theta)^* \alpha\alpha' l(\omega;\ \theta)] I(\omega) d\omega \right\}$$

where γ and l are as described in note (i) above. Since the matrix multiplying $I(\omega)$ on the right of this expression is uniformly continuous in $(\omega;\ \theta)$ it follows, as in Lemma 2.2 *(b)*, that this expression converges a.s. to $\frac{1}{\gamma} \cdot \alpha'\left\{\frac{1}{2\pi}\int l(\omega;\ \theta)f_0(\omega)l(\omega;\ \theta)^* d\omega\right\}\alpha$ uniformly in $\theta \in \overline{\Theta}$. However lf_0l^* is the spectrum of a stationary process with prediction covariance matrix K_0 (since $l(0;\ \theta) \equiv I_s$), and if $y(n)$ is a vector process with this spectrum and P is an orthogonal matrix with α as first row then the first element of $Py(n)$ has spectrum $\alpha' lf_0l^*\alpha$. Hence $(2\pi)^{-1}\int \alpha' lf_0l^*\alpha d\omega$ is the variance of a scalar stationary process with prediction variance which is larger than the first element of the prediction covariance matrix for $Py(n)$. This prediction matrix is PK_0P' with first element $\alpha' K_0\alpha \geq \lambda_1(K_0) > 0$, $\forall \alpha'\alpha = 1$. Hence

$$(2\pi)^{-1}\int \alpha' l(\theta)f_0 l(\theta)^*\alpha d\omega \geq \lambda_1(K_0) \qquad \forall\theta \in \overline{\Theta}.$$

Therefore

$$\alpha'\left(\frac{1}{2\pi}\int k^{-1}(\theta)I(\omega)k^{-1}(\theta)^* d\omega\right) \geq \frac{1}{\gamma}\left(\lambda_1(K_0)+\varepsilon_N\right)$$

where ε_N converges to zero uniformly in $\overline{\Theta}$. From this it follows that one can find a $\delta > 0$ and $N_0 < \infty$ such that the lemma holds. The proof for the summation (over t) version of the lemma follows in the same way.

Proof of Theorem 2.1. For $\overline{\theta}_N$ we have, using part *(a)* of Lemma 2.2 and C2.6, the result

$$\overline{\lim_{N\to\infty}}\, \overline{L}_N(\overline{\theta}_N) \leq \inf_{\theta\in\Theta_0} \overline{\lim_{N\to\infty}}\, \overline{L}_N(\theta) = \inf_{\theta\in\Theta_0} L(\theta) = \log\det K(\theta_0) + s < \infty. \qquad (2.5)$$

From this it follows that, for N sufficiently large, $\det K(\overline{\theta}_N)$ is uniformly bounded (since $\overline{Q}_N(\theta)$ in (1.8) is nonnegative) and in order to show that $\overline{\theta}_N$ eventually enters $\Theta_{\delta_1,\delta_2}$ for some $0 < \delta_1 < \delta_2 < \infty$ we require only to show that $\lambda_1(K(\overline{\theta}_N))$ remains bounded away from zero for N sufficiently large. For this consider

$$\bar{L}_N(\bar{\theta}_N) \geq s \log \lambda_1(K(\bar{\theta}_N)) + \lambda_1^{-1}(K(\bar{\theta}_N)).\lambda_1\left[\frac{1}{2\pi}\int k^{-1}(\bar{\theta}_N)I(\omega)k^{-1}(\bar{\theta}_N)^*d\omega\right]$$

where we have used the fact that if A, B are positive definite matrices then $\text{tr}(AB) \geq \lambda_s(A).\lambda_1(B)$ and $\lambda_s(A) = \lambda_1^{-1}(A^{-1})$. But

$$\lambda_1\left[\frac{1}{2\pi}\int k^{-1}(\theta)I(\omega)k^{-1}(\theta)^*d\omega\right] = \inf_{\alpha'\alpha=1} \alpha'\left[\frac{1}{2\pi}\int k^{-1}(\theta)I(\omega)k^{-1}(\theta)^*d\omega\right]\alpha$$

$$\geq \delta\ , \quad \delta > 0\ , \text{ all } \ \theta \in \bar{\Theta}$$

the inequality following from Lemma 2.3. Hence

$$\bar{L}_N(\bar{\theta}_N) \geq s \log \lambda_1(K(\bar{\theta}_N)) + \lambda_1^{-1}(K(\bar{\theta}_N)).\delta\ , \tag{2.6}$$

so that if $\lambda_1(K(\bar{\theta}_N))$ is not bounded away from zero for all sufficiently large N the r.h.s. of (2.6) will not be bounded above contradicting (2.5). Hence it may be assumed that if $\bar{\theta}_N$ does not converge to θ_0 there is a subsequence $\bar{\theta}_{N(M)}$ converging to $\theta' \in \Theta_{\delta_1,\delta_2}$ where for M large $\bar{\theta}_{N(M)} \in \Theta_{\delta_1,\delta_2}$ for some $0 < \delta_1 < \delta_2 < \infty$. Now

$$\varliminf_{M\to\infty} \bar{L}_{N(M)}(\bar{\theta}_{N(M)})$$

$$\geq \sup_{\eta>0} \varliminf_{M\to\infty} \{\log\det K(\bar{\theta}_{N(M)}) + \bar{Q}_{N,\eta}(\bar{\theta}_{N(M)})\}$$

$$= \log\det K(\theta') + \sup_{\eta>0} Q_\eta(\theta') \qquad \text{(by Lemma 2.2, part } (b))$$

$$= \log\det K(\theta') + (2\pi)^{-1}\int \text{tr}\left\{f^{-1}(\omega;\ \theta')f(\omega;\ \theta_0)\right\}d\omega$$

$$= \log\det K(\theta') + \text{tr}\left\{K(\theta')^{-\frac{1}{2}}(2\pi)^{-1}\int k(\theta')^{-1}k(\theta_0)K(\theta_0)k(\theta_0)^*k(\theta')^{-1*}d\omega.\right.$$

$$\left. .K(\theta')^{-\frac{1}{2}}\right\}\ . \tag{2.7}$$

If the second term in this last line is not finite then a contradiction ensues immediately. If it is finite then $k(\theta')^{-1}k(\theta_0)K(\theta_0)k(\theta_0)^*k\ (\theta')^{-1*}$ may be taken as the spectral density matrix of a stationary vector process.

Writing $k^{-1}\left(e^{i\omega};\ \theta'\right)k\left(e^{i\omega};\ \theta_0\right) = \sum_{j=0}^{\infty} D(j)e^{ij\omega}$, where $D(0) = I_s$, it follows that

$$\log\det K(\theta') + \mathrm{tr}\left\{(2\pi)^{-1}\int f^{-1}(\omega;\ \theta')f\left(\omega;\ \theta_0\right)d\omega\right\}$$

$$= \log\det K(\theta') + \mathrm{tr}\left\{K(\theta')^{-\frac{1}{2}}K\left(\theta_0\right)K(\theta')^{-\frac{1}{2}}\right\}$$

$$+ \mathrm{tr}\left\{K(\theta')^{-\frac{1}{2}}\sum_{j=1}^{\infty} D(j)K_0D(j)'K(\theta')^{-\frac{1}{2}}\right\} . \quad (2.8)$$

But the third term on the r.h.s. of (2.8) is non-negative since $K(\theta')^{-\frac{1}{2}}\sum_{j=1}^{\infty} D(j)K_0D(j)'K(\theta')^{-\frac{1}{2}}$ is non-negative definite. Let $\lambda_1, \ldots, \lambda_s$ denote the s eigenvalues of $K(\theta')^{-\frac{1}{2}}K\left(\theta_0\right)K(\theta')^{-\frac{1}{2}}$. Then

$$\mathrm{tr}\left\{K(\theta')^{-\frac{1}{2}}K\left(\theta_0\right)K(\theta')^{-\frac{1}{2}}\right\} = \sum_{j=1}^{s} \lambda_j$$

$$\geq s\left(\prod_{j=1}^{s} \lambda_j\right)^{1/s}$$ (using the arithmetic-geometric mean inequality)

$$= s\left(\det K\left(\theta_0\right)/\det K(\theta')\right)^{1/s} . \quad (2.9)$$

Using (2.9) in (2.8) gives

$$\log\det K(\theta') + \mathrm{tr}\left\{(2\pi)^{-1}\int f^{-1}(\omega;\ \theta')f\left(\omega;\ \theta_0\right)d\omega\right\}$$

$$\geq \log\det K(\theta') + s\left(\det K\left(\theta_0\right)/\det K(\theta')\right)^{1/s}$$

$$\geq \log\det K\left(\theta_0\right) + s \quad (2.10)$$

with equality in the last line only if $\det K(\theta') = \det K\left(\theta_0\right)$. However if this is the case and equality holds throughout (2.10), and thus (2.9), then

$$\mathrm{tr}\left\{K(\theta')^{-\frac{1}{2}}\sum_{j=0}^{\infty} D(j)K_0D(j)'K(\theta')^{-\frac{1}{2}}\right\} = s \quad (2.11)$$

and

$$\mathrm{tr}\left\{K(\theta')^{-\frac{1}{2}}K(\theta_0)K(\theta')^{-\frac{1}{2}}\right\} = s\left(\det K_0/\det K(\theta')\right)^{1/s} = s \ . \qquad (2.12)$$

But (2.12) implies that all eigenvalues of $K^{-\frac{1}{2}}(\theta')K(\theta_0)K^{-\frac{1}{2}}(\theta')$ are equal to unity and thus that $K(\theta') = K(\theta_0)$, which when used in (2.11) implies that $D(j) = 0$, $j \geq 1$. Hence $k^{-1}(e^{i\omega};\ \theta')k\left(e^{i\omega};\ \theta_0\right) = I_s$ a.e. $(d\omega)$ which contradicts C2.3. Hence $\theta' = \theta_0$ and the proof for $\bar{\theta}_N$ is complete. The proof of the theorem for $\tilde{\theta}_N$ is almost the same by using the appropriate parts of Lemmas 2.2 and 2.3. □

The SLLN for the estimator $\hat{\theta}_N$ obtained by minimisation of $\hat{L}_N(\theta)$ will now be discussed. In order to prove this result a further condition has so far been unavoidable. Recall that in the proof for $\bar{\theta}_N$ (say) it was shown that $\bar{\theta}_N$ eventually entered the set $\Theta_{\delta_1,\delta_2}$, for some $\delta_1 > 0$, $\delta_2 < \infty$. However in the present case such a proof is not available. The approach taken therefore is to assume that the true value θ_0 lies in a subset of our previous Θ defined by the additional requirement that $\lambda_1(K(\theta)) \geq \lambda > 0$ where λ is known. Thus the minimisation of $\hat{L}_N(\theta)$ is carried out subject to this requirement. We will refer to $\bar{\Theta}^\lambda$, Θ_K^λ, Θ^λ, $\Theta^\lambda_{\delta_1,\delta_2}$, Θ^λ_0 as our previous $\bar{\Theta}$, Θ_K, Θ, $\Theta_{\delta_1,\delta_2}$, Θ_0 subject to this requirement.

COROLLARY 2.4. *Under the conditions of Theorem* 2.1, *and the additional condition that* $\theta_0 \in \Theta^\lambda$ *where* $\lambda > 0$ *is known,* $\hat{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$.

Again we proceed to prove this result by way of some lemmas. These do not necessarily have to be considered for the new sets $\bar{\Theta}^\lambda$, etc. but will apply more generally as the statement of the lemmas will show. Before stating the first lemma we mention a fact that is used frequently. If f and g are $s \times s$ non-negative definite matrix functions of ω with $f(\omega) \leq g(\omega)$, $\omega \in [-\pi, \pi]$ in the usual ordering of non-negative definite matrices then $\Gamma_N(f) \leq \Gamma_N(g)$ for all N again the ordering being as for non-negative definite matrices, where $\Gamma_N(f)$, $\Gamma_N(g)$ are defined by (1.5).

LEMMA 2.5. *(a) For* $\theta \in \Theta_0$, $\hat{Q}_N(\theta) \xrightarrow{\text{a.s.}} Q(\theta)$.

(b) For any $\eta > 0$, $0 < \delta_1 < \delta_2 < \infty$, $\hat{Q}_{N,\eta}(\theta) \xrightarrow{\text{a.s.}} Q_\eta(\theta)$ *where the convergence is uniform in* $\theta \in \Theta_{\delta_1,\delta_2}$.

Proof. Consider the proof of *(a)*. The proof of this lemma also resembles that of the corresponding result for $s = 1$ given in Hannan (1973a). We again approximate f^{-1} by a matrix trigonometric polynomial P , of degree M , so that $\varepsilon I_s \geq P - f^{-1} \geq 0$. Thus $0 \leq \Gamma_N^{-1}(P^{-1}) - \Gamma_N^{-1}(\theta) \leq C.\varepsilon.I_{Ns}$ where C is a finite constant. Now writing $P = \frac{1}{2\pi} B(e^{i\omega})\, K_P^{-1} B(e^{i\omega})^*$ where $B(e^{i\omega}) = \sum_{l=0}^{M} B(l)e^{il\omega}$, $B(0) = I_s$, we may define an $Ns \times Ns$ matrix A as

$$A = \begin{bmatrix} I_{Ms} & \vdots & 0 & \cdots & 0 \\ \cdots & \cdots & \ddots & & \vdots \\ B(M) & \cdots & B(0) & \ddots & \vdots \\ 0 & \ddots & & \ddots & 0 \\ 0 & \cdots\; 0 & B(M) & \cdots & B(0) \end{bmatrix}$$

then

$$\Gamma_N^{-1}(P^{-1}) = A' \begin{bmatrix} \Gamma_M^{-1}(P^{-1}) & \vdots & & \\ \cdots & \cdots & & \\ & & K_P^{-1} & & \\ & & & \ddots & \\ & & & & K_P^{-1} \end{bmatrix} A \qquad (2.13)$$

so that

$$\hat{Q}_N(P^{-1}) = N^{-1}x_M'\Gamma_M^{-1}(P^{-1})x_M + N^{-1}\sum_{M+1}^{N} e(n)'K_P^{-1}e(n) \tag{2.14}$$

where $e(n) = \sum_{j=0}^{M} B(j)x(n-j)$ for $n \geq M+1$ and for $n \leq M$ we calculate $e(n)$ by extending $x(n)$ cyclically after an interval of $N' - N$ zeros. Then

$$\tilde{Q}_N(P^{-1}) = N^{-1} \sum_{N-N'+1}^{N} e(n)'K_P^{-1}e(n)$$

which differs from $\hat{Q}_N(P^{-1})$ by a fixed number of terms, which, by ergodicity, converge to zero as N tends to ∞. Since

$$\hat{Q}_N(P^{-1}) \geq \hat{Q}_N(\theta) \geq \hat{Q}_N(P^{-1}) - \varepsilon.C.N^{-1}x_N'I_{Ns}x_N ,$$

$\hat{Q}_N(p^{-1})$ converges to $Q(P)$, $Q(P)$ differs from $Q(\theta)$ only in the replacement of f^{-1} by P, and ε is arbitrary, the result follows. The corresponding result for $\hat{Q}_{N,\eta}$ follows in the same way, the approximation by P now being uniform for $\theta \in \Theta_{\delta_1,\delta_2}$. □

LEMMA 2.6. *(a) For all* $\theta \in \Theta$, $N^{-1} \log \det\{\Gamma_N(\theta)\} \geq \log \det\{K(\theta)\}$.

(b) If $\theta \in \Theta_0$, $\lim_{N\to\infty} N^{-1} \log \det\{\Gamma_N(\theta)\} = \log \det\{K(\theta)\}$.

Proof. For $\theta \in \Theta$ let $P(\omega)$ be a matrix of trigonometric polynomials such that $(f+\zeta I_s)^{-1} \leq P \leq (f+\zeta I_s)^{-1} + \varepsilon I_s$, $\varepsilon > 0$, $\zeta > 0$. Then P^{-1} may be taken as the spectrum of an autoregression

$$\sum_0^M B(j)y(n-j) = \xi(n) , \quad E(\xi(n)\xi(n)') = \delta_{mn}K_P .$$

Let $\hat{y}(j)$ be the best linear predictor of $y(j)$ using observations

$y(0), \ldots, y(j-1)$ and put $\hat{\xi}(j) = y(j) - \hat{y}(j)$. Then $\hat{\xi}(j) = \xi(j)$ for $j > M$ and $E\left(\hat{\xi}(j)\hat{\xi}(k)'\right)$ is null for $j \neq k$. Now

$$y' = \left(y(1)' \vdots y(2)' \vdots \ldots \vdots y(N)'\right)$$

$$= \left(y(1)' \vdots \hat{\xi}(2)' \vdots \ldots \vdots \hat{\xi}(N)'\right)A'^{-1} = \hat{y}'A'^{-1}$$

where A is lower triangular with $B(0) = I_s$ along the diagonal blocks. Taking expectations in $Ayy'A' = \hat{y}\hat{y}'$ we obtain

$$\det\left\{\Gamma_N\left(P^{-1}\right)\right\} = \det\{\Gamma_P(0)\}\det\{\hat{\Gamma}_2(0)\} \ldots \det\{\hat{\Gamma}_M(0)\}\left[\det\{K_P\}\right]^{N-M}, \quad N > M$$

$$= \det\{\Gamma_P(0)\}\det\{\hat{\Gamma}_2(0)\} \ldots \det\{\hat{\Gamma}_N(0)\}, \quad N \leq M,$$

where $K_P \leq \hat{\Gamma}_j(0) = E\left(\hat{\xi}(j)\hat{\xi}(j)'\right) \leq \Gamma_P(0)$. Thus

$$N^{-1} \log \det\{\Gamma_N(\theta)+\zeta I_{Ns}\} \geq N^{-1} \log \det\left\{\Gamma_N\left(P^{-1}\right)\right\}$$

$$\geq \log \det\{K_P\} \geq \log\left(\det\{K_\zeta\}.[1+\varepsilon(C+\zeta)]^{-s}\right)$$

where C is an upper bound for the eigenvalues of $f(\omega)$ and K_ζ corresponds to $\left(f+\zeta I_s\right)$. The last term is not less than $\log\left(\det K(\theta).[1+\varepsilon(C+\zeta)]^{-s}\right)$ and since ε, ζ are arbitrary we obtain the first part of the lemma. For $\theta \in \Theta_0$ we may let $\zeta = 0$ and moreover approximate f^{-1} both above and below by matrices of trigonometric polynomials that are positive definite and arbitrarily close. The same type of argument as that just used gives the second part of the lemma. □

Proof of Corollary 2.4. By Lemma 2.5 *(a)*, Lemma 2.6 *(b)* and C2.6 it follows that

$$\overline{\lim_{N\to\infty}} \hat{L}_N\left(\hat{\theta}_N\right) \leq \log \det K\left(\theta_0\right) + s. \tag{2.15}$$

In case $\hat{\theta}_{N(M)} \in \overline{\Theta}^\lambda \backslash \Theta^\lambda$ where $\hat{\theta}_{N(M)} \to \theta'$ we let $\theta_{N(M)}$ be a sequence of points in Θ^λ for which $\theta_{N(M)} \to \theta'$ and

$$|\inf_{\theta\in\Theta^\lambda} \hat{L}_{N(M)}(\theta)-\hat{L}_{N(M)}(\theta_{N(M)})| \le \varepsilon_{N(M)} ,$$

$\varepsilon_{N(M)} \to 0$ as $M \to \infty$. Then by Lemma 2.6 *(a)*,

$$\underline{\lim}_{M\to\infty} \hat{L}_{N(M)}(\hat{\theta}_{N(M)}) = \underline{\lim}_{M\to\infty} \hat{L}_{N(M)}(\theta_{N(M)}) \ge \underline{\lim}_{M\to\infty} \log\det K(\theta_{N(M)}) . \qquad (2.16)$$

But $\lambda_1(K(\theta)) \ge \lambda > 0$, $\forall\theta \in \overline{\Theta}^\lambda$ so that, combining (2.15) and (2.16), $\lambda_s(K(\theta_{N(M)}))$ must be uniformly bounded for M sufficiently large. Thus $\|K(\theta_{N(M)})\|$ remains so and the sequence $\theta_{N(M)}$ eventually enters $\Theta^\lambda_{\delta_1,\delta_2}$ for $\delta_1 \ge \lambda$ and $\delta_2 < \infty$. Applying Lemma 2.5 *(b)* reduces the proof to that for Theorem 2.1. □

Before closing this section some final remarks are relevant.

C(i) We could consider replacing $N^{-1}\log\det\Gamma_N(\theta)$ in $\hat{L}_N(\theta)$ by $\log\det K(\theta)$ to give a new criterion

$$\hat{\hat{L}}_N(\theta) = \log\det K(\theta) + N^{-1}x_N'\Gamma_N^{-1}(\theta)x_N .$$

Corollary 2.4 will continue to hold in this case. However this replacement does not allow the extra condition that $\theta_0 \in \Theta^\lambda$ to be dropped. The difficulty is to obtain a lemma like Lemma 2.3 for $\hat{Q}_N(\theta)$. However one alternative specification to the requirement that $\lambda_1(K(\theta_0)) \ge \lambda > 0$ is to require θ_0 to be such that $\left|\det h\left(e^{i\omega};\, \theta_0\right)\right| \ge \delta > 0$ and this δ is known. Then it is easy to show that

$$\frac{1}{N}\, x_N'\Gamma_N^{-1}(\theta)x_N \ge b^{-1}\delta^2\, \mathrm{tr}\left(K^{-1}(\theta)G(0)\right) , \text{ all } \theta \in \Theta^\delta$$

where $b < \infty$ is derived from the boundedness of $\|g\|$ and $\|\mathrm{adj}\, h\|$ on $\overline{\Theta}$, and Θ^δ is Θ subject to the requirement concerning $|\det h|$. Then Corollary 2.4 holds, by using the same proof as for

Theorem 2.1, and Lemma 2.5 *(a)* with the set Θ^{λ} being replaced by Θ^{δ} . Other specifications can be given for special cases. However as will be shown in the next chapter none of these additional conditions are required to prove that $\hat{\theta}_N \to \theta_0$ in the ARMA case.

C(ii) Kang (1973) has indicated that the above replacement (i.e.: the use of $\hat{\hat{L}}_N$ instead of $\hat{L}_N$) can adversely effect the estimation procedure for N not large in the sense that the inclusion of $N^{-1} \log \det \Gamma_N(\theta)$ may help to keep the estimates away from the boundary of the region under consideration. To illustrate consider the ARMA model with $p = q = s = 1$,

$$x(n) + \beta x(n-1) = \varepsilon(n) + \alpha\varepsilon(n-1) \ , \quad E\varepsilon(n)^2 = \sigma^2 \ . \tag{2.17}$$

Here, using results of Grenander and Szegö (1958, pp. 77-79) it follows that

$$N^{-1} \log \det \Gamma_N(\theta) = \log \sigma^2 + N^{-1} \log\left\{\frac{(1-\alpha\beta)^2}{(1-\alpha^2)(1-\beta^2)} - \frac{\alpha^{2N}(\alpha-\beta)^2}{(1-\alpha^2)(1-\beta^2)}\right\} . \tag{2.18}$$

The inclusion of the second term in this (which has been dropped from $\hat{\hat{L}}_N(\theta)$) would have the tendency to keep $|\alpha|$ and $|\beta|$ away from the value 1 . In this special case a better approximation to $N^{-1} \log \det \Gamma_N(\theta)$ than $\log \det \sigma^2$ may be obtained by dropping the second term in braces in (2.18). To obtain (2.18) for general (p, q) in the ARMA model is difficult however, since such an expression will not always be directly derivable in terms of the coefficients $\alpha(j)$, $\beta(j)$.

An alternative to $\tilde{L}_N$ which is sometimes considered is to replace $\log \det K(\theta)$ by

$$(N')^{-1} \sum_t \log \det 2\pi f(\omega_t; \theta) \ . \tag{2.19}$$

In the case where $q = 0$ in (2.17) (i.e. $x(n)$ is a first order moving average) then this term is evaluated as

$$\log \sigma^2 + 2(N')^{-1} \log\left(1-(-\alpha)^{N'}\right)$$

if $|\alpha| \neq 1$. The effect of the second term in this on keeping $|\alpha|$ away from unity depends on whether N' is even or odd. For example if N' is even the second term tends to $-\infty$ as $|\alpha|$ tends to 1 and would thus be detrimental to the estimation procedure.

C(iii) If K is parameterised separately from θ then some conditions on the parameters μ specifying K may be weakened. For example if μ consists of the $s(s+1)/2$ freely varying elements of K then in Theorem 2.1 it would not be required that μ lie in a compact subset of $\mathbb{R}^{s(s+1)/2}$ as the proof of that theorem shows. In general special facets of the parameterisation of any special model under consideration will have to be used in order to determine the minimal conditions required for Theorem 2.1 and Corollary 2.4 to hold. However it is felt that the methods used to prove corresponding results for special cases may not depart substantially from those used in the above proofs.

C(iv) It is of interest to compare the above treatment with previous consistency results for the case $s = 1$. In the remainder of this note we will refer to K as σ^2 since the latter is a more commonly used notation when $s = 1$. The only significant difference in the treatment here and that of Hannan (1973a) is that σ^2 was taken to depend on θ also in the above. Walker (1964) treats the case where $x(n)$ is a weakly stationary non-deterministic process of the type (1.1) where θ specifies only $C(j)$ (and not σ^2) and θ is estimated via what we have called $\overline{L}_N$. (The cases $\tilde{L}_N$ and $\hat{L}_N$, the correct likelihood when the data are Gaussian, are not treated.) Walker establishes the convergence in probability of $\overline{\theta}_N$ to θ_0 and $\overline{\sigma}_N^2$ to σ_0^2 under stronger conditions than those imposed in C2 above. The main additional conditions required in Walker's treatment are as follows:

1. The fourth moment of $\varepsilon(m)$ is finite;
2. $\theta_0 \in \Theta$ where Θ is a bounded closed set contained in an open set S in R^u ;
3. $k(e^{i\omega}; \theta)$ is a continuous function of $\omega \in [-\pi, \pi]$ whenever $\theta \in S$;
4. $k^{-1}(e^{i\omega}; \theta)$ is continuous in $\omega \in [-\pi, \pi]$. The derivatives $\partial k^{-1}(e^{i\omega}; \theta)/\partial\theta_j$ are continuous in $(\omega; \theta) \in [-\pi, \pi] \times S$.

The conditions 2,3,4 are much stronger than those required in C2.1, C2.2, C2.4 and C2.5. For example 2, 3, 4 mean in relation to the ARMA model (2.17) $(p = q = s = 1)$ that $|\alpha| \leq 1-\delta$, $|\beta| \leq 1-\delta$ where $\delta > 0$ is known. This is not required in Theorem 2.1 above. Note also that when 2, 3, 4 hold our condition C2.6 automatically follows since in this case $Q(\theta)$ is a continuous function of $\theta \in \Theta$ and Θ is a compact set so that $Q(\theta)$ attains its infimum in Θ (actually at $\theta = \theta_0$ by C2.3 which is also assumed by Walker). As Walker points out his proofs are based on methods similar to those indicated (without detail) in Whittle (1954). Whittle (1961) also supplies some detailed proof for the case $s = 1$ and $\overline{L}_N$. This treatment differs from that of Walker in several ways. For example there is no requirement that σ^2 be treated as a separate parameter from θ for the proof of the convergence in probability of $\overline{\theta}_N$ to θ_0 . However the conditions required to establish this are not the same as Walker's and only partially overlap with these. In relation to the ARMA model (2.17) ($p = q = s = 1$) it would appear that again $|\alpha| \leq 1-\delta$ and $|\beta| \leq 1-\delta$ are required in order that Whittle's conditions hold. (Note that the proof of Theorem 3 in Whittle (1961) is incorrect since it is not true that the variance of $NG(l)$ is $O(N)$ (see equation (8), p. 5 of Whittle) when the spectral density has reciprocal with absolutely convergent Fourier series unless for example the original spectral density is square integrable.) Whittle's article is concerned however, with more general situations than the use of the criteria $\overline{L}_N$ for estimation in stationary time series.

C(v) The results of this section have all been strong convergence results and these were based on the property that $G(n) \xrightarrow{a.s.} \Gamma(n)$ and certain "end effects" in quantities like $G(n)$ converge to zero a.s. The strong convergence of the $G(n)$ is due to the assumption that $x(n)$ is ergodic. This assumption cannot be checked if only (part of) one realisation of the process is available (see Hannan (1970), p. 201) and so is a costless assumption to make. However it may be of interest to extend the weak convergence results of Walker (1964) to the cases excluded by his conditions but included in C2 above. We have not investigated this problem.

C(vi) It may be of interest to indicate how the conditions C2 apply to the examples presented at the end of §1. In Example 1, K, which is denoted by σ^2, is not dependent on the parameters θ specifying k. For this example $h(e^{i\omega};\theta)$ in the above may be taken as identically 1 and $g(e^{i\omega};\theta) \equiv k(e^{i\omega};\theta)$. If, for example, $\overline{\Theta}$ is taken as a closed and bounded subset of $\mathbb{R}^p$ then, since

$k(e^{i\omega};\ \theta)$ is never zero for $\theta \in \overline{\Theta}$ it follows that $\Theta_0 = \Theta = \overline{\Theta}$. The conditions C2 clearly hold in this case and there are no identification problems (i.e. each point in $\overline{\Theta}$ yields a different function $k(e^{i\omega};\ \theta)$ and thus a different $f(\omega;\ \theta)$). It is not necessary to take σ^2 as belonging to a compact subset of R here since σ^2 is independent of θ . The condition that $\overline{\Theta}$ be a closed and bounded subset of R^p could possibly be relaxed for this example.

The second example at the end of §1 is not easily put in the above framework in general, since, for the above treatment it would be necessary to factorise

$$f(\omega;\ \theta) = \sum_0^p \theta_k g_k(e^{i\omega}) \tag{2.20}$$

as $\sigma^2(\theta)|k(e^{i\omega};\ \theta)|^2$ (when $s = 1$, for example), where $\sigma^2(\theta)$ and $k(e^{i\omega};\ \theta)$ possess the properties required above. Thus it may be more straightforward in cases such as (2.20) to proceed with the asymptotic theory directly and not through the factorisation just mentioned. Such an approach would probably require different conditions on the parameter space Θ and on the function $f(\omega;\ \theta)$.

3. The Central Limit Theorem

Since the central limit theorems for $N^{\frac{1}{2}}(\overline{\theta}_N-\theta_0)$, $N^{\frac{1}{2}}(\tilde{\theta}_N-\theta_0)$ and $N^{\frac{1}{2}}(\hat{\theta}_N-\theta_0)$ are proved by expanding the derivative (w.r.t. θ) of $\overline{L}_N$, $\tilde{L}_N$ and $\hat{L}_N$ respectively about the true parameter point θ_0 it is necessary to introduce coordinates for θ . Also in order that the CLT may be derived without assuming the existence of moments higher than the second for $\varepsilon(m)$

it has been necessary to consider the restriction of model (1.1) to the case when K depends upon the parameter μ, k depends on the parameter θ where θ and μ may vary independently of each other. (We return to the case when this is not so in Chapter 4.) We proceed as follows. Assume that $(\theta_0, \mu_0) \in M_\theta \times M_\mu$ where M_θ and M_μ are C_m-manifolds, $m \geq 2$ (see (Matsushima (1972)). For example M_θ may be taken to be the open subset Θ_0 of Θ as described in the last section (note this means that cases where $\det g(z; \theta)$ has a zero on the unit circle are excluded from the following treatment) and M_μ may be taken to be an open set in a topological space for which $K(\mu)$ is strictly positive definite and has finite norm. Since, for example, $(\overline{\theta}_N, \overline{\mu}_N)$ converges a.s. to (θ_0, μ_0), from some sample size onwards $(\overline{\theta}_N, \overline{\mu}_N)$ will be in some specified neighbourhood of (θ_0, μ_0) which may be taken as the product of a neighbourhood N_{θ_0} of $\theta_0 \in M_\theta$ and a neighbourhood N_{μ_0} of $\mu_0 \in M_\mu$. Note that N_{θ_0} may be taken to be diffeomorphic to an open set in Euclidean space where the diffeomorphism establishes a coordinate system for N_{θ_0}. Similarly for N_{μ_0}. This means that N_{θ_0} may be coordinatised by u parameters $\theta_1, \ldots, \theta_u$ and N_{μ_0} may be coordinatised by v parameters $\mu_1, \ldots, \mu_v$. (For the ARMA model a full description of the appropriate manifolds will be given in Chapter 3.)

In addition to the conditions C2 the following will be assumed throughout this section.

C3.1. The elements of $k(e^{i\omega}; \theta)$ are twice differentiable functions of θ for $\theta \in M_\theta$. These derivatives are continuous functions of $(\omega; \theta) \in [-\pi, \pi] \times M_\theta$. Similarly $K(\mu)$ has elements which are twice continuously differentiable for $\mu \in M_\mu$.

C3.2. $\sum_{j=0}^{\infty} j\|C(j;\ \theta)\|^2 < \infty$ (which for example is satisfied for the ARMA model discussed in the next chapter).

C3.3. (a) $E\left(\varepsilon(n) \mid F_{n-1}\right) = 0$ a.s.

(b) $E\left(\varepsilon(m)\varepsilon(m)' \mid F_{n-1}\right) = K\left(\mu_0\right)$ a.s.

Here F_n is the Borel sub-field of all events generated by the history of $\varepsilon(m)$ (i.e. $x(m)$) for $m \le n$. As explained in Hannan and Heyde (1972) C3.3 (a) is equivalent to the assertion that for $x(n)$ the best predictor is the best linear predictor (both best in the least squares sense). To emphasise this fact we repeat the argument of Hannan and Heyde (1972, p. 2059). We now let $F_x(n)$ be the σ-field generated by the $x(n)$, $m \le n$, and similarly let $F_\varepsilon(n)$ correspond to $\varepsilon(n)$. Let $\dot{x}(n)$ be the best linear predictor given $x(m)$, $m \le n-1$ and let $\hat{x}(n)$ be the best predictor. Then, since $x(n)$ is generated by (1.1), $F_x(n) \subseteq F_\varepsilon(n)$ for all n. Also when $\varepsilon(m)$ are the linear prediction errors $\varepsilon(n) = x(n) - \dot{x}(n)$ so that $F_\varepsilon(n) \subseteq F_x(n)$ all n. Thus $F_x(n) \equiv F_\varepsilon(n)$. Now

$$\hat{x}(n) = E\left(x(n) \mid F_x(n-1)\right) \quad \text{a.s.} \tag{3.1}$$

Thus when $\dot{x}(n) = \hat{x}(n)$ it follows that

$$E\left(\varepsilon(n) \mid F_\varepsilon(n-1)\right) = E\left(x(n)-\hat{x}(n) \mid F_\varepsilon(n-1)\right) = 0 \quad \text{a.s.} \tag{3.2}$$

by (3.1) and the equivalence of $F_x(n)$ and $F_\varepsilon(n)$. That is C3.3 (a) is satisfied. Conversely if this is so then, since $\dot{x}(n)$ is $F_\varepsilon(n-1)$ measurable,

$$\begin{aligned} 0 &= E\left(x(n)-\dot{x}(n) \mid F_\varepsilon(n-1)\right) \\ &= E\left(x(n) \mid F_\varepsilon(n-1)\right) - E\left(x(n) \mid F_\varepsilon(n-1)\right) \\ &= E\left(x(n) \mid F_x(n-1)\right) - \dot{x}(n) \quad \text{a.s.} \end{aligned}$$

Hence $\dot{x}(n) = \hat{x}(n)$. This establishes the required equivalence. This fact illustrates that the condition C3.3 (a) is a natural condition when models of the type (1.1) are believed to describe $x(n)$. The second part of C3.3, i.e. C3.3 (b), is required in order that simple formulae for the covariances of the estimators in their limiting distribution should result.

The CLT for $\overline{\theta}_N$ holds under conditions C2 and C3. For $\tilde{\theta}_N$, $\hat{\theta}_N$ additional regularity conditions are required for $k(e^{i\omega};\ \theta)$ which will be introduced later. The above formulation means that derivatives (up to second order) of the elements of $k(e^{i\omega};\ \theta)$ with respect to θ and $K(\mu)$ with respect to μ are well defined in the above neighbourhoods. With regard to notation we will write, for example, $\partial k_0^{-1}(e^{i\omega})/\partial\theta_j$ for the derivative evaluated at θ_0 . For notational convenience *throughout this section we will let* $w(\omega)$ *and* $I(\omega)$ *denote what were previously* $\sqrt{2\pi}\, w(\omega)$ $2\pi I(\omega)$ *respectively*. Also we define

$$\overline{K}_N = (2\pi)^{-1} \int_{-\pi}^{\pi} k^{-1}\left(e^{i\omega};\ \overline{\theta}_N\right) I(\omega) k^{-1}\left(e^{i\omega};\ \overline{\theta}_N\right)^* d\omega\ .$$

THEOREM 3.1. *For the general linear model* (1.1) *let conditions* C2 *and* C3 *hold and let* $(\theta_0,\ \mu_0) \in M_\theta \times M_\mu$ *which is a twice differentiable manifold. Then*

(a) $N^{\frac{1}{2}}\left(\overline{\theta}_N - \theta_0\right)$ *is asymptotically normal with zero mean vector and covariance matrix*

$$\Omega^{-1} = \left[\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left\{f(\omega;\ \theta_0)^{-1}\frac{\partial f(\omega;\theta_0)}{\partial\theta_j} f(\omega;\ \theta_0)^{-1}\frac{\partial f(\omega;\theta_0)}{\partial\theta_k}\right\}d\omega\right]^{-1}$$

(b) $N^{\frac{1}{2}}\left(\overline{K}_N - \dot{K}_N\right) \xrightarrow{p} 0$, *where* $\dot{K}_N = N^{-1}\sum_{1}^{N} \varepsilon(m)\varepsilon(m)'$;

(c) if it is further assumed that the $\varepsilon(m)$ *have finite fourth moment then* $N^{\frac{1}{2}}\left(\overline{\mu}_N - \mu_0\right)$ *is asymptotically normally distributed*

with zero mean and covariance matrix

$$\Pi = B^{-1}AB^{-1}$$

where

$$B = \left[\operatorname{tr}\left(K_0^{-1}\frac{\partial K_0}{\partial \mu_j}K_0^{-1}\frac{\partial K_0}{\partial \mu_k}\right)\right]$$

and

$$A = \left[\operatorname{Vec}\left[K_0^{-1}\frac{\partial K_0}{\partial \mu_j}K_0^{-1}\right]' \cdot \textstyle\sum \operatorname{Vec}\left[K_0^{-1}\frac{\partial K_0}{\partial \mu_k}K_0^{-1}\right]\right]$$

with $\sum$ *having* (a, b)*th element of the* (c, d)*th block of* $s \times s$ *elements as* σ_{acdb} *where*

$$\sigma_{abcd} = E\left(\varepsilon_a(m)\varepsilon_b(m)-(K_0)_{ab}\right)\left(\varepsilon_c(m)\varepsilon_d(m)-(K_0)_{cd}\right) .$$

For the general finite parameter model (1.1), Ω easily reduces to

$$\Omega = (2\pi)^{-1}\int_{-\pi}^{\pi}\operatorname{tr}\left\{\left(k_0^{-1}(\omega)\frac{\partial k_0^{-1}(\omega)}{\partial \theta_j}\right)K_0\left(k_0^{-1}(\omega)\frac{\partial k_0^{-1}(\omega)}{\partial \theta_k}\right)^* K_0^{-1}\right\}d\omega$$

where K_0 does not cancel (*cf.* the scalar case). Also some simplification of A and B above is possible when μ consists of the $s(s+1)/2$ elements in the upper triangular part of $K(\mu)$. Note that Π does not depend upon θ_0 .

Proof of *(a)*. Since $(\overline{\theta}_N, \overline{\mu}_N)$ minimises $\overline{L}_N(\theta, \mu)$ we have

$$0 = N^{\frac{1}{2}}\frac{\partial}{\partial \theta_j}\overline{L}_N(\theta, \mu)\Big|_{(\dot{\theta}_N, \overline{\mu}_N)} , \quad 1 \le j \le u ,$$

$$= N^{\frac{1}{2}}\frac{\partial}{\partial \theta_j}\overline{L}_N(\theta, \mu)\Big|_{(\theta_0, \overline{\mu}_N)} + \sum_{k=1}^{u}\frac{\partial^2}{\partial \theta_j \partial \theta_k}\overline{L}_N(\theta, \mu)\Big|_{(\dot{\theta}_N, \overline{\mu}_N)} \cdot \left(N^{\frac{1}{2}}(\overline{\theta}_N-\theta_0)\right)$$

where $\|\overline{\theta}_N-\theta_0\| \ge \|\dot{\theta}_N-\theta_0\|$. Since $\overline{\mu}_N - \mu_0$ and $\overline{\theta}_N - \theta_0$ both converge a.s. to zero the $u \times u$ matrix with typical element

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k}\overline{L}_N(\theta, \mu)\Big|_{(\dot{\theta}_N, \overline{\mu}_N)}$$

converges a.s. to Ω . Hence the CLT for the vector $N^{\frac{1}{2}}\left(\overline{\theta}_N-\theta_0\right)$ reduces to that for the u quantities

$$N^{\frac{1}{2}}\frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{tr}\left\{\frac{\partial}{\partial\theta}\left[k^{-1}(\theta)^{*}K\left(\overline{\mu}_N\right)^{-1}k^{-1}(\theta)\right]\Big|_{\theta_0}I(\omega)\right\}d\omega\ ,\quad 1\le j\le u\ . \tag{3.4}$$

Let $\psi_N(\omega)=\frac{\partial}{\partial\theta}\left[k^{-1}(\theta)^{*}K\left(\overline{\mu}_N\right)^{-1}k^{-1}(\theta)\right]\Big|_{\theta_0}$ and when $K\left(\overline{\mu}_N\right)^{-1}$ is replaced by $K\left(\mu_0\right)^{-1}$ we call the result $\psi_0(\omega)$. Since the proof of the CLT for the quantities in (3.4) is lengthy we will outline it as follows:-

(i) $N^{\frac{1}{2}}\left\{(2\pi)^{-1}\int\operatorname{tr}\left(\psi_N(\omega)k_0\dot{K}_Nk_0^{*}\right)d\omega\right\}=0$. This follows since the l.h.s. is

$$-N^{\frac{1}{2}}\operatorname{tr}\left\{\dot{K}_NK\left(\overline{\mu}_N\right)^{-1}\frac{1}{2\pi}\int k_0^{-1}\frac{\partial k_0}{\partial\theta_j}d\omega+K\left(\overline{\mu}_N\right)^{-1}\dot{K}_N\frac{1}{2\pi}\int\left(k_0^{-1}\frac{\partial k_0}{\partial\theta_j}\right)^{*}d\omega\right\}\ .$$

But $\partial k_0\left(e^{i\omega}\right)/\partial\theta_j$ has Fourier coefficients which are zero for indices $j\le 0$ and thus $k_0^{-1}\partial k_0/\partial\theta_j$ also has this property. Thus each integral in this last expression is the null matrix. Hence we may modify (3.4) and consider in its place

$$N^{\frac{1}{2}}(2\pi)^{-1}\int_{-\pi}^{\pi}\operatorname{tr}\left\{\psi_N(\omega)\left[I(\omega)-k_0\left(e^{i\omega}\right)\dot{K}_Nk_0\left(e^{i\omega}\right)^{*}\right]\right\}d\omega\ . \tag{3.5}$$

In fact it is difficult to prove the next step of the proof without this modification and this modification is not possible unless (i) holds, or holds asymptotically in probability. This in turn depends on the form of the matrix function $\psi_N(\omega)$ which is as it is because $K(\mu)$ is parameterised separately from $k(\theta)$.

(ii) We will show, at the end of this outline proof, that *the contribution to* (3.5) *from the squared terms* (i.e. those involving an $\varepsilon(m)\varepsilon(m)'$ arising from the second factor under the trace sign) *converges in* *probability to zero.* The proof we give also applies if ψ_N is replaced by ψ_0 and since (i) also holds with ψ_0 it then will follow that $N^{\frac{1}{2}}\frac{1}{2\pi}\int \mathrm{tr}\left(\psi_N-\psi_0\right)\tilde{I}(\omega)d\omega$ converges in probability to zero, where $\tilde{I}(\omega)$ is $I(\omega)$ with the squared terms removed.

Indeed

$$N^{\frac{1}{2}}\frac{1}{2\pi}\int \mathrm{tr}\left[\left(\psi_N-\psi_0\right)\tilde{I}(\omega)\right]d\omega = N^{\frac{1}{2}}\,\mathrm{tr}\left\{\left(K\left(\overline{\mu}_N\right)^{-1}-K_0^{-1}\right)\cdot\frac{1}{2\pi}\int\frac{\partial k_0^{-1}}{\partial\theta_j}\tilde{I}(\omega)k_0^{-1*}d\omega\right\}$$

$$+\;N^{\frac{1}{2}}\,\mathrm{tr}\left\{\left(K\left(\overline{\mu}_N\right)^{-1}-K_0^{-1}\right)\frac{1}{2\pi}\int k_0^{-1}\tilde{I}(\omega)\left(\frac{\partial k_0^{-1}}{\partial\theta_j}\right)^{*}d\omega\right\}. \qquad (3.6)$$

Both terms in the r.h.s. of (3.6) are handled in the same way. Now the first, say, is equal to

$$\sum_{a}\sum_{b}\sum_{c}\sum_{d}^{s}{}_{=1}\left(K\left(\overline{\mu}_N\right)^{-1}-K_0^{-1}\right)_{ab}\left[\frac{N^{\frac{1}{2}}}{2\pi}\int\left(\frac{\partial k_0^{-1}}{\partial\theta_j}\right)_{bd}\left(k_0^{-1}\right)_{ca}\tilde{I}_{dc}(\omega)d\omega\right]. \qquad (3.7)$$

It will now be established that the variance of the factor in square brackets in the typical summand of (3.7) is finite. Then (3.7) and hence (3.6), converge in probability to zero, since $K\left(\overline{\mu}_N\right)\xrightarrow{\text{a.s.}}K_0$. Thus letting $\phi(\omega)$ denote $\left(\frac{\partial k_0^{-1}}{\partial\theta_j}\right)_{bd}\left(k_0^{-1}\right)_{ca}$ the typical factor we need to consider in (3.7) is

$$\frac{N^{\frac{1}{2}}}{2\pi}\int\phi(\omega)\tilde{I}_{dc}(\omega)d\omega$$

$$=\sum_{\substack{p,q\\=1}}^{s}\sum\frac{N^{-\frac{1}{2}}}{2\pi}\int\phi(\omega)\sum_{\substack{m,n=1\\m-j\neq n-k}}^{N}\sum\;\sum_{j,k=0}^{\infty}\sum C_{dp}(j)C_{cq}(k)\varepsilon_p(m-j)\varepsilon_q(n-k)e^{i(m-n)\omega}d\omega\,. \qquad (3.8)$$

But, since $m-j\neq n-k$, the expected value of this expression is zero and hence its variance is

$$E\left(N^{-\frac{1}{2}}\sum_{p,q=1}^{s}\sum_{\substack{m}}\sum_{n}\sum_{j}\sum_{k}\,_{m-j\neq n-k} C_{dp}(j)C_{cq}(k).\frac{1}{2\pi}\int \phi(\omega)e^{i(m-n)\omega}d\omega\ .\ \varepsilon_p(m-j)\varepsilon_q(n-k)\right)^2$$

$$= N^{-1}\sum_p\sum_q\sum_{p'}\sum_{q'}\sum_{\substack{m\ n\ j\ k \\ m-j\neq n-k}}\sum_{\substack{m'\ n'\ j'\ k' \\ m'-j'\neq n'-k'}}\Big\{C_{dp}(j)C_{cq}(k)C_{dp'}(j')C_{cq'}(k')\ .$$

$$.\ \left[\frac{1}{2\pi}\int \phi(\omega)e^{i(m-n)\omega}d\omega\right]\left[\frac{1}{2\pi}\int \phi(\omega)e^{i(m'-n')\omega}d\omega\right]\ .$$

$$.\ E\left(\varepsilon_p(m-j)\varepsilon_q(n-k)\varepsilon_{p'}(m'-j')\varepsilon_{q'}(n'-k')\right)\Big\}\ . \qquad (3.9)$$

But since $m-j \neq n-k$ and $m'-j' \neq n'-k'$ the value of $E\left(\varepsilon_p(m-j)\varepsilon_q(n-k)\varepsilon_p(m'-j')\varepsilon_q(n'-k')\right)$ when the $\varepsilon(m)$ satisfy C3.3 is precisely the same as when the $\varepsilon(m)$ are Gaussian. Therefore, letting Var_G denote the variance when the $\varepsilon(m)$ are Gaussian we have

$$\mathrm{Var}\left\{\frac{N^{\frac{1}{2}}}{2\pi}\int \phi(\omega)\tilde{I}_{dc}(\omega)d\omega\right\} = \mathrm{Var}_G\left\{\frac{N^{\frac{1}{2}}}{2\pi}\int \phi(\omega)\tilde{I}_{dc}(\omega)d\omega\right\}\ . \qquad (3.10)$$

But letting $\dot{I}(\omega)$ denote the contribution to $I(\omega)$ from the squared terms we have $I(\omega) = \tilde{I}(\omega) + \dot{I}(\omega)$ so that

$$\mathrm{Var}_G\left\{N^{\frac{1}{2}}\ .\ \frac{1}{2\pi}\int \phi(\omega)I_{dc}(\omega)d\omega\right\}$$

$$= \mathrm{Var}_G\left\{N^{\frac{1}{2}}\ .\ \frac{1}{2\pi}\int \phi(\omega)\tilde{I}_{dc}(\omega)d\omega\right\} + \mathrm{Var}_G\left\{N^{\frac{1}{2}}\ .\ \frac{1}{2\pi}\int \phi(\omega)\dot{I}_{dc}(\omega)d\omega\right\}$$

$$+ E_G\left\{\left[N^{\frac{1}{2}}\ .\ \frac{1}{2\pi}\int \phi(\omega)\tilde{I}_{dc}(\omega)d\omega\right]\left[N^{\frac{1}{2}}\frac{1}{2\pi}\int \phi(\omega)\left(\dot{I}_{dc}(\omega)-E_G\dot{I}_{dc}(\omega)\right)d\omega\right]\right\}.$$

Using a similar expansion as given in (3.9) it follows, in a straightforward way, that the last term in this expression is zero so that

$$\mathrm{Var}_G\left\{N^{\frac{1}{2}}\frac{1}{2\pi}\int \phi(\omega)\tilde{I}_{dc}(\omega)d\omega\right\} \le \mathrm{Var}_G\left\{N^{\frac{1}{2}}\frac{1}{2\pi}\int \phi(\omega)I_{dc}(\omega)d\omega\right\} \qquad (3.11)$$

and hence, by (3.10), that

$$\mathrm{Var}\left\{N^{\frac{1}{2}}\frac{1}{2\pi}\int \phi(\omega)\tilde{I}_{dc}(\omega)d\omega\right\} \le \mathrm{Var}_G\left\{N^{\frac{1}{2}}\frac{1}{2\pi}\int \phi(\omega)I_{dc}(\omega)d\omega\right\}\ . \qquad (3.12)$$

Now

$$\mathrm{Var}_G\left\{N^{\frac{1}{2}}\,\frac{1}{2\pi}\int\phi(\omega)I_{dc}(\omega)d\omega\right\}$$

$$= NE_G\left\{\left(\frac{1}{2\pi}\int\phi(\omega)I_{dc}(\omega)d\omega - N^{-1}\frac{1}{2\pi}\int\phi(\omega)\sum_1^N\sum\gamma_{dc}(n-m)e^{i(m-n)\omega}d\omega\right)^2\right\}$$

$$= N^{-1}\sum_{m}\sum_{n}\sum_{m'}\sum_{n'=1}^{N}\left(\frac{1}{2\pi}\int\phi(\omega)e^{i(m-n)\omega}d\omega\right)\left(\frac{1}{2\pi}\int\phi(\omega)e^{i(m'-n')\omega}d\omega\right)$$

$$\cdot\, E\left(x_d(m)x_c(n)x_d(m')x_c(n')\right) - N^{-1}\left(\frac{1}{2\pi}\int\phi(\omega)\sum_1^N\sum\gamma_{dc}(n-m)e^{i(m-n)\omega}d\omega\right)^2$$

$$= N^{-1}\sum_{m}\sum_{n}\sum_{m'}\sum_{n'=1}^{N}\left(\left(\frac{1}{2\pi}\right)^2\iint\phi(\omega)\phi(\lambda)e^{i(m-n)\omega}e^{i(m'-n')\lambda}d\omega d\lambda\right)\cdot$$

$$\cdot\left(\gamma_{cd}(m'-m)\gamma_{cc}(n'-n)+\gamma_{dd}(n'-m)\gamma_{dc}(m'-n)\right) \qquad (3.13)$$

using the expression for $E\left(x_d(m)x_c(n)x_d(m')x_c(n')\right)$ on p. 23 of Hannan (1970) and the fact that when $x(n)$ is Gaussian the fourth cumulant vanishes. Both the term arising from $\gamma_{cd}(n'-m)\gamma_{dc}(m'-n)$ and the term arising from $\gamma_{dd}(m'-m)\gamma_{cc}(n'-n)$ on the r.h.s. of (3.13) may be treated in the same way. Therefore consider, say,

$$N^{-1}\sum_{m}\sum_{n}\sum_{m'}\sum_{n'=1}^{N}\left(\frac{1}{2\pi}\right)^2\iint\phi(\omega)\phi(\lambda)e^{i(m-n)\omega}e^{i(m'-n')\lambda}d\omega d\lambda\,\cdot\,\gamma_{dd}(m'-m)\gamma_{cc}(n'-n)$$

$$= \left(\frac{1}{2\pi}\right)^2\iint\phi(\omega)\phi(\lambda)\sum_{u}\sum_{v=-N+1}^{N-1}\gamma_{dd}(u)\gamma_{cc}(v)e^{i(v-u)\omega}\,\cdot$$

$$\cdot\left\{\frac{1}{\sqrt{N}}\sum_u\!\!{}' \, e^{im(\omega+\lambda)}\,\frac{1}{\sqrt{N}}\sum_v\!\!{}' \, e^{-im(\omega+\lambda)}\right\}d\lambda d\omega \qquad (3.14)$$

where $\sum_u'$ is the sum over m for $1 \le m \le N$ *and* $1 \le m+u \le N$. Thus the modulus of the r.h.s. of (3.14) is dominated by (where $\sup|\phi(\omega)| \le b$)

$$b^2 \cdot \left(\frac{1}{2\pi}\right)^2 \iint \left| \sum_{v=-N+1}^{N-1} \gamma_{dd}(v) e^{iv\omega} \frac{1}{\sqrt{N}} \sum_v e^{im(\omega+\lambda)} \right|^2 d\omega d\lambda$$

$$= b^2 \cdot \left(\frac{1}{2\pi}\right) \int \sum_{u=-N+1}^{N-1} \sum_{v=-N+1}^{N-1} \gamma_{dd}(v) e^{iv\omega} \gamma_{dd}(u) e^{-iu\omega} \left(1 - \frac{\min(|u|,|v|)}{N}\right) d\omega$$

$$= b^2 \cdot \sum_{v=-N+1}^{N-1} \gamma_{dd}(v)^2 \left(1 - \frac{|v|}{N}\right)$$

$$< \infty \,.$$

Hence (3.6) converges in probability to zero. The basic argument used here is as in Hannan and Heyde (1972, p. 2065). Thus we may consider the CLT for

$$N^{\frac{1}{2}}(2\pi)^{-1} \int \mathrm{tr}[\psi_0(\omega)\tilde{I}(\omega)]\, d\omega \,. \tag{3.15}$$

(iii) If $\psi_0^{(M)}$ is the Cesaro sum to M terms of the Fourier series for ψ_0 where $\sup_\omega \left\|\psi_0(\omega)-\psi_0^{(M)}(\omega)\right\| \le \varepsilon$ we may replace ψ_0 by $\psi_0^{(M)}$ in (3.15) where the error in this replacement may be made arbitrarily small in probability, using the argument given in (ii) to replace ψ_N by ψ_0 .

(iv) With $\psi_0^{(M)}$ inserted, (3.15) reduces to a finite linear combination of autocovariances $G_{rs}(n)$, from which the contributions of the squared terms have been removed, and whose removal makes asymptotically no difference to the distribution of $N^{\frac{1}{2}}\left(G_{rs}(n)-\Gamma_{rs}(n)\right)$. By essentially the same argument given in Hannan and Heyde (1972) the proof of the theorem reduces to that for a finite number of expressions of the form $N^{-\frac{1}{2}} \sum_{m=1}^{N} \varepsilon_r(m)\varepsilon_s(m+n)$, $n > 0$. Since the summands are stationary, ergodic, martingale differences with finite variance, the result follows from the theorem of Billingsley (1961).

To complete the proof of *(a)* we must now prove the assertion made in (ii) concerning the elimination of squared terms. This is achieved as follows. The contribution to (3.5) from the squared terms is

$$N^{-\frac{1}{2}} \frac{1}{2\pi} \int \operatorname{tr}\left[\psi_N(\omega)\left\{\sum_j \sum_k C_0(j)\left(\sum_{jk} - \sum\right) C_0(k)^* e^{i(j-k)\omega}\right\}\right] d\omega\ , \tag{3.16}$$

where $\sum$ and $\sum_{jk}$ are sums of $\varepsilon(n)\varepsilon(n)'$, the former for $1 \le n \le N$ and the latter for $1-k \le n \le N-j$, $j \ge k$, or $1-j \le n \le N-k$, $k \ge j$. We decompose $\left(\sum_{jk} - \sum\right)$ as $\left(\sum'_{jk} - \sum''_{jk}\right)$ where $\sum'_{jk}$ is a sum over n for $n+1 \le j,\ k \le N+n$, $n \ge 0$, while $\sum''_{jk}$ contains $\varepsilon(n)\varepsilon(n)'$, $1 \le n \le N$, if either j or k is not less than $N-n+1$. Now $\sum_j \sum_k C_0(j) \sum'_{jk} C_0(k)^* e^{i(j-k)\omega} \ge 0$ and its contribution to (3.16) is dominated by (with $\|\psi_N(\omega)\| \le c < \infty$ for N large since $\psi_N(\omega) \xrightarrow{a.s.} \psi_0(\omega)$)

$$cN^{-\frac{1}{2}} \frac{1}{2\pi} \int \operatorname{tr}\left\{\sum_j \sum_k C_0(j) \sum'_{jk} C_0(k)^* e^{i(j-k)\omega}\right\} d\omega \ge 0\ ,$$

which has expectation dominated by

$$cN^{-\frac{1}{2}} \operatorname{tr} \sum_{j=0}^{\infty} j\ C_0(j) K_0 C_0(j)^*$$

which converges to zero using C3.2. The contribution to (3.16) from $\sum''_{jk}$ is

$$N^{-\frac{1}{2}} \sum_0^{N-1} \frac{1}{2\pi} \int \operatorname{tr}\Bigg[\psi_N(\omega)\Bigg\{k_0 \varepsilon(N-n)\varepsilon(N-n)' \sum_{n+1}^{\infty} C_0(k)^* e^{-ik\omega}$$

$$+ \sum_{n+1}^{\infty} C_0(j) e^{ij\omega} \varepsilon(N-n)\varepsilon(N-n)' k_0^*$$

$$+ \sum_{n+1}^{\infty} C_0(j) e^{ij\omega} \varepsilon(N-n)\varepsilon(N-n)' \sum_{n+1}^{\infty} C_0(k)^* e^{-ik\omega}\Bigg\}\Bigg] d\omega\ . \tag{3.17}$$

To show that (3.17) converges in probability to zero we will proceed by demonstrating that ψ_N in (3.17) may be replaced by the Cesaro sum to M (finite) terms of the Fourier series for $(2\pi)^{-1} \partial f^{-1}(\omega;\ \theta_0)/\partial\theta_j$ (which is the a.s. limit of $\psi_N(\omega)$). Consider then (3.17) with $\psi_N(\omega)$ replaced by $\psi_N(\omega) - \frac{\partial}{\partial\theta_j} f^{-1}(\omega;\theta_0)/2\pi = \phi_N(\omega)$ (say). The last term in (3.17) after this replacement is dominated by the positive quantity

$$\sup_\omega \|\phi_N(\omega)\| \cdot \left[N^{-\frac{1}{2}} \sum_0^{N-1} \sum_{j=n+1}^{\infty} \operatorname{tr}\{C_0(j)\varepsilon(N-n)\varepsilon(N-n)'C_0(j)^*\}\right]$$

in which the first factor converges to zero a.s. and the second factor (in square brackets) has expectation dominated by

$$\|K_0\|.N^{-\frac{1}{2}} \sum_0^{\infty} \min(j,\ N)\|C_0(j)\|^2 \to 0\ .$$

The first and second terms in (3.17) are handled in the same way as follows. The modulus of the first term (say) is bounded by

$$b\ .\ \sup_\omega \|\phi_N(\omega)\| \cdot \left[N^{-\frac{1}{2}} \sum_{n=0}^{N-1} (2\pi)^{-1} \int \left\|\sum_{n+1}^{\infty} C_0(j)e^{-ij\omega}\right\| \operatorname{tr}\left(\varepsilon(N-n)\varepsilon(N-n)'\right)d\omega\right]$$

(where b is a bound for $\|k_0(e^{i\omega})\|$). Again $\sup_\omega \|\phi_N(\omega)\| \xrightarrow{\text{a.s.}} 0$ and the expectation of the factor in square brackets is bounded by

$$N^{-\frac{1}{2}}\ .\ \|K_0\| \sum_{n=0}^{N-1} \left[(2\pi)^{-1} \int \left\|\sum_{n+1}^{\infty} C_0(j)e^{ij\omega}\right\|^2 d\omega\right]^{\frac{1}{2}} \tag{3.18}$$

using the Cauchy-Schwarz inequality. Using this inequality again with respect to the sum over n gives an upper bound for (3.18) as

$$\|K_0\|\ .\ \left\{\sum_{n=0}^{N-1} \|C_0(j)\|^2 \min(j,\ N)\right\}^{\frac{1}{2}}$$

which is finite by C3.2. Repeating the above argument with $\phi_N(\omega)$ replaced by $\frac{1}{2\pi}\left\{\frac{\partial}{\partial\theta_j} f^{-1}(\omega;\ \theta_0) - P_M(\omega)\right\}$ where $P_M(\omega)$ is the Cesaro sum corresponding to $\frac{\partial}{\partial\theta_j} f^{-1}(\omega;\ \theta_0)$ the terms in (3.17) may be made arbitrarily small in probability by choosing M large. This replacement of $\frac{\partial}{\partial\theta_j} f^{-1}(\omega;\ \theta_0)$ by its Cesaro sum reduces consideration of the contribution to (3.16) from

$\sum_{jk}^{n}$ with ψ_N replaced by $\exp(iu\omega)$, $|u| \leq M$. The proof that this converges to zero in probability is as follows:

$$E\left[N^{-\frac{1}{2}} \sum_{n=0}^{N-1} (2\pi)^{-1} \int \operatorname{tr}\left[e^{iu\omega}\left\{k_0 \varepsilon(N-n)\varepsilon(N-n)' \sum_{j=n+1}^{\infty} C_0(j)^* e^{-ij\omega}\right\}\right] d\omega\right]$$

$$= E\left[N^{-\frac{1}{2}} \sum_{n=0}^{N-1} \varepsilon(N-n)' \sum_{j=n+1}^{\infty} C_0(j) C_0(j+u)^* \varepsilon(N-n)\right]$$

$$\leq c.N^{-\frac{1}{2}} \sum_{n=0}^{N-1} \left\| \sum_{j=n+1}^{\infty} C_0(j) C_0(j+u)^* \right\| \qquad \left(\|K_0\| \leq c < \infty\right)$$

$$\leq c.N^{-\frac{1}{2}} \left[\sum_{n=0}^{N-1} \sum_{j=n+1}^{\infty} \|C_0(j)\|^2\right]^{\frac{1}{2}} \left[\sum_{n=0}^{N-1} \sum_{j=n+1}^{\infty} \|C_0(j+u)\|^2\right]^{\frac{1}{2}}$$

$$= c \cdot \left[N^{-\frac{1}{2}} \sum_{j=0}^{\infty} \|C_0(j)\|^2 \min(j, N)\right]^{\frac{1}{2}} \left[N^{-\frac{1}{2}} \sum_{j=0}^{\infty} \|C_0(j)\|^2 \min(j+u, N)\right]^{\frac{1}{2}}$$

$\to 0$, since u is finite.

This completes the proof of part *(a)*.

Proof of *(b)*. $N^{\frac{1}{2}}\left(\overline{K}_N - \dot{K}_N\right) = N^{\frac{1}{2}}\left(\overline{K}_N - \overline{K}_0\right) + N^{\frac{1}{2}}\left(\overline{K}_0 - \dot{K}_N\right)$, where $\overline{K}_0$ is defined as is $\overline{K}_N$ but $\overline{\theta}_N$ is replaced by θ_0 . First of all consider

$$N^{\frac{1}{2}}\left(\overline{K}_N - \overline{K}_0\right) = N^{\frac{1}{2}}(2\pi)^{-1} \int \left[k\left(\overline{\theta}_N\right)^{-1} I(\omega) k^*\left(\overline{\theta}_N\right)^{-1} - k_0^{-1} I(\omega) k_0^{-1*}\right] d\omega \,. \tag{3.19}$$

By expanding $k\left(\overline{\theta}_N\right)^{-1}$ in the first two terms of a Taylor series expansion about θ_0 we may re-write (3.19) as

$$N^{\frac{1}{2}}\left(\overline{K}_N - \overline{K}_0\right) = \left\{N^{\frac{1}{2}}\left(\overline{\theta}_{Nj} - \theta_{0j}\right)\right\}\left\{(2\pi)^{-1} \int k^{-1}\left(\phi_N\right) \frac{\partial k\left(\phi_N\right)}{\partial \theta_j} k^{-1}\left(\phi_N\right) I(\omega) k^*\left(\overline{\theta}_N\right) d\omega\right\}$$

$$+ \left\{N^{\frac{1}{2}}\left(\overline{\theta}_{Nj} - \theta_{0j}\right)\right\}\left\{(2\pi)^{-1} \int k_0^{-1} I(\omega) \left[k^{-1}\left(\phi_N\right) \frac{\partial k\left(\phi_N\right)}{\partial \theta_j} k^{-1}\left(\phi_N\right)\right]^* d\omega\right\} \tag{3.20}$$

where $\|\phi_N - \theta_0\| \leq \|\overline{\theta}_N - \theta_0\|$. The second factor in the first term on the r.h.s. of (3.20) converges a.s. to $(2\pi)^{-1} \int k_0^{-1} \frac{\partial k_0}{\partial \theta_j} K_0 d\omega$ which, as in (i) above, is null. Similarly the second factor in the second term as the r.h.s.

of (3.20) converges to the null matrix. Thus since $N^{\frac{1}{2}}(\overline{\theta}_N - \theta_0)$ converges in distribution to normality, by part *(a)* above, it follows that $N^{\frac{1}{2}}(\overline{K}_N - K_0) \xrightarrow{P} 0$. Next consider, where $w_\varepsilon(\omega) = N^{-\frac{1}{2}} \sum_1^N \varepsilon(m) e^{im\omega}$ and $I_\varepsilon(\omega) = w_\varepsilon(\omega) w_\varepsilon(\omega)^*$,

$$N^{\frac{1}{2}}(\overline{K}_0 - \dot{K}_N) = N^{\frac{1}{2}}(2\pi)^{-1} \int k_0^{-1} I(\omega) k_0^{-1} - I_\varepsilon(\omega) d\omega$$

$$= N^{\frac{1}{2}}(2\pi)^{-1} \int k_0^{-1}\{(w - k_0 w_\varepsilon) w_\varepsilon^* k_0^* + k_0 w_\varepsilon (w - k_0 w_\varepsilon)^*$$

$$+ (w - k_0 w_\varepsilon)(w - k_0 w_\varepsilon)^*\} k_0^{-1*} d\omega \ . \quad (3.21)$$

Hence

$$N^{\frac{1}{2}}(\overline{K}_0 - \dot{K}_N) = (2\pi)^{-1} \int \left\{k_0^{-1} \chi_0 w_\varepsilon^* + w_\varepsilon \chi_0^* k_0^{-1*} + N^{-\frac{1}{2}} \ k_0^{-1} \chi_0 \chi_0^* k_0^{-1*}\right\} d\omega \quad (3.22)$$

where $\frac{1}{\sqrt{N}} \chi_0(\omega) = w(\omega) - k_0 w_\varepsilon(\omega)$. But

$$\chi_0(\omega) = \sum_0^\infty C_0(j) e^{ij\omega} R_{j,N}(\omega) \ ,$$

and $R_{j,N}(\omega)$ is defined on p. 246 of Hannan (1970) as

$$R_{j,N}(\omega) = \left\{\sum_{n=-j+1}^{0} \varepsilon(n) e^{in\omega} - \sum_{m=N-j+1}^{N} \varepsilon(n) e^{in\omega}\right\} \qquad 0 \le j \le N$$

$$= \left\{- \sum_{1}^{-j} \varepsilon(n) e^{in\omega} + \sum_{N+1}^{N-j} \varepsilon(n) e^{in\omega}\right\} \qquad -N \le j \le 0$$

$$= \left\{\sum_{-j+1}^{N-j} \varepsilon(n) e^{in\omega} - \sum_{1}^{N} \varepsilon(n) e^{in\omega}\right\} \qquad N \le j < \infty$$

$$= \left\{- \sum_{1}^{N} \varepsilon(n) e^{in\omega} - \sum_{-j+1}^{N-j} \varepsilon(n) e^{in\omega}\right\} \qquad -\infty < j \le -N \ .$$

Now the last term in (3.22) is a non-negative definite matrix whose trace has expectation dominated by

$$c.N^{-\frac{1}{2}} \cdot \sup_{\omega} E\left(\chi_0^*(\omega)\chi_0(\omega)\right) \leq c.N^{-\frac{1}{2}} \sum_0^{\infty} \|C_0(j)\| \sup_{\omega} E\left(R_{j,N}(\omega)^* R_{j,N}(\omega)\right)$$

$$\leq c.N^{-\frac{1}{2}} \sum_0^{\infty} \|C_0(j)\| \min(j, N)$$

(see Hannan (1970, p. 247)),

where c is a finite constant, not always the same. Since this converges to zero the last term in (3.22) may be neglected.

Since $E \operatorname{tr} \int \chi_0 \chi_0^* d\omega < \infty$ and $E \operatorname{tr} \int w_\varepsilon w_\varepsilon^* d\omega < \infty$ we may replace k_0^{-1} in (say) the second term on the right of (3.22) by a trigonometric polynomial and reduce ourselves to consideration of

$$\frac{1}{2\pi} \int e^{iu\omega} \chi_0 w_\varepsilon^* d\omega = \frac{1}{\sqrt{N}} \left[\sum_0^{\infty} C_0(j) \sum\nolimits_j \varepsilon(m-u-j)\varepsilon(m)'\right] \tag{3.23}$$

where $\sum_j$ is a sum over $1+u \leq m \leq \min(u+j,\ u+N)$. An upper bound for the expectation of the squared norm of (3.23) is thus

$cN^{-1} \sum_j \|C_0(j)\|^2 \min(j, N) \to 0$. Hence $N^{\frac{1}{2}}\left(\overline{K}_0 - \dot{K}_N\right) \xrightarrow{P} 0$, and *(b)* is proved.

Proof of *(c)*. Similarly to the proof of *(a)* we are led to consider

$$N^{\frac{1}{2}}\left(\overline{\mu}_N - \mu_0\right) = -B_N^{-1} N^{\frac{1}{2}} \left.\frac{\partial \overline{L}_N(\overline{\theta}_N, \mu)}{\partial \mu}\right|_{\mu_0}$$

where

$$B_N = \left[\frac{\partial^2}{\partial \mu_j \partial \mu_k} \overline{L}_N\left(\overline{\theta}_N,\ \mu\right)\Big|_{\phi_N}\right], \quad \|\phi_N - \mu_0\| \leq \|\overline{\mu}_N - \mu_0\| .$$

Now

$$B_N \xrightarrow{\text{a.s.}} B = \left[\operatorname{tr}\left(K^{-1}(\mu_0) \frac{\partial K(\mu_0)}{\partial \mu_j} K^{-1}(\mu_0) \frac{\partial K(\mu_0)}{\partial \mu_k}\right)\right]$$

which is non singular by virtue of C2.6 since

$$-N^{\frac{1}{2}}\left(\left.\frac{\partial \overline{L}_N(\overline{\theta}_N, \mu)}{\partial \mu_j}\right|_{\mu_0}\right) = \operatorname{tr}\left[K^{-1}(\mu_0) \frac{\partial K(\mu_0)}{\partial \mu_j} K^{-1}(\mu_0) \left(N^{\frac{1}{2}}\left(\overline{K}_N - K(\mu_0)\right)\right)\right] .$$

By part *(b)*, $N^{\frac{1}{2}}[(\overline{K}_N-K(\mu_0)) - (\dot{K}_N-K(\mu_0))] \xrightarrow{P} 0$ and so the CLT reduces to that for the elements of $N^{\frac{1}{2}}(\dot{K}_N-K(\mu_0))$. If the $\varepsilon_j(m)$ have finite fourth moment, the fact that $\varepsilon(m)\varepsilon(m)' - K(\mu_0)$ is a sequence of matrices of martingale differences completes the proof of *(c)*. □

REMARK. It is *not* true that $\overline{K}_N$ would, in general, be the same as $K(\overline{\mu}_N)$ although asymptotically they must agree. However in the special case when M_μ is a subset of $\mathrm{R}^{s(s+1)/2}$ and K is parameterised by its upper triangular elements without restrictions, then $\overline{L}_N$ may be easily concentrated with respect to K to obtain $\overline{K}_N = K(\overline{\mu}_N)$. Thus since $N^{\frac{1}{2}}(\overline{K}_N-\dot{K}_N) \xrightarrow{P} 0$, $\overline{K}_N$ would be a very satisfactory estimator of K_0 even when moments higher than the second do not exist since $\dot{K}_N$ would obviously be satisfactory if it were available.

For the CLT for $N^{\frac{1}{2}}(\tilde{\theta}_N-\theta_0)$, $N^{\frac{1}{2}}(\tilde{\mu}_N-\mu_0)$ we have been unable to avoid imposing "smoothness" conditions on $k\left(e^{i\omega};\ \theta_0\right)$ and its derivative with respect to θ . These conditions are satisfied, for example, by the ARMA model to be discussed in Chapter 3. Denoting the Lipschitz class of degree α by Λ_α (see Zygmund, 1959, p. 42, for the definition) then

COROLLARY 3.2. *If the elements of* $k\left(e^{i\omega};\ \theta_0\right)$ *and* $\partial k\left(e^{i\omega};\ \theta_0\right)/\partial\theta_j$, $1 \le j \le u$ *, belong to* Λ_α *,* $\frac{1}{2} < \alpha \le 1$ *, then all the results of Theorem* 3.1 *hold for* $\tilde{\theta}_N, \tilde{\mu}_N, \tilde{K}_N$.

Note that here $\tilde{K}_N$ is the analogue of $\overline{K}_N$ defined as

$$\tilde{K}_N = (N')^{-1} \sum_t k^{-1}\left(e^{i\omega t};\ \tilde{\theta}_N\right) I(\omega_t) k^{-1}\left(e^{i\omega t};\ \tilde{\theta}_N\right)^* .$$

Of course if $\alpha > 1$ in the statement of the corollary the elements of k_0

and $\frac{\partial k_0}{\partial\theta}$ are constants (in fact $k_0 = I_s$ and $\frac{\partial k_0}{\partial\theta} = 0$). In this case the corollary is easy to establish.

Proof. The proof of this corollary is very much like that of Theorem 3.1 so that only the parts which are substantially different will be given. As in the proof for $\overline{\theta}_N$ the CLT for $\tilde{\theta}_N$ reduces to that for

$$N^{\frac{1}{2}}(N')^{-1}\sum_t \operatorname{tr}\left\{\frac{\partial}{\partial\theta_j}\left[k^{-1}(\theta)^*K(\overline{\mu}_N)^{-1}k^{-1}(\theta)\right]\bigg|_{\theta_0} I(\omega_t)\right\}, \quad 1 \le j \le u . \tag{3.24}$$

We define ψ_N and ψ_0 as in the proof for $\overline{\theta}_N$. By analogy with that proof we first show that

(i) $N^{\frac{1}{2}}(N')^{-1}\sum_t \operatorname{tr}\left(\psi_N(\omega)k_0\dot{K}_N k_0^*\right) \xrightarrow{P} 0$.

(Exact equality does not hold here.) Now the l.h.s. of this reduces to

$$-N^{\frac{1}{2}}\operatorname{tr}\left\{\dot{K}_N K(\tilde{\mu}_N)^{-1}(N')^{-1}\sum_t\left[\frac{\partial k_0^*}{\partial\theta_j}k_0^{-1*}+k_0^{-1}\frac{\partial k_0}{\partial\theta_j}\right]\right\} . \tag{3.25}$$

But the matrix $A(\omega)$, say, in square brackets satisfies $A(\omega) = A(-\omega)$ and has elements which belong to Λ_α, $\alpha > \frac{1}{2}$. Hence if $P_{N'/2}$ is the Cesaro sum of the Fourier series for A then

$$\sup_\omega \|A(\omega)-P_{N'/2}(\omega)\| = O\left[\left(\frac{N'}{2}\right)^{-\alpha}\right]$$

(see Zygmund, 1959, Theorem (3.15), p. 91). Hence the error in replacing A by $P_{N'/2}$ in (3.25) is dominated by

$$N^{\frac{1}{2}}\left\|\dot{K}_N K(\tilde{\mu}_N)^{-1}\left(\frac{N'}{2}\right)^{-\alpha}\right\| \to 0$$

since $\dot{K}_N \xrightarrow{a.s.} K_0$ and $K(\tilde{\mu}_N) \xrightarrow{a.s.} K_0$. But if we evaluate (3.25) with A replaced by $P_{N'/2}$ we obtain

$$-N^{\frac{1}{2}}\operatorname{tr}\left(\dot{K}_N\tilde{K}_N^{-1}P_{N'/2}(0)\right) = 0 ,$$

since $P_{N'/2}(0)$ is the zero-th Fourier coefficient of A which is null (see the proof of Theorem 3.1, (i)).

(ii) The discussion in (ii), (iii) and (iv) of the proof of Theorem 3.1 is substantially the same for the present case. The major difference is to show that we may disregard the squared terms in

$$N^{-\frac{1}{2}}(N')^{-1}\sum_t \mathrm{tr}\{\psi_N(\omega_t)\,[I(\omega_t)-k_0\dot{k}_N k_0^*]\} \tag{3.26}$$

which are given by

$$N^{-\frac{1}{2}}(N')^{-1}\sum_t \mathrm{tr}\left\{\psi_N(\omega_t)\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} C_0(j)\left[\Sigma'_{jk}-\Sigma''_{jk}\right]C_0(k)'.e^{i(j-k)\omega_t}\right]\right\} \tag{3.27}$$

where Σ'_{jk}, Σ''_{jk} are as before. The contribution to (3.27) from the term involving Σ'_{jk} can be shown to converge in probability in the same way as for $\overline{\theta}_N$.

Thus we need only consider

$$N^{-\frac{1}{2}}(N')^{-1}\sum_t \mathrm{tr}\left\{\psi_N(\omega_t)\ \sum_0^{\infty}\sum C_0(j)\ \Sigma''_{jk}C_0(k)'e^{i(j-k)\omega_t}\right\} = 2A + B \tag{3.28}$$

where

$$A = N^{-\frac{1}{2}}\sum_{n=0}^{N-1}(N')^{-1}\sum_t \mathrm{tr}\left\{\psi_N(\omega_t)\left[k_0\varepsilon(N-n)\varepsilon(N-n)'\sum_{n+1}^{\infty}C_0(k)^*e^{-ik\omega_t}\right]\right\}, \tag{3.29}$$

$$B = N^{-\frac{1}{2}}\sum_{n=0}^{N-1}(N')^{-1}\sum_t \mathrm{tr}\left\{\psi_N(\omega_t)\ .\right.$$

$$\left. .\left[\sum_{n+1}^{\infty}C_0(j)e^{ij\omega_t}\varepsilon(N-n)\varepsilon(N-n)'\sum_{n+1}^{\infty}C_0(k)^*e^{-ik\omega_t}\right]\right\}. \tag{3.30}$$

Now the expected value of the modulus of B is bounded above by

$$cN^{-\frac{1}{2}}\sum_0^{N-1}(N')^{-1}\sum_t \mathrm{tr}\big[E(n,\ N')K_0E(n,\ N')^*+E(n,\ N')K_0E(N',\ \infty)^*$$

$$+E(N',\ \infty)K_0E(n,\ N')^*+E(N',\ \infty)K_0E(N',\ \infty)^*\big] \tag{3.31}$$

where c is an upper bound for $\|\psi_N(\omega)\|$ and

$$E(a, b) = \sum_{j=a+1}^{b} C_0(j)e^{ij\omega_t} .$$

Now the term corresponding to $E(n, N')K_0E(n, N')^*$ in (3.31) converges to zero as in the proof for $\overline{\theta}_N$. The term corresponding to $E(n, N')K_0E(N', \infty)^*$ has modulus bounded above by

$$c.N^{-\frac{1}{2}}.\|K_0\| . \sup_{\omega} \left\|\sum_{N'}^{\infty} C_0(k)e^{ik\omega_t}\right\|^2 . \sum_{0}^{N-1} \left(\sum_{n+1}^{N'} \|C_0(j)\|^2\right)^{\frac{1}{2}}$$

$$\leq c.\|K_0\| . \sup_{\omega} \left\|\sum_{N'}^{\infty} C_0(k)e^{ik\omega_t}\right\|^2 . \left(\sum_{0}^{N-1} \sum_{n+1}^{\infty} \|C_0(j)\|^2\right)^{\frac{1}{2}} . \qquad (3.32)$$

But since k_0 has elements in Λ_α, $\alpha > \frac{1}{2}$, it follows by Theorem (10.8), p. 64 of Zygmund (1959) that

$$\sup_{\omega} \left\|\sum_{N'}^{\infty} C_0(k)e^{ik\omega_t}\right\|^2 = O\left((N')^{-\alpha} \log N'\right) .$$

Hence the r.h.s. of (3.32) converges to zero. The third term in (3.31) is exactly the same. The fourth term, corresponding to $E(N', \infty)K_0E(N', \infty)^*$ is handled by a similar technique. Hence $B \xrightarrow{P} 0$. The same methods are used to show that A may be considered with $\psi_N(\omega_t)$ replaced by $\exp(iu\omega_t)$, u finite. The proof that the resulting quantity converges to zero is almost the same as that given in the proof for $\overline{\theta}_N$. This completes the sketch proof for $\tilde{\theta}_N$. The proof that $N^{\frac{1}{2}}(\tilde{K}_N-\dot{K}_N) \xrightarrow{P} 0$ and the proof of the CLT for $N^{\frac{1}{2}}(\tilde{\mu}_N-\mu_0)$ are similar to those given in Theorem 3.1. □

REMARK. An alternative specification of the conditions of Corollary 3.2 is as follows. We could allow the rate at which N' increases with N to be, for example, of the form $N' = O(N^\beta)$ where $\beta \geq 1$. Then it is only required that $k(\theta_0)$ and $\partial k(\theta_0)/\partial\theta_j$ belong to Λ_α for $\alpha\beta > \frac{1}{2}$. We note however that in practice one would normally choose N' to be the greatest power of 2 not less than N . Then $N'/N \leq 2$, i.e. $\beta = 1$.

We finally discuss the procedure $\hat{L}_N$. Further smoothness conditions on $k\left(e^{i\omega};\, \theta_0\right)$ and $\partial k\left(e^{i\omega};\, \theta_0\right)/\partial\theta_j$ have been required in order that our proof holds. *Again these conditions are satisfied for the* ARMA *model.* Since $\hat{L}_N$ is difficult (if not impossible) to concentrate w.r.t. K even when μ is just the vector of upper triangular elements of K , we have given no corresponding result to *(b)* of Theorem 3.1. Obviously, however, if we let $\hat{K}_N$ be $\overline{K}_N$ where $\overline{\theta}_N$ is replaced by $\hat{\theta}_N$, then $\hat{K}_N$ satisfies *(b)* of Theorem 3.1.

COROLLARY 3.3. *Let the elements of* $k\left(e^{i\omega};\, \theta\right)$, $\partial k\left(e^{i\omega};\, \theta\right)/\partial\theta_j$, $1 \leq j \leq u$ *be differentiable with respect to* ω *with these derivatives belonging to* Λ_α , $\alpha > 0$. *Then* $N^{\frac{1}{2}}\left(\hat{\theta}_N-\theta_0\right)$, *and* $N^{\frac{1}{2}}\left(\hat{\mu}_N-\mu_0\right)$ *have the same asymptotic distributions as given in Theorem* 3.1.

Proof. To prove the result for $\hat{\theta}_N$ we are led to consider $N^{\frac{1}{2}} \left.\dfrac{\partial L_N(\theta,\hat{\mu}_N)}{\partial\theta_j}\right|_{\theta_0}$, $1 \leq j \leq u$. The proof consists in showing that

$$N^{\frac{1}{2}} \left.\frac{\partial \hat{L}_N(\theta,\hat{\mu}_N)}{\partial\theta_j}\right|_{\theta_0} - N^{\frac{1}{2}} \left.\frac{\partial \tilde{L}_N(\theta,\hat{\mu}_N)}{\partial\theta_j}\right|_{\theta_0} \xrightarrow{P} 0\ , \quad 1 \leq j \leq u\ . \tag{3.33}$$

Thus since the conditions of Corollary 3.3 are stronger than those of Corollary 3.2 and $\hat{\mu}_N \to \mu_0$ then

$$N^{\frac{1}{2}} \left.\frac{\partial \hat{L}_N(\theta,\hat{\mu}_N)}{\partial\theta_j}\right|_{\theta_0} \quad \text{and} \quad N^{\frac{1}{2}} \left.\frac{\partial \tilde{L}_N(\theta,\tilde{\mu}_N)}{\partial\theta_j}\right|_{\theta_0}$$

have the same asymptotic distribution as given by Corollary 3.2.

To establish (3.33) we write $\sigma_M\left(e^{i\omega};\, \theta_0\right)$ as the Cesaro sum to M terms of the Fourier series for $k^{-1}\left(e^{i\omega};\, \theta_0\right)$ and write

$\phi_M = 2\sigma_{2M-1} - \sigma_{M-1}$. It then follows from Theorems 13.5, 13.6, Zygmund (1959 p. 115) that if $M = N^{\beta}$, $\beta < \frac{1}{2}$, $\beta(1+\alpha) = \frac{1}{2} + \delta$, $\delta > 0$ we have

$$\sup_{\omega} \left\| k^{-1}\left(e^{i\omega};\ \theta_0\right) - \phi_M\left(e^{i\omega};\ \theta_0\right) \right\| = O\left(N^{-(\frac{1}{2}+\delta)}\right) \tag{3.34}$$

and

$$\sup_{\omega} \left\| \frac{\partial k^{-1}\left(e^{i\omega};\theta_0\right)}{\partial\theta_j} - \frac{\partial\phi_M\left(e^{i\omega};\theta_0\right)}{\partial\theta_j} \right\| = O\left(N^{-(\frac{1}{2}+\delta)}\right)\ . \tag{3.35}$$

We will write

$$\psi_N\left(e^{i\omega};\ \theta,\ \mu\right) = (2\pi)^{-1}\phi^{-1}_{N^{\beta}}\left(e^{i\omega};\ \theta\right)K^{-1}(\mu)\phi^{-1}_{N^{\beta}}\left(e^{i\omega};\ \theta\right)^{*}$$

and $\Gamma_N\left(\psi_N\right)$ as the $Ns \times Ns$ matrix with $\int e^{i(m-n)\omega}\psi_N d\omega$ in the $(n,\ m)$th block. The proof proceeds in four steps.

(i).

$$N^{\frac{1}{2}} \frac{\partial}{\partial\theta_j}\left[N^{-1} \log\det\Gamma_N\left(\theta,\ \hat{\mu}_N\right) - N^{-1} \log\det\Gamma_N\left(\psi_N\left(\theta,\ \hat{\mu}_N\right)\right)\right]\Big|_{\theta_0} \xrightarrow{\text{a.s.}} 0\ .$$

To show this we consider the modulus of the l.h.s. of (i) which equals

$$N^{-\frac{1}{2}}\ \mathrm{tr}\left\{\left[\Gamma_N^{-1}\left(\theta_0,\ \hat{\mu}_N\right) - \Gamma_N^{-1}\left(\psi_N\left(\theta_0,\ \hat{\mu}_N\right)\right)\right] \frac{\partial\Gamma_N\left(\theta,\hat{\mu}_N\right)}{\partial\theta_j}\Big|_{\theta_0}\right\}$$

$$+ N^{-\frac{1}{2}}\ \mathrm{tr}\left\{\Gamma_N^{-1}\left(\psi_N\left(\theta_0,\ \hat{\mu}_N\right)\right)\left[\frac{\partial\Gamma_N\left(\theta,\hat{\mu}_N\right)}{\partial\theta_j}\Big|_{\theta_0} - \frac{\partial\Gamma_N\left(\psi_N\left(\theta,\hat{\mu}_N\right)\right)}{\partial\theta_j}\Big|_{\theta_0}\right]\right\}.. \tag{3.36}$$

The first term in (3.36) is equal to

$$N^{-\frac{1}{2}}\ \mathrm{tr}\left\{\left[\Gamma_N^{-1}\left(\theta_0,\ \hat{\mu}_N\right) - \Gamma_N^{-1}\left(\psi_N\left(\theta_0,\ \hat{\mu}_N\right)\right) + bN^{-(\frac{1}{2}+\delta)}.I_{Ns}\right] \frac{\partial\Gamma_N\left(\theta,\hat{\mu}_N\right)}{\partial\theta_j}\Big|_{\theta_0}\right\}$$

$$- b.N^{-(1+\delta)}\mathrm{tr}\left\{\frac{\partial\Gamma_N\left(\theta,\hat{\mu}_N\right)}{\partial\theta_j}\Big|_{\theta_0}\right\} \tag{3.37}$$

where $0 \leq b < \infty$ is chosen so that

$$\Gamma_N^{-1}\left(\theta_0,\ \hat{\mu}_N\right) - \Gamma_N^{-1}\left(\psi_N\left(\theta_0,\ \hat{\mu}_N\right)\right) + b.N^{-(\frac{1}{2}+\delta)}I_{Ns} \tag{3.38}$$

is a positive definite matrix. This may always be done since

$$\Gamma_N^{-1}(\theta_0, \hat{\mu}_N) - \Gamma_N^{-1}(\psi_N(\theta_0, \hat{\mu}_N)) + c.N^{-(\frac{1}{2}+\delta)}\Gamma_N^{-1}(\theta_0, \hat{\mu}_N)\Gamma_N^{-1}(\psi_N(\theta_0, \hat{\mu}_N))$$

$$= \Gamma_N^{-1}(\theta_0, \hat{\mu}_N)\left[\Gamma_N(\psi_N(\theta_0, \hat{\mu}_N))-\Gamma_N(\theta_0, \hat{\mu}_N)+N^{-(\frac{1}{2}+\delta)}.c.I_{Ns}\right]\Gamma_N^{-1}(\psi_N(\theta_0, \hat{\mu}_N)) .$$

But the matrix in square brackets is positive definite for c appropriately chosen (by (3.34)) and

$$cN^{-(\frac{1}{2}+\delta)}\Gamma_N^{-1}(\theta_0, \hat{\mu}_N)\Gamma_N^{-1}(\psi_N(\theta_0, \hat{\mu}_N)) \le c.b_1^{-1}.b_2^{-1}.N^{-(\frac{1}{2}+\delta)}.I_{Ns}$$

where b_1 is a lower bound for the smallest eigenvalue of $\Gamma_N(\theta_0, \hat{\mu}_N)$ and b_2 is similarly for $\Gamma_N(\psi_N(\theta_0, \hat{\mu}_N))$. Since for N sufficiently large $b_1 > 0$ and $b_2 > 0$ (a.s.) we may choose b (in (3.38)) as $c.b_1^{-1}.b_2^{-1}$. Now using the positive definiteness of (3.38) in the first term of (3.34) a bound for this term is obtained as $\left(b_3.N^{-(1+\delta)}.N\right) \to 0$. The second term in (3.37) converges to zero since the norm of $\left.\frac{\partial\Gamma_N(\theta,\hat{\mu}_N)}{\partial\theta_j}\right|_{\theta_0}$ is $O(N)$.

Hence the first term in (3.36) converges a.s. to zero. The second term in (3.36) may be similarly treated but now b (similar to in (3.38)) is such that

$$\left.\frac{\partial\Gamma_N(\theta,\hat{\mu}_N)}{\partial\theta_j}\right|_{\theta_0} - \left.\frac{\partial\Gamma_N(\psi_N(\theta,\hat{\mu}_N))}{\partial\theta_j}\right|_{\theta_0} + bN^{-(\frac{1}{2}+\delta)}I_{Ns}$$

is positive definite. Again such a bound may be chosen by use of (3.35). Hence (i) is established.

(ii).

$$N^{\frac{1}{2}}\frac{\partial}{\partial\theta_j}\left(N^{-1}\log\det\Gamma_N(\psi_N(\theta, \hat{\mu}_N))-\log\det K(\hat{\mu}_N)\right)\Big|_{\theta_0} \xrightarrow{a.s.} 0 .$$

To establish this the relationship derived in the proof of Lemma 2.6 for $\det\Gamma_N(P)$ where P is the spectrum of an autoregression is used.

Differentiating this result we have

$$N^{-\frac{1}{2}} \frac{\partial}{\partial\theta_j} \left[\log \det \Gamma_N\left(\psi_N^{-1}(\theta, \hat{\mu}_N)\right)\right]\Big|_{\theta_0}$$

$$= N^{\frac{1}{2}} \frac{\partial}{\partial\theta_j} \log \det K\left(\psi_N^{-1}\right)\Big|_{\theta_0} + N^{-\frac{1}{2}}\left[\frac{\partial}{\partial\theta_j} \log \det \Gamma_{\psi_N^{-1}}(0)\Big|_{\theta_0}\right.$$

$$\left. + \frac{\partial}{\partial\theta_j} \log \det \hat{\Gamma}_2(0)\Big|_{\theta_0} + \ldots + \frac{\partial}{\partial\theta_j} \log \det \hat{\Gamma}_{N^\beta}(0)\Big|_{\theta_0}\right] . \qquad (3.39)$$

Now $\frac{\partial}{\partial\theta_j} \log \det K\left(\psi_N^{-1}\right)\Big|_{\theta_0} = \frac{\partial}{\partial\theta_j} \log \det K\left(\hat{\mu}_N\right)\Big|_{\theta_0} = 0$, and the second term on the r.h.s. of (3.39) has modulus which is $O\left(N^{-\frac{1}{2}}.N^\beta\right)$ (since there are N^β terms each of which has finite modulus). But $\beta < \frac{1}{2}$ so that (ii) is established.

(iii).

$$N^{\frac{1}{2}} \frac{\partial}{\partial\theta_j} \left\{N^{-1}x_N'\Gamma_N^{-1}(\theta, \hat{\mu}_N)x_N - N^{-1}x_N'\Gamma_N^{-1}(\psi_N(\theta, \hat{\mu}_N))x_N\right\}\Big|_{\theta_0} \xrightarrow{\text{a.s.}} 0 \,.$$

The l.h.s. of (iii) has modulus bounded by

$$N^{-\frac{1}{2}}\left\|\Gamma_N^{-1}(\theta_0, \hat{\mu}_N)\right\|.\left\|\Gamma_N^{-1}\left\{\psi_N^{-1}(\theta_0, \hat{\mu}_N)\right\}\right\|.\left|x_N'\left(\frac{\partial\Gamma_N(\theta,\hat{\mu}_N)}{\partial\theta_j}\Big|_{\theta_0} - \frac{\partial\Gamma_N(\psi_N(\theta,\hat{\mu}_N))}{\partial\theta_j}\Big|_{\theta_0}\right)x_N\right|$$

$$\le b_1.b_2.\left(N^{\frac{1}{2}} \operatorname{tr} G(0)\right) . \sup_\omega \left\|\frac{\partial f(\theta_0,\hat{\mu}_N)}{\partial\theta_j} - \frac{\partial\psi_N(\theta_0,\hat{\mu}_N)}{\partial\theta_j}\right\|$$

$$\le b_1.b_2.N^{-\delta} . \operatorname{tr} G(0) \xrightarrow{\text{a.s.}} 0 ,$$

again using (3.34) and (3.35).

(iv).

$$N^{\frac{1}{2}} \frac{\partial}{\partial\theta_j} \left\{N^{-1}x_N'\Gamma_N^{-1}(\psi_N(\theta, \hat{\mu}_N))x_N - (N')^{-1} \sum_t \operatorname{tr}\left(\psi_N(\theta, \hat{\mu}_N)I(\omega_t)\right)\right\}\Big|_{\theta_0} \xrightarrow{P} 0 .$$

The proof of this final step proceeds by use of the relationship for $x_N'\Gamma_N^{-1}(P)x_N$ (where P is the spectral density matrix for an autoregression) established in the proof of Lemma 2.5. Steps (i), (ii), (iii) and (iv) combined give the result (3.33) and thus the proof is complete. □

To complete this section a brief discussion of some previous treatments of the CLT for the general finite parameter model (1.1) will be given. In the first place the CLT's for $N^{\frac{1}{2}}(\hat{\theta}_N-\theta_0)$, $N^{\frac{1}{2}}(\tilde{\theta}_N-\theta_0)$ and $N^{\frac{1}{2}}(\overline{\theta}_N-\theta_0)$ given above are extensions to the vector case of the corresponding results in Hannan (1973a) for the scalar case. In the scalar case μ will usually be taken as σ^2 (used to denote K when $s = 1$) whereas in the vector case it may be that the elements of K are specified by a smaller set of parameters given by μ. In Walker (1964) the CLT for $N^{\frac{1}{2}}(\overline{\theta}_N-\theta_0)$ and $N^{\frac{1}{2}}\left(\overline{\sigma}_N^2-\sigma_0^2\right)$ is established (when $s = 1$) under stronger conditions than C3 above. In addition to the conditions imposed (see note C(iv) in §2) for the convergence of $\overline{\theta}_N$ to θ_0 in probability Walker assumes that derivatives with respect to θ of $k^{-1}\left(e^{i\omega};\ \theta\right)$ up to the third order exist and are continuous in $(\omega;\ \theta)$, that $\sum_{j=0}^{\infty} j\|C(j;\ \theta)\| < \infty$ (which implies C3.2) and that the $\varepsilon(m)$ are i.i.d. with finite fourth moment (in contrast to C3.3).

A quite different treatment to those just mentioned and that given above is given in Davies (1973). That article is concerned with establishing the asymptotic differentiability conditions of Le Cam (see Davies (1973) for the details of this theory) for the family of distributions associated with a stationary Gaussian vector time series in which the covariance structure is specified up to a finite number of parameters θ. We will discuss, albeit briefly, only a part of Davies' treatment which relates to the treatment given above. Included in that article is a discussion of estimation based on $\hat{L}_N$ and a variation to $\tilde{L}_N$. There the parameter θ is taken to specify $f(\omega;\ \theta)$ and the decomposition $(2\pi)^{-1}k\left(e^{i\omega};\ \theta\right)K(\mu)k\left(e^{i\omega};\ \theta\right)^*$ required by us has not been assumed. Letting

$\Delta_N(\theta)$ (in Davies' notation) be the vector of derivatives with respect to θ of $(2\sqrt{N})^{-1}\left\{\log\det\Gamma_N(\theta)+x_N'\Gamma_N^{-1}(\theta)x_N\right\}$ (i.e. $-N^{-\frac{1}{2}}\frac{\partial}{\partial\theta}\hat{L}_N(\theta)$ in our notation) then given a preliminary $\sqrt{N}$- consistent estimator $\dot{\theta}_N$ (i.e. $\sqrt{N}(\dot{\theta}_N-\theta_0)$ is bounded in probability) and letting $\Gamma_N = \dot{\theta}_N + N^{-\frac{1}{2}}\Omega^{-1}(\dot{\theta}_N)\Delta_N(\dot{\theta}_N)$ (where $\Omega(\dot{\theta}_N)$ is as in Theorem 3.1 with θ_0 replaced by $\dot{\theta}_N$) then Davies shows that $N^{\frac{1}{2}}(\Gamma_N-\theta_0)$ has the asymptotic distribution of Theorem 3.1 above (see part (a)). In particular when $\dot{\theta}_N = \hat{\theta}_N$ (i.e. $\Delta_N(\hat{\theta}_N) = 0$) then $\Gamma_N = \hat{\theta}_N$ so that $N^{\frac{1}{2}}(\hat{\theta}_N-\theta_0)$ also satisfies this asymptotic distribution provided $\hat{\theta}_N$ is $\sqrt{N}$-consistent (the $\sqrt{N}$-consistency is not established in Davies (1973)). These results hold when $x(n)$ is Gaussian and in particular no assumptions are made ooncerning the existence of derivatives of order higher than the first of $f(\theta)$ with respect to θ. Note that when the $x(n)$ are Gaussian the CLT for θ specifying both $K(\theta)$ and $k(e^{i\omega};\theta)$ given in Chapter 4 follows as a corollary to Theorem 4.2.1. Davies also gives some discussion to indicate that this normality assumption may not be required for his treatment when $x(n)$ is in the form (1.1) with θ specifying $C(j;\theta)$.

CHAPTER 3

THE STRONG LAW OF LARGE NUMBERS AND CENTRAL LIMIT THEOREM FOR ARMA MODELS

1. Introduction

In the preceding chapter the asymptotic theory was presented for estimation in a very general class of stationary time series for which the parametric structure was not of special form. In this chapter we will be solely concerned with applying and extending the theory of Chapter 2 to the important class of time series models with rational spectral density matrices corresponding to the ARMA models introduced in §1.5. This model was given in that section as

$$\sum_{j=0}^{q} B(j)x(n-j) = \sum_{j=0}^{p} A(j)\varepsilon(n-j) \ , \quad B(0) = A(0) = I_s \ , \tag{1.1}$$

$$E\varepsilon(m) = 0 \ , \quad E\varepsilon(m)\varepsilon(n)' = \delta_{mn}K \ .$$

Here $A(j)$, $B(j)$ and K are $s \times s$ matrices while $x(n)$ and $\varepsilon(n)$ are s-vectors. It will be convenient to introduce the notation

$$\theta = \begin{bmatrix} \tau \\ \cdots \\ \mu \end{bmatrix} \tag{1.2}$$

where

$$\tau = \text{vec}\{[B(1) \vdots \cdots \vdots B(q) \vdots A(1) \vdots \cdots \vdots A(p)]'\}$$

and

$$\mu' = (K_{11}, \ldots, K_{1s}, K_{22}, \ldots, K_{2s}, \ldots, K_{ss})$$

(μ is the listing of the upper triangular elements of K). Note then that $\theta \in \mathrm{R}^{(p+q)s^2} \times \mathcal{K}$ where $\mathcal{K}$ is the space of all $s \times s$ positive definite matrices with finite norm.

There are several reasons for inclusion of this separate chapter dealing with the asymptotic theory of estimation for (1.1). The first is that it is necessary to indicate how the abstract theory of Chapter 2 specialises to this case. In this regard the question of the choice of suitable topologies for the parameter space arises and this will be discussed in §3. Closely allied to this is the question of identification - i.e. the choice of a unique representative from the many structures of the form (1.1) which yield the same spectral density matrix. This will be discussed in §2. A second reason for this separate discussion of the ARMA model is that the conditions on the parameter space Θ of Chapter 2 (see conditions 2.C2) are sometimes stronger than required for the theory of (1.1). That is, the treatment of Chapter 2 would require that τ in (1.2) belongs to B, a sphere of known bounded radius in $R^{(p+q)s^2}$. This may be avoided in certain cases of the model (1.1). In this sense the treatment given here, while using very similar methods of proof, is more general than that of Chapter 2. However, in other ways, the opposite is true.

In §4 the strong law of large numbers will be presented. The result of this section establishes the convergence of (k_N, K_N) to (k_0, K_0) in a suitable pointwise sense (to be defined in §3). A result of §3 relates this mode of convergence to that of equivalence classes of θ-points. In §5 the central limit theorem for the model (1.1) is related to the CLT established in the last chapter. For this use is made of the result that a suitably identified parameter space is a manifold. The final section of this chapter, §6, is concerned with extending the results of §4 and §5 to the case where initial values in (1.1), i.e. $x(0), \ldots, x(-q+1)$, $\varepsilon(0), \ldots, \varepsilon(-p+1)$, are specified as fixed and known, for example at zero, or are taken to be parameters to be estimated. These "pre-period" specifications can lead to simplification of estimation, in some cases.

2. Some methods of identification in ARMA models

The spectral density matrix corresponding to (1.1) is (as introduced in §1.5)

$$f(\omega) = (2\pi)^{-1}k\left(e^{i\omega}\right)Kk\left(e^{i\omega}\right)^* \tag{2.1}$$

where k is the rational function of complex argument

$$k(z) = h^{-1}(z)g(z) \tag{2.2}$$

and the polynomials h and g are given by

$$h(z) = \sum_{j=0}^{q} B(j)z^j \ , \quad g(z) = \sum_{j=0}^{p} A(j)z^j \ . \tag{2.3}$$

To every rational spectral density matrix corresponding to a stationary process $x(n)$ there is a unique decomposition in the form (2.1) where

C2.1 $k(0) = I_s$, $\det k(z) \neq 0$ for $|z| < 1$, and $k(z)$ is holomorphic for $|z| \leq 1$.

(See Rozanov, 1967, pp. 47-48.) Then the $\varepsilon(n)$ are the one step prediction errors for predicting $x(n)$ from its past. In general there are many different structures of the type (1.1) giving the same $k(z)$ and thus for K fixed the same f . Thus since it is f , and hence (k, K) , that may be determined by the likelihood procedures $\hat{L}_N$, $\bar{L}_N$, $\tilde{L}_N$ of previous chapters, some method of choosing a canonical representation of h, g , and thus of (1.1), is required. We will discuss this in the remainder of this section.

The first form of redundancy that can arise is if h and g have a common left divisor that disappears when $h^{-1}g$ is formed. That is if e is a matrix of polynomials such that $h = eh_1$, $g = eg_1$ and h_1, g_1 are matrices of polynomials then e is termed a common left divisor of h and g . Then h is said to be a right multiple of e . The matrix e is said to be a greatest common left divisor (g.c.l.d.) for h and g if it is a common left divisor and any other common left divisor d has e as right multiple, i.e. $e = d.d_1$ (see MacDuffee, 1956, p. 35). Then any other g.c.l.d. of h and g, say e_1, is of the form $e_1 = eu$ where u is unimodular (i.e. is a matrix of polynomials in z with determinant a constant independent of z). Thus the first requirement for identifiability is

C2.2 g and h have I_s as g.c.l.d.

In this case g and h are said to be "left prime". The imposition of this condition is not sufficient to identify (1.1). Before proceeding however we will relate the condition C2.1 to

C2.3 $\det h(z) \neq 0$, $|z| \leq 1$, $\det g(z) \neq 0$, $|z| < 1$.

(We will refer to C2.3' when $\det g(z)$ is additionally required to be non-zero for $|z| = 1$.)

LEMMA 2.1. *When h and g have I_s as g.c.l.d. the condition* C2.1 *is equivalent to* C2.3.

Proof. That C2.3 implies C2.1 in the presence of C2.2 is clear. For the reverse implication we proceed as follows. By way of contradiction assume that there exists a zero, z_0 say, of $\det h(z)$ where $|z_0| \leq 1$. (The same argument below applies when $\det g(z)$ has such a zero.) By Smith's normal form (MacDuffee, 1956, p. 41) we may write $h = uDv$ where

$$D = \operatorname{diag}\left[d_{11}, \ldots, d_{ss}\right] , \quad d_{jj} \mid d_{j+1,j+1} \tag{2.4}$$

and u, v are unimodular. Now since $\det u$, $\det v$ are constant, if z_0 is a zero of $\det h(z)$ then $(z-z_0)^{p_i}$ occurs as a factor of d_{ii} where $p_i \leq p_{i+1}$ (and if $p_i = 0$, $(z-z_0)^{p_i} = 1$). It then follows that

$$D = D_1 D_2 \quad \text{where} \quad D_1 = \operatorname{diag}\left[(z-z_0)^{p_1}, \ldots, (z-z_0)^{p_s}\right] . \tag{2.5}$$

Hence $k = h^{-1}g = v^{-1}D_2^{-1}D_1^{-1}(u^{-1}g)$ so that

$$D_2 v k = D_1^{-1}(u^{-1}g) . \tag{2.6}$$

But k is holomorphic near z_0, by hypothesis, and $D_2 vk$ is thus also holomorphic near z_0. Hence from (2.6) it follows that $(z-z_0)^{p_i}$ divides the ith row of $u^{-1}g$ and hence there exists a matrix of polynomials g_1 such that $u^{-1}g = D_1 g_1$ and hence $g = (uD_1)g_1$. But also $h = (uD_1)(D_2 v) = (uD_1)h_1$ (say) so that h, g are not left prime because $e = uD_1$ is not unimodular. This is the required contradiction. □

The conditions C2.2 and C2.3 (equivalently C2.2 and C2.1) will be referred to as the "elementary conditions" for identification (Hannan (1975a)) and would appear to be the minimum conditions required. They also appear to fit any natural interpretation of the model (1.1) in that C2.2 eliminates redundancies while C2.3 (or C2.1) ensures that $x(n)$ is stationary, depends only on $\varepsilon(m)$, $m \leq n$, and these $\varepsilon(m)$ are the prediction errors. However more is needed for a canonical choice of a structure (1.1).

By the identification of the ARMA systems, (1.1), is meant a description of all equivalence classes of structures (1.1), two structures being equivalent if they give rise to the same likelihood on Gaussian assumptions (i.e. the same spectral density matrix $f(\omega)$, i.e. the same pair (k, K)). Identification is usually achieved by exhibiting a canonically chosen member of each equivalence class. Of course one canonical choice may be better than another for reasons connected with over identifying constraints for example. (For a discussion see Hannan (1975a).) By a slight abuse of language we will say that a particular structure, (1.1), is identified if it is so constrained by requirements (such as C2.2, C2.3) that there is no other structure satisfying these same requirements giving rise to the same k, K. In this regard consider the condition

C2.4 $[A(p) \vdots B(q)]$ is of rank s.

DEFINITION 2.1. *The structure* (1.1) *is said to be "simply identified" if* C2.2, C2.3 *and* C2.4 *are satisfied.*

Hannan (1969) shows that for p, q fixed a simply identified structure *is* identified in the above sense. This means that Definition 2.1 is reasonable. Simple identification is "over identifying" in the sense that there are equivalence classes with no elements satisfying the simple identification requirements. (In general we will use the terminology "the model (1.1) is over identified" when certain equivalence classes are excluded.) In connection with Definition 2.1 we may consider the following method of identifying the class of *all* ARMA structures (with p, q varying but s fixed). First consider all structures with degrees of h, g prescribed as q, p respectively. Divide these into equivalence classes. The same general equivalence class may be represented more than once (i.e. for different p, q). Choose the representative for which p is smallest and q is smallest for that p . This clearly chooses a unique representative of each overall equivalence class. The method seems to correspond to what might be done in practice since usually for estimation it is desirable to keep p (i.e. the number of moving average matrices) small. The method also appears to have, at first sight, virtues. The first is that only two integer parameters occur, namely p and q , as compared to up to $(s+1)$ in other methods (see, e.g. "triangular identification" below). Since the physical meaning of the model may lead to a specification other than, say, in triangular form (again see below) it will be best to avoid considering transformations to such forms. Also over identifying constraints may become unrecognisable when a special form for the model is required. The second virtue of the method being considered is that it does not force a special structure on h and g . The first virtue (although not the second) is somewhat illusory because the structure of the equivalence classes for *fixed* p, q is so complicated (see Examples 3.3, 3.4 below). The space of all such equivalence classes for fixed p, q is *not* a manifold but rather a very complicated union of manifolds of different dimensionality. Integer parameters (other than p, q) implicitly occur which delineate the manifold that the true structure belongs to.

However it will be shown in §5 that the interior Θ_{01} (where also $\det g(z) \neq 0$, $|z| = 1$) of Θ_1 (the set of all simply identified structures for p, q fixed) is a manifold and constitutes "almost all" of the space of equivalence classes for fixed p, q in the sense that all of the finite number of other manifolds are of lower dimension. (Again see Examples 3.3, 3.4.) *A priori* the true parameter point may be expected to lie in Θ_1 , but more importantly it is not likely that the maximum of the likelihood will lie outside Θ_1 . Thus it may not be unreasonable to assume that $\theta \in \Theta_1$ for any fixed p, q . In that case the "natural" parameters which are the elements of the $A(j)$, $B(j)$, K (see $\theta' = (\tau', \mu')$ in (1.2)) can be used as coordinates. Because of this and because this seems the most likely coordinatisation of the problem to be used in practice a large proportion of the results to follow in this chapter will be concerned with Θ_1 (and Θ_{01}).

We shall also discuss estimation corresponding to other methods of identification , one of which we define as follows (Hannan, 1971b).

DEFINITION 2.2. *The structure* (1.1) *is said to be "triangularly identified" if* C2.2 *and* C2.3 *and*

C2.5 $h(z)$ *is lower triangular with* $h(0) = I_s$ *and the elements below the main diagonal have no higher degree than those on the diagonal in the same column*

are satisfied.

The virtue of this method of identification is that it chooses a unique member from each equivalence class. Its vice is that more integer parameters have to be specified - that is, the degrees q_j of each diagonal element in h as well as the degree p of g have to be considered. A second difficulty with triangular identification is that constraints on the original coefficient matrices $A(j)$, $B(k)$ which correspond to a physical meaning may become unrecognisable in the canonical form. Simple identification does not suffer so much in this way.

A further method that has been proposed to identify (1.1) is that given in Zellner and Palm (1974). In this $h(z)$ is taken to be scalar. (This may be achieved by factoring k as $(\det h)^{-1}(Hg)$ where H is the adjoint of h.) Then it must be required that any common factor in $\det h(z)$ with all elements of g are cancelled out. This method will be referred to as "scalar identification". The defects of this method are that again the physical meaning of constraints on the original structure (which is not in this scalar form) may be obscured and secondly that this method of identification introduces a greater number of parameters than, e.g. simple identification. Hannan (1976a) discusses all three types of identification.

The model (1.1) as it stands exhibits no constraints on the elements of $A(j)$, $B(k)$, and K. However as mentioned in Chapter 1 such constraints will arise naturally in many situations. For example it is often the case in econometrics that elements in the $B(j)$ may be required *a priori* to be zero. Another example is the familiar econometric model of the form (1.1) wherein $B(0) \neq I_s$. These restrictions may be replaced by other restrictions on the $B(1), \ldots, B(q)$ and $A(1), \ldots, A(p)$ when the model (1.1) with $B(0) \neq I_s$ is replaced by (1.1) for which $B(0) = I_s$. Further examples where constraints naturally arise occur when a discrete time approximation to processes arising from systems of differential equations or when seasonal ARMA models are used in which some of the $B(j)$ at certain lags j are required to be null. For such situations we may often put the restrictions in the form

$$\Phi_l(\tau, \mu) = 0 \ , \quad 1 \le l \le r \ , \tag{2.7}$$

and it is this type of equality constraint which will be mainly discussed below. Inequality constraints of the form

$$\Phi_l(\tau, \mu) > 0 \ , \quad 1 \le l \le r \tag{2.8}$$

may also be dealt with in the following treatment. The restrictions (2.7) may be over identifying restrictions since they may exclude certain equivalence classes. The functions Φ_l may be quite general and requirements, such as continuity of Φ_l, will be introduced when needed below.

3. Topological considerations in parameterising the ARMA model

In this section suitable parameter spaces and associated topologies will be introduced in order that the asymptotic theory of estimation of (1.1) may be discussed. Initially the unconstrained situation will be discussed with constraints of the type (2.7) being introduced later. Since for estimation the degree of h and g must be specified as being no greater than q and p respectively we initially introduce the space

$$\Theta = \{\theta \in R^{(p+q)s^2} \times K : \det h(z;\, \theta) \neq 0,\ |z| \le 1 \text{ and } \det g(z;\, \theta) \neq 0,\ |z| < 1\} \tag{3.1}$$

where we have used the notation $h(z;\, \theta)$, $g(z;\, \theta)$ for the matrices of polynomials in (2.3) corresponding to the point $\theta' = (\tau', \mu')$ defined in (1.2). (It is convenient to have such a notation for the constrained situation to be introduced below.) The points in Θ giving rise to the same spectral density matrix can now be grouped into equivalence classes. (In the unconstrained case the points θ for μ fixed in (1.2) giving rise to the same matrix function $k(z;\, \theta) = h(z;\, \theta)^{-1} g(z;\, \theta)$ are grouped into equivalence classes.) The set of all equivalence classes will be denoted by $[\Theta]$ and a typical element of this set as $[\theta]$. The notation $k(z;\ \theta)$ will mean the matrix $k(z)$ evaluated at any member of $[\theta]$. The subset of Θ (and $[\Theta]$) on which $\det g(z;\ \theta) \neq 0$ for $|z| \leq 1$ will be referred to as Θ_0 (and $[\Theta_0]$). In order that the convergence of estimators of θ may be discussed it is necessary to define ways in which a sequence in $[\Theta]$ may converge.

DEFINITION 3.1. $[\theta_n]$ *is said to converge pointwise to* $[\theta]$ (*written* $[\theta_n] \xrightarrow{p_t} [\theta]$) *if* $K(\theta_n) \to K(\theta)$ *and* $k(z;\ \theta_n) \to k(z;\ \theta)$ *for all* $|z| < 1$.

The notation $(k_n,\ K_n) \xrightarrow{p_t} (k,\ K)$ will also be used when referring to the convergence just defined.

A second mode of convergence is related to the "quotient" or "identification" topology induced as follows. The equivalence relation (of points in Θ giving the same f) induces a mapping $\chi : \Theta \mapsto [\Theta]$ called the canonical map where $\chi(\theta) = [\theta]$ and $[\theta]$ is the equivalence class of points in Θ for which θ is a representative. Then a subset Y of $[\Theta]$ is defined as open if $\chi^{-1}(Y)$ is open in Θ (see Ward, 1972, p. 62, for example). The quotient topology is the topology of convergence of equivalence classes and may be interpreted as follows.

DEFINITION 3.2. $[\theta_n]$ *converges to* $[\theta]$ *in the quotient topology* (*written* $[\theta_n] \xrightarrow{Q} [\theta]$) *whenever there exists a point* $\dot{\theta}_n \in \chi^{-1}[\theta_n]$ *such that* $k(z;\ \theta_n) = h^{-1}(z;\ \dot{\theta}_n)g(z;\ \dot{\theta}_n)$ *and a point* $\ddot{\theta}_n \in \chi^{-1}[\theta]$ *such that* $k(z;\ \theta) = h^{-1}(z;\ \ddot{\theta}_n)g(z;\ \ddot{\theta}_n)$ *for which* $\dot{\theta}_n - \ddot{\theta}_n \to 0$.

Note that the representatives of both $\chi^{-1}[\theta_n]$ *and* $\chi^{-1}[\theta]$ are chosen to depend on n. To illustrate Definition 3.2 consider the following.

EXAMPLE 3.1. Let $p = q = s = 1$. Let α, β, σ^2 be as described in c(ii) of Chapter 2 (Equation 2.17) with $\sigma^2 = 1$. Then an open neighbourhood of $[(0, 0)]$ (the equivalence class of structures for which $\alpha = \beta$) is $Y = \{(\alpha, \beta) : |\alpha-\beta| < \varepsilon, \alpha \neq \beta\} \cup \{(0, 0)\}$. This is so because the inverse image $\chi^{-1}(Y) = \{(\alpha, \beta) : |\alpha-\beta| < \varepsilon\}$ is an open subset of $\Theta = \{(\alpha, \beta) : |\alpha| \leq 1, |\beta| < 1\}$. A sequence of points θ_n in Θ may search along the line $\alpha = \beta$ as $[\theta_n] \xrightarrow{Q} [(0, 0)]$.

It is of interest to relate the two modes of convergence afforded by Definitions 3.1 and 3.2. Before doing so the following notation will be used. When A_n is a matrix then we will write $A_n \to A$ to mean $\|A_n-A\| \to 0$ where the norm is any matrix norm. When x_n is a vector in R^t we will write $x_n \to x$ to mean $x_n \to x$ in the Euclidean topology on R^t. The notation $\xrightarrow{P_t}$ and $\xrightarrow{Q}$ will always be used when referring to convergence in the sense of Definitions 3.1 and 3.2 respectively.

THEOREM 3.1. *Let* $[\theta_n]$ *be a sequence of points in* $[\Theta]$ *and let* $[\theta]$ *belong to* $[\Theta]$. *Then* $[\theta_n] \xrightarrow{P_t} [\theta]$ *implies* $[\theta_n] \xrightarrow{Q} [\theta]$.

Proof. The proof proceeds by constructing preimages $\dot{\theta}_n$ of $[\theta_n]$ and $\ddot{\theta}_n$ of $[\theta]$ belonging to Θ for which $\dot{\theta}_n - \ddot{\theta}_n \to 0$ and for which $k(z; \theta_n) = h^{-1}(z; \dot{\theta}_n)g(z; \dot{\theta}_n)$, $k(z; \theta) = h^{-1}(z; \ddot{\theta}_n)g(z; \ddot{\theta}_n)$. For brevity it will be convenient to let $k_n(z) = k(z; \theta_n)$, $k(z) = k(z; \theta)$, $\dot{h}_n(z) = h(z; \dot{\theta}_n)$, $\dot{g}_n(z) = g(z; \dot{\theta}_n)$, $\ddot{h}_n(z) = h(z; \ddot{\theta}_n)$, $\ddot{g}_n(z) = g(z; \ddot{\theta}_n)$, $K_n = K(\theta_n)$ and $K = K(\theta)$. Now, for $|z| < 1$, $k_n(z)$ and $k(z)$ have the one-sided representations

$$k_n(z) = \sum_{j=0}^{\infty} C_n(j)z^j \ , \quad k(z) = \sum_{j=0}^{\infty} C(j)z^j$$

where for $j \geq 0$ and $\rho < 1$, $C_n(j) = \rho^{-j} \int k_n(\rho e^{i\omega})e^{-ij\omega}d\omega$,

$C(j) = \rho^{-j} \int k(\rho e^{i\omega})e^{-ij\omega}d\omega$ and $C_n(j) = C(j) = 0$ if $j < 0$. But

$k_n(z) \to k(z)$ for $|z| < 1$ so that $k_n(\rho e^{i\omega}) \to k(\rho e^{i\omega})$ uniformly in

$\omega \in [-\pi, \pi]$ for $\rho < 1$ and hence $C_n(j) \to C(j)$ for each j. Now to construct decompositions of a given k we may consider solving the following equations for $\check{B}(u)$, $u = 1, 2, \ldots, q$, and $\check{A}(u)$, $u = 1, \ldots, p$:

$$\sum_{u=0}^{q} \check{B}(u)C(v-u) = \check{A}(v) \ , \quad v = 1, 2, \ldots, p+qs \tag{3.2}$$

where for $v - u < 0$, $C(v-u) = 0$ and for $v < 0$ or $v > p+qs$, $\check{A}(v) = 0$. Of course $\check{B}(0)$ and $\check{A}(0)$ are taken to be I_s. Now equations (3.2) have at least one solution for each $[\theta] \in [\Theta]$ since $k(z;\ \theta) = h^{-1}(z;\ \theta)g(z;\ \theta)$ for at least one point $\theta \in \chi^{-1}[\theta]$. However (3.2) may have many solutions. In order that any particular solution gives a decomposition of k as $\check{h}^{-1}\check{g}$ where $\check{h}(z) = \sum_0^q \check{B}(u)z^u$, $\check{g}(z) = \sum_0^p \check{A}(u)z^u$ it is necessary that the relation (3.2) holds for *all* v in the range $-\infty < v < \infty$. (Then $\check{h}k = \check{g}$.) Now (3.2) clearly holds for $v < 0$ since the only $C(j)$ appearing on the l.h.s. are null. Similarly $\check{A}(v) = 0$, $v < 0$. For $v = 0$ the l.h.s. gives $\check{B}(0)C(0) = I_s = \check{A}(0)$ (the r.h.s.). To extend (3.2) to all $v > p+qs$ consider first of all $v = p + qs + 1$. For this v the l.h.s. of (3.2) is

$$\sum_{u=0}^{q} \check{B}(u)C(p+qs+1-u) = \sum_{j=0}^{qs} \sum_{u=0}^{q} \check{B}(u)C(p+qs+1-u-j)b(j) \tag{3.3}$$

where $\det h(z) = \sum_0^{qs} b(j)z^j$ (hence $b(0) = 1$) where $h(z)$ corresponds to any preimage of $[\theta]$. This last line follows since

$$\sum_{u=0}^{q} \check{B}(u)C(p+qs+1-u-j) = 0$$

for $j = 0, \ldots, qs$ by equations (3.2) for $v = p+1, \ldots, p+qs$. But the r.h.s. of (3.3) is, where $\rho < 1$,

$$\rho^{-(p+qs+1)} \int \check{h}\left(\rho e^{i\omega}\right) H\left(\rho e^{i\omega}\right) g\left(\rho e^{i\omega}\right) e^{-i(p+qs+1)} d\omega \tag{3.4}$$

where $H(z)$ is the adjoint of $h(z)$ which is of degree $q(s-1)$ at most. But then $\check{h}Hg$ is of degree $q + q(s-1) + p = p + qs$ at most and therefore the integral in (3.4) vanishes. Hence (3.2) extends to $v = p + qs + 1$. Repeating the above steps for successively increased v establishes the required extension of (3.2) to all v . Now the equations (3.2) may be rewritten in the form

$$R\check{\tau} = r \tag{3.5}$$

where $\check{\tau}$ is as defined in (1.2) and

$$R = I_s \otimes \left[\begin{array}{ccc:c} C(0)' & \cdots & C(-q+1)' & \\ \vdots & (ps\times qs) & \vdots & -I_{ps} \\ C(p-1)' & \cdots & C(p-q)' & \\ \hdashline C(p)' & \cdots & C(p-q+1)' & \\ \vdots & \left(qs^2\times qs\right) & \vdots & 0_{qs^2\times ps} \\ C(p+qs-1) & \cdots & C(p)' & \end{array}\right], \quad r = \text{vec}\begin{bmatrix} -C(1)' \\ \vdots \\ -C(p+qs)' \end{bmatrix}$$

where $0_{qs^2\times ps}$ denotes the $qs^2 \times ps$ null matrix. Note that R is of dimension $(p+qs)s^2 \times (p+q)s^2$ and r is of dimension $(p+qs)s^2$. Let R_n, r_n and $\dot{\tau}_n$ correspond to (3.5) when the $C_n(j)$ are obtained from k_n and let R, r and $\ddot{\tau}_n$ correspond to (3.5) when k is used. To continue with the proof it will be convenient to reduce R_n and R to canonical form. This reduction is facilitated by the following lemma the proof of which will be omitted since it is relatively straight forward.

LEMMA 3.2. *Let* A_n *be a sequence of* $t \times t$ *symmetric non-negative* *definite matrices such that* $\|A_n - A\| \to 0$. *Then there exist* $t \times t$ *orthogonal matrices* $\dot{P}_n$ *and* $\ddot{P}_n$ *such that* $\|\dot{P}_n - \ddot{P}_n\| \to 0$ *and for which* $\dot{P}_n A_n \dot{P}_n = \Lambda_n$, $\ddot{P}_n A \ddot{P}_n = \Lambda$ *where* Λ_n *is the diagonal matrix of eigenvalues of* A_n *ranked in non-increasing order and* Λ *similarly corresponds to* A .

Note that some of the eigenvalues of A may be zero and since $A_n \to A$ the number of non-zero eigenvalues of A_n must eventually (i.e. for n large) be no less than the number of non-zero eigenvalues of A . Now returning to construction of solutions $\dot{\tau}_n$ and $\ddot{\tau}_n$ satisfying

$$R_n \dot{\tau}_n = r_n \;, \quad R\ddot{\tau}_n = r \tag{3.6}$$

consider first of all $R_n R_n^*$ which converges to RR^* and $R_n^* R_n$ which converges to R^*R . Then by Lemma 3.2 there exist orthogonal $\dot{U}_n$, $\ddot{U}_n$, $\dot{V}_n$, $\ddot{V}_n$ such that $\dot{U}_n - \ddot{U}_n \to 0$, $\dot{V}_n - \ddot{V}_n \to 0$ and for which all of $\dot{U}_n^* R_n R_n^* \dot{U}_n$, $\ddot{U}_n^* RR^* \ddot{U}_n$, $\dot{V}_n^* R_n^* R_n \dot{V}_n$ and $\ddot{V}_n^* R^* R \ddot{V}_n$ are diagonal. If the jth non-zero diagonal element of $\dot{U}_n^* R_n R_n^* \dot{U}_n$ is $\mu_n^2(j)$ for example then

$$\dot{U}_n^* R_n \dot{V}_n = \left[\begin{array}{ccc:c} \mu_n(1) & & & \\ & \ddots & & 0 \\ & & \mu_n(m_n) & \\ \hdashline & 0 & & (p+qs)s^2-m_n \\ & & & \times(p+q)s^2 \end{array}\right], \quad \ddot{U}_n^* R \ddot{V}_n = \left[\begin{array}{ccc:c} \mu(1) & & & \\ & \ddots & & 0 \\ & & \mu(m) & \\ \hdashline & 0 & & (p+qs)s^2-m \\ & & & \times(p+q)s^2 \end{array}\right]$$

where $m_n \geq m$ for n sufficiently large and, for example, $\mu_n(j)$ is taken to be the positive square root of $\mu_n^2(j)$. Then equations (3.6) may be rewritten

$$\begin{bmatrix} \begin{matrix}\mu_n(1) & & \\ & \ddots & \\ & & \mu_n(m_n)\end{matrix} & \vdots \; 0 \\ \cdots\cdots\cdots\cdots & \cdots \\ 0 & \end{bmatrix} \left(\dot{V}_n^* \dot{\tau}_n\right) = \left(\dot{U}_n^* r_n\right)$$

$$\begin{bmatrix} \begin{matrix}\mu(1) & & \\ & \ddots & \\ & & \mu(m)\end{matrix} & \vdots \; 0 \\ \cdots\cdots\cdots\cdots & \cdots \\ 0 & \end{bmatrix} \left(\ddot{V}_n^* \ddot{\tau}_n\right) = \left(\ddot{U}_n^* r\right) . \tag{3.7}$$

Now the (m_n+1)th to $(p+q)s^2$ elements of both $\dot{U}_n^* r_n$ and $\ddot{U}_n^* r$ are zero so that the corresponding elements of $\dot{V}_n^* \dot{\tau}_n$ and $\ddot{V}_n^* \ddot{\tau}_n$ may be chosen as zero also. Furthermore the $(m+1)$th to m_nth elements $(\text{when } m_n > m)$ of $\ddot{V}_n^* \ddot{\tau}_n$ are arbitrary and these may be chosen therefore as the corresponding elements of $\dot{V}_n^* \dot{\tau}_n$. Finally the first m elements of both $\dot{V}_n^* \dot{\tau}_n$ and $\ddot{V}_n^* \ddot{\tau}_n$ are uniquely determined by (3.7) and since $\mu_n(j) \to \mu(j)$ (because $R_n \to R$) it follows that $\dot{V}_n^* \dot{\tau}_n - \ddot{V}_n^* \ddot{\tau}_n \to 0$. Hence

$$\begin{aligned} \dot{\tau}_n - \ddot{\tau}_n &= \dot{V}_n\left(\dot{V}_n^* \dot{\tau}_n\right) - \ddot{V}_n\left(\ddot{V}_n^* \ddot{\tau}_n\right) \\ &= \dot{V}_n\left[\dot{V}_n^* \dot{\tau}_n - \left(\dot{V}_n^* \ddot{V}_n\right)\left(\ddot{V}_n^* \ddot{\tau}_n\right)\right] \end{aligned}$$

converges to zero since $\dot{V}_n^* \ddot{V}_n \to I_{(p+q)s^2}$. This completes the proof. □

The converse of Theorem 3.1 is not in general true as the following example illustrates.

EXAMPLE 3.2. Let $s = 2$, $p = q = 1$. Call $A(1) = A$, $B(1) = B$ for convenience. Let $[\theta]$ be the equivalence class of points yielding

$k(z;\,\theta) = I_2$ and take $K = I_2$ to avoid unnecessary complication. Also let $K_n = I_2$ and define

$$A_n = \begin{bmatrix} -n & -n \\ n & n \end{bmatrix} \qquad \text{so that} \quad g(z;\,\theta_n) = \begin{bmatrix} 1-nz & -nz \\ nz & 1+nz \end{bmatrix}$$

$$B_n = \begin{bmatrix} -(n+a_n) & -n \\ n & (n+b_n) \end{bmatrix} \qquad \text{so that} \quad h(z;\,\theta_n) = \begin{bmatrix} 1-(n+a_n)z & -nz \\ nz & 1+(n+b_n)z \end{bmatrix}$$

where $a_n \to 0$, $na_n \to a \neq 0$, $b_n \to 0$, $nb_n \to b \neq 0$ and a, b are chosen so that $1 - (a+b)z^2 \neq 0$ for $|z| \le 1$. (For example if $a_n = \frac{1}{2n}$, $b_n = \frac{1}{4n}$ then $(a+b) = 3/4$ and $1 - 3/4\, z^2 = 0$ when $z = \pm\sqrt{4/3}$.) Now

$$k(z;\,\theta_n) = \frac{\begin{bmatrix} 1+(n+b_n)z & nz \\ -nz & 1-(n+a_n)z \end{bmatrix}\begin{bmatrix} 1-nz & -nz \\ nz & 1+nz \end{bmatrix}}{\left\{1+(b_n-a_n)z-(na_n+nb_n+a_nb_n)z^2\right\}}$$

$$= \frac{\begin{bmatrix} 1+b_nz-nb_nz^2 & -nb_nz^2 \\ -na_nz^2 & 1-a_nz-na_nz^2 \end{bmatrix}}{\left\{1+(b_n-a_n)z-(na_n+nb_n+a_nb_n)z^2\right\}}$$

$$\to \frac{\begin{bmatrix} 1-bz^2 & -bz^2 \\ -az^2 & 1-az^2 \end{bmatrix}}{1-(a+b)z^2} \quad \text{as} \quad n \to \infty .$$

But there exist the following decompositions of $k(z;\,\theta)$:

$$h(z;\,\dot{\theta}_n) = g(z;\,\dot{\theta}_n) = \begin{bmatrix} 1-nz & -nz \\ nz & 1+nz \end{bmatrix}$$

and clearly $h(z;\,\theta_n) - h(z;\,\dot{\theta}_n) \to 0$, $g(z;\,\theta_n) - g(z;\,\dot{\theta}_n) \to 0$ so that $[\theta_n] \xrightarrow{Q} [\theta]$. But clearly, from the above, $k(z;\,\theta_n) \not\to I_2 = k(z;\,\theta)$.

That is $[\theta_n] \overset{P_t}{\not\longrightarrow} [\theta]$.

This example, although not the simplest of its type has been used because it will be referred to in the next section in the context of the strong consistency of the estimators. Note also that the sequence $[\theta_n]$ is a sequence of "one-point" equivalence classes since conditions C2.2, C2.3 and C2.4 are satisfied. For example, in C2.4, the matrix (for $p = q = 1$)

$$[A(p) \vdots B(q)] = \begin{bmatrix} -n & -n & \vdots & -(n+a_n) & -n \\ & & \vdots & & \\ n & n & \vdots & n & (n+b_n) \end{bmatrix}$$

is of full rank for $a_n = \frac{1}{2n}$, $b_n = \frac{1}{4n}$, as in the above. The problem with obtaining the implication: $[\theta_n] \xrightarrow{Q} [\theta]$ implies $[\theta_n] \xrightarrow{p_t} [\theta]$: arises in the above example from the fact that, although all points $[\theta_n]$ belong to $[\Theta]$ they correspond to $A(\theta_n)$, $B(\theta_n)$ which become unbounded as $n \to \infty$. Thus if we denote by $\Theta_{h,g}$ (and $[\Theta_{h,g}]$) the subset of Θ (and $[\Theta]$) such that the elements of $A(1), \ldots, A(p)$ and $B(1), \ldots, B(q)$ are uniformly bounded (i.e. τ in (1.2) is constrained to be in a ball of finite radius in $R^{(p+q)s^2}$) then

COROLLARY 3.3. *When* $[\theta_n]$ *is a sequence in* $[\Theta_{h,g}]$ *and* $[\theta] \in [\Theta_{h,g}]$ *then* $[\theta_n] \xrightarrow{Q} [\theta]$ *is equivalent to* $[\theta_n] \xrightarrow{p_t} [\theta]$.

Proof. All that needs to be demonstrated is that $[\theta_n] \xrightarrow{Q} [\theta]$ implies $[\theta_n] \xrightarrow{p_t} [\theta]$. Now for any z such that $|z| < 1$, and $h^{-1}(z;\, \ddot{\theta}_n)g(z;\, \ddot{\theta}_n)$ any decomposition of $k(z;\, \theta)$ satisfying Definition 3.2 we have

$$k(z;\, \dot{\theta}_n) - k(z;\, \ddot{\theta}_n)$$

$$= h^{-1}(z;\, \dot{\theta}_n)g(z;\, \dot{\theta}_n) - h^{-1}(z;\, \ddot{\theta}_n)g(z;\, \ddot{\theta}_n)$$

$$= \left[h^{-1}(z;\, \dot{\theta}_n)-h^{-1}(z;\, \ddot{\theta}_n)\right]g(z;\, \dot{\theta}_n) + h^{-1}(z;\, \ddot{\theta}_n)\left(g(z;\, \dot{\theta}_n)-g(z;\, \ddot{\theta}_n)\right) . \quad (3.8)$$

Now $\|g(z;\, \dot{\theta}_n)\| \le \sum_{j=0}^{p} \|A(j;\, \dot{\theta}_n)\| \le b < \infty$ by assumption. Similarly $\|h(z;\, \dot{\theta}_n)\| \le b < \infty$. Also

$$\left\|h^{-1}(z;\, \dot{\theta}_n)-h^{-1}(z;\, \ddot{\theta}_n)\right\| \le \left\|h^{-1}(z;\, \dot{\theta}_n)\right\| \|h(z;\, \ddot{\theta}_n)-h(z;\, \dot{\theta}_n)\| \left\|h^{-1}(z;\, \ddot{\theta}_n)\right\| .$$

But $\left\|h^{-1}(z;\, \dot{\theta}_n)\right\|$ is uniformly bounded (for z fixed) since $h^{-1}(z;\, \dot{\theta}_n) = H(z;\, \dot{\theta}_n)/\det\left(h(z;\, \dot{\theta}_n)\right)$ where H is the adjoint of h . Now the elements of H are bounded since they are determined as continuous functions of the elements in the coefficient matrices of h . Also $|\det h(z;\, \dot{\theta}_n)|$ is bounded away from zero since for any $\theta \in \Theta$ $\det h(z;\, \theta)$ has all zeros outside the unit circle and the point z is strictly inside the unit circle. Similarly $\left\|h^{-1}(z;\, \ddot{\theta}_n)\right\|$ is bounded in n . Since $h(z;\, \ddot{\theta}_n) - h(z;\, \dot{\theta}_n)$ converges to the null matrix the first term on the r.h.s. of (3.8) converges to the null matrix. Similarly the second term of (3.8) converges to the null matrix. □

In particular when $s = 1$ (i.e. the scalar case) the quotient convergence and pointwise convergence are equivalent since the condition 2C.2 (concerning the zeros of $h(z)$ and $g(z)$) ensures that the coefficients in h and g lie in a compact subset of $R^{(p+q)}$.

A further space of interest is that in which the simple identification conditions are satisfied. Define

$$\Theta_1 = \{\theta \in \Theta : \text{C2.2, C2.3, C2.4 are satisfied}\} . \quad (3.9)$$

Then each point in Θ_1 is not equivalent to any other point in Θ_1 so that the notation $[\Theta_1]$ is redundant. This set will be of importance in our discussion of the SLLN in the next section. The interior of Θ_1 is the intersection of Θ_0 with Θ_1 and we denote this by Θ_{01} (recall that this is the set of all points for which the simple identification conditions hold and in particular for which C2.3 is strengthened to C2.3'). We will show in §5 that Θ_{01} is a manifold,a requirement that must be met for the CLT of Chapter 2 to be applied to this case. Some examples may clarify the above.

EXAMPLE 3.3. For the space $[\Theta]$ of Example 3.1 $(p = q = s = 1)$ the subset Θ_1 is just $\{(\alpha, \beta) : |\alpha| \leq 1, |\beta| < 1, \alpha \neq \beta\}$ while Θ_{01} excludes points at which $|\alpha| = 1$.

EXAMPLE 3.4. $p = q = 1$, $s = 2$. Here $M = R^8 \times K$ and Θ may be decomposed into the union of three subsets of $R^8 \times K$. Each subset is a product of a subset of R^8 with K . The first is the trivial one $A(1) = B(1) = 0$. The second is of the form

$$[A(1) \vdots B(1)] = \alpha\begin{pmatrix}\cos\psi\\ \sin\psi\end{pmatrix}(\sin\psi, -\cos\psi, \sin\psi, -\cos\psi) + \begin{pmatrix}\cos\psi\\ \sin\psi\end{pmatrix}(d_{11}, d_{12}, d_{21}, d_{22})$$

where $-\pi < \psi \leq \pi$ and $|d_{j1}\cos\psi + d_{j2}\sin\psi| < 1$, $j = 1, 2$. As α varies an equivalence class of points is obtained and of course $k(z)$ does not depend on α . Letting $\alpha = 0$ the resulting space is 5-dimensional. The third subset is the set Θ_1 of $A(1)$, $B(1)$ for which the simple identification holds and is an 8-dimensional space. The remaining points in Θ are boundary points for which $\det g(z)$ can be zero (i.e. for which $|d_{11}\cos\psi + d_{12}\sin\psi| = 1$). Thus M is decomposed into 3, 8, 11 dimensional subsets.

To close this section some discussion of the constrained situation will be given. When constraints of the type (2.7) are imposed the subset of Θ satisfying these constraints will be referred to as Θ_Φ for the present discussion. Again points in Θ_Φ may be grouped into equivalence classes of points and the notation $[\Theta_\Phi]$ will be used for the space of these equivalence classes. Again also the definitions of pointwise convergence and quotient convergence may be applied. However now it is necessary to allow as preimages of certain equivalence classes certain additional points in the boundary of $[\Theta_\Phi]$. We will let $\Theta_{\Phi,h,g} = \Theta_\Phi \cap \Theta_{h,g}$ and call its closure in the Euclidean topology $\overline{\Theta}_{\Phi,h,g}$. Any point in $\overline{\Theta}_{\Phi,h,g} \backslash \Theta_{\Phi,h,g}$ which is equivalent to a point $\theta \in \Theta_{\Phi,h,g}$ will also be considered as a preimage of $[\theta]$. This will be clarified after the next corollary. The constraints Φ_j in (2.7) will be required to be continuous for the following corollary and to keep the discussion simple we will consider only constraints which act on τ alone in (1.2) and not μ.

COROLLARY 3.4. *When* $[\theta_n]$ *is a sequence of points in* $[\Theta_{\Phi,h,g}]$ *and* $[\theta] \in [\Theta_{\Phi,h,g}]$ *then* $[\theta_n] \xrightarrow{Q} [\theta]$ *is equivalent to* $[\theta_n] \xrightarrow{p_t} [\theta]$.

Proof. That $[\theta_n] \xrightarrow{Q} [\theta]$ implies $[\theta_n] \xrightarrow{p_t} [\theta]$ follows exactly as in the proof of Corollary 3.3. For the converse consider the following. In the proof of Theorem 3.1, $\dot{\tau}_n$ (corresponding to $[\theta_n]$) may be chosen to satisfy the constraints (2.7). However what needs to be shown is that $\ddot{\tau}_n$ of that proof may also be chosen to satisfy the constraints. Assume, by way of contradiction, that there is an open neighbourhood G in $[\Theta_{\Phi,h,g}]$ of $[\theta]$ for which $[\theta_n]$ has an infinite number of terms not belonging to

G . That is the inverse image $\chi^{-1}(G)$ excludes an infinite number of $\dot{\tau}_n$ points in $\Theta_{\Phi,h,g}$. Call these excluded points $\dot{\tau}'_n$. But we know from the proof of Theorem 3.1 that there exist $\check{\tau}_n$ (corresponding to k) in $\Theta_{h,g}$ (*without* restrictions) for which $\dot{\tau}_n - \check{\tau}_n \to 0$. Since $\dot{\tau}'_n$ belongs to a closed and bounded subset of R^t (for some t) then there is a subsequence $\dot{\tau}''_n$ converging to the limit τ (in the Euclidean topology). But then the corresponding terms of $\check{\tau}_n$, $\check{\tau}''_n$ say, also converge to τ since $\check{\tau}''_n - \tau = \left(\check{\tau}''_n - \dot{\tau}''_n\right) + \left(\dot{\tau}''_n - \tau\right)$. But then, by the continuity of Φ_j , $\Phi_j\left(\check{\tau}''_n\right) \to \Phi_j(\tau) = \lim_{n\to\infty} \Phi_j\left(\dot{\tau}''_n\right) = 0$ for $j = 1, 2, \ldots, r$. Thus, since $k\left(z; \check{\tau}''_n\right) \equiv k(z; \theta)$ and k is a continuous function of τ_n for z fixed and $|z| < 1$, it follows that $k(z; \tau) \equiv k(z; \theta)$. That is $\check{\tau}''_n$ converges to the point τ which satisfies the constraints and is thus a preimage of $[\theta] \in \left[\Theta_{\Phi,h,g}\right]$. Hence $\dot{\tau}''_n$ converges to the point τ contradicting the assumption that the $\dot{\tau}'_n$ points are excluded from a neighbourhood of $\chi^{-1}[\theta]$. □

The continuity of the Φ_j is necessary for Corollary 3.4 as the following example illustrates. Take Example 3.1. Define a constraint function on the region $|\alpha| \le 1$, $|\beta| \le 1$ by $\Phi(0, 0) = 0$, $\Phi(\frac{1}{2}, \beta) = 0$ for $\beta < \frac{1}{2}$ and $\Phi(\alpha, \beta) = 1$ elsewhere. Then, in particular, there is a discontinuity in Φ at the point $\alpha = \beta = \frac{1}{2}$. Then the parameter space for which the constraints apply is

$$\Theta_\Phi = \{(0, 0)\} \cup \{(\alpha, \beta) : \alpha = \tfrac{1}{2}, \beta < \tfrac{1}{2}\}$$

and $\left[\Theta_\Phi\right] = \Theta_\Phi$ here. Now let τ_n be the sequence defined as $\tau'_n = \left(\frac{1}{2}, \frac{1}{2}-(1/n)\right)$. Then $\tau'_n \to (\frac{1}{2}, \frac{1}{2})$ so that for $|z| < 1$,

$k(z;\ \tau_n) \to 1 = k(z;\ \tau)$ where $\tau' = (0,\ 0)$. Hence $\tau_n \xrightarrow{p_t} \tau$. Yet there do *not* exist any points in Θ_Φ equivalent to $(0,\ 0)$ which are arbitrarily close to $\left(\tfrac{1}{2},\ \tfrac{1}{2}-(1/n)\right)$ and therefore $\theta_n \overset{Q}{\not\to} \theta$.

The condition that attention be restricted to $\Theta_{\Phi,h,g}$ does not appear to be easily replaced by Θ_Φ . It is likely that, in a similar way to Example 3.2, one may find sequences $k(z;\ \theta_n)$ which have unique decompositions which become unbounded as $n \to \infty$ but for which in the unconstrained space there exist suitable preimages (as constructed in the proof of Theorem 3.1). However there may not exist suitable preimages of θ in the constrained space. This problem might be overcome by stronger assumptions concerning the functions Φ_j .

That one needs to include certain boundary points for Corollary 3.4 to have a meaning may be illustrated as follows. Consider Example 3.1 again. Let $\Phi(\alpha,\ \beta) = \beta - \alpha^2$ so that when $\Phi(\alpha,\ \beta) = 0$, $\beta - \alpha^2$ describes a parabola cutting the line $\alpha = \beta$ at $(\alpha,\ \beta) = (0,\ 0)$ and $(1,\ 1)$. If $\theta = (0,\ 0)$ but $\theta_n = \left(1-(1/n),\ \left(1-(1/n)\right)^2\right) \to (1,\ 1)$ then in order that Corollary 3.4 has a meaning it is required that *both* $(0,\ 0)$ and $(1,\ 1)$ be included as preimages of $[(0,\ 0)]$. This type of example illustrates the discussion preceding the statement of Corollary 3.1.

The continuity of the constraints is not needed for the proofs of Theorem 3.1, Corollary 3.2, Corollary 3.3 when θ_n is a sequence of points in Θ_1 and θ , the limit point, is in Θ_1 also, since to each θ_n and to θ there correspond *unique* decompositions of $k(z;\ \theta_n)$ for each n and a *unique* decomposition of $k(z;\ \theta)$.

When triangular identification via C2.2, C2.3 and C2.5 is used we may proceed as follows. Now let Θ be the set of all triangularly identified structures for which the jth diagonal element of h is of degree q_j at most while the degree of g is of p at most. Again the space $[\Theta]$ of equivalence classes of structures of this type is introduced, and the quotient topology may be defined as before. In regard to Theorem 3.1 solutions to (3.2) are considered of the form $\check{B}(j)$ lower triangular of which at least one solution in this form exists corresponding to $[\theta] \in [\dot{\Theta}]$. The proof for the analogue for triangular structures of Theorem 3.1 is very similar to that given above. The remaining results of this section also apply to triangular identification where now $\Theta_1 = [\Theta_1]$ is defined by requiring that the q_j, specifying the degrees of the diagonal elements of h are the true degrees (i.e. the coefficients of z^{q_j} in these elements is not zero) or $A(p)$ in $g(z)$ is of full rank.

4. The strong law of large numbers in the ARMA model

The first result of this section is concerned with the strong consistency of the estimators of the parameters specifying (1.1) when it is known *a priori* that the true value, θ_0, belongs to the space Θ_1 where now this will also denote the previous set $\Theta_{1,\Phi}$ if constraints apply. Then the points θ in Θ_1 are in one-to-one correspondence with the parameters τ, μ defined in (1.2) subject to the constraints (2.7). Now Θ_1 is not a closed space and its closure contains points for which the simple identification conditions do not hold, points for which $\det h(z; \theta)$ has a zero on the unit circle and points at infinity. This remark is relevant when an attempt is made to obtain the minimum of each of the "likelihoods" $\bar{L}_N$, $\tilde{L}_N$, $\hat{L}_N$ over Θ_1 so that it is conceivable, for finite N, the minimum may be obtained at a point outside Θ_1. However as will be shown below this will cease to occur for N sufficiently large (but finite with probability one). Let $\overline{\theta}_N$, $\hat{\theta}_N$, $\tilde{\theta}_N$ denote the estimators obtained by minimising $\overline{L}_N$, $\hat{L}_N$, $\tilde{L}_N$ over Θ_1 .

THEOREM 4.1. *For the* ARMA *model* (1.1) *with* θ_0 *known to belong to* Θ_1 *the estimators* $\overline{\theta}_N$, $\hat{\theta}_N$, $\tilde{\theta}_N$ *all converge to* θ_0 *in the Euclidean topology.*

Proof. Consider the case of $\overline{\theta}_N$ first. It is convenient to let $k_N(z) = k\left(z;\ \overline{\theta}_N\right)$, $K_N = K\left(\overline{\theta}_N\right)$. As in the proof of Theorem 2.2.1 it follows by part (a) of Lemma 2.2.2 (which continues to hold here) that

$$\varlimsup_{N\to\infty} \inf_{\theta\in\Theta_1} \overline{L}_N(\theta) \le \inf_{\theta\in\Theta_{01}} \varlimsup_{N\to\infty} \overline{L}_N(\theta) = \inf_{\theta\in\Theta_{01}} L(\theta) \ . \tag{4.1}$$

Now whereas condition 2.C2.6 of Chapter 2 was used to obtain the infimum on the r.h.s. of (4.1) as $\log\det K\left(\theta_0\right) + s$ this condition may be shown to hold in the ARMA case as follows.

LEMMA 4.2. *For the* ARMA *model* (1.1) *when* Θ_0, Θ *are the unconstrained parameter spaces described in Section* 3 *then*

$$\inf_{\theta\in\Theta_0} L(\theta) = L\left(\theta_0\right) = \log\det K\left(\theta_0\right) + s$$

where $\theta_0 \in \Theta$.

Proof. First we assert that $L(\theta)$, for $\theta \in \Theta_0$ (i.e. the space where $\det g(z;\ \theta)$, $\det h(z;\ \theta)$ have no zeros within or on the unit circle) cannot be less than $L\left(\theta_0\right)$ since $Q(\theta)$ is the variance of a stationary process with corresponding prediction variance

$$\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left[2\pi\ \mathrm{tr}\left(f(\omega;\ \theta)^{-1}f\left(\omega;\ \theta_0\right)\right)\right]d\omega\right\}$$

$$\ge s\ \exp\left\{\frac{1}{2\pi s}\int_{-\pi}^{\pi} \log\det\left[f(\omega;\ \theta)^{-1}f\left(\omega;\ \theta_0\right)\right]d\omega\right\}$$

by Jensen's inequality. This last expression reduces to

$s\left[\det\left[K^{-1}(\theta)K(\theta_0)\right]\right]^{1/s}$, (see Hannan, 1970, Chapter III) and since

$\log(x) + s(c/x)^{1/s}$, $c > 0$, is minimized at $c = x$ the assertion follows. Using a zero subscript for h, g, K at θ_0 we may choose $h = h_0$, $K = K_0$ and retain $f(\omega;\ \theta) > 0$, $\omega \in [-\pi,\ \pi]$. Also (MacDuffee, 1956, p. 41) $g_0 = u_0P_0v_0$ where u_0, v_0 are unimodular matrices of polynomials while P_0 is diagonal. Choosing $\tilde{P}$ as P_0 except for zeros of $\det(P_0)$ on the unit circle, which we slightly remove outside the unit circle, the result easily follows by putting $g = u_0\tilde{P}v_0$. □

When Θ_0 and Θ are the spaces corresponding to the case where constraints of the type (2.7) are applied there is sometimes a need to impose (2.C2.6) when $\theta_0 \in \Theta$ is such that $\det g(z;\ \theta_0)$ has a zero on the unit circle. However if θ_0 actually belongs to Θ_0 there is no need for this imposition since then the latter part of the proof of Lemma 4.2 is not needed. Also when this is true (4.1) may be replaced by the statement $\overline{\lim_{N\to\infty}} \inf_{\theta\in\Theta} \bar{L}_N(\theta) \le \overline{\lim_{N\to\infty}} \bar{L}_N(\theta_0) = \log\det K(\theta_0) + s$.

Hence (4.1) becomes

$$\overline{\lim_{N\to\infty}} \inf_{\theta\in\Theta_1} \bar{L}_N(\theta) \le \log\det K(\theta_0) + s\ . \tag{4.2}$$

Before proceeding with the proof of Theorem 4.1 it will be convenient to establish a result of use in what follows.

LEMMA 4.3. *If f is a spectral density matrix with continuous elements and* $\det f(\omega) = 0$ *at a finite number of points in* $[-\pi,\ \pi]$ *then the* $Vs \times Vs$ *covariance matrix* $\Gamma_V(f)$ *defined in* (1.3.2) *is strictly positive definite.*

Proof. We will show that the smallest eigenvalue of $\Gamma_V(f)$ is positive. Thus consider

$$\lambda_1\left(\Gamma_V(f)\right) = \inf_{x'x=1} x'\Gamma_V(f)x$$

$$= \inf_{x'x=1} \int \left(\sum_1^V x(v)e^{iv\omega}\right)^* f(\omega)\left(\sum_1^V x(v)e^{iv\omega}\right)d\omega \qquad (4.3)$$

(where $x' = \left(x(1)', \ldots, x(V)'\right)$ and $x(j)$ is an s-vector). Now defining $A_\varepsilon = \{\omega \in [-\pi, \pi] : \lambda_1(f) \geq \varepsilon > 0\}$ and ω_j to be the jth of the R (finite) ω-values at which $\det f(\omega)$ is zero, then $\varepsilon > 0$ may be chosen so that $[-\pi, \pi]\backslash A_\varepsilon$ has arbitrarily small Lebesgue measure δ. (This follows since $\lambda_1(f)$ is a continuous function of ω.) Then the r.h.s. of (4.3) is not less than

$$\varepsilon \,.\, \inf_{x'x=1} \int_{A_\varepsilon} \left\|\sum_1^V x(v)e^{iv\omega}\right\|^2 d\omega$$

$$\geq \varepsilon \,.\, \inf_{x'x=1} \left(1 - 2 \,.\, \delta \,.\, R \,.\, \sup_{\substack{\omega\in\omega_j\pm\delta \\ j=1,2,\ldots,R}} \left\|\sum_1^V x(v)e^{iv\omega}\right\|^2\right)$$

$$\geq \varepsilon \,.\, (1-2.\delta.R.V) \quad \text{since} \quad \left\|\sum_1^V x(v)e^{iv\omega}\right\|^2 \leq V \,.$$

Thus by choosing δ s.t. $1 - 2.\delta.R.V > 0$ the result follows. □

Returning to the proof of Theorem 4.1 let $\theta \in \Theta_1$ and denote by $K_\theta, h_\theta, g_\theta, k_\theta$ the matrix functions K, h, g, k at θ. Then

$$\overline{L}_N(\theta) = \log \det K_\theta + \int \mathrm{tr}\left\{h_\theta^* g_\theta^{*-1} K_\theta^{-1} g_\theta^{-1} h_\theta I(\omega)\right\} d\omega$$

$$\geq s \log \lambda_1\left(K_\theta\right) + \lambda_1^{-1}\left(K_\theta\right).\gamma^{-2}.\lambda_1\left(\int P_\theta I(\omega)P_\theta^* d\omega\right) \qquad (4.4)$$

where $P_\theta = \left(\text{adjoint } g_\theta\right).h_\theta$, $|\det g_\theta| \leq 2^{ps} = \gamma$ (say) for all $\theta \in \Theta_1$ since $\det g_\theta$ has no zeros for $|z| < 1$. Then P_θ is a matrix of polynomials of degree $V = q + p(s-1)$ at most and $P_\theta(0) = I_s$. Now

$$\int P_\theta I(\omega) P_\theta^* d\omega = \sum_{j=0}^{V} \sum_{k=0}^{V} p_\theta(j) G(j-k) p_\theta(k)' \quad \text{where} \quad P_\theta(z) = \sum_0^V p_\theta(j) z^j$$

$$= \sum_{j=0}^{V} \sum_{k=0}^{V} p_\theta(j) \left[\Gamma_0(j-k) + E_N(j-k)\right] p_\theta(k)'$$

where $E_N(l)$ is the matrix of errors in replacing $G(l)$ by $\Gamma_0(l)$. Then defining the Vs-row vector $q_\theta' = [\alpha' p_\theta(0), \ldots, \alpha' p_\theta(V)]$ and $E_{V,N}$ to be the $Vs \times Vs$ matrix with $E_N(j-k)$ in the (j, k)th block of s^2 elements we obtain

$$\inf_{\alpha'\alpha=1} \alpha' \left(\int P_\theta I(\omega) P_\theta^* d\omega \right) \alpha$$

$$= \inf_{\alpha'\alpha=1} q_\theta' [\Gamma_V(\theta_0) + E_{V,N}] q_\theta$$

$$\geq \lambda_1\left(\Gamma_V(\theta_0) + E_{V,N}\right) \cdot \inf_{\alpha'\alpha=1} \left\{ \sum_0^V \alpha' p_\theta(j) p_\theta(j)' \alpha \right\}$$

$$\geq \lambda_1\left(\Gamma_V(\theta_0) + E_{V,N}\right), \text{ since } p_\theta(0) = I_s \text{ for all } \theta \in \Theta_1 .$$

Now since $\det f(\omega; \theta_0)$ has only finitely many fixed zeros for $\omega \in [-\pi, \pi]$ it follows from Lemma 4.3 that $\Gamma_V(\theta_0) \geq 2\delta I_{Vs}$ for some $\delta > 0$. Also, by ergodicity, given $\delta > 0$ there exists an N_0, a.s. finite, such that $\sup_{|l| \leq V} \|E_N(l)\| \leq \delta/V$ and hence $\|E_{V,N}\| \leq \delta$ for all $N \geq N_0$. Thus for $N \geq N_0$ the l.h.s. of (4.4) is bounded below by

$$\overline{L}_N(\theta) \geq s \log \lambda_1(K_\theta) + \frac{b}{\lambda_1(K_\theta)}, \quad b = \frac{\delta}{\gamma^2} > 0, \tag{4.5}$$

for all $\theta \in \Theta_1$ and $N \geq N_0$. Hence for N fixed the infimum of $\overline{L}_N(\theta)$ must be obtained when $\lambda_1(K_\theta) > 0$ or else the upper bound in (4.2) is violated. Similarly if K_N denotes the value of K_θ at which $\overline{L}_N(\theta)$ is minimised then $\lambda_1(K_N)$ *cannot* converge to zero or again the upper bound in (4.2) is violated. Thus, when minimising $\overline{L}_N(\theta)$ attention may be restricted to that subset Θ_K, say, of Θ_1 for which $\lambda_1(K_\theta) \geq b_1 > 0$ and $\lambda_s(K_\theta) \leq b_2 < \infty$. Then, returning to (4.4) it follows that for $\theta \in \Theta_K$,

$$\overline{L}_N(\theta) \geq s.\log(b_1) + b_2^{-1}.\gamma^{-2}.\lambda_1\left(\Gamma_V(\theta_0)+E_{N,V}\right).\mathrm{tr}\left[\sum_0^V p_\theta(j)p_\theta(j)'\right] .$$

Hence if $N \geq N_0$ then $\mathrm{tr}\left[\sum_0^V p_\theta(j)p_\theta(j)'\right]$ must remain uniformly bounded or else (4.2) cannot hold. Hence the minimum of $\overline{L}_N(\theta)$ for $N \geq N_0$ may be found by minimising of θ belonging to that subset of Θ_K, $\Theta_{K,P}$ say, for which $\mathrm{tr}\left[\sum_0^V p_\theta(j)p_\theta(j)'\right] \leq b_3 < \infty$. Of course, as mentioned before the coefficients in $d_\theta = \det g(\theta)$ are uniformly bounded. Denote by P_N, d_N and K_N the values of P, d and K at which $\overline{L}_N(\theta)$ is minimised. These may not correspond to a point in $\Theta_{K,P}$ since this is not a closed set, because Θ_1 is not closed.

We will denote by $\overline{\theta}_N$ the minimising θ value whether or not this $\overline{\theta}_N$ belongs to Θ_1. As will be shown below $\overline{\theta}_N$ will eventually belong to Θ_1 for all N sufficiently large so that this notation is not inconvenient. Now, by way of contradiction, assume that $\overline{\theta}_N \not\to \theta_0$. When this is so either $\overline{\theta}_N$ is in Θ_1 for all sufficiently large N or $\overline{\theta}_N$ lies outside Θ_1 infinitely often. In this latter case we may find $\dot{\theta}_N$ in $\Theta_{K,P}$ where $\dot{\theta}_N$ also does not converge to θ_0 and for which

$$|\overline{L}_N(\overline{\theta}_N)-\overline{L}_N(\dot{\theta}_N)| \leq \varepsilon_N , \quad \varepsilon_N \to 0 \text{ as } N \to \infty . \tag{4.6}$$

This $\dot{\theta}_N$ may be constructed as follows. For N fixed we may find a sequence $\theta_{R,N}$ along which $P(z;\,\theta_{R,N}) \to P_N(z)$, $d(z;\,\theta_{R,N}) \to d_N(z)$ and $K(\theta_{R,N}) \to K_N$ as $R \to \infty$. Choose R large enough so that (4.6) holds for each N where $\varepsilon_N \to 0$ as $N \to \infty$ and call this $\theta_{R,N}$, $\dot{\theta}_N$. This may always be done since $\overline{L}_N(\theta)$ is bounded below on $\Theta_{K,P}$ by $s \log (b_1)$. (When $\overline{\theta}_N$ belongs to Θ_1 for all N large $\dot{\theta}_N$ will be taken as $\overline{\theta}_N$.) Now since $\overline{\theta}_N \not\to \theta_0$ and $\dot{\theta}_N \not\to \theta_0$ it follows that either $K(\dot{\theta}_N) \not\to K(\theta_0)$ or $k(z;\,\dot{\theta}_N) \overset{p_t}{\not\longrightarrow} k(z;\,\theta_0)$ or both. (If $\dot{\theta}_N \not\to \theta_0$ and $K(\dot{\theta}_N) \to K(\theta_0)$ then $k(\dot{\theta}_N) \overset{p_t}{\not\longrightarrow} k(\theta_0)$ since if $k(\dot{\theta}_N) \overset{p_t}{\longrightarrow} k(\theta_0)$ then the unique factorisation $(\dot{h}_N,\, \dot{g}_N)$ of $k(\dot{\theta}_N)$ and $(h_0,\, g_0)$ of $k(\theta_0)$ satisfy $\dot{h}_N \to h_0$, $\dot{g}_N \to g_0$ by the proof of Theorem 3.1; i.e. $\dot{\theta}_N \to \theta_0$ contradicting $\dot{\theta}_N \not\to \theta_0$.) Now since $(k(\dot{\theta}_N),\, K(\dot{\theta}_N)) \overset{p_t}{\not\longrightarrow} (k(\theta_0),\, K(\theta_0))$ there exists a subsequence $N(M)$ along which $(\dot{k}_{N(M)},\, \dot{K}_{N(M)}) \overset{p_t}{\longrightarrow} (k,\, K) \neq (k(\theta_0),\, K(\theta_0))$ and, by the uniform boundedness of the coefficient matrices in $\dot{P}_N$ and the coefficients in $\dot{d}_N$, for which $\dot{P}_{N(M)} \to P$, $\dot{d}_{N(M)} \to d$ where $k(z) = P^{-1}(z).d(z)$ for $|z| < 1$. Now, using (4.6),

$$\varliminf_{M\to\infty} \inf_{\theta\in\Theta_1} \overline{L}_{N(M)}(\theta) = \varliminf_{M\to\infty} \overline{L}_{N(M)}(\dot{\theta}_{N(M)})$$

$$\geq \sup_{\eta>0} \varliminf_{M\to\infty} \left[\log\det \dot{K}_{N(M)} + \int \mathrm{tr}\left\{\left(|\dot{d}_{N(M)}|^2+\eta\right)^{-1} . \; . \dot{P}_{N(M)}\dot{K}^{-1}_{N(M)}\dot{P}_{N(M)}.I(\omega)\right\}d\omega\right]$$

which, by using the boundedness properties proved above and a similar result to Lemma 2.2.2, becomes

$$\sup_{\eta>0}\left[\log\det K \quad + \quad \int \mathrm{tr}\left\{\left(|d|^2+\eta\right)^{-1}\left(P^*K^{-1}P\right)f\left(\theta_0\right)\right\}d\omega\right]$$

$$= \log\det K + (2\pi)^{-1}\int \mathrm{tr}\left[f^{-1}f\left(\theta_0\right)\right]d\omega$$

where $f = \frac{1}{2\pi}kKk^*$. But as was shown in the proof of Theorem 2.2.1 this last line is strictly greater than $\log\det\left(K_0\right) + s$ unless $k\left(e^{i\omega}\right) = k\left(e^{i\omega};\ \theta_0\right)$ a.e. (ω) and $K = K_0$. Hence there is a contradiction and $\overline{\theta}_N \to \theta_0$.

(ii) Proof for $\tilde{\theta}_N$. We need only show that the argument leading to the uniform boundedness of $\|K_\theta\|$ and of the coefficient matrices in $P_\theta\left(e^{i\omega}\right)$ for all sufficiently large N also applies here. For this consider

$$\lambda_1\left[(N')^{-1}\sum_t \mathrm{tr}\left\{P_\theta\left(e^{i\omega_t}\right)I\left(\omega_t\right)P_\theta\left(e^{i\omega_t}\right)^*\right\}\right]$$

$$= \frac{1}{2\pi}\lambda_1\left(\sum_0^V\sum_0^V p_\theta(k)\left[G(k-l)+D_N(k-l)\right]p_\theta(l)'\right)$$

where $D_N(l) = 0$ if $|l| \le N'-N$, $N^{-1}\sum_1^{N-N'-j} x(m)x(N'+m-j)$ if $-V \le j \le -(N'-N+1)$ with a similar expression for $N'-N+1 \le j \le V$. Then since each $D_N(l)$ is composed of at most V terms of the form $N^{-1}x(m)x(N'+l)$ they can be made arbitrarily small for N large. The result proceeds as for $\overline{\theta}_N$, where now $E_N(l)$ is replaced by $E_N(l) + D_N(l)$. The remainder of the proof for $\tilde{\theta}_N$ follows that for $\overline{\theta}_N$ using Lemma 2.2.2.

(iii) Proof for $\hat{\theta}_N$. By Lemma 2.2.6,

$$N^{-1}\log\det\Gamma_N(\theta) \ge \log\det K(\theta) ,$$

at least for $\theta \in \Theta_1$, so that we consider $N^{-1}x_N'\Gamma_N^{-1}(\theta)x_N$ for $\theta \in \Theta_1$ as follows. Now, if $\theta \in \Theta_1$,

$$P_\theta f_\theta P_\theta^* = G_\theta h_\theta f_\theta h_\theta^* G_\theta^* = \left(\det\left(g_\theta g_\theta^*\right)\right)K_\theta \le \gamma^2 K_\theta \ .$$

Thus if A is the $(N-V)s \times Ns$ matrix (where V is the degree of P and is thus $\le q + p(s-1)$) with the (m, l)th block of s^2 elements having $A(m, l) = P(V+m-l)$, $m \le l \le m+V$ and zeros elsewhere then

$$A\Gamma_N(\theta)A' = \Gamma_{N-V}\left(P_\theta f_\theta P_\theta^*\right) \le \gamma^2\left(I_{N-V} \otimes K_\theta\right) \ .$$

In general $u'B'Bu \le \left(u'\Gamma^{-1}u\right)(v'B'\Gamma Bv)$, $v'v = 1$ so that $u'B'Bu \le u'\Gamma^{-1}u$ if $B'\Gamma B \le I$. Choosing $B = \left(I_{N-V} \otimes K_\theta^{-\frac{1}{2}}\right)A$ and $\Gamma = \Gamma_N$ it follows in our particular case that

$$N^{-1}x_N'\Gamma_N^{-1}(\theta)x_N \ge \gamma^{-2}N^{-1}x_N\left(A'\left(I_{N-V} \otimes K_\theta^{-1}\right)A\right)x_N$$

$$= \gamma^{-2} \operatorname{tr}\left\{K_\theta^{-1} \sum_0^V \sum_0^V p_\theta(j)\left(N^{-1} \sum_{n=V+1}^{N} x(n-j)x(n-k)'p_\theta(k)'\right\}\right.$$

$$= \gamma^{-2} \operatorname{tr}\left\{K_\theta^{-1} \sum_0^V \sum_0^V p_\theta(j)\left(\Gamma_0(j-k)+E_N(j-k)\right)p_\theta(k)'\right\}$$

where now $E_N(j-k)$ is the error in replacing $N^{-1} \sum_{n=V+1}^{N} x(n-j)x(n-k)'$ by $\Gamma_0(j-k)$. Again this error may be made arbitrarily small for N sufficiently large and the remainder of the proof of the boundedness of $\|K_\theta\|$ and the coefficient matrices $P_\theta(j)$ is exactly as before. By using a result similar to Lemma 2.2.5 the result proceeds as before. This completes the proof. □

Note that continuity of the Φ_l in (2.7) is not required for the above proof.

The extension of Theorem 4.1 to the case when θ_0 belongs to the larger space Θ is not possible for the following reasons. While it is still possible to show that the K_N, P_N, d_N in the proof of Theorem 4.1 satisfy the uniform boundedness properties established in that proof and it is still possible to show that $(k_N, K_N) \xrightarrow{p_t} (k_0, K_0)$, the result may be meaningless in terms of convergence of points $[\theta_n]$ to $[\theta_0]$ in the senses defined above (i.e. the quotient or pointwise) when $[\theta_0]$ has more than one preimage. To illustrate the difficulties consider Example 3.2. Assume that $[\theta_0]$ is the equivalence class of points for which $k(z;\ \theta_0) = I_2$, $K(\theta_0) = I_2$. Then it may be possible that in attempting to minimise $\overline{L}_N(\theta)$ the estimate of k_N and K_N is obtained for which there is *no* h_N, g_N with *finite* coefficient matrices A_N and B_N . For example let $k(z;\ \theta_n)$ (for N fixed) converge to

$$\frac{\begin{bmatrix} 1-b_N z^2 & -b_N z^2 \\ -a_N z^2 & 1-a_N z^2 \end{bmatrix}}{1-(a_N+b_N)z^2}$$

where $b_N \to 0$, $a_N \to 0$ as $N \to \infty$ so that $k_N \to k_0$. But since there is *no* decomposition of k_N as $h_N^{-1} g_N$ where $h_N = (I+B_n z)$, $g_N = (I+A_N z)$ with A_N and B_N having finite elements, it is impossible to conclude any sort of convergence of $[\theta_N]$ to $[\theta_0]$ in this example. The above proof does not exclude such a minimising sequence, unless $\theta_0 \in \Theta_1$, and in fact it is difficult to see how this could be done.

However if $[\theta_0] \in [\Theta_{h,g}]$ then one may use the proof of Theorem 4.1 to conclude that $[\overline{\theta}_n] \xrightarrow{Q} [\theta_0]$ *and* that $[\overline{\theta}_n] \xrightarrow{p_t} [\theta_0]$. Alternatively one may proceed without the use of P_N, d_N in that proof by noting that the uniform boundedness of K_N may be established as before and then that

$$\Phi_\eta(\omega;\ \theta) = h_\theta^*\left(g_\theta K_\theta g_\theta^* + \eta I_s\right)^{-1} h_\theta$$

is uniformly continuous in $\theta \in \Theta_{h,g}$. The proof of Theorem 2.2.1 may then be used. *In particular for the case* $s = 1$ *there is no need to exclude* $[\theta_0]$ *belonging to* $[\Theta]$ (*and not just* Θ_1) *since in this case the coefficients in* $h(z;\ \theta)$ *and* $g(z;\ \theta)$ *belong to a compact subset* $R^{(p+q)}$. This case was not covered in the treatment of Hannan (1973a). The case when *a priori* it is assumed that $\theta \in [\Theta_{h,g}]$ is not restrictive in practice since any conceivable method of minimisation of $\overline{L}_N(\theta)$ would require that estimates which are "too large" be excluded.

Alternatively one could adopt a convention for estimation when it is not known whether $\theta \in \Theta_1$ and could therefore belong to $[\Theta]$. Such a convention is: take $\overline{\theta}_N$ as having finite coefficients and such that $\overline{L}_N(\overline{\theta}_N)$ is arbitrarily close to $\inf_{\theta \in [\Theta]} \overline{L}_N(\theta)$. Then Theorem 3.1 will hold.

Before closing this section we would like to discuss the application of Theorem 4.1 to cases where we have in mind triangular identification. Now Θ_1 is taken as defined at the end of §3.

THEOREM 4.4. *For the* ARMA *model* (1.1) *triangularly identified with* $\theta_0 \in \Theta_1$ (*defined for triangular structures*) *then* $\overline{\theta}_N, \tilde{\theta}_N, \hat{\theta}_N$ *converge to* θ_0 *in the Euclidean topology.*

The proof of this result is not markedly different from that of Theorem 4.1.

5. The central limit theorem for ARMA models

We will indicate in this section how the central limit theorem of §2.3

may be applied in the ARMA model (1.1). Recall that Θ_{01} is that subset of Θ for which the simple identification conditions hold with C2.3 strengthened to C2.3'. It is the set Θ_{01} that is of importance for the CLT. To begin with it constitutes almost all of Θ. In truth the data will not be generated by the model and it is most unlikely that the inclusion of the remainder of the space Θ will increase the likelihood. In the second place if the true point θ_0 lies in one of the lower dimensional spaces (e.g., the first and second subsets of Example 3.4) and *this is not known* then the CLT cannot hold, for the sequence, $\hat{\theta}_N$ say, of estimates will approach a surface in $R^u \times K$, will search that surface in a haphazard way and will not approach $\theta_0 \in R^t \times K$. (The estimate $\hat{\theta}_N$ could be restricted *a priori*, to one of the lower dimensional spaces in Θ so that the CLT holds.) When $\theta_0 \in \Theta_{01}$, Theorem 4.1 shows that $\hat{\theta}_N$, e.g., converges to θ_0 in the Euclidean topology on $\Theta_{01} \times K$. The main thing that must be done to apply the CLT, Theorem 2.3.1 of §2.3, is to show that Θ_{01} is a twice differentiable manifold. First, the space K of all $s \times s$ positive definite symmetric matrices may be topologised as an analytic manifold of dimension $s(s+1)/2$ for example by regarding it as a coset space of the group of all $s \times s$ nonsingular matrices with positive determinant modulo the proper rotation group in s-dimensional space. Thus in the proof of the following theorem we will concentrate on showing that the elements of $A(j)$, $B(k)$ for which the simple identification requirements hold is an open subset of R^u, $u = (p+q)s^2$.

THEOREM 5.1. Θ_{01} *is an analytic manifold of dimension* $u + s(s+1)/2$.

Proof. We show in fact that Θ_1 is an open subset of $R^u \times K$. Let

$\theta_0' \in R^u$ correspond to $\theta_0 \in \Theta_1$ in the sense that its coordinates are the elements of $A_0(j)$, $B_0(k)$, $j = 1, \ldots, p$; $k = 1, \ldots, q$, taken in a certain order. We can evidently find a ball B containing θ_0' at which conditions C2.3', C2.4 are satisfied and $\det g(z) \neq 0$, $|z| = 1$. It remains to show that B may be chosen so that C2.2 holds also. Let $h(z, \theta')$, $g(z, \theta')$ correspond to θ' . If the theorem is not true then $h(z, \theta') = e(z, \theta')h_1(z, \theta')$, $g(z, \theta') = e(z, \theta')g_1(z, \theta')$ where e, g_1, h_1 are polynomials and, for θ' arbitrarily near to θ_0' , $e(z, \theta')$ has non constant determinant.

We may put $h^{-1}g = w^{-1}Q^{-1}Pv$ where w, v are unimodular, P, Q are diagonal and p_i divides p_{i+1}, q_{i+1} divides q_i where p_i, q_i are the ith diagonal elements of P, Q and are relatively prime (see Hannan, 1975b). Then v, w, P, Q are uniquely determined and, since this reduction to Smith's normal form is accomplished by elementary transformations the matrices v, w, P, Q , are continuous in θ' uniformly for $|z| \leq 1$ (see MacDuffee, 1956, p. 41). Put $h_1 = Qw$, $g_1 = Pv$. Then h_1, g_1 are continuous in θ' and thus so is e . Now we may further reduce $e(z)$ to upper triangular form, with elements off the diagonal of lower degree than the diagonal element in the same row, by right multiplication by a unimodular factor. Again this reduction to Hermite's normal form (MacDuffee, 1956, p. 32) is unique and the factors are continuous in θ' uniformly in $|z| \leq 1$. Thus $e(z, \theta')$ converges to a constant, non singular, matrix (which we may choose to be diagonal) as $\theta' \to \theta_0'$. Put $R_1(\theta') = [A_1(p, \theta') \vdots B_1(q, \theta')]$, where $A_1(p, \theta')$, $B_1(q, \theta')$ are the p, qth (respectively) coefficient matrices in g_1, h_1 . We will show that $e(0, \theta')R_1(\theta')$ differs from

$$R(\theta_0') = [A(p, \theta_0'), B(q, \theta_0')]$$

corresponding to g, h by a quantity which converges to zero as $\theta' \to \theta_0'$. Then since $e(0, \theta')$ converges to a non singular matrix as $\theta' \to \theta_0'$ and $R(\theta_0')$ is of full rank it follows that we may choose $\mathcal{B}$ so that $R_1(\theta')$ is of full rank for $\theta' \in \mathcal{B}$. Now the coefficient of $[z^p, z^q]$ in $e[g_1, h_1]$ is of the form (for $q > p$ for example)

$$e(0, \theta')[A_1(p, \theta'), B_1(q, \theta')] + \ldots + e(q, \theta')[0, B_1(0, \theta')] .$$

The result we seek will follow if we show that

$$[0, B_1(0, \theta')], \ldots, [A_1(p-1, \theta'), B_1(q-1, \theta')]$$

have bounded elements since each of $e(q, \theta'), \ldots, e(1, \theta')$ converge to 0. Firstly $B_1(0, \theta') = I_s$. Next the coefficient of z^1 in $e(g_1, h_1)$ is

$$e(0, \theta')[A_1(0,\theta'), B_1(1, \theta')] + e(1, \theta')[0, B_1(0, \theta')] .$$

(when $q = p + 1$ for example; if q is greater than $p + 1$ the argument is obviously the same). This converges to a finite limit as $\theta' \to \theta_0'$ so that since $e(1, \theta') \to 0$, $B_1(0, \theta') = I_s$ and $e(0, \theta')$ converges to a constant matrix it follows that $B_1(1, \theta')$ converges to a matrix with finite elements as $\theta' \to \theta_0'$. The coefficient of z^2 in $e(g_1, h_1)$ is

$$e(0, \theta')[A_1(1, \theta'), B_1(2, \theta')] + e(1, \theta')[A_0(0, \theta'), B_1(1, \theta')] + e(2, \theta')[0, B_1(0, \theta')] .$$

By repeating the above argument we get $[A_1(1, \theta'), B_1(2, \theta')]$ converging to a finite limit. Repeating the argument q times gives the result we seek. Since now we may assume that $\mathcal{B}$ is such that $R_1(\theta')$ is of full rank it follows that the last row of $R_1(\theta')$ is not null. This shows that the last diagonal element of $e(z, \theta')$ is constant in $\mathcal{B}$ and the last row of $g_1(z, \theta')$ is of degree p and no higher and that row of $h_1(z, \theta')$ is of degree q and no higher, and for one the degree is attained. Otherwise either the last row of $g(z, \theta')$ is of higher degree than p or that row of $h(z, \theta')$ is of higher degree than q, which is impossible. Remembering that the off diagonal element to the right of the second last diagonal element of $e(z, \theta')$ is of lower degree than the diagonal element, that the second last row of $R_1(\theta')$ is not null and the result established for the last row of $[g_1, h_1]$ we see that the second last diagonal element is constant and the off diagonal element is zero for $\theta' \in \mathcal{B}$. Continuing in this way the result follows. □

When constraints of the type (2.7) are considered for the central limit theorem the following requirement will be made. The $\Phi_l(\theta)$ are to be twice continuously differentiable and for $\theta \in \Theta_{01}$ the matrix with (l, j)th element $\partial\Phi_l(\theta)/\partial\theta_j$ is of rank r (the number of constraints). Then Θ_{01} may be replaced by a new manifold which is now twice differentiable, instead of analytic, of dimension $t + s(s+1)/2 - r$ (see Auslander and MacKenszie, 1963, or Matsushima, 1972). As required in the central limit theorem of §2.3 this new manifold must be of the form $M_\theta \times M_\mu$ where $M_\theta = \Theta'_{01}$ is of dimension $u - r_1$ and $M_\mu = K'$ is of dimension $s(s-1)/2 - r_2$ and Θ'_{01} specifies g, h while K' specifies K. Thus no constraints tie the $A(j)$, $B(j)$ to K and the first r_1 functions Φ_l involve only $A(j)$, $B(k)$ and the last $r_2 = r - r_1$ involve only K.

The other conditions required for the central limit theorem for $\overline{\theta}_N$ in §2.3 are 2.C3.1, 2.C3.2 and 2.C3.3. The first two are clearly satisfied for the ARMA model with constraints and the M_θ and M_μ as just described. The last, concerning the $\varepsilon(m)$, is assumed here also. Hence Theorem 2.3.1 holds for $\overline{\theta}_N$ in this ARMA case. For the other estimators $(\tilde{\theta}_N, \tilde{\mu}_N)$ and $(\hat{\theta}_N, \hat{\mu}_N)$ the Lipschitz conditions of Corollaries 2.3.2 and 2.3.3 are also satisfied for the model (1.1). The covariance matrix Ω in the asymptotic distribution of Theorem 2.3.1 may be greatly simplified for the ARMA model (1.1) when no constraints are imposed. This is given in Hannan (1975a).

When no constraints are imposed on (1.1) and $\theta_0 \in \Theta_{01}$ then initial estimates of $A(j)$, $B(k)$, K may be obtained in closed form (i.e. not involving iterative calculations) which are strongly consistent (see Hannan, 1975b).

To complete this section we will briefly discuss the case of triangular identification via C2.2, C2.3 and C2.4. For the CLT it is now required that we take Θ_{01} to be the space of structures with h triangular for which C2.3 is strengthened to C2.3' (i.e. additionally $\det g(z) \neq 0$, $|z| \leq 1$) and that the q_j are the precise degrees of the diagonal elements of h. Again to apply the CLT to this case the basic thing that must be done is to show that this new Θ_{01} is a manifold in the same manner as Theorem 4.1. Again (as in the proof of that theorem) it is clear that a ball B in R^u, $u = \sum_1^s (s-j+1)q_j + s^2p$ may be chosen so that C2.3' is preserved in that ball and the degrees of the diagonal elements of h are not decreased. The remaining thing to be done is to show that C2.2 holds and the proof of this fact follows much the same steps as in the proof for the corresponding result for simply identified structures, but is somewhat simpler.

6. Further asymptotic theory for the ARMA model

In this section we wish to discuss how the results of the previous sections extend to the case where

$$x'_{-q} = \left(x(-q+1)', \ldots, x(0)'\right) , \tag{6.1}$$

$$\varepsilon'_{-p} = \left(\varepsilon(-p+1)', \ldots, \varepsilon(0)'\right) \tag{6.2}$$

are taken as additional parameters in the model (1.1) or are given specified initial values, say for example, $x_{-q} = 0$, $\varepsilon_{-p} = 0$. The main advantage of proceeding in this way is that in some cases the equations for maximum likelihood are simplified. A second is that such a specification may be useful in modelling an ARMA process which has not yet settled into stationarity but is nonetheless stable. However this last point is of minor importance.

We will consider only the extension of the estimation results of §4 and §5 to the case of $\hat{L}_N$ since it is for this procedure that the above specification will most likely be used. If again there are N observations from (1.1) and we assume that x_{-q} and ε_{-p} are either known or are to be estimated we may write equations (1.1) for $n = 1, \ldots, N$ as

$$B_N x_N + \begin{bmatrix} bx_{-q} \\ 0 \end{bmatrix} = A_N \varepsilon_N + \begin{bmatrix} a\varepsilon_{-p} \\ 0 \end{bmatrix} \tag{6.3}$$

wherein

$$B_N = \sum_{j=0}^{q} L^j \otimes B(j) , \quad A_N = \sum_{j=0}^{p} L^j \otimes A(j) \tag{6.4}$$

and L is the $s \times s$ matrix with the elements in the first lower diagonal equal to unity and zeros elsewhere (by convention L^0 is taken to be I_s) and

$$b = \begin{bmatrix} B(q) & \cdots & B(1) \\ \cdot & \cdot & \vdots \\ 0 & & B(q) \end{bmatrix}, \quad a = \begin{bmatrix} A(p) & \cdots & A(1) \\ \cdot & \cdot & \vdots \\ 0 & & A(p) \end{bmatrix}. \tag{6.5}$$

Recall that $x_N' = \left(x(1)', \ldots, x(N)'\right)$ and now also $\varepsilon_N' = \left(\varepsilon(1)', \ldots, \varepsilon(N)'\right)$. When the $\varepsilon(j)$ are distributed as independent normal $(0, K)$ vectors, the vector

$$\varepsilon_N = A_N^{-1}\left(B_N x_N + \begin{bmatrix} bx_{-q} \\ 0 \end{bmatrix} - \begin{bmatrix} a\varepsilon_{-p} \\ 0 \end{bmatrix}\right) \tag{6.6}$$

has a normal $\left(0, I_N \otimes K\right)$ distribution so that $(-2/N)$ times the log likelihood based on this assumption is

$$\hat{\hat{L}}_N\left(A, B, K, x_{-q}, \varepsilon_{-p}\right) = \log\det K + N^{-1} y_N' \mathsf{A}_N y_N \tag{6.7}$$

where

$$y_N = B_N x_N + \begin{bmatrix} bx_{-q} \\ 0 \end{bmatrix} - \begin{bmatrix} a\varepsilon_{-p} \\ 0 \end{bmatrix}$$

and

$$\mathsf{A}_N = \left[A_N\left(I_N \otimes K\right)A_N'\right]^{-1} .$$

In equation (6.7) only "log det K" appears because $\det A_N = \det B_N = 1$ for all N.

Constraints may also be introduced of the type (2.7). However there appears to be little point in introducing constraints that also act on x_{-q}, ε_{-p} since the asymptotic limit of (6.7) is independent of $x_{-q} \in \mathrm{R}^{sq}$ and $\varepsilon_{-p} \in \mathrm{R}^{sp}$. That is, these parameters (if treated as such) cannot be identified. In finite samples the estimation of x_{-q}, ε_{-p} may increase the efficiency of the estimation of the structural parameters $A(j)$, $B(k)$, K. However as we will show in this section asymptotically this does not occur. Because the case where x_{-q}, ε_{-p} are treated as parameters is a little more difficult than when they are given fixed values we will consider only that former case from now on. We will also only need to sketch the connection between the proof of the SLLN for this present case with that of Theorem 4.1. It will be useful in what follows to write Γ_N for the covariance matrix of a sample of size N from a vector moving average with spectral density matrix

$$f = (2\pi)^{-1} gKg^* , \quad g(z) = \sum_0^p A(j)z^j .$$

Then

LEMMA 6.1. *If* Γ_N *is as just defined then*

$$A_N = \Gamma_N^{-1} + \Gamma_N^{-1}\begin{pmatrix} a \\ \cdots \\ 0 \end{pmatrix}\left[I_p \otimes K^{-1} - (a' \,\vdots\, 0)\Gamma_N^{-1}\begin{pmatrix} a \\ \cdots \\ 0 \end{pmatrix}\right]^{-1}(a' \,\vdots\, 0)\Gamma_N^{-1} .$$

Proof.

$$\Gamma_N = \begin{bmatrix} a & \vdots & A_N \\ 0 & \vdots & \end{bmatrix}(I_{N+p} \otimes K)\begin{bmatrix} a' \,\vdots\, 0 \\ \cdots\cdots \\ A_N' \end{bmatrix} .$$

Hence,

$$A_N^{-1} = \Gamma_N - \begin{pmatrix} a \\ 0 \end{pmatrix}(I_p \otimes K)(a' \,\vdots\, 0) .$$

But, by a result in Rao (1965, p. 29) applied to the r.h.s. of this, the result follows. □

Note that it follows from the proof of this lemma that $\Gamma_N - A_N(I_N \otimes K)A_N'$ is non-negative definite a fact we will use below. We will let Θ_1 and Θ_{01} denote the previously defined spaces (see §3).

LEMMA 6.2. *For all* $\theta \in \Theta_0$ *and all* $(x_{-q}', \varepsilon_{-p}')' \in R^{(q+p)s}$ *then* $\hat{\hat{L}}_N(\theta, x_{-q}, \varepsilon_{-p}) \xrightarrow{\text{a.s.}} L(\theta)$ *as* $N \to \infty$.

Proof. Let $y = \begin{bmatrix} x_{-q} \\ \cdots \\ \varepsilon_{-p} \end{bmatrix}$, $c = \begin{bmatrix} b & \vdots & a \\ & \vdots & 0 \end{bmatrix}$ (where if $q < p$ we append rows with qs zeros below b in c and not a). Then the quadratic form in $\hat{\hat{L}}_N$ is

$$\hat{\hat{Q}}_N(A,\ B,\ y) = N^{-1}x_N'B_N'A_NB_Nx_N + 2N^{-1}[(cy)' \vdots 0]A_NB_Nx_N$$

$$+ N^{-1}[(cy)' \vdots 0]A_N\begin{bmatrix} cy \\ 0 \end{bmatrix} . \quad (6.8)$$

Letting Γ_N be the $Ns \times Ns$ matrix of Lemma 6.1 we may use that lemma to express the third term in (6.8) as

$$N^{-1}((cy)' \vdots 0)\Gamma_N^{-1}\begin{pmatrix} cy \\ 0 \end{pmatrix}$$

$$+ N^{-1}((cy)' \vdots 0)\Gamma_N^{-1}\left[\begin{pmatrix} a \\ 0 \end{pmatrix}\left[\left(I_p \otimes K^{-1}\right) - (a' \vdots 0)\Gamma_N^{-1}\begin{pmatrix} a \\ 0 \end{pmatrix}\right]^{-1}(a' \vdots 0)\right]\Gamma_N^{-1} . \quad (6.9)$$

But the upper left hand $ps \times ps$ matrix of Γ_N^{-1} has bounded norm so that the first term of (6.9) converges to zero as $N \to \infty$. Also

$$\Gamma_N^{-1} \cdot \begin{pmatrix} a \\ \vdots \\ 0 \end{pmatrix}\left[I_p \otimes K^{-1} - (a' \vdots 0)\Gamma_N^{-1}\begin{pmatrix} a \\ \vdots \\ 0 \end{pmatrix}\right]^{-1}(a' \vdots 0)\Gamma_N^{-1} \le \begin{bmatrix} b_1^2 . b_2 I_{ps} & 0 \\ 0 & 0 \end{bmatrix}$$

where $\Gamma_N^{-1} \le b_1 I_{Ns}$ and

$$\begin{pmatrix} a \\ 0 \end{pmatrix}\left[I_p \otimes K^{-1} - (a' \vdots 0)\Gamma_N^{-1}\begin{pmatrix} a \\ 0 \end{pmatrix}\right]^{-1}(a' \vdots 0) \le \begin{bmatrix} b_2 I_{ps} & 0 \\ 0 & 0 \end{bmatrix} .$$

Hence also the second term in (6.9) converges to zero. The second term in (6.8) has squared modulus bounded by

$$4 . \left(N^{-1}[(cy)' \vdots 0]A_N\begin{bmatrix} cy \\ 0 \end{bmatrix}\right)\left(N^{-1}x_N'B_N'A_NB_Nx_N\right) . \quad (6.10)$$

The first factor in (6.10) converges to zero since it is just (6.9). The second factor is just the first term in (6.8) with which we now deal. Again by Lemma (6.1) we have

$$N^{-1}x_N'B_N'A_NB_Nx_N = N^{-1}x_N'B_N'\Gamma_N^{-1}B_Nx_N$$

$$+ N^{-1}x_N'B_N'\Gamma_N^{-1}\begin{bmatrix} a \\ 0 \end{bmatrix}\left[I_p \otimes K^{-1} - (a'\ 0)\Gamma_N^{-1}\begin{pmatrix} a \\ 0 \end{pmatrix}\right]^{-1}(a'\ 0)\Gamma_N^{-1} \quad (6.11)$$

in which the second term is bounded by

$$b_3 . N^{-1}x_N'B_N'\Gamma_N^{-1}\begin{pmatrix} a \\ 0 \end{pmatrix}(a'\ 0)\Gamma_N^{-1}B_Nx_N \quad (6.12)$$

where $b_3 < \infty$ is a bound for $\left[I_p \otimes K^{-1} - (a' \; 0)\Gamma_N^{-1}\begin{pmatrix} a \\ 0 \end{pmatrix}\right]^{-1}$. Now (6.12) is $b_3 N^{-1}$ times the norm of the ps-vector which has typical element $\left(a_j' \; \vdots \; 0\right)\Gamma_N^{-1} B_N x_N$. Thus using the Cauchy-Schwarz inequality (6.12) is bounded by

$$b_3 \cdot \left(N^{-1} x_N' B_N' \Gamma_N^{-1} B_N x_N\right)^{\frac{1}{2}} \left(N^{-1}\left(a_j' \; \vdots \; 0\right)\Gamma_N^{-1}\begin{pmatrix} a_j \\ 0 \end{pmatrix}\right)^{\frac{1}{2}} . \tag{6.13}$$

The second factor of this converges to zero and the first factor is just the square root of the first term in (6.11) which is now considered. We will finally show that

$$N^{-1} x_N' B_N' \Gamma_N^{-1} B_N x_N - N^{-1} x_N' \Gamma_N^{-1}(\theta) x_N \xrightarrow{\text{a.s.}} 0 \tag{6.14}$$

where $\Gamma_N(\theta)$ is the covariance matrix corresponding to the spectral density matrix $f(\omega;\,\theta) = \frac{1}{2\pi} h^{-1}\left(e^{i\omega};\,\theta\right) g\left(e^{i\omega};\,\theta\right) K(\theta) g\left(e^{i\omega};\,\theta\right)^* h^{-1}\left(e^{i\omega};\,\theta\right)^*$.

The conclusion of the lemma will then hold by application of Lemma 2.2.5, part (a). To prove (6.14) we proceed as follows:

$$\begin{aligned} & N^{-1} x_N' \left(B_N' \Gamma_N^{-1} B_N - \Gamma_N^{-1}(\theta)\right) x_N \\ & = N^{-1} x_N' B_N' \left(\Gamma_N^{-1} - \left(B_N \Gamma_N(\theta) B_N'\right)^{-1}\right) B_N x_N \\ & \le \left(N^{-1} x_N' B_N' \Gamma_N^{-1} \left(B_N \Gamma_N(\theta) B_N' - \Gamma_N\right) \Gamma_N^{-1} B_N x_N\right)^{\frac{1}{2}} \\ & \qquad \cdot \left(N^{-1} x_N' \left(B_N \Gamma_N(\theta) B_N'\right)^{-1} \left(B_N \Gamma_N(\theta) B_N' - \Gamma_N\right) \left(B_N \Gamma_N(\theta) B_N'\right)^{-1} x_N\right)^{\frac{1}{2}} . \end{aligned} \tag{6.15}$$

Now

$$\left|N^{-1} y_N \left(B_N \Gamma_N(\theta) B_N' - \Gamma_N\right) y_N\right| = \left|N^{-1} y_N' \left\{\begin{bmatrix} b \\ 0 \end{bmatrix} \Gamma_{N+q}(\theta) [b' \; 0] + 2 [b \; \vdots \; 0] \Gamma_{N+q}(\theta) \begin{bmatrix} 0 \\ \cdots \\ B_N' \end{bmatrix}\right\} y_N\right|$$

which by previous arguments converges to zero if $N^{-1} y_N' y_N$ is finite. Applying this to both factors of (6.15) the result (6.14) follows. □

Before proving the SLLN for the estimator $\hat{\hat{\theta}}_N$ obtained by minimising

$\hat{\hat{L}}_N(\theta, y)$ over $\Theta \times R^{(q+p)s}$ where $y' = \left(x'_{-q} \,\vdots\, \varepsilon'_{-p}\right)$ we will give a final lemma.

LEMMA 6.3. *If* Γ_N *is as defined in Lemma* 6.1 *then*

$$\Gamma_N^{-1} = Z - Y'X^{-1}Y$$

where A_{N+p} *is partitioned as* $\begin{bmatrix} X & \vdots & Y \\ \cdots & \cdots & \cdots \\ Y' & \vdots & Z \end{bmatrix}$, X *is* $ps \times ps$, Z *is* $Ns \times Ns$.

Proof.

$$A_{N+p}^{-1} = A_{N+p}\left(I_{N+p} \otimes K\right)A'_{N+p} \quad \text{(by definition)}$$

$$= \begin{bmatrix} a\left(I_p \otimes K\right)a' & \left[a\left(I_p \otimes K\right) \,\vdots\, 0\right]\alpha'_N \\ \alpha_N\begin{bmatrix}\left(I_p \otimes K\right)a' \\ 0\end{bmatrix} & \alpha_N\left(I_N \otimes K\right)\alpha'_N \end{bmatrix}$$

where

$$\alpha_N = \begin{bmatrix} A(p) & \cdots & A(0) & & & 0 \\ & \ddots & & \ddots & & \\ 0 & & A(p) & \cdots & A(0) \end{bmatrix},$$

has N block rows.

But $\alpha_N\left(I_N \otimes K\right)\alpha'_N = \Gamma_N$. Thus the matrix in the lower right hand corner of $Ns \times Ns$ elements in A_{N+p}^{-1} is Γ_N. But by the standard rule for inverting partitioned matrices

$$A_{N+p}^{-1} = \begin{bmatrix} X^{-1}+X^{-1}Y\left(Z-Y'X^{-1}Y\right)^{-1}Y'X^{-1} & -X^{-1}Y\left(Z-Y'X^{-1}Y\right)^{-1} \\ -\left(Z-Y'X^{-1}Y\right)^{-1}Y'X^{-1} & \left(Z-Y'X^{-1}Y\right)^{-1} \end{bmatrix}.$$

Hence equating $\Gamma_N = \left(Z-Y'X^{-1}Y\right)$ and inverting gives the result. □

We remark in passing that this gives a quick way to invert Γ_N for a vector MA process. That is, invert A_{N+p} which is easily done, form

A_{N+p} and then form $Z - Y'X^{-1}Y$. The largest matrix to be inverted is X which is only $ps \times ps$. The following theorem is the analogue of Theorem 4.1 for the present situation.

THEOREM 6.4. *For the* ARMA *model* (1.1) *with* θ_0 *known to belong to* Θ_1 *and the elements of* x_{-q} *and* ε_{-p} *treated as parameters to be estimated then* $\hat{\hat{\theta}}_N$ *converges a.s. to* θ_0 *in the Euclidean topology.*

Proof. All that needs to be done is to relate the proof for this theorem to that for $\hat{\theta}_N$ in Theorem 4.1. By Lemmas 6.2 and 4.2 and the fact that the limit in Lemma 6.2 does not depend on $y \in \mathrm{R}^{(p+q)s}$ it follows that

$$\log \det K(\theta_0) + s \geq \overline{\lim_{N\to\infty}} \inf_{\substack{\theta\in\Theta \\ y\in\mathrm{R}^{(p+q)s}}} \hat{\hat{L}}_N(\theta, y) .$$

But by the proof of Lemma 6.3, (where y_N is defined below (6.7))

$$\left(y_N+\begin{pmatrix} z\\0\end{pmatrix}\right)' A_N\left(y_N+\begin{pmatrix} z\\0\end{pmatrix}\right) = \left(y_N+\begin{pmatrix} z\\0\end{pmatrix}\right)'\begin{bmatrix} X\\ \cdots\\ Y'\end{bmatrix}[X^{-1}][X \;\vdots\; Y]\left(y_N+\begin{pmatrix} z\\0\end{pmatrix}\right)$$

$$+ \left(y_N+\begin{pmatrix} z\\0\end{pmatrix}\right)'\begin{bmatrix} 0 & 0\\ 0 & Z-Y'X^{-1}Y\end{bmatrix}\left(y_N+\begin{pmatrix} z\\0\end{pmatrix}\right)$$

$$\geq y_N'\begin{bmatrix} 0 & 0\\ 0 & Z-Y'X^{-1}Y\end{bmatrix}y_N .$$

Hence applying this to $\hat{\hat{L}}_N(\theta, y)$ gives for $\theta \in \Theta_1$ and $y \in \mathrm{R}^{(p+q)s}$,

$$\hat{\hat{L}}_N(\theta, y) \geq \log \det K(\theta) + N^{-1}\left(B_N(\theta)x_N\right)'\begin{bmatrix} 0 & \vdots & \cdots\cdots\\ 0 & \vdots & \Gamma^{-1}_{N-q-p}(\theta)\end{bmatrix}\left(B_N(\theta)x_N\right) .$$

The quadratic form of this expression may be re-written as

$$x'_{p,N}B_{N-p-q}(\theta)\Gamma^{-1}_{N-(p+q)}(\theta)B_{N-p-q}(\theta)x_{p,N} \tag{6.16}$$

where

$$x'_{p,N} = [x(p+1)', \ldots, x(N)']$$

and

$$\underset{(Ns\times(N+q)s)}{B_N} = \begin{bmatrix} B(q;\,\theta) & \dots & B(0;\,\theta) & & 0 \\ \cdot & & \cdot & & \\ & \cdot & & \cdot & \\ & & \cdot & & \cdot \\ 0 & & \cdot & & \cdot \\ & & B(q;\,\theta) & \dots & B(0;\,\theta) \end{bmatrix} .$$

We may now apply the proof for $\hat{\theta}_N$ in Theorem 4.1 to complete the present proof. □

As remarked earlier the case where x_{-q}, ε_{-p} are given fixed values may be covered by much the same type of proofs. The Central Limit Theorem for θ is an obvious modification of the earlier results.

To close this chapter a brief discussion of the results contained in Caines and Rissanen (1974a, 1974b) will be given. In these articles the model (1.1) is considered where $K = I_s$ and $A(0)$ is upper triangular with positive diagonal elements. The results (to be discussed below) are stated in Caines and Rissanen (1974a) with the proofs appearing in Caines and Rissanen (1974b). Their approach is to assume that the process $x(n)$ in (1.1) is Gaussian and they estimate the coefficients $A(j)$, $B(j)$ by use of the Gaussian likelihood constructed in terms of the prediction errors for predicting $x(m)$ from $x(0), \dots, x(m-1)$. This is asymptotically equivalent to use of the procedures $\hat{L}_N$, $\bar{L}_N$ or $\tilde{L}_N$. The result they give is in terms of the strong convergence of $\Phi_N(z) = h_N(z)^{-1}g_N(z)$ to $\Phi_0(z) = h_0(z)^{-1}g_0(z)$ where $h_N(z)$, $g_N(z)$ correspond to any estimates of $A_N(j)$, $0 \le j \le n$, $B_N(j)$, $1 \le j \le n$ (where $p = q = n$) belonging to the parameter space (see below) which give the same $\Phi_N(z)$, ($\Phi_0(z)$ corresponds to the true structure). Caines and Rissanen point out that convergence of $\Phi_N(z)$ to $\Phi_0(z)$ in the l_1-metric (i.e.:- $\sum_0^\infty \|\phi_N(j)-\phi_0(j)\| \to 0$ as $N \to \infty$, where $\Phi(z) = \sum_0^\infty \phi(j)z^j$) does not ensure the convergence of a canonical choice of the $A_N(j)$, $B_N(j)$ to a canonical choice of $A_0(j)$, $B_0(j)$ in the Euclidean topology. They suggest the use of "Kronecker indices" to obtain such canonical choices and note that the convergence of the $A_N(j)$, $B_N(j)$ to $A_0(j)$, $B_0(j)$ cannot be ensured unless there is *a priori* knowledge of the Kronecker indices. This problem is analogous to that discussed in §2.4 (see Theorem 4.1 and the associated discussion of Θ_1 and the triangularily identified space). To establish the a.s. convergence of $\Phi_N(z)$ to $\Phi(z)$ in the l_1-metric Caines and Rissanen assume that (apart from the assumption of Gaussianity) the elements of the $A(j)$ and $B(j)$ are uniformly bounded and that $\det h(z) \neq 0$, $\det g(z) \neq 0$ for $|z| \leq 1+\varepsilon$, $\varepsilon > 0$ (i.e. the zeros of these determinants are bounded away from the unit circle). In the treatment given above (see the discussion after the proof of Theorem 4.1) it was shown that $k_N(z) \to k(z)$ for all $|z| < 1$. When the more stringent conditions, used by Caines and Rissanen, just discussed are imposed in our treatment then it is possible to conclude that $k_N(z) \to k_0(z)$ uniformly for $|z| \leq 1+\varepsilon$, where ε is some positive number. Then $k_N(z) \to k_0(z)$ in the l_1-metric also. However l_1-metric convergence is not required in order that $[\theta_N]$ (corresponding to $k_N(z)$) should converge to $[\theta_0]$ (corresponding to $k_0(z)$) in the quotient topology (if $\theta_0 \in \Theta_{h,g}$) or in the Euclidean topology (if $\theta_0 \in \Theta_1$) since all that is required for Theorem 3.1 is that $k_N(z) \to k(z)$ for $|z| < 1$.

CHAPTER 4

EXTENSIONS TO THE CLT AND ESTIMATION IN MODELS FOR A SIGNAL OBSERVED WITH NOISE

1. Introduction

In Chapter 2 the Central Limit Theorem for the parameters θ specifying $k\left(e^{i\omega};\ \theta\right)$ and μ specifying $K(\mu)$ in the general finite parameter model (2.1.1) was established. The two most important requirements in order that the CLT for $\overline{\theta}_N$ (say) holds without, for example, the existence of moments higher than the second for the $\varepsilon(m)$ in (2.1.1) were

(i) the parameters θ specifying k and the parameters μ specifying K vary independently, and

(ii) the $\varepsilon(m)$ in (2.1.1) satisfy the martingale condition 2.C3.3.

Either one or both of these requirements will not hold for many naturally occurring models. In the first place constraints may be imposed in the ARMA model of Chapter 3 so that (i) above may not hold. In such cases (ii) will still be appropriate. There is thus a need to establish a CLT for θ (now specifying *both* k *and* K) in the general finite parameter model along the lines of Theorem 2.3.1. This will be done in §2 below when, for example, higher moment conditions are imposed on the $\varepsilon(m)$ (see condition 2C.3 of the next section).

A second situation in which (i) or (ii) may not apply naturally arises in connection with models for a stationary signal observed with noise. Consider

$$z(n) = y(n) + x(n) \tag{1.1}$$

where the "signal", $y(n)$, and the "noise", $x(n)$, are zero mean, stationary, purely non-deterministic processes and for the present, at least, will be taken to be incoherent one to another (i.e. $E(y(m)x(n)') = 0$, all m, n). Only $z(m)$ is observed at $m = 1, 2, \ldots, N$. Such "signal plus noise" processes were introduced in §1.6. If $y(n)$ and $x(n)$ are specified by the parameters θ_y and θ_x respectively and have spectral density matrices $f_y(\omega;\ \theta_y)$ and $f_x(\omega;\ \theta_x)$ respectively (see the general formula (1.4.2)) then the spectral density matrix of $z(n)$ is

$$f_z(\omega;\ \theta_z) = f_y(\omega;\ \theta_y) + f_x(\omega;\ \theta_x) \tag{1.2}$$

where θ_z will be used to denote $\theta_z' = (\theta_y', \theta_x')$ or a suitably transformed parameter set. Since $z(n)$ is also zero mean, stationary and purely nondeterministic it also has a one-sided representation (see §1.2) as

$$z(n) = \sum_{j=0}^{\infty} C_z(j;\ \theta_z)\varepsilon_z(n-j)\ ,\quad E\varepsilon_z(n) = 0\ ,$$

$$E\varepsilon_z(m)\varepsilon_z(n)' = \delta_{mn}K_z(\theta_z)\ , \tag{1.3}$$

where the $\varepsilon_z(n)$ are the vectors of one step prediction errors for $z(n)$. Then f_z may also be written as

$$f_z(\omega;\ \theta_z) = (2\pi)^{-1}k_z\left[e^{i\omega};\ \theta_z\right]K_z(\theta_z)k_z\left[e^{i\omega};\ \theta_z\right]^* \tag{1.4}$$

where $k_z\left[e^{i\omega};\ \theta_z\right]$ is given by formula (1.2.10) in terms of the $C_z(j;\ \theta_z)$. Now the minimisation criteria $\hat{L}_N, \bar{L}_N, \tilde{L}_N$ of previous chapters will be based on $z(1), \ldots, z(N)$ and the spectral density matrix $f_z(\omega;\ \theta_z)$. By these means θ_z may be estimated. In regard to the SLLN of §2.2 there, apart from the regularity conditions 2.C2 it was only required that the $\varepsilon_z(n)$ in (1.3) be uncorrelated (which is assured by the Wold decomposition (1.3)) and that the $\varepsilon_z(n)$ be ergodic (which will be assumed). Moreover it is not necessary that K_z and k_z be separately parameterised for the SLLN of §2.2 to apply. Since it is the main concern of this chapter to discuss the CLT and not the SLLN applied to estimation in (1.1) and since it is likely that the conditions 2.C2 may be easily checked for the commonly used models for $y(n)$ and $x(n)$ (i.e. ARMA models) we will assume that $\overline{\theta}_{z,N} \xrightarrow{a.s.} \theta_{z,0}$, for example, throughout this chapter. Also, only the procedure $\overline{L}_N$ will be discussed for simplicity of presentation.

For the CLT of §2.3 to apply to $N^{\frac{1}{2}}(\overline{\theta}_{z,N}-\theta_{z,0})$ it is required that (i) and (ii) hold. However it is not always true that K_z and k_z appearing in (1.4) may be parameterised separately because of the complicated way in which the factorisation of f_z in (1.4) is arrived at commencing from the original parameters θ_y and θ_x . However when the $\varepsilon_z(m)$ satisfy the martingale requirement (ii) above and the additional higher moment requirements of C2.3 (in the next section) the CLT for $\overline{\theta}_{z,N}$ may be established as in Theorem 2.1 below. On the other hand the condition that (at least) the $\varepsilon_z(m)$ are martingale difference with respect to their past may not be appropriate since it is not always the prediction of $z(n)$ on its past which is required but rather the prediction of $y(n)$ on the past of $z(n)$. However, when the $\varepsilon_y(m)$ and $\varepsilon_x(m)$ (in the Wold decomposition of $y(m)$ and $x(m)$) are taken together as a sequence of vectors $\eta(m)$ (where $\eta(m)' = (\varepsilon_y(m)', \varepsilon_x(m)')$) satisfying the condition C2.3 of the next section then the CLT to follow in §2 may also be applied to $\overline{\theta}_{z,N}$ specifying both k_z and K_z .

In §3 some prediction theory for $y(n)$ given the past of $z(n)$ is discussed and conditions under which the $\varepsilon_z(m)$ discussed above may reasonably be taken to satisfy the conditions needed for the CLT of the next section. In §4 the special case where $z(n)$ is scalar, $y(n)$ is generated by an autoregressive process and $x(n)$ is white noise is discussed. There it will be shown that the discussion of §3 applies and there exists a new parameterisation of the model for $z(n) = y(n) + x(n)$ for which the subset of parameters, specifying the transfer function giving the best linear predictor of $y(n)$ based on $z(m)$, $m \leq n-1$, may be efficiently estimated. For this subset of parameters the CLT of Chapter 2 (see §2.3) may be applied.

2. Extensions to the Central Limit Theorem

For the initial part of this section $x(n)$ will be taken to be a stationary, ergodic purely nondeterministic time series generated by the general finite parameter model (see (2.1.1))

$$x(n) = \sum_{j=0}^{\infty} C\left(j;\ \theta_0\right)\varepsilon(n-j) \ , \quad E\varepsilon(m) = 0 \ , \quad E\varepsilon(m)\varepsilon(m)' = K\left(\theta_0\right) \ ,$$

where θ_0 now specifies both $k\left(e^{i\omega};\ \theta_0\right)$ and $K\left(\theta_0\right)$. $x(n)$ will be assumed to satisfy conditions 2.C2 of Chapter 2, so that, for example, $\overline{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$. Throughout this section only estimation via $\overline{L}_N(\theta)$ will be discussed for reasons of space and also because the proof of the CLT (to follow) for this method contains the essential ideas. As in §2.3, θ_0 will be taken as belonging to a twice differentiable manifold M_θ of dimension u_0. The CLT for $\overline{\theta}_N$ will be established under the following conditions which replace conditions 2.C3 of §2.3.

C2.1 k and K have elements which are twice continuously differentiable functions of $\theta \in M_\theta$. The second derivitives of the elements of k with respect to θ are also continuous in $\omega \in [-\pi, \pi]$.

C2.2 $f(\omega; \theta) = \frac{1}{2\pi} k\left(e^{i\omega}; \theta\right)K(\theta)k\left(e^{i\omega}; \theta\right)^*$ has elements belonging to Λ_α , the Lipschitz class of degree α , where $\alpha > \frac{1}{2}$.

C2.3 For all $1 \le a, b, c, d \le s$ and $-\infty < n < \infty$,

(a) $E\left(\varepsilon_a(n) \mid F_{n-1}\right) = 0$ a.s.,

(b) $E\left(\varepsilon_a(n)\varepsilon_b(n) \mid F_{n-1}\right) = K_{ab}\left(\theta_0\right)$ a.s.,

(c) $E\left(\varepsilon_a(n)\varepsilon_b(n)\varepsilon_c(n) \mid F_{n-1}\right) = \beta_{abc}$ a.s.,

(d) $E\left(\varepsilon_a(n)\varepsilon_b(n)\varepsilon_c(n)\varepsilon_d(n) \mid F_{n-1}\right) = \kappa_{abcd}$ a.s.,

where $K_{ab}\left(\theta_0\right)$, β_{abc}, κ_{abcd} are constants and F_n is the sub σ-algebra generated by the elements of $\varepsilon(m)$ for $m \le n$. The last condition, which is similar to independence up to fourth moments for the $\varepsilon(m)$, ensures that the covariance matrix in the CLT below is of a reasonably simple form.

THEOREM 2.1. *Under conditions* C2 *the vector* $N^{\frac{1}{2}}\left(\overline{\theta}_N-\theta_0\right)$ *has an asymptotic normal distribution with zero mean vector and covariance matrix*

$$\Omega^{-1}(2\Omega+\Pi)\Omega^{-1} \tag{2.1}$$

where

$$\Omega_{jl} = \frac{1}{2\pi} \int \operatorname{tr}\left[f_0^{-1}(\omega) \frac{\partial f_0(\omega)}{\partial \theta_j} f_0^{-1}(\omega) \frac{\partial f_0(\omega)}{\partial \theta_l}\right] d\omega$$

and

$$\Pi_{jl} = \sum_{a}\sum_{b=1}^{s}\sum_{c}\sum_{d} \kappa_{abcd}\left[\sum_{p}\sum_{q=1}^{s}\sum_{u}\sum_{v} \left\{(2\pi)^{-1} \int \overline{k_{qb}\left(e^{i\omega}\right)}k_{pa}\left(e^{i\omega}\right)\phi_{pq}^{(j)}(\omega)d\omega\right\}\right.$$
$$\left. \cdot \left\{(2\pi)^{-1} \int \overline{k_{uc}\left(e^{i\omega}\right)}k_{vd}\left(e^{i\omega}\right)\phi_{uv}^{(l)}(\omega)d\omega\right\}\right]$$

where $\phi_{pq}^{(j)}(\omega)$ *is the* (p, q)th *element of* $\partial f_0^{-1}(\omega)/\partial\theta_j$ *and* $k_{pq}\left(e^{i\omega}\right)$ *is the* (p, q)th *element of* $k\left(e^{i\omega}; \theta_0\right)$.

Again the notation $f_0(\omega)$, $\partial f_0(\omega)/\partial\theta_j$, etc. has been used to denote $f(\omega;\,\theta)$, $\partial f(\omega;\,\theta)/\partial\theta_j$ evaluated at θ_0 , etc. Note that Ω is twice the Ω of Theorem 2.3.1 while Π_{jl} simplifies to

$$\Pi_{jl} = \sum_a \sum_b \sum_c \sum_d{}_{=1}^{s} \left[\kappa_{abcd}\left(K_0^{-1}\frac{\partial K_0}{\partial\theta_j}K_0^{-1}\right)_{ab}\left(K_0^{-1}\frac{\partial K_0}{\partial\theta_l}K_0^{-1}\right)_{cd}\right]$$

$$\left(\text{since } \frac{1}{2\pi}\int k_0^{-1}\frac{\partial k_0}{\partial\theta_j}\,d\omega = 0\ ,\ \text{for example}\right)$$

in this case. (Π_{jl} has been written in the form given in the statement of the theorem for later reference - see the latter part of this section.)

Proof of Theorem 2.1. Since $\overline{\theta}_N$ eventually enters an arbitrary neighbourhood of θ_0 we are led to consider, as in the proof of Theorem 2.3.1,

$$0 = N^{\frac{1}{2}}\frac{\partial}{\partial\theta}\overline{L}_N\left(\overline{\theta}_N\right) = N^{\frac{1}{2}}\frac{\partial}{\partial\theta}\overline{L}_N\left(\theta_0\right) + \left[\frac{\partial^2}{\partial\theta^2}\overline{L}_N\left(\phi_N\right)\right]N^{\frac{1}{2}}\left(\overline{\theta}_N-\theta_0\right)$$

where $\|\phi_N-\theta_0\| \le \|\overline{\theta}_N-\theta_0\|$ and by $\frac{\partial}{\partial\theta}\overline{L}_N\left(\theta_0\right)$, for instance, is meant the derivative evaluated at θ_0 . Thus the central limit theorem for $N^{\frac{1}{2}}\left(\overline{\theta}_N-\theta_0\right)$ reduces to that for $N^{\frac{1}{2}}\left[\frac{\partial^2}{\partial\theta^2}\overline{L}_N\left(\phi_N\right)\right]^{-1}\left[\frac{\partial}{\partial\theta}\overline{L}_N\left(\theta_0\right)\right]$. As in the of Theorem 2.3.1 the $(j,\,k)$th element of $\frac{\partial^2}{\partial\theta^2}\overline{L}_N\left(\phi_N\right)$ converges almost surely to

$$\Omega_{jk} = \frac{1}{2\pi}\int \mathrm{tr}\left\{\left[\frac{\partial}{\partial\theta_j}\log f_0(\omega)\right]\left[\frac{\partial}{\partial\theta_k}\log f_0(\omega)\right]\right\}d\omega\ .$$

Further the jth element of $N^{\frac{1}{2}}\frac{\partial\overline{L}_N\left(\theta_0\right)}{\partial\theta}$ can be written

$$N^{\frac{1}{2}}\frac{\partial\overline{L}_N\left(\theta_0\right)}{\partial\theta_j} = \frac{N^{\frac{1}{2}}}{2\pi}\int \mathrm{tr}\left\{\left[I(\omega)-f_0(\omega)\right]\phi^{(j)}(\omega)\right\}d\omega \tag{2.2}$$

where $\phi^{(j)}(\omega) = \partial f_0^{-1}(\omega)/\partial\theta_j$. Writing $f_N(\omega)$ for the Nth order Cesaro sum of the Fourier series for $f_0(\omega)$ we have, since $f_N(\omega) = EI(\omega)$,

$$\frac{N^{\frac{1}{2}}}{2\pi}\int \mathrm{tr}\left\{\left[EI(\omega)-f_0(\omega)\right]\phi^{(j)}(\omega)\right\}d\omega = \frac{N^{\frac{1}{2}}}{2\pi}\int \mathrm{tr}\left\{\left[f_N(\omega)-f_0(\omega)\right]\phi^{(j)}(\omega)\right\}d\omega$$

$$= O\left(N^{\left(\frac{1}{2}-\alpha\right)}\right)$$

By Theorem 13.14 of Zygmund (1959). Hence $f_0(\omega)$ may be replaced by $EI(\omega)$ in (2.2) to arrive at

$$A_N\left(\phi^{(j)}\right) = \frac{N^{\frac{1}{2}}}{2\pi}\int \mathrm{tr}\{[I(\omega)-EI(\omega)]\phi^{(j)}(\omega)\}d\omega\ , \quad 1 \le j \le u_0\ . \tag{2.3}$$

To see that $\phi^{(j)}(\omega)$ may be replaced by its Cesaro sum to a finite number of terms, M , consider $A_N(\delta)$ where $\delta(\omega) = \phi(\omega) - \phi_M(\omega)$, $\sup_\omega |\delta(\omega)| \le \varepsilon$ and ϕ_M is the Cesaro sum corresponding to ϕ . Then

$$A_N(\delta) = \sum_{u}\sum_{v=1}^{s} \frac{N^{\frac{1}{2}}}{2\pi}\int \left[I_{uv}(\omega)-EI_{uv}(\omega)\right]\delta_{vu}(\omega)d\omega$$

$$= \sum_u\sum_v \left\{N^{-\frac{1}{2}}\sum_m\sum_{n=1}^{N}\left[x_u(m)x_v(n)-\gamma_{uv}(n-m)\right]\delta_{vu}(m-n)\right\}\ .$$

Calling the summand in braces D_{uv} we need to consider, in evaluating the variance of $A_N(\delta)$ (see Hannan, 1970, pp. 209-211)

$$E\left[D_{rt}D_{uv}\right] = N^{-1}\sum_m\sum_n\sum_p\sum_{q=1}^{N}\left\{\gamma_{ru}(p-m)\gamma_{tv}(q-n)+\gamma_{rv}(q-m)\gamma_{tu}(p-n)\right.$$

$$\left.+\sum_b\sum_c\sum_d\sum_{e=1}^{s}\kappa_{bcde}\sum_l c_{rb}(l)c_{tc}(l+n-m)c_{ud}(t+p-m)c_{ve}(l+q-m)\right\}\delta_{tr}(m-n)\delta_{vu}(p-q)\Big\} \tag{2.4}$$

where $c_{rb}(l)$ denotes the (r, b)th element of $C\left(l, \theta_0\right)$ and $\gamma_{ru}(p-m) = E\left\{x_r(m)x_u(p)\right\}$. First consider

$$\left| N^{-1} \sum_m \sum_n \sum_p \sum_q \gamma_{ru}(p-m)\gamma_{tv}(q-n)\delta_{tr}(m-n)\delta_{vu}(p-q) \right|$$

$$= \left| N^{-1}\left(\frac{1}{2\pi}\right)^2 \iint \left[\sum_m \sum_n \delta_{tr}(m-n)e^{-i(m-n)\lambda}e^{-in(\omega+\lambda)}f_{ru}(\lambda)\right]\right.$$

$$\left.\left[\sum_p \sum_q \delta_{vu}(p-q)e^{-i(p-q)\lambda}e^{ip(\omega+\lambda)}f_{tv}(\omega)\right] d\omega d\lambda \right|$$

$$\leq B^2 \cdot \left\{\frac{N^{-1}}{(2\pi)^2} \iint \left|\sum_m \sum_n \delta_{tr}(m-n)e^{-i(m-n)\lambda}e^{-in(\omega+\lambda)}\right|^2 d\omega d\lambda\right\}^{\frac{1}{2}}$$

$$\cdot \left\{\frac{N^{-1}}{(2\pi)^2} \iint \left|\sum_p \sum_q \delta_{vu}(p-q)e^{-i(p-q)\lambda}e^{ip(\omega+\lambda)}\right|^2 d\omega d\lambda\right\}^{\frac{1}{2}}$$

where each typical element, $f_{ru}(\omega)$, of $f_0(\omega)$ has modulus bounded by B because $f_0(\omega)$ is continuous. Now the square of the first factor can be written

$$\frac{1}{2\pi} \int L_N(\omega) \int \delta_{tr}(\lambda)\delta_{tr}(\omega-\lambda) d\lambda d\omega \leq \frac{1}{2\pi} \int L_N(\omega)\left[\frac{1}{2\pi} \int |\delta_{tr}(\lambda)|^2 d\lambda\right] d\omega \leq \varepsilon^2$$

where $L_N(\omega)$ is Fejer's kernel. The second factor is handled in exactly the same way so that the contribution to $E\, D_{rt} D_{uv}$ from the first term under the braces in (2.4) may be made arbitrarily small. Similarly for the second term. Finally consider the term arising from the fourth cumulant:

$$\sum_b \sum_c \sum_d \sum_e \kappa_{bcde}\left\{N^{-1} \sum_m \sum_n \sum_p \sum_q \sum_l c_{rb}(l)c_{tc}(l+n-m)c_{ud}(l+p-m)\right.$$

$$\left. c_{ve}(l+q-m)\delta_{tr}(m-n)\delta_{vu}(p-q)\right\} .$$

The modulus of the factor in braces can be written

$$\left|\frac{N^{-1}}{2\pi}\int \frac{1}{2\pi}\int \sum_m\sum_n \delta_{rt}(n-m)e^{i(n-m)\lambda}e^{-in\omega}k_{rb}(e^{i\lambda})k_{tc}(e^{i(\omega-\lambda)})d\lambda\ .\right.$$

$$\left.\cdot\ \frac{1}{2\pi}\int \sum_p\sum_q \delta_{uv}(q-p)e^{i(q-p)\lambda}e^{-iq\omega}k_{ud}(e^{i\lambda})k_{ve}(e^{i(\omega-\lambda)})d\lambda . d\omega\right|$$

$$\leq B^2\left(\frac{N^{-1}}{2\pi}\int \cdot\frac{1}{2\pi}\int \left|\sum_m\sum_n \delta_{rt}(n-m)e^{i(n-m)\lambda}e^{-in\omega}\right|^2 d\lambda d\omega\right)^{\frac{1}{2}}\ .$$

$$\cdot\ \left(\frac{N^{-1}}{2\pi}\int \frac{1}{2\pi}\int \left|\sum_p\sum_q \delta_{uv}(q-p)e^{i(q-p)\lambda}e^{-iq\omega}\right|^2 d\lambda d\omega\right)^{\frac{1}{2}}$$

where $k_{mb}(e^{i\omega})$ has modulus bounded by $B < \infty$. Again taking the square of the first factor it follows that

$$\frac{1}{2\pi}\int L_N(\omega)\int \delta_{rt}(\lambda)\delta_{rt}(\omega-\lambda)d\lambda d\omega \leq \varepsilon^2$$

as before. Therefore the contribution to (2.4) from the fourth cumulant term may be made arbitrarily small. Hence the CLT for $A_N(\phi^{(j)})$ in (2.3) may be established by considering the CLT for

$$A_N\left(\phi_M^{(j)}\right) = \sum_r\sum_t\sum_{l=-M}^{M}\left(1-\frac{|l|}{M}\right)\phi_{rt}^{(j)}(l)N^{\frac{1}{2}}\left\{g_{rt}(l) - \frac{N-|l|}{N}\gamma_{rt}(l)\right\}\ ,\quad 1 \leq j \leq u_0$$

where $g_{rt}(l)$ is the (r, t)th element of $G(l)$ (see (1.4.3)). Now $A_N\left(\phi_M^{(j)}\right)$ is asymptotically equivalent to

$$B_N\left(\phi_M^{(j)}\right) = \sum_r\sum_t\sum_{l=-M}^{M}\left(1-\frac{|l|}{M}\right)\phi_{rt}^{(j)}(l)\tilde{\tau}_{rt}(l)\ ,\quad 1 \leq j \leq u_0 \qquad (2.5)$$

where

$$\tilde{\tau}_{rt}(l) = N^{-\frac{1}{2}}\sum_{m=1}^{N}\left(x_r(m)x_t(m+l)-\gamma_{rt}(l)\right)\ .$$

In Hannan (1976b) it is proved that when C2.3 is satisfied the necessary and sufficient condition that any finite set of the $\tilde{\tau}_{rt}(l)$ is asymptotically jointly normal is that the diagonal elements of $f_0(\omega)$ be square integrable. The condition C2.2 certainly ensures this so that the theorem of Hannan (1976b) may be applied to $B_N\left(\phi_M^{(j)}\right)$ in (2.5) since M is finite. All that remains is to evaluate the asymptotic covariance between $B_N\left(\phi_M^{(j)}\right)$ and $B_N\left(\phi_M^{(k)}\right)$. Now, applying formula (3) of Hannan (1976b),

$$E\left\{B_N\left(\phi_M^{(j)}\right)B_N\left(\phi_M^{(k)}\right)\right\}$$

$$= \sum_r\sum_t\sum_u\sum_v \sum_{l,m=-M}^{M} \left(1-\frac{|l|}{M}\right)\left(1-\frac{|m|}{M}\right)\phi_{rt}^{(j)}(l)\phi_{uv}^{(k)}(m)E\left[\tilde{\tau}_{rt}(l)\tilde{\tau}_{uv}(m)\right]$$

$$\to \sum_r\sum_t\sum_u\sum_v\sum_{l,m} \left(1-\frac{|l|}{M}\right)\left(1-\frac{|m|}{M}\right)\phi_{rt}^{(j)}(l)\phi_{uv}^{(k)}(m)\left\{\frac{1}{2\pi}\int \overline{f_{tv}(\omega)}f_{ru}(\omega)e^{-i(m-l)\omega}d\omega\right.$$

$$+\frac{1}{2\pi}\int f_{rv}(\omega)\overline{f_{tu}(\omega)}e^{i(m+l)\omega}d\omega + \sum_a\sum_b\sum_c\sum_d \kappa_{abcd}\frac{1}{2\pi}\int \overline{k_{tb}\left(e^{i\omega}\right)k_{ra}\left(e^{i\omega}\right)}e^{il\omega}d\omega$$

$$\left.\cdot\frac{1}{2\pi}\int \overline{k_{uc}\left(e^{i\omega}\right)}k_{vd}\left(e^{i\omega}\right)e^{-im\omega}d\omega\right\}$$

as $N\to\infty$. By taking M large this last line is easily seen to be arbitrarily close to

$$\sum_r\sum_t\sum_u\sum_{v=1}^{s}\left\{\frac{1}{2\pi}\int \overline{f_{tv}(\omega)}f_{ru}(\omega)\overline{\phi_{rt}^{(j)}(\omega)}\phi_{uv}^{(k)}(\omega)d\omega\right.$$

$$+\frac{1}{2\pi}\int f_{rv}(\omega)\overline{f_{tu}(\omega)}\overline{\phi_{rt}^{(j)}(\omega)}\ \overline{\phi_{uv}^{(k)}(\omega)}d\omega + \sum_a\sum_b\sum_c\sum_{d=1}^{s}\kappa_{abcd}\frac{1}{2\pi}\int \overline{k_{tb}\left(e^{i\omega}\right)k_{ra}\left(e^{i\omega}\right)}\ .$$

$$\left.\cdot\ \overline{\phi_{rt}^{(j)}(\omega)}d\omega\ \frac{1}{2\pi}\int \overline{k_{uc}\left(e^{i\omega}\right)}k_{vd}\left(e^{i\omega}\right)\phi_{uv}^{(k)}(\omega)d\omega\right\}.$$

The sum over the first two terms is $2\Omega_{jk}$ and the sum over the last term Π_{jk}. This completes the proof. □

The application of the above theorem to the signal plus noise model (1.1) will now be discussed. In the first place the parameters θ_z specifying $f_z(\omega;\ \theta_z)$ given in (1.4) may be estimated by $\bar{L}_N(\theta_z)$ using the observations $z(1), \ldots, z(N)$. Then, if the $\varepsilon_z(m)$ in (1.3) satisfy C2.3 and f_z, k_z, K_z satisfy C2.1 and C2.2 the CLT for $N^{\frac{1}{2}}\left(\theta_{z,N}-\theta_{z,0}\right)$ is given by Theorem 2.1. Conditions under which at least C2.3(a) may reasonably be assumed are discussed in the next section. However not all signal plus noise cases may be treated this way. One situation in which C2.3 is not assumed to hold for $\varepsilon_z(m)$ but for which the CLT of Theorem 2.1 still applies is as follows. Consider

$$y(n) = \sum_0^\infty C_y(j;\ \theta_y)\varepsilon_y(n-j) \ , \quad x(n) = \sum_0^\infty C_x(j;\ \theta_x)\varepsilon_x(n-j) \ . \tag{2.6}$$

Let $\eta(n)' = (\varepsilon_y(n)',\ \varepsilon_x(n)')$ and let $F_\eta(n)$ be the σ-algebra generated by $\eta(m)$, $m \le n$. We will assume that $\eta(n)$ satisfies C2.3 but will now refer to the constants in this as $K_{ab}^{(\eta)}$, $\beta_{abc}^{(\eta)}$ and $\kappa_{abcd}^{(\eta)}$ respectively. However it need not be true that the $\varepsilon_z(m)$ in the representation (1.3) satisfy C2.3 in this case. An alternative representation of $z(n)$ is

$$z(n) = \sum_{j=0}^\infty D(j;\ \theta_z)\eta(n-j) \tag{2.7}$$

where $D(j;\ \theta_z) = [C_y(j;\ \theta_y) \ \vdots \ C_x(j;\ \theta_x)]$. If now C2.1 and C2.2 are assumed (and hence it is implicitly assumed that the factorisation (1.4) of f_z exists with $\theta_z' = (\theta_y',\ \theta_x')$ or θ_z is an equivalent parameter set) and C2.3 is assumed for $\eta(n)$ then Theorem 2.1 continues to hold. To see that this is plausible consider the following. The proof of Theorem 2.1 up to equation (2.4) is not changed by this new specification since only C2.1 and C2.2 are required in the argument to this point.

Now the fourth cumulant term in (2.4) needs to be modified by increasing the upper limit of summation to $2s$, replacing κ_{bcde} by $\kappa_{bcde}^{(\eta)}$ and using $c_{rb}(l)$ to denote the $(r,\ b)$th element of the $s \times 2s$ matrix $D(l)$ given above. This replacement does not effect the validity of the steps in the proof that $A_N(\delta)$ has negligible variance. Thus the proof up to (2.5) is valid here also. By examining the proof of the theorem in Hannan

(1976b) the same replacements as described above may be made and that theorem still applies. (That is the proof of the theorem in Hannan (1976b) does not require that $D(j)$ be square.) Thus the CLT for $\overline{\theta}_{z,N}$ holds for this case. The asymptotic covariance matrix (2.1) may be obtained (see, in particular, Π_{jl}) in the form stated but with the elements of $k(e^{i\omega})$ replaced by those of the $2s \times s$ matrix $\sum_0^\infty D(j)e^{ij\omega}$, the range of summation increased to $2s$ and κ_{abcd} replaced by $\kappa_{abcd}^{(\eta)}$.

When k_z and K_z are separately parameterised but C2.3 only holds for $\eta(n)$ and not $\varepsilon_z(n)$ then the discussion just given still applies. However if now the previous θ_z is partitioned as $\theta_z^{(1)}$ (to specify k_z) and $\theta_z^{(2)}$ (to specify K_z) the asymptotic covariance matrix of Theorem 2.1 easily simplifies to

$$\begin{bmatrix} 2\Omega^{(1)-1} & 0 \\ 0 & \Omega^{(2)-1}\left(2\Omega^{(2)}+\Pi^{(2)}\right)\Omega^{(2)-1} \end{bmatrix}$$

where, in an obvious notation, $2\Omega^{(1)-1}$ is the asymptotic covariance corresponding to $\theta_z^{(1)}$. This means that the asymptotic covariance for $\theta_z^{(1)}$ does not depend on fourth moments (i.e. does not depend on $\kappa_{abcd}^{(\eta)}$). It is probably true that in this case the CLT for $N^{\frac{1}{2}}\left(\overline{\theta}_{z,N}^{(1)}-\theta_{z,0}^{(1)}\right)$ may be established along similar lines to that given in §2.3 without the assumption that moments higher than the second exist.

Finally, if the $\varepsilon_x(n)$ and $\varepsilon_y(n)$ are Gaussian and independent the fourth cumulant term Π vanishes from the asymptotic covariance (2.1). Also in this Gaussian case the $\varepsilon_z(n)$ satisfy C2.3 directly so that there is no need for the modification to the proof of Theorem 2.1 mentioned above.

3. Some prediction theory for the signal plus noise model

Initially in this section $x(n)$ in (1.1) will be taken to be white noise. By this is meant that $x(n)$ are zero mean, serially uncorrelated vectors with $Ex(n)x(n)' = K_x$. The signal $y(n)$ will be taken to be a stationary nondeterministic process with one sided representation

$$y(n) = \sum_{j=0}^{\infty} C_y(j)\varepsilon_y(n-j) \, , \quad C_y(0) = I_s \tag{3.1}$$

in which the $\varepsilon_y(n)$ form a white noise sequence with covariance matrix K_y. If $x(n)$ and $y(n)$ are incoherent then (see (1.2))

$$f_z(\omega) = (2\pi)^{-1}\left\{k_y\left(e^{i\omega}\right)K_y k_y\left(e^{i\omega}\right)^* + K_x\right\} \tag{3.2}$$

which may also be written in the following form (see (1.4))

$$f_z(\omega) = (2\pi)^{-1}k_z\left(e^{i\omega}\right)K_z k_z\left(e^{i\omega}\right)^* \, . \tag{3.3}$$

Reference to θ_x, θ_y, θ_z has been suppressed in the above because these parameters play no part in the following discussion.

Since only $z(n)$ is observed in (1.1) then prediction of the signal $y(n)$ must be based on $z(m)$, $m \le n-1$. Once the $C_y(j)$ in (3.1) are known the optimal *linear* filter for predicting $y(n)$ in terms of $z(m)$, $m \le n-1$, may be constructed using only these $C_y(j)$ (see the proof of the theorem to follow). It is therefore relevant, if the above model for $z(n)$ (with $y(n)$ as in (3.1)) is correct, to impose the condition that the best linear predictor of $y(n)$ is the best predictor of $y(n)$ (both based on $z(m)$, $m \le n-1$). To discuss this the following notation is convenient. Let

$$\left.\begin{aligned}
\hat{y}(n) &= \text{the best predictor of } y(n) \text{ given } z(m),\, m \le n-1 \\
\dot{y}(n) &= \text{the best } \textit{linear} \text{ predictor of } y(n) \text{ given } z(m),\, m \le n-1 \\
\hat{z}(n) &= \text{the best predictor of } z(n) \text{ given } z(m),\, m \le n-1 \\
\dot{z}(n) &= \text{the best } \textit{linear} \text{ predictor of } z(n) \text{ given } z(m),\, m \le n-1 \, .
\end{aligned}\right\} \tag{3.4}$$

Let $F_z(n)$ be the σ-algebra generated by the elements of $z(m)$ (equivalently by the elements of $\varepsilon_z(m)$ given in (1.3)) for $m \le n$ and let $F_x(n)$, $F_y(n)$ be similarly defined for $x(m)$, $y(m)$ respectively. Then $\hat{y}(n) = E\big(y(n) \mid F_z(n-1)\big)$ and $\hat{z}(n) = E\big(z(n) \mid F_z(n-1)\big)$.

THEOREM 3.1. *If* $x(n)$, $y(n)$ *and* $z(n) = y(n) + x(n)$ *are as described above and if*

C3.1: $E\big(x(m) \mid F_z(m-1)\big) = 0$ *a.s., all* m ,

then $\hat{z}(n) = \dot{z}(n)$ *if and only if* $\hat{y}(n) = \dot{y}(n)$.

Proof. The result will be established in two steps. The first shows that $\dot{y}(n) = \dot{z}(n)$, the second that $\hat{y}(n) = \hat{z}(n)$.

(i) $\dot{y}(n) = \dot{z}(n)$. By Theorem 10', p. 173, Hannan (1970), the response function of the optimal linear filter for $y(n)$ given $z(m)$, $m \le n-1$ is

$$h\left(e^{i\omega}\right) = e^{i\omega}\left[e^{-i\omega} f_y\left(e^{i\omega}\right) k_z^*\left(e^{i\omega}\right)^{-1}\right]_+ K_z^{-1} k_z\left(e^{i\omega}\right)^{-1}$$

where $\left[g\left(e^{i\omega}\right)\right]_+$ denotes that only non-negative powers of $e^{i\omega}$ are to be taken in the (matrix) series expansion of the (matrix) function $g\left(e^{i\omega}\right)$. Using (2.2) f_y may be replaced in $\left[e^{-i\omega} f_y\left(e^{i\omega}\right) k_z^*\left(e^{i\omega}\right)^{-1}\right]_+$ to reach

$$h\left(e^{i\omega}\right) = e^{i\omega}\left[e^{-i\omega} k_z\left(e^{i\omega}\right) + e^{-i\omega} K_x k_z^*\left(e^{i\omega}\right)^{-1} K_z^{-1}\right]_+ k_z\left(e^{i\omega}\right)^{-1} .$$

But the second term in $[\ ]_+$ of this expression has only negative powers of $e^{i\omega}$ so that it makes no contribution to $h\left(e^{i\omega}\right)$. Thus

$$h\left(e^{i\omega}\right) = e^{i\omega}\left(\sum_{j=1}^{\infty} C_z(j) e^{i(j-1)\omega}\right) k_z\left(e^{i\omega}\right)^{-1}$$

$$= I_s - k_z\left(e^{i\omega}\right)^{-1} .$$

On the other hand, the transfer function giving $\dot{z}(n)$ is easily obtained (see Theorem 1, p. 129 and Theorem 1", p. 163, Hannan (1970)) as this

$h(e^{i\omega})$. Thus $\dot{z}(n) = \dot{y}(n)$.

(ii) $\hat{y}(n) = \hat{z}(n)$. Now

$$\begin{aligned}\hat{z}(n) &= E\big(z(n) \mid F_z(n-1)\big)\\ &= E\big(y(n) \mid F_z(n-1)\big) + E\big(x(n) \mid F_z(n-1)\big)\\ &= \hat{y}(n) \quad \text{a.s.,}\end{aligned}$$

since by assumption C3.1 the second term in the previous expression is null. □

Note that $\dot{y}(n) = \dot{z}(n)$ even when C3.1 does not hold. Insofar as the requirement $\dot{y}(n) = \hat{y}(n)$ is natural (for linear modelling to be appropriate) then the additional condition C3.1 ensures that $\dot{z}(n) = \hat{z}(n)$. But $\dot{z}(n) = \hat{z}(n)$ if and only if

$$E\big(\varepsilon_z(n) \mid F_z(n-1)\big) = 0 \quad \text{a.s., all } m \ , \tag{3.6}$$

(see §2.3). This means that part (a) of C2.3 is satisfied for $\varepsilon_z(m)$.

It would be of interest to determine the minimal conditions under which the remainder of C2.3 hold for $\varepsilon_z(m)$ but we have not done this. Some conditions on $y(n)$ and $x(n)$ which ensure C3.1 are as follows:

(a) If the $x(n)$ are serially independent and independent of $y(n)$ then C3.1 holds.

(b) If $\eta(n)' = \big(\varepsilon_y(n)', x(n)'\big)$ and $F_\eta(n)$ is the σ-algebra generated by $\eta(m)$, $m \le n$ then if $E\big(\eta(m) \mid F_\eta(m-1)\big) = 0$ a.s., all m , the condition C3.1 holds.

Other conditions are no doubt available to ensure C3.1.

4. The model for an autoregressive signal observed with noise

Recently Pagano (1974) has considered the model (1.1), for $s = 1$, in which $y(n)$ satisfies the autoregression

$$\sum_{j=0}^{q} \beta(j)y(n-j) = \varepsilon_y(n) \ , \quad \beta(0) = 1 \ , \quad E\varepsilon_y(m)\varepsilon_y(n) = \delta_{mn}\sigma_y^2 \ , \tag{4.1}$$

where $h(\zeta) = \sum_{j=0}^{q} \beta(j)\zeta^j$ has all zeros outside the unit circle (i.e. for $|\zeta| > 1$) and $x(n)$ is white noise with $Ex(m)x(n) = \delta_{mn}\sigma_x^2$. In the above $x(n)$ and $y(n)$ are taken to be at least incoherent. Pagano (1974) considers the estimation of the parameters $\beta(1), \ldots, \beta(q), \sigma_y^2, \sigma_x^2$ and establishes the asymptotic consistency and efficiency of the estimators when it is assumed that $x(n)$ and $y(n)$ are independent one to another and are Gaussian. These estimators are obtained by non-linear least squares using consistent estimators as starting values. In the following we will take a different approach and base estimation on $\hat{L}_N, \tilde{L}_N, \bar{L}_N$. It will be convenient to define the vector of parameters

$$\theta' = \left\{\beta(1), \ \ldots, \ \beta(q), \ \sigma_y^2, \ \sigma_x^2\right\} . \tag{4.2}$$

The vector θ will be taken to belong to the set

$$\Theta = \left\{\theta \in R^{(q+2)} : h(\zeta) \neq 0, \ |\zeta| \leq 1; \ \beta(q) \neq 0; \ \sigma_x^2 > 0, \ \sigma_y^2 > 0\right\} . \tag{4.3}$$

(The assumption that $\beta(q) \neq 0$ pre-supposes that the true degree, q, of the autoregression is known. This is an identification requirement for the above autoregressive signal plus noise model.)

In the model just described the spectral density of $z(n)$ (given in (1.2)) becomes

$$f_z(\omega; \theta) = \frac{1}{2\pi} \left\{\frac{\sigma_y^2}{\left|\sum_0^q \beta(j)e^{ij\omega}\right|^2} + \sigma_x^2\right\} . \tag{4.4}$$

As is well known (see Pagano (1974), for example) this may be rewritten (in the form (1.4)) as

$$f_z(\omega;\ \theta) = \frac{\sigma_z^2}{2\pi} \cdot \frac{\left|\sum_0^q \alpha(j)e^{ij\omega}\right|^2}{\left|\sum_0^q \beta(j)e^{ij\omega}\right|^2}, \tag{4.5}$$

where σ_z^2 and $\alpha(1), \ldots, \alpha(q)$ depend on θ. In (4.5) it is always possible to choose the $\alpha(j)$ so that $g(\zeta) = \sum_{j=0}^{q} \alpha(j)\zeta^j$ has all zeros outside the unit circle.

The procedure $\overline{L}_N(\theta)$ (which will be the only one discussed below) may be used to obtain the estimator of θ defined in (4.2) as $\overline{\theta}_N$. Since it is our intention to discuss only the CLT for $\overline{\theta}_N$ below and not the SLLN we will assume that $\overline{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$. (If for example θ_0 belongs to that subset of Θ for which $0 < c \le \sigma_x^2$, $\sigma_y^2 \le b < \infty$ then the conditions for the SLLN of Chapters 2 or 3 will apply.) For the parameterisation of the problem chosen (i.e. θ in (4.2)) the CLT of Chapter 2 will not apply since both σ_z^2 and $k\left(e^{i\omega}\right) = \left(\sum_0^q \beta(j)e^{ij\omega}\right)^{-1}\left(\sum_0^q \alpha(j)e^{ij\omega}\right)$ depend on the same vector of parameters θ. (That this joint dependence on θ is not vacuous may be seen in the simplest example of the above type, i.e. when $q = 1$, as discussed on p. 171 of Hannan (1970).) However the CLT of §2 may be applied in this case to $N^{\frac{1}{2}}\left(\overline{\theta}_N-\theta_0\right)$ as follows. In the first place conditions C2.1 and C2.2 may be shown to hold on the parameter space Θ defined in (4.3) - see also the discussion below. Secondly when the $x(n)$ are taken as serially independent and independent of $y(n)$, Theorem 3.1 applies so that part (a) of C2.3 is not an unnatural requirement here. If further the remainder of C2.3 are assumed then the conditions of Theorem 2.1 hold so that $N^{\frac{1}{2}}\left(\overline{\theta}_N-\theta_0\right)$ has the asymptotic distribution of that theorem.

However, as will now be discussed, there is a different parameterisation of the above model which has certain advantages. We propose to take as parameters

$$\tau' = \{\beta(1), \ldots, \beta(q), \alpha(q)\} \tag{4.6}$$

and σ_z^2 . The appropriate space to consider for τ and σ_z^2 is $T \times R^+$, where R^+ is the positive real line and

$$T = \left\{\tau \in R^{q+1} : h(\zeta) \neq 0,\ |\zeta| \leq 1;\ 0 < \alpha(q)/\beta(q) < 1;\ \beta(q) \neq 0\right\} . \tag{4.7}$$

To relate the different parameterisations, θ and $\left(\tau, \sigma_z^2\right)$, and the corresponding spaces, Θ and $T \times R^+$, we proceed as follows. Equating (4.4) and (4.5), multiplying through by $\left|\sum_0^q \beta(j)e^{ij\omega}\right|^2$ and taking Fourier coefficients gives

$$\sigma_z^2 \sum_{j=0}^{q-l} \alpha(j)\alpha(j+l) = \delta_{0l}\cdot\sigma_y^2 + \sigma_x^2 \sum_{j=0}^{q-l} \beta(j)\beta(j+l) , \quad 0 \leq l \leq q . \tag{4.8}$$

Now, given any $\theta \in \Theta$ there exists a unique set of $\alpha(1), \ldots, \alpha(q), \sigma_z^2$ with $g(\zeta) = \sum_0^q \alpha(j)\zeta^j$ having all zeros outside the unit circle. (That this is a unique set follows from the fact that there is only one such factorisation of the spectral density $\sigma_y^2 + \sigma_x^2\left|\sum_0^q \beta(j)e^{ij\omega}\right|^2$ which corresponds to a moving average with all zeros outside the unit circle.) However if $\alpha(q)$ and $\beta(1), \ldots, \beta(q)$ are uniquely specified then $\alpha(1), \ldots, \alpha(q-1)$ may be obtained by solving

$$\sum_{j=0}^{q-l} \alpha(j)\alpha(j+l) - \frac{\alpha(q)}{\beta(q)} \sum_{j=0}^{q-l} \beta(j)\beta(j+l) = 0 , \quad 1 \leq l \leq q-1 . \tag{4.9}$$

(These are just (4.8), for $1 \leq l \leq q-1$, with $\sigma_z^2\alpha(q)/\beta(q) = \sigma_x^2$, from the equation (4.8) for $l = q$, substituted.) The solution to (4.9) for

$\alpha(1), \ldots, \alpha(q-1)$ may be shown to be unique as follows. Call each l.h.s. in (4.9) $\Phi_l\big(\beta(1), \ldots, \beta(q), \alpha(q); \alpha(1), \ldots, \alpha(q-1)\big)$ for $1 \le l \le q-1$. Then by the implicit function theorem there is a unique solution $\big(\alpha(1), \ldots, \alpha(q-1)\big)$ in terms of $\big(\beta(1), \ldots, \beta(q), \alpha(q)\big)$ to (4.9) provided the $(q-1) \times (q-1)$ matrix with (l, k)th element $\partial\Phi_l/\partial\alpha(k)$, for $1 \le k$, $l \le q-1$, has full rank. This is just the condition that the matrix

$$\begin{bmatrix} 1 & & & \\ \alpha(1) & \ddots & & 0 \\ \vdots & \ddots & \ddots & \\ \alpha(q-2) & \cdots & \alpha(1) & 1 \end{bmatrix} + \begin{bmatrix} \alpha(2) & \cdots & \alpha(q-1) & \alpha(q) \\ \vdots & \iddots & \iddots & \\ \alpha(q-1) & \iddots & & 0 \\ \alpha(q) & & & \end{bmatrix} \qquad (4.10)$$

be of full rank. But by the results in Marden (1949, pp. 152-155) (4.10) has non-zero determinant if the zeros of $g(\zeta) = \sum_0^q \alpha(j)\zeta^j$ are all outside the unit circle which is true here. Hence the $\alpha(j)$, for $1 \le j \le q-1$, in (4.5) may be written as functions only of τ and we will write $\alpha(j, \tau)$, $1 \le j \le q-1$, to emphasize this. Furthermore, since the Φ_l defined above are (at least) twice continuously differentiable in $\alpha(1), \ldots, \alpha(q-1)$ and (4.10) is of full rank the functions $\alpha(j; \tau)$ solving (4.9) are also twice continuously differentiable functions of τ (see Matsushima (1972, p. 24), for example). This remark is of relevance to the CLT to be discussed shortly. For the given $\theta \in \Theta$ there exist $\sigma_z^2, \alpha(1), \ldots, \alpha(q)$ such that (4.8) holds. But, as we have just seen $\alpha(1), \ldots, \alpha(q-1)$ may be obtained in terms of $\tau' = \big(\beta(1), \ldots, \beta(q), \alpha(q)\big)$. Furthermore this τ belongs to T since by equation (4.8) for $l = q$, since $\sigma_z^2 > \sigma_x^2 > 0$, it follows that $\alpha(q)$ have the same sign, i.e. that $\alpha(q)/\beta(q) > 0$ and that $\alpha(q)/\beta(q) < 1$. (Note, since $|\beta(q)| < 1$ then $|\alpha(q)| < 1$ also.) If $\bar{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$ then $\bar{\tau}_N, \bar{\sigma}^2_{z,N}$ (corresponding to

$\overline{\theta}_N$) will converge a.s. to τ_0, $\sigma^2_{z,0}$ (corresponding to θ_0). Now take a small open neighbourhood of $\left(\tau_0, \sigma^2_{z,0}\right) \in T \times R^+$. For τ fixed σ^2_z may vary in this neighbourhood and as it varies there will be a signal plus noise model of the above type (i.e. a parameter $\theta \in \Theta$) corresponding to each σ^2_z. On the other hand, fix σ^2_z and let τ vary in this neighbourhood. Then there is also a $\theta \in \Theta$ corresponding to this τ. This means that for N large (at least) $\overline{L}_N$ may be minimised as a function of τ and σ^2_z with τ specifying $k\left(e^{i\omega};\ \tau\right) = \left\{\sum_0^q \beta(j)e^{ij\omega}\right\}^{-1}\left\{\sum_0^q \alpha(j;\ \tau)e^{ij\omega}\right\}$ alone and σ^2_z may vary freely from τ. Since $k\left(e^{i\omega};\ \tau\right)$ is a twice continuously differentiable function of τ (see the argument below (4.10)) condition 2.C3.1 is satisfied. We have already noted that when $x(n)$ are serially independent and independent of $y(n)$ part (a) of C2.3 for $\varepsilon_z(n)$ in (using the new notation)

$$z(n) = \sum_0^\infty c(j;\ \tau)\varepsilon_z(n-j)\ ,\quad k\left(e^{i\omega};\ \tau\right) = \sum_0^\infty c(j;\ \tau)e^{ij\omega}\ , \tag{4.11}$$

is not unreasonable. But C2.3 (a) is just 2.C3.3(a). Hence assuming also 2.C3.3(b) the CLT of Chapter 2 (Theorem 2.3.1) may be applied to yield the asymptotic normality of $N^{\frac{1}{2}}\left(\overline{\tau}_N-\tau_0\right)$ without, in particular, the extra moment conditions (see C2.3) of Theorem 2.1. The CLT of Theorem 2.3.1 for $N^{\frac{1}{2}}\left(\overline{\mu}_N-\mu_0\right)$ applies also to $N^{\frac{1}{2}}\left(\overline{\sigma}^2_{z,N}-\sigma^2_{z,0}\right)$.

One advantage of the above parameterisation in terms of τ and σ^2_z is that the CLT for the vector τ may be established under more general conditions than can the CLT for θ in (4.2). Since the transfer function giving the best linear predictor of $y(n)$ based on $z(m)$, $m \le n-1$ is completely specified by $\alpha(1), \ldots, \alpha(q)$, $\beta(1), \ldots, \beta(q)$ (see §3), that is by τ , the efficient estimation of τ may be of principal interest. Also it would appear no easier to estimate θ than $\left(\tau, \sigma_z^2\right)$ via $\bar{L}_N$, $\tilde{L}_N$ or $\hat{L}_N$.

CHAPTER 5

ASYMPTOTIC THEORY FOR REGRESSION MODELS

1. Introduction

In this chapter we will discuss the application of the previous methods $\hat{L}_N$, $\tilde{L}_N$, $\bar{L}_N$ to the linear regression model

$$z(n) = \Delta y(n) + x(n) \tag{1.1}$$

where $z(n)$, $x(n)$ are s-vectors, $y(n)$ is a t-vector and Δ is $s \times t$. We assume that $z(1), \ldots, z(N)$ and $y(1), \ldots, y(N)$ are observed and we wish to estimate Δ and the parameters θ specifying the spectrum of $x(n)$. For this $x(n)$ will be taken to be a stationary ergodic and nondeterministic[1] process of the type discussed in Chapter 2 (see (2.1.1)). The $y(n)$ sequence is taken to be independent of $x(n)$ and in this chapter will be considered to be a deterministic sequence with elements satisfying the Grenander conditions (see Hannan (1970, p. 77):

C1.1 (a) $\lim_{N\to\infty} d_j(N)^2 = \infty$, $d_j(N)^2 = \sum_{m=1}^{N} y_j(m)^2$, $1 \le j \le t$;

(b) $\lim_{N\to\infty} \dfrac{y_j(N)^2}{d_j(N)^2} = 0$, $1 \le j \le t$;

(c) $\lim_{N\to\infty} G_{jk}^{(Y)}(n) = \Gamma_{jk}^{(Y)}(n)$, $|\Gamma_{jk}^{(Y)}(n)| \le 1$

where $G_{jk}^{(Y)}(n) = d_j(N)^{-1} d_k(N)^{-1} \sum_{m=1}^{N} y_j(m) y_k(m+n)$ for $1 \le j \le t$, $1 \le k \le t$, $-\infty < n < \infty$. We let $\Gamma^{(Y)}(n)$ be the matrix with $\Gamma_{jk}^{(Y)}(n)$ as (j, k)th element.

As a consequence of C1.1 (c) there exists an Hermitian nondecreasing matrix valued function of ω, $F_Y(\omega)$ such that

[1] In this chapter we will use a stronger form of nondeterminism than that used previously - see §2 below.

$$\Gamma^{(Y)}(n) = \int_{-\pi}^{\pi} e^{in\omega} F_Y(d\omega) \; . \tag{1.2}$$

It will usually be assumed below that

C1.2 $\Gamma^{(Y)}(0)$ is positive definite.

Also, the following condition limiting the "rate of growth" of $d_j(N)^2$ will sometimes be assumed below.

C1.3 $\lim_{N\to\infty} N^{-\beta_j} \log\left(d_j(N)^2\right) = 0$ where $0 \le \beta_j < \frac{1}{2}$.

Among more recent articles concerning the estimation of (1.1) are Hannan (1971a) (for $s = 1$) and Robinson (1972) (for $s > 1$) where also $y(n)$ is taken to be a nonlinear function of a further parameter, τ say. In these treatments the estimation of θ specifying $x(n)$ is only briefly discussed. However we will not consider the nonlinear regression situation below but will be content with establishing some asymptotic theory for the joint estimation of Δ and θ by similar methods to those of previous sections. In Hannan (1973a) this case for $s = 1$, $t \ge 1$ is discussed in a general manner. The following treatment is aimed at extending this discussion to $s > 1$, $t \ge 1$. General treatments of the asymptotic theory for the estimation of Δ alone in (1.1) (by frequency domain methods, i.e. via $\tilde{L}_N$) are given in Hannan (1973b) and (1973c).

More generally the $y(n)$ sequence is sometimes taken to depend upon N in which case the notation $y^{(N)}(n)$ is used. Much of the treatment in this chapter does not cover this case. However when results apply equally well to sequences of the type $y^{(N)}(n)$ as to $y(n)$ it will be pointed out.

By analogy with the method $\hat{L}_N$ of previous chapters we consider

$$\hat{L}_N(\theta, \Delta) = N^{-1} \log \det \Gamma_N(\theta) + \hat{S}_N(\theta, \Delta) \tag{1.3a}$$

where

$$\hat{S}_N(\theta, \Delta) = N^{-1}\left(z_N - (I_s \otimes \Delta) y_N\right)' \Gamma_N^{-1}(\theta) \left(z_N - (I_s \otimes \Delta) y_N\right)$$

and in this, $z'_N = (z(1)', \ldots, z(N)')$, $y'_N = (y(1)', \ldots, y(N)')$ and $\Gamma_N(\theta)$ is as in (2.1.4). The approximations $\tilde{L}_N$, $\overline{L}_N$ are analogously defined as

$$\tilde{L}_N(\theta, \Delta) = \log \det K(\theta) + \tilde{S}_N(\theta, \Delta) \qquad (1.3b)$$

where

$$\tilde{S}_N(\theta, \Delta) = (N')^{-1} \sum_u \mathrm{tr}\left\{f^{-1}(\theta)\left[I_Z - I^*_{YZ}\Delta^* - \Delta I_{YZ} + \Delta I_Y \Delta^*\right]\right\} ;$$

$$\overline{L}_N(\theta, \Delta) = \log \det K(\theta) + \overline{S}_N(\theta, \Delta) \qquad (1.3c)$$

where

$$\overline{S}_N(\theta, \Delta) = (2\pi)^{-1} \int \mathrm{tr}\left\{f^{-1}(\theta)\left[I_Z - I^*_{YZ}\Delta^* - \Delta I_{YZ} + \Delta I_Y \Delta^*\right]\right\} d\omega .$$

In $\tilde{S}_N$ and $\overline{S}_N$, $I_Z = W_Z W_Z^*$, $I_{YZ} = W_Y W_Z^*$ and W_Z, W_Y are the DFT's corresponding to $z(n)$ and $y(n)$ given by (1.3.7). We will call $(\hat{\theta}_N, \hat{\Delta}_N)$, $(\tilde{\theta}_N, \tilde{\Delta}_N)$, $(\overline{\theta}_N, \overline{\Delta}_N)$ the values of (θ, Δ) minimising $\hat{L}_N$, $\tilde{L}_N$, $\overline{L}_N$ (respectively) over $\Theta^* \times R^{ts}$ where Θ^* is a suitably defined (see the next section) subset of $\overline{\Theta}$ (see 2.C2.2) and R^{ts} corresponds to Δ . By fixing θ in (1.3a, b, c) and minimising these with respect to Δ gives

$$\hat{\delta}_N(\theta) = \mathrm{vec}\left(\hat{\Delta}_N(\theta)\right) = \hat{M}_N^{-1}(\theta).\left[N^{-1}\left(Y_N \otimes I_s\right)'\Gamma_N^{-1}(\theta) z_N\right] \qquad (1.4a)$$

where

$$\hat{M}_N(\theta) = N^{-1}\left(Y_N \otimes I_s\right)'\Gamma_N^{-1}(\theta)\left(Y_N \otimes I_s\right)$$

and Y_N is the $N \times t$ matrix with $y(n)'$ in the nth row;

$$\tilde{\delta}_N(\theta) = \mathrm{vec}\left(\tilde{\Delta}_N(\theta)\right) = \tilde{M}_N^{-1}(\theta).\left[(N')^{-1} \sum_u \left[I_t \otimes f^{-1}(\theta)\right] \mathrm{vec}\, I_{ZY}\right] \qquad (1.4b)$$

where[2]

[2] Again we have used the overbar to mean complex conjugation as well as the notation for $\overline{L}_N$. This will not cause confusion since it is clear when the complex conjugation is intended.

$$\tilde{M}_N(\theta) = (N')^{-1} \sum_u \bar{I}_Y \otimes f^{-1}(\theta) \; ;$$

$$\bar{\delta}_N(\theta) = \text{vec}\left(\bar{\Delta}_N(\theta)\right) = \bar{M}_N^{-1}(\theta)\left[\frac{1}{2\pi} \int \left(I_t \otimes f^{-1}(\theta)\right) \text{vec } I_{ZY} d\omega\right] \qquad (1.4c)$$

where

$$\bar{M}_N(\theta) = \frac{1}{2\pi} \int \bar{I}_Y \otimes f^{-1}(\theta) d\omega \; .$$

Back substitution of (1.4a, b, c) in (1.3a, b, c) yields the "likelihoods" concentrated with respect to $\delta = \text{Vec}(\Delta)$ as

$$\hat{L}_N(\theta) = N^{-1} \log \det \Gamma_N(\theta) + \hat{Q}_N(\theta) - \hat{R}_N(\theta) \qquad (1.5a)$$

where

$$\hat{R}_N(\theta) = \hat{m}_N(\theta)^* \hat{M}_N^{-1}(\theta) \hat{m}_N(\theta)$$

and

$$\hat{m}_N(\theta) = \left[N^{-1}\left(Y_N \otimes I_s\right)' . \Gamma_N^{-1}(\theta) . x_N\right] \; ,$$

$$\tilde{L}_N(\theta) = \log \det K(\theta) + \tilde{Q}_N(\theta) - \tilde{R}_N(\theta) \qquad (1.5b)$$

where

$$\tilde{R}_N(\theta) = \tilde{m}_N(\theta)^* \tilde{M}_N^{-1}(\theta) \tilde{m}_N(\theta)$$

and

$$\tilde{m}_N(\theta) = (N')^{-1} \sum_u \left(I_t \otimes f^{-1}(\theta)\right) \text{vec } I_{XY} \; ,$$

$$\bar{L}_N(\theta) = \log \det K(\theta) + \bar{Q}_N(\theta) - \bar{R}_N(\theta) \qquad (1.5c)$$

where

$$\bar{R}_N(\theta) = \bar{m}_N(\theta)^* \bar{M}_N^{-1}(\theta) \bar{m}_N(\theta)$$

and

$$\bar{m}_N(\theta) = (2\pi)^{-1} \int \left(I_t \otimes f^{-1}(\theta)\right) \text{vec } I_{XY} d\omega \; .$$

Note that $\hat{Q}_N$, $\tilde{Q}_N$, $\bar{Q}_N$ are precisely as defined in (2.1.3), (2.1.9), (2.1.8) respectively. We note that, in terms of the above definition,

$$\hat{\delta}_N(\theta) - \delta_0 = \hat{M}_N^{-1}(\theta)\hat{m}_N(\theta) , \tag{1.6a}$$

$$\tilde{\delta}_N(\theta) - \delta_0 = \tilde{M}_N^{-1}(\theta)\tilde{m}_N(\theta) , \tag{1.6b}$$

$$\overline{\delta}_N(\theta) - \delta_0 = \overline{M}_N^{-1}(\theta)\overline{m}_N(\theta) . \tag{1.6c}$$

The estimates obtained by minimising (1.5a, b, c) over Θ^* will again be referred to as $\hat{\theta}_N$, $\tilde{\theta}_N$, $\overline{\theta}_N$ and when these are substituted in (1.4a, b, c). The results will be called $\hat{\delta}_N = \hat{\delta}_N(\hat{\theta}_N)$, $\tilde{\delta}_N = \tilde{\delta}_N(\tilde{\theta}_N)$, $\overline{\delta}_N = \overline{\delta}_N(\overline{\theta}_N)$. The convergence of the estimators $(\hat{\theta}_N, \hat{\delta}_N)$, $(\overline{\theta}_N, \overline{\delta}_N)$ and $(\tilde{\theta}_N, \tilde{\delta}_N)$ are discussed in the next section.

To close this section we would like to introduce a useful convention of reference to the above quantities by δ_N, θ_N, M_N, m_N, S_N, L_N, Q_N, etc., where δ_N , for example, represents any of $\tilde{\delta}_N$, $\hat{\delta}_N$, $\overline{\delta}_N$ as the case may be.

2. The strong law of large numbers for general stationary residuals

Before discussing the consistency of $(\hat{\theta}_N, \hat{\delta}_N)$, $(\overline{\theta}_N, \overline{\delta}_N)$, $(\tilde{\theta}_N, \tilde{\delta}_N)$ some lemmas concerning the a.s. convergence to zero of $\hat{R}_N$, $\tilde{R}_N$, $\overline{R}_N$ will be given. These will be needed for the proof of the main result. In order to state these lemmas in a convenient form we will let $\hat{R}_N(\Phi)$, $\tilde{R}_N(\Phi)$, $\overline{R}_N(\Phi)$ denote the quantities given by (1.5) when f^{-1} is replaced by Φ , an Hermitian non-negative definite matrix function of ω . If R_N denotes any of $\hat{R}_N$, $\overline{R}_N$, $\tilde{R}_N$ the proof that R_N converges a.s. to zero will be accomplished by writing

$$R_N = \left(N^{\frac{1}{2}}D_N^{-1}m_N\right)^*\left(ND_N^{-1}M_ND_N^{-1}\right)^{-1}\left(N^{\frac{1}{2}}D_N^{-1}m_N\right)$$

and considering the limits of $\left(ND_N^{-1}M_ND_N^{-1}\right)$ and $\left(N^{\frac{1}{2}}D_N^{-1}m_N\right)$ where

$$D_N = d(N) \otimes I_s$$

and

$$d(N) = \mathrm{diag}\big(d_1(N), \ldots, d_t(N)\big) .$$

We will, for the strong law of large numbers of this chapter, require $x(n)$ (in addition to being zero mean, stationary and ergodic with finite variance) to be purely nondeterministic in the strict sense of nonlinear prediction as compared with our previous requirement that $x(n)$ be *linearly* purely nondeterministic. (This latter is the sense in which we have used purely nondeterministic in previous chapters - see §1.2). To clarify this consider the following. Let F_n be the σ-algebra generated by the components of $x(m)$, $m \le n$, and let $F_{-\infty} = \bigcap_{n=-\infty}^{\infty} F_n$. Let H_n be the Hilbert space of all real functions, measurable with respect to F_n , of finite mean square and let S_n be the orthogonal complement of H_{n-1} in H_n . If $\zeta(n, u)$ is the projection of $x(n)$ onto S_u , $u \le n$, and $\zeta(n, -\infty)$ the projection onto $H_{-\infty}$ then

$$x(n) = \sum_{u=0}^{\infty} \zeta(n, n-u) + \zeta(n, -\infty)$$

where $\sum_{u=0}^{\infty} E\big(\zeta(n, n-u)'\zeta(n, n-u)\big) < \infty$. When $\zeta(n, -\infty)$ is a.s. null $x(n)$ is said to be *strictly* purely nondeterministic (and it is in this sense we will use purely nondeterministic in this chapter) and when this is so $x(n)$ is also linearly purely nondeterministic (but not conversely). The assumption that $x(n)$ is strictly purely nondeterministic allows us to use a neat result (see Lemma 2.3 below) when establishing the SLLN for the various estimators. This lemma requires only that $y(n)$ satisfy the mild "rate of growth" condition C1.3. In relation to the CLT of §5 below it will be assumed that the best linear predictor for $x(n)$ is also the best predictor. Then, when $x(n)$ is linearly purely nondeterministic, $C(j)\varepsilon(n-j) = \zeta(n, n-j)$ where the $C(j)$, $\varepsilon(j)$ are as in the representation (2.1.1) for $x(n)$. Also, as has been discussed in §2.3 the $\varepsilon(n)$ are martingale differences with respect to F_n in this case so that in the proof of Lemma 2.3 below the representation $x(n) = \sum_{j=0}^{\infty} C(j)\varepsilon(n-j)$ could be used. For the SLLN the assumption about the best linear predictor is not required.

LEMMA 2.1. *Let* $x(n)$ *and* $y(n)$ *be as in* §1 *with* $y(n)$ *satisfying* Cl.1. *Then*

(a) $ND_N^{-1}M_N(\theta)D_N^{-1} \xrightarrow{\text{a.s.}} \frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes f^{-1}(\omega;\ \theta)$, *for all* $\theta \in \Theta_0$, *where* Θ_0 *is defined in condition* (2.C2.5); *and*

(b) $ND_N^{-1}M_N(\Phi(\theta))D_N^{-1} \xrightarrow{\text{a.s.}} \frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes \Phi(\omega;\ \theta)$ *uniformly in* $\theta \in \Theta_c$ *where* Θ_c *is a compact subset of* $\overline{\Theta}$ *and* Φ *is continuous in* $(\omega;\ \theta) \in [-\pi,\ \pi] \times \Theta_c$.

The conditions under which the limit matrices in (a) and (b) are strictly positive definite will be discussed later.

Proof of (a) for $M_N = \overline{M}_N$. Let P_M be the Cesaro sum of the Fourier series for f^{-1} where M may be chosen to make the error $\sup_{\omega}\left\|P_M-f^{-1}\right\| \le \varepsilon$, $\varepsilon > 0$ arbitrary (Note that for any $\theta \in \Theta_0$, $f^{-1}(\omega;\ \theta)$ is continuous in $\omega \in [-\pi\ \ \pi]$.) Then

$$\left\|ND_N^{-1}\left(M_N(\theta)-M_N\left(P_M^{-1}\right)\right)D_N^{-1}\right\|$$

$$= \left\|ND_N^{-1}\left[\frac{1}{2\pi}\int \overline{I}_Y \otimes \left(f^{-1}-P_M\right)d\omega\right]D_N^{-1}\right\|$$

$$= \left\|N\,\frac{1}{2\pi}\int \left(d(N)^{-1}\overline{I}_Y d(N)^{-1}\right)\otimes\left(f^{-1}-P_M\right)d\omega\right\| \quad \left(\text{since } (A \otimes B)(C \otimes D) = AC \otimes BD\right)$$

$$\le s.\left\|\frac{1}{2\pi}\int Nd(N)^{-1}\overline{I}_Y d(N)^{-1}d\omega\right\| \ .\ \sup_{\omega}\left\|f^{-1}-P_M\right\|$$

$$= \left(\frac{s}{2\pi}\right)\|G^{(Y)}(0)\| \ .\ \sup_{\omega}\left\|f^{-1}-P_M\right\|$$

$$\le \varepsilon.b$$

where $b < \infty$ for N sufficiently large since $\|G^{(Y)}(0)\| \to \|\Gamma^{(Y)}(0)\| < \infty$ by Cl.1(c). Now we may consider, where $P_M(\omega) = \sum_{-M}^{M} p(l)e^{il\omega}$,

$$N \cdot \frac{1}{2\pi}\int \left[d(N)^{-1}\bar{I}_Y d(N)^{-1}\right] \otimes P_M d\omega = \frac{1}{2\pi}\sum_{l=0}^{M}\left|d(N)^{-1}\sum_{1}^{N-l} y(m+l)y(m)'d(N)^{-1}\right|$$

$$\otimes\, p(l) + \frac{1}{2\pi}\sum_{l=-M}^{-1}\left|d(N)^{-1}\sum_{1-l}^{N} y(m)y(m+l)'\right| \otimes p(l) \; .$$

Now C1.1(b) may be used to eliminate end effects in each of the terms in square brackets. Hence this last line may be replaced by

$$\frac{1}{2\pi}\sum_{l=0}^{M} G^{(Y)}(l)' \otimes p(l) + \frac{1}{2\pi}\sum_{l=-M}^{-1} G^{(Y)}(l) \otimes p(l)$$

which converges, by C1.1(c) to

$$\frac{1}{2\pi}\sum_{l=0}^{M} \Gamma^{(Y)}(l)' \otimes p(l) + \frac{1}{2\pi}\sum_{l=-M}^{-1} \Gamma^{(Y)}(l) \otimes p(l) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \bar{F}_Y(d\omega) \otimes P_M(\omega) \; .$$

This last line may be made arbitrarily close to the limit in statement (a) of the lemma by taking M sufficiently large.

Proof of (a) for $M_N = \tilde{M}_N$. The proof that f^{-1} may be replaced by the Cesaro sum P_M is exactly the same as the proof for $\bar{M}_N$. Now

$$N.(N')^{-1}\sum_{u} d(N)^{-1}\bar{I}_Y(\omega_u)d(N)^{-1} \otimes P_M(\omega_u)$$

$$= N \cdot \frac{1}{2\pi}\int d(N)^{-1}\bar{I}_Y(\omega)d(N)^{-1} \otimes P_M(\omega)d\omega + \frac{1}{2\pi}\sum_{-M}^{M} E(l) \otimes p(l)$$

where

$$E(l) = \begin{cases} 0 & \text{if } N'-N \geq M \, , \\ d(N)^{-1}\displaystyle\sum_{n=1}^{-N'+N+l} y(n)y(N'+n-l)'d(N)^{-1} & \text{if } N'-N \leq l \leq M \, , \\ d(N)^{-1}\displaystyle\sum_{n=N'+1+l}^{N} y(n)y(n-l-N')d(N)^{-1} & \text{if } -M \leq l \leq N-N' \, . \end{cases}$$

Now, for example, for $N'-N \leq l \leq M$ there are at most M terms of the form $d(N)^{-1}y(j)y(N'-l+j)'d(N)^{-1}$ in each $E(l)$, where $j = 1, \ldots, -N'+N+l$. But $y(j)/d(N)$ is bounded and $y(N-l)/d(N)$ converges to zero (by C1.1(b)) $0 \leq l \leq M$, M finite. Hence the term $\left\|\frac{1}{2\pi}\sum_{-M}^{M} E(l) \otimes p(l)\right\|$ converges to zero and the proof for the present case proceeds as for $\overline{M}_N(\theta)$.

Proof of (a) for $M_N = \hat{M}_N$. Let P_M be the Cesaro sum to M terms for f , where $\sup_{\omega} \|f-P_M\| \le \varepsilon$ and let $\Gamma_N\left(P_M^{-1}\right)$ be $\Gamma_N(\theta)$ with f^{-1} replaced by P_M^{-1}. then

$$\left\|ND_N^{-1}\hat{M}_N(\theta)D_N^{-1}-ND_N^{-1}\hat{M}_N\left(P_M^{-1}\right)D_N^{-1}\right\|$$

$$= \left\|D_N^{-1}\left(Y_N \otimes I_s\right)'\left(\Gamma_N^{-1}(\theta)-\Gamma_N^{-1}\left(P_M^{-1}\right)\right)\left(Y_N \otimes I_s\right)D_N^{-1}\right\|$$

$$= \left\|\left(Y_N d(N)^{-1} \otimes I_s\right)'\Gamma_N^{-1}(\theta)\left(\Gamma_N\left(P_M^{-1}\right)-\Gamma_N(\theta)\right)\Gamma_N^{-1}\left(P_M^{-1}\right)\left(Y_N d(N)^{-1} \otimes I_s\right)\right\|$$

$$\le \varepsilon . b_1 . b_2 . \left\|\left(Y_N d(N)^{-1} \otimes I_s\right)'\left(Y_N d(N)^{-1} \otimes I_s\right)\right\|$$

where ε is as above and $\left\|\Gamma_N^{-1}(\theta)\right\| \le b_1$, $\left\|\Gamma_N^{-1}\left(P_M^{-1}\right)\right\| \le b_2$ and $b_1 < \infty$, $b_2 < \infty$ since $\|f^{-1}(\theta)\|$, $\left\|P_M^{-1}\right\|$ are bounded for M sufficiently large.

Now

$$\left\|\left(Y_N d(N)^{-1} \otimes I_s\right)'\left(Y_N d(N)^{-1} \otimes I_s\right)\right\|$$

$$= \left\|d(N)^{-1}Y_N'Y_N d(N)^{-1} \otimes I_s\right\|$$

$$\le s . \|G^{(Y)}(0)\| \xrightarrow{\text{a.s.}} \varepsilon . \|\Gamma^{(Y)}(0)\| < \infty \text{ by Cl.1.}$$

Thus $f^{-1}(\theta)$ may be replaced by P_M^{-1} in $ND_N^{-1}\hat{M}_N(\theta)D_N^{-1}$. Using the relationship for $\Gamma_N^{-1}\left(P_M^{-1}\right)$ given in the proof of Lemma 2.2.5 and the remainder of that argument together with the conditions Cl.1(a), (b), (c) the proof for $M_N = \hat{M}_N$ may be established by reducing consideration to the case when $M_N = \tilde{M}_N$ (see the proof of Lemma 2.2.5).

Proofs of (b). The proofs for (b) are the same as the above except now $\Phi(\omega;\, \theta)$ may be approximated by P_M uniformly in $\theta \in \Theta_c$ since Φ is uniformly continuous in θ on Θ_c^* . □

LEMMA 2.2. *Let* $x(n)$ *and* $y(n)$ *be as in* §1 *with* $y(n)$ *satisfying* C1.3 *and let* m_N *denote either of* $\hat{m}_N$, $\bar{m}_N$, $\tilde{m}_N$. *Then*

(a) $N^{\frac{1}{2}}D_N^{-1}m_N(\theta) \xrightarrow{\text{a.s.}} 0$ *for all* $\theta \in \Theta_0$ *where* Θ_0 *is defined in conditions* (2.C2.5); *and*

(b) $N^{\frac{1}{2}}D_N^{-1}m_N(\Phi) \xrightarrow{\text{a.s.}} 0$ *uniformly in* $\theta \in \Theta_c$ *where* Θ_c *and* Φ *are as in Lemma* 2.1.

Proof of (a) for $m_N = \bar{m}_N$. Let P_M be the Cesaro sum to M terms for $f^{-1} + \varepsilon I_s$ chosen so that $P_M - f^{-1} \geq 0$ and $\sup_\omega \left\|f^{-1}+\varepsilon I_s - P_M\right\| \leq \varepsilon$.

Then P_M is an "upper" approximation to f^{-1} . Hence

$$\left\|N^{\frac{1}{2}}D_N^{-1}\left(m_N(P^{-1})-m_N(\theta)\right)\right\| = \left\|N^{\frac{1}{2}} \cdot \frac{1}{2\pi}\int \text{Vec}\left\{\left(P_M-f^{-1}\right)w_X\left(d(N)^{-1}w_Y\right)^*\right\}d\omega\right\|$$

$$= \left\|N^{\frac{1}{2}} \cdot \frac{1}{2\pi}\int \left(P_M-f^{-1}\right)w_X\left(d(N)^{-1}w_Y\right)^* d\omega\right\|$$

$$\leq N^{\frac{1}{2}} \cdot \frac{1}{2\pi}\int \left\|P_M-f^{-1}\right\|\left(w_X^*w_X\right)^{\frac{1}{2}}\left(d(N)^{-1}w_Y\right)^*\left(d(N)^{-1}w_Y\right)d\omega$$

$$\leq \varepsilon.\left[\frac{1}{2\pi}\int w_X^*w_X d\omega\right]^{\frac{1}{2}}\left[\frac{1}{2\pi}\int \left(d(N)^{-1}w_Y\right)^*\left(d(N)^{-1}w_Y\right)d\omega\right]^{\frac{1}{2}}$$

$\leq \varepsilon.b_1.b_2$ for N sufficiently large,

where $\text{tr}\,\Gamma^{(X)}(0) < b_1 < \infty$, $\text{tr}\,\Gamma^{(Y)}(0) < b_2 < \infty$. Thus we may consider

$$N^{\frac{1}{2}}D_N^{-1}m_N\left(P_M^{-1}\right) = \text{Vec}\left\{N^{\frac{1}{2}} \cdot \frac{1}{2\pi}\int P_M(\omega).(2\pi N)^{-1}\sum_1^N\sum x(m)\left(d(N)^{-1}y(n)\right)'e^{i(n-m)\omega}d\omega\right.$$

$$= \text{Vec}\left\{\frac{1}{2\pi}\sum_{l=-M}^{M} p(l)\left(N^{-\frac{1}{2}}\sum_1^{N-l} x(m)\left(d(N)^{-1}y(n\pm l)\right)'\right\}.$$

Now since each element of each $p(l)$ is finite and since there are only a finite number of terms in this last line the result we seek will follow if

$$N^{-\frac{1}{2}} \cdot d_k(N)^{-1} \cdot \sum_{m=1}^{N-l} x_j(m)y_k(m+l) \xrightarrow{\text{a.s.}} 0 \qquad (2.1)$$

for $1 \leq j \leq s$, $1 \leq k \leq t$, $-M \leq l \leq M$. This will be demonstrated in the next lemma. The proof of (a) for $\hat{m}_N$, $\tilde{m}_N$ and (b) for $\bar{m}_N$, $\hat{m}_N$, $\tilde{m}_N$ may be accomplished by using the *same* approximations P_M as used for the respective cases in the proof of Lemma 2.1. □

Before turning to the next lemma we note that the condition that Φ be continuous on a compact subset of Θ is stronger than required for Lemmas 2.1 and 2.2. All that is required is that there exist a matrix of trigonometric polynomials of degree M , P_M say, such that $\sup_{\theta\in\Theta^*} \sup_{\omega} \|\Phi(\omega;\ \theta)-P_M(\omega)\| \leq \varepsilon$ where $\Theta^* \subseteq \overline{\Theta}$. Also as far as Lemma 2.1 is concerned the sequence $y(n)$ could be taken as depending on N , i.e. $y^{(N)}(n)$, for which Cl.1 holds with

$$\lim_{N\to\infty} \max_{1\leq n\leq N} \frac{y_j^{(N)}(n)^2}{d_j(N)^2} = 0$$

replacing Cl.1(b). This remark also applies to the proof of Lemma 2.2 preceding equation (2.2).

The next lemma needed to complete the proof of the last lemma is of general interest. Since (2.1) is stated in terms of scalar sequences $x_j(m)$ and $y_k(m)$ the following result is sufficient for our purposes.

LEMMA 2.3. *Let* $x(n)$ *be a scalar zero mean, ergodic, purely non-deterministic process and let* $y(n)$ *be a scalar sequence satisfying* Cl.3. *Then*

$$N^{-\frac{1}{2}}d(N)^{-1}\sum_{1}^{N-l} x(m)y(m+l) \xrightarrow{\text{a.s.}} 0 \tag{2.2}$$

for each finite l .

Proof[3]. Let F_n be the space of events determined by the history of $x(m)$ to time n (i.e. generated by $x(m)$, $m \leq n$) and let H_n be the Hilbert space of square integrable functions that are measurable F_n .

[3] The proof of this result is due to Professor E.J. Hannan.

Let $\xi(n, n-j)$ be the projection of $x(n)$ onto the part of H_n orthogonal to H_{n-1} then

$$x(n) = \sum_{j=0}^{\infty} \xi(n, n-j)$$

where $\xi(n, -\infty)$. (the projection on $H_{-\infty}$ this being the Hilbert space corresponding to $F_{-\infty} = \bigcap_{-\infty}^{\infty} F_n$) is zero because $x(n)$ is purely non-deterministic. Also $E(\xi(n, n-j) \mid F_{n-j-1}) = 0$ for each fixed j so that for j fixed $\{\xi(n, n-j), F_{n-j-1}\}$ is a sequence of martingale differences and thus $\{x(n), F_{n-j-1}\}$ is a martingale. Now let

$$x^{(M)}(n) = \sum_{j=0}^{M} \xi(n, n-j) \; , \quad M < \infty \; .$$

Then

$$N^{-\frac{1}{2}} d(N)^{-1} \sum_{n=1}^{N-l} \left(x(n) - x^{(M)}(n)\right) y(n+l)$$

$$\leq \left\{N^{-1} \sum_{n=1}^{N-l} \left(x(n) - x^{(M)}(n)\right)^2\right\}^{\frac{1}{2}} \cdot \left\{d(N)^{-2} \sum_{n=1}^{N-l} y(n+l)^2\right\}^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality. But, by ergodicity,

$$N^{-1} \sum_{1}^{N-l} \left(x(n) - x^{(M)}(n)\right)^2 \xrightarrow{\text{a.s.}} E\left(\sum_{M+1}^{\infty} \xi(0, -j)^2\right) < \varepsilon$$

for M sufficiently large where $\varepsilon > 0$ is arbitrary. Also $d(N)^{-2} \sum_{n=1}^{N-l} y(n+l)^2 \leq d(N)^{-2} \sum_{n=l+1}^{N} y(n)^2 \leq 1$ so that the r.h.s. of (2.2) may be replaced by M (finite) terms of the form

$$N^{-\frac{1}{2}} d(N)^{-1} \sum_{n=1}^{N-l} y(n+l)\xi(n, n-j) = N^{-\frac{1}{2}} d(N)^{-1} . X_N$$

where $X_N = \sum_{1}^{N-l} y(n+l)\xi(n, n-j)$. For simplicity of discussion l will be taken as zero (the case $l \neq 0$ is similar). Then

$$E\left(X_N \mid F_{N-j-1}\right) = \sum_1^N y(n)E\left(\xi(n, n-j) \mid F_{N-j-1}\right)$$

$$= \sum_1^{N-1} y(n)\xi(n, n-j) + y(N).E\left(\xi(N, N-j) \mid F_{N-j-1}\right)$$

$$= X_{N-1} \text{ a.s. by the martingale property above.}$$

Thus $\{X_N, F_{N-j-1}\}$ is a martingale. Now Neveu (1965, p. 150) may be used to show that for $0 < \varepsilon = (2\beta)^{-1} - 1$

$$d(N)^{-1}\left(\log d^2(N)\right)^{-\frac{1}{2}-\varepsilon}X_N \xrightarrow{\text{a.s.}} 0 \ .$$

But (2.2) may be re-written (for $l = 0$) as

$$\{N^{-\frac{1}{2}}\left(\log d^2(N)\right)^{\frac{1}{2}+\varepsilon}\}.\left\{d(N)^{-1}\left(\log d^2(N)\right)^{-1-\varepsilon}.X_N\right\}$$

in which the second factor converges to zero by what has just been said and the first factor converges to zero by Cl.3. □

Note that this lemma does not require assumptions on the $y(n)$ other than Cl.3. (That is, Cl.1 is not required for example.) The condition of ergodicity in Lemma 2.3 is difficult to avoid since if, for example, $y(n) = 1$ for all n (i.e. (2.2) is just the mean of $x(1), \ldots, x(N)$ when $l = 0$) then (2.2) may not converge a.s. to zero unless these is no jump in the spectrum of $x(n)$ at the origin (see Corollary 1, p. 205, Hannan (1970)). The condition of pure nondeterminism is also hard to avoid since if $x(n) = \cos(n\omega+\Phi)$, $\omega/2\pi$ irrational and Φ is a random variable uniformly distributed in $[-\pi, \pi]$ then for $y(n) = \cos n\omega$, (2.2) becomes

$$N^{-\frac{1}{2}}\left[\sum_1^N \cos^2(n\omega)\right]^{-\frac{1}{2}} \sum_1^N \cos(n\omega+\Phi)\cos n\omega$$

$$= \tfrac{1}{2}\left[N^{-1}\sum_1^N \cos^2(n\omega)\right]^{-\frac{1}{2}}\left[N^{-1}\sum_1^N \cos^2(n\omega)\ .\ \cos\Phi - N^{-1}\sum_1^N \cos n\omega \sin n\omega . \sin\Phi\right]$$

$$\xrightarrow{\text{a.s.}} \begin{cases} \tfrac{1}{4}\cos\Phi & \text{if } \omega \neq 0, \pi \\ \tfrac{1}{2}\cos\Phi & \text{if } \omega = 0, \pi , \end{cases}$$

and therefore (2.2) does not converge a.s. to zero.

The next result is concerned with the strong consistency of $\bar{\theta}_N$, $N^{-\frac{1}{2}}D_N(\bar{\delta}_N-\delta_0)$ and $\tilde{\theta}_N$, $N^{-\frac{1}{2}}D_N(\tilde{\delta}_N-\delta_0)$. The result for $\hat{\theta}_N$, $N^{-\frac{1}{2}}D_N(\hat{\delta}_N-\delta_0)$ will be given later. In order to establish the following theorem for the regression model (1.1) when $x(n)$ is specified by the general finite parameter model discussed in Chapter 2 further conditions to those required in Theorem 2.2.1 (i.e. Conditions 2.C2) have been imposed in the parameter space associated with $x(m)$. The notation of Chapter 2 will be used below (for definitions of $h(e^{i\omega};\ \theta)$, $g(e^{i\omega};\ \theta)$, $l(e^{i\omega};\ \theta)$, $d(e^{i\omega};\ \theta)$ see §2.2).

THEOREM 2.4. *For the regression model* (1.1) *let* $y(n)$ *satisfy* C1.1, C1.2 *and* C1.3. *Let* $x(n)$ *satisfy* 2.C2 *and in addition assume that* θ_0 *belongs to the subset of* $\bar{\Theta}$ *defined by*

$$\Theta^* = \left\{\theta : \lambda_s\big(K(\theta)\big)/\lambda_1\big(K(\theta)\big) \le b\ , \quad \frac{1}{2\pi}\int \bar{F}_Y(d\omega) \otimes h^*(e^{i\omega};\ \theta)h(e^{i\omega};\ \theta) \ge cI_{st}\right\} \qquad (2.3)$$

where $b < \infty$ *and* $c > 0$. *Furthermore let* $(\bar{\theta}_N, \bar{\delta}_N)$, $(\tilde{\theta}_N, \tilde{\delta}_N)$ *respectively minimise* $\bar{L}_N(\theta, \delta)$, $\tilde{L}_N(\theta, \delta)$ *over* (θ, δ) *in* $\Theta^* \times R^{ts}$. *Then* $\bar{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$, $\tilde{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$, $N^{-\frac{1}{2}}D_N(\bar{\delta}_N-\delta_0) \xrightarrow{\text{a.s.}} 0$ *and* $N^{-\frac{1}{2}}D_N(\tilde{\delta}_N-\delta_0) \xrightarrow{\text{a.s.}} 0$.

Note that C1.2 will always hold on Θ^* .

Proof. Only the proof for $\overline{\theta}_N$, $\overline{\delta}_N$ will be given since the other case is proved in a similar way using results of Chapter 2 with Lemmas 2.1 and 2.2. Throughout the following proof, therefore, all quantities defined in §1 will be written without the overbar so that θ_N will denote $\overline{\theta}_N$ for example. It will be convenient to break the proof into parts for later reference.

(i) Since θ_N minimizes $L_N(\theta)$ (see (1.5c)) and since $R_N(\theta) \geq 0$ it follows that

$$\overline{\lim_{N\to\infty}} L_N(\theta_N) \leq \overline{\lim_{N\to\infty}} L_N(\theta) \quad \text{for all} \quad \theta \in \Theta_0 \cap \Theta^*$$

$$\leq \overline{\lim_{N\to\infty}} \left\{\log \det K(\theta) + \frac{1}{2\pi} \int \operatorname{tr}\left[I_X f^{-1}(\omega;\ \theta)\right] d\omega\right\}$$

$$= \log \det K(\theta) + \frac{1}{2\pi} \int \operatorname{tr}\left[f(\omega;\ \theta_0) f^{-1}(\omega;\ \theta)\right] d\omega \tag{2.4}$$

using Lemma 2.2.2. Just as in the proof of Theorem 2.2.1 it then follows that

$$\overline{\lim_{N\to\infty}} L_N(\theta_N) \leq \log \det K(\theta_0) + s\ . \tag{2.5}$$

(ii) Let $K_N = K(\theta_N)$, $h_N(e^{i\omega}) = h\left(e^{i\omega};\ \theta_N\right)$, $g_N(e^{i\omega}) = g\left(e^{i\omega};\ \theta_N\right)$.

Then

$$L_N(\theta_N,\ \delta_N) \geq s \log \lambda_1(K_N) + \lambda_s^{-1}(K_N) . \dot{S}_N(\theta_N,\ \delta_N) \tag{2.6}$$

where

$$\dot{S}_N(\theta_N,\ \delta_N) = \frac{1}{2\pi} \int \left[(\Delta_N-\Delta_0)w_Y+w_X\right]^* h_N^* (g_N g_N^*)^{-1} h_N \left[(\Delta_N-\Delta_0)w_Y+w_X\right] d\omega$$

$$\geq \frac{1}{\gamma}\left(\frac{1}{2\pi} \int \left[(\Delta_N-\Delta_0)w_Y+w_X\right]^* h_N^* h_N \left[(\Delta_N-\Delta_0)w_Y+w_X\right] d\omega\right)$$

$$\left(\text{where } g(e^{i\omega};\ \theta) g(e^{i\omega};\ \theta)^* \leq \gamma I_s \text{ for all } \theta \in \overline{\Theta}\right)$$

$$\geq \frac{1}{\gamma} \cdot \left(\frac{1}{2\pi} \int \left[(\dot{\Delta}_N-\Delta_0)w_Y+w_X\right]^* h_N^* h_N \left[(\dot{\Delta}_N-\Delta_0)w_Y+w_X\right] d\omega\right) \tag{2.7}$$

where $\dot{\Delta}_N$ is the value of Δ minimising the second factor in the previous line. But the second factor in the last line of (2.7) may be written as (see the argument leading to (1.5))

$$\operatorname{tr}\left(\frac{1}{2\pi}\int h_N^* h_N I_X d\omega\right) - R_N\left(h_N^* h_N\right)$$

where

$$R_N\left(h_N^* h_N\right) = m_N\left(h_N^* h_N\right)^* M_N^{-1}\left(h_N^* h_N\right) m_N\left(h_N^* h_N\right)$$

with $m_N\left(h_N^* h_N\right)$ defined as in (1.5c) where f^{-1} is replaced by $h_N^* h_N$ and $M_N\left(h_N^* h_N\right)$ is defined as in (1.4c) similarly. Now by a similar result to Lemma 2.1 it follows that

$$ND_N^{-1} M_N\left(h_N^* h_N\right) D_N^{-1} - \frac{1}{2\pi}\int dF_Y \otimes h_N^* h_N \xrightarrow{\text{a.s.}} 0 \tag{2.8}$$

and $\frac{1}{2\pi}\int dF_Y \otimes h_N^* h_N \geq cI_{ts}$ (by the definition of Θ^* in (2.3)) so that $\left\|M_N^{-1}\left(h_N^* h_N\right)\right\|$ is uniformly bounded. Also by a similar result to Lemma 2.2,

$$N^{\frac{1}{2}} D_N^{-1} m_N\left(h_N^* h_N\right) \xrightarrow{\text{a.s.}} 0 . \tag{2.9}$$

Combining (2.7), (2.8) and (2.9) gives

$$\dot{S}_N\left(\theta_N, \delta_N\right) \geq \operatorname{tr}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} h_N^* h_N I_X d\omega\right) - \varepsilon_N , \quad \varepsilon_N \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad N \to \infty .$$

But Lemma 2.2.4 may be applied to the first term of this expression to get $\dot{S}_N\left(\theta_N, \delta_N\right) \geq \varepsilon$ for N sufficiently large where $\varepsilon > 0$. Using this in (2.6) gives

$$L_N\left(\theta_N, \delta_N\right) \geq s \log \lambda_1\left(K_N\right) + \lambda_1^{-1}\left(K_N\right) . \left(\frac{\lambda_s\left(K_N\right)}{\lambda_1\left(K_N\right)}\right) . \varepsilon$$

$$\geq s \log \lambda_1\left(K_N\right) + \lambda_1^{-1}\left(K_N\right) . b^{-1} . \varepsilon$$

where b is as in the definition of Θ^* (see (2.3)). Hence, using the same argument as presented in the proof of Theorem 2.2.1 it follows that for N sufficiently large $\lambda_1\left(K_N\right)$ is bounded away from zero and $\|K_N\|$ is uniformly bounded. That is, for N sufficiently large

$\theta_N \in \Theta^* \cap \Theta_{\varepsilon_1,\varepsilon_2}$ for some $0 < \varepsilon_1 < \varepsilon_2 < \infty$ where $\Theta_{\varepsilon_1,\varepsilon_2}$ is as in §2.2.

(iii) Now defining

$$\psi_\eta = (2\pi)h^*\left(gKg^*+\eta I_s\right)^{-1}h \ , \quad \eta > 0 \tag{2.10}$$

it follows that ψ_η is uniformly continuous in θ belonging to $\Theta^* \cap \Theta_{\varepsilon_1,\varepsilon_2}$ for each $\eta > 0$. Also $f^{-1} \geq \psi_\eta$ so that, putting $\psi_{\eta,N}(\omega) = \psi_\eta\left(\omega;\ \theta_N\right)$,

$$L_N\left(\theta_N\right) \geq \log\det K_N + \frac{1}{2\pi}\int \operatorname{tr}\left[\psi_{\eta,N}I_X\right]d\omega - R_N\left(\psi_{\eta,N}\right) . \tag{2.11}$$

But

$$R_N\left(\psi_{\eta,N}\right) = \left[N^{\frac{1}{2}}D_N^{-1}m_N\left(\psi_{\eta,N}\right)\right]^*\left[ND_N^{-1}M_N\left(\psi_{\eta,N}\right)D_N^{-1}\right]^{-1}\left[N^{\frac{1}{2}}D_N^{-1}m_N\left(\psi_{\eta,N}\right)\right] . \tag{2.12}$$

Therefore, by Lemma 2.2b, $N^{\frac{1}{2}}D_N^{-1}m_N\left(\psi_{\eta,N}\right) \xrightarrow{\text{a.s.}} 0$ and by Lemma 2.1b,

$$\left[ND_N^{-1}M_N\left(\psi_{\eta,N}\right)D_N^{-1} - \frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes \psi_{\eta,N}\right] \xrightarrow{\text{a.s.}} 0 .$$

But $\left(gKg^*+\eta I_s\right)^{-1} \geq \left(b_1+\eta\right)^{-1}I_s$ where $gKg^* \leq b_1 I_s$ and $b_1 < \infty$ for all $\theta \in \Theta^* \cap \Theta_{\varepsilon_1,\varepsilon_2}$, so that

$$\frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes \psi_{\eta,N} \geq \left(b_1+\eta\right)^{-1} \cdot \frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes h_N^*h_N \geq \frac{c}{\left(b_1+\eta\right)}$$

by the definition of Θ^* above. Hence $\left\|\left(ND_N^{-1}M_N\left(\psi_{\eta,N}\right)D_N^{-1}\right)^{-1}\right\|$ is uniformly bounded and consequently $R_N\left(\psi_{\eta,N}\right)$ converges a.s. to zero. Using this in (2.11) gives

$$\varliminf_{N\to\infty} L_N\left(\theta_N\right) \geq \sup_{\eta>0}\ \varliminf_{N\to\infty}\left\{\log\det K_N + \frac{1}{2\pi}\int \operatorname{tr}\left[\psi_{\eta,N}I_X\right]d\omega\right\} \tag{2.13}$$

and the remainder of the proof of Theorem 2.2.1 may be used in exactly the same way to obtain $\overline{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$ (see equation (2.2.7) and beyond).

(iv) To show that $N^{-\frac{1}{2}}D_N\left(\overline{\delta}_N-\delta_0\right) \xrightarrow{\text{a.s.}} 0$ we consider

$$S_N(\theta_N) = \frac{1}{2\pi}\int \operatorname{tr}\left[\{(\Delta_N-\Delta_0)I_Y(\Delta_N-\Delta_0)'-(\Delta_N-\Delta_0)I_{YX}-I_{XY}(\Delta_N-\Delta_0)'+I_X\}f^{-1}(\omega;\,\overline{\theta}_N)\right]d\omega$$

$$\geq \frac{1}{2\pi}\int \operatorname{tr}[\{(\Delta_N-\Delta_0)I_Y(\Delta_N-\Delta_0)'-(\Delta_N-\Delta_0)I_{YX}-I_{XY}(\Delta_N-\Delta_0)'+I_X\}\psi_\eta(\omega;\,\theta_N)]d\omega$$

$$= \left[\left(N^{-\frac{1}{2}}D_N(\delta_N-\delta_0)\right)' \,\vdots\, (\operatorname{Vec} I_s)'\right] .$$

$$\cdot \left\{\frac{1}{2\pi}\int \begin{bmatrix} ND_N^{-1}\overline{I}_Y D_N^{-1} & \vdots & N^{\frac{1}{2}}D_N^{-1}\overline{I}_{YX} \\ \cdots & & \cdots \\ \left(N^{\frac{1}{2}}D_N^{-1}\overline{I}_{YX}\right)' & \vdots & \overline{I}_{XX}\end{bmatrix} \otimes \psi_\eta(\overline{\theta}_N)\,d\omega\right\}\begin{bmatrix} N^{-\frac{1}{2}}D_N(\delta_N-\delta_0) \\ \cdots \\ (\operatorname{Vec} I_s)\end{bmatrix} . \quad (2.14)$$

But

$$\frac{1}{2\pi}\int \begin{bmatrix} ND_N^{-1}\overline{I}_Y D_N^{-1} & \vdots & N^{\frac{1}{2}}D_N^{-1}\overline{I}_{YX} \\ \cdots & & \cdots \\ \left(N^{\frac{1}{2}}D_N^{-1}\overline{I}_{YX}\right)' & \vdots & \overline{I}_X\end{bmatrix} \otimes \psi_\eta(\overline{\theta}_N)\,d\omega$$

$$\geq \begin{bmatrix} \frac{1}{2\pi}\int \overline{F}_Y(d\omega)\otimes\psi_\eta(\theta_0) & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & \frac{1}{2\pi}\int \overline{f}_0\otimes\psi_\eta(\theta_0)\,d\omega\end{bmatrix} - \varepsilon_N I_{(t+s)s} \quad (2.15)$$

where $\varepsilon_N \to 0$ as $N \to \infty$ since the matrix in the l.h.s. of (2.15) converges to the first matrix on the r.h.s. of (2.15). Hence the last line of (2.14) is not less than

$$\left[N^{-\frac{1}{2}}D_N(\delta_N-\delta_0)\right]'\left[N^{-\frac{1}{2}}D_N(\delta_N-\delta_0)\right] . \varepsilon + \operatorname{tr}\left(\frac{1}{2\pi}\int \psi_\eta(\theta_0)\, f(\theta_0)\,d\omega\right) - s.\varepsilon_N \quad (2.16)$$

where-in $\left(\frac{1}{2\pi}\int \overline{F}_Y(d\omega)\otimes\psi_\eta(\theta_0)\right) - \varepsilon_N \geq \varepsilon > 0$ for N sufficiently large.

Combining (2.14) and (2.16) and taking $\overline{\lim}_{N\to\infty}$ throughout gives, since $S_N(\theta_N) \to 0$ by parts (i), (ii) and (iii) above,

$$s \geq \varepsilon . \overline{\lim_{N\to\infty}}\left\|N^{-\frac{1}{2}}D_N(\delta_N-\delta_0)\right\|^2 + \frac{1}{2\pi}\int \operatorname{tr}(\psi_\eta(\theta_0)f(\theta_0))\,d\omega . \quad (2.17)$$

Taking the supremum over $\eta > 0$ yields s for the second term on the r.h.s. of (2.17). Thus since $\varepsilon > 0$, $\left\|N^{-\frac{1}{2}}D_N(\overline{\delta}_N-\delta_0)\right\|^2 \xrightarrow{a.s.} 0$. □

For the full likelihood procedure $\hat{L}_N$ stronger conditions are required in order that Theorem 2.4 holds.

COROLLARY 2.5. *In addition to the conditions of Theorem 2.4 assume that* θ_0 *belongs to that subset of* Θ *defined as*

$$\Theta^{**} = \{\theta : \lambda_s(K(\theta))/\lambda_1(K(\theta)) \le b,\ |\det h| \ge c > 0\} \tag{2.18}$$

and let $(\hat{\theta}_N, \hat{\delta}_N)$ *minimise* $\hat{L}_N(\theta, \delta)$ *over* $\Theta^{**} \times \mathrm{R}^{ts}$ *. Then*

$\hat{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$, $N^{-\frac{1}{2}}D_N(\hat{\delta}_N - \delta_0) \xrightarrow{\text{a.s.}} 0$.

Two remarks concerning Θ^* and Θ^{**} are relevant.

(i) When $\Gamma^{(Y)}(0) > 0$ and $|\det h|^2 \ge c > 0$ then $\frac{1}{2\pi}\int \bar{F}_Y(d\omega) \otimes h^*h \ge c' > 0$ so that the conditions of Corollary 2.5 imply those defining Θ^* of Theorem 2.4.

(ii) The condition $|\det h|^2 \ge c > 0$ corresponds, in the case of a pure autoregressive process, to bounding the zeros of $\det\left\{\sum_0^q B(j)z^j\right\}$ away from the unit circle. On the other hand the condition of Theorem 2.4 that $\frac{1}{2\pi}\int d\bar{F}_Y \otimes h^*h \ge c > 0$ only bounds the zeros of $\det h$ which occur on the unit circle away from jumps of F_Y . To illustrate consider $h(e^{i\omega}) = (1+\beta e^{i\omega})$ (i.e. $p = 0$, $q = s = 1$ in the ARMA residual). If F_Y has a single jump at the origin (corresponding to, for example, $y(n) \equiv 1$) then in order that $\int d\bar{F}_Y \otimes h^*h \ge c > 0$ the parameter β would need to be taken as lying in the range $-1 + \varepsilon \le \beta \le 1$ where $\varepsilon > 0$. The ARMA case will be discussed more fully in the next section.

Proof of Corollary 2.5. Part (i) of the proof of Theorem 2.4 follows in the same way here by use of Lemmas 2.1, 2.2, 2.2.5 and 2.2.6. The argument corresponding to part (ii) of the previous proof is as follows. Since

$$\Gamma_N(\theta) = \Gamma_N\left(\frac{1}{2\pi} h^{-1} gKg^* h^{-1*}\right) \leq \lambda_s(K) \;.\; c' \;.\; c^{-1} \tag{2.19}$$

where $\left\|\frac{1}{2\pi} Hgg^*H^*\right\| \leq c' < \infty$ and $\det hh^* \geq c > 0$ and H is the adjoint of h, then, by Lemma 2.2.6,

$$L_N(\theta_N, \delta_N) \geq \log\det K(\theta_N) + S_N(\theta_N, \Delta_N)$$

where S_N is defined by (1.3a). But, by (2.19),

$$S_N(\theta_N, \Delta_N) \geq \lambda_s(K) \;.\; (c'/c) \;.\; N^{-1}\left(x_N - \left(I_s \otimes (\Delta_N - \Delta_0)\right)y_N\right)' \;.$$
$$.\; \left(x_N - \left(I_s \otimes (\Delta_N - \Delta_0)\right)y_N\right) , \tag{2.20}$$

and $\left(x_N - \left(I_s \otimes (\Delta - \Delta_0)y_N\right)\right)'\left(x_N - \left(I_s \otimes (\Delta - \Delta_0)\right)y_N\right)$ is minimised by choosing δ as $\dot{\delta}_N$ where

$$\dot{\delta}_N - \delta_0 = \left[N^{-1}(Y_N \otimes I_s)'(Y_N \otimes I_s)\right]^{-1}\left[N^{-1}(Y_N \otimes I_s)'x_N\right] .$$

Back substitution in (2.20) gives

$$S_N(\theta_N, \Delta_N) \geq \lambda_1^{-1}(K_N) \;.\; \left(\frac{\lambda_1(K_N)}{\lambda_s(K_N)}\right) \;.\; \left(\frac{c'}{c}\right) \;.$$
$$.\; \left\{N^{-1}x_N'x_N - N^{-1}\left((Y_N \otimes I_s)'x_N\right)'\left(Y_N'Y_N \otimes I_s\right)^{-1}\left((Y_N \otimes I_s)'x_N\right)\right\} . \tag{2.21}$$

But the term in braces converges a.s. to zero (as in Lemmas 2.1 and 2.2) while $N^{-1}x_N'x_N$ converges a.s. to $\mathrm{tr}\left(\Gamma^{(X)}(0)\right) > \varepsilon > 0$. Hence

$$S_N(\theta_N, \Delta_N) \geq \lambda_1^{-1}(K_N) \;.\; \varepsilon' \tag{2.22}$$

where ε' is positive since $\lambda_1/\lambda_s > 0$ and $c'/c > 0$ so that the factor on the r.h.s. of (2.21) is positive for N large. Hence $L_N(\theta_N, \delta_N) \geq s \log \lambda_1(K_N) + \lambda_1^{-1}(K_N) \;.\; \varepsilon$ and (ii) of the proof of Theorem 2.4 is established here also. Parts (iii) and (iv) of that proof follow easily here using ψ_η as precisely defined and Lemmas 2.1 and 2.2. □

Note that in the proofs of Theorem 2.4 and Corollary 2.5 it is always possible to show that $\log \det K(\theta_N)$ is uniformly bounded. Then, given this, the condition that $\lambda_s(K)/\lambda_1(K) \le b < \infty$ defining Θ^* and Θ^{**} is implied by the condition that $\lambda_1(K) \ge c > 0$, but not conversely.

There are many alternative specifications of the type of conditions required in order that Theorem 2.4 and Corollary 2.5 hold. Individual special models for the residual process, $x(n)$, will facilitate the imposition of different conditions. In particular for ARMA residuals some theory is given in the next section. One alternative specification of some generality is to assume that δ_0 belongs to a sphere in R^{ts} of *known* finite radius. When this assumption is made it is possible to dispense with the assumption that $\theta_0 \in \Theta^*$ in Theorem 2.4 for the case when $d_j(N)^2$ (see C1.1) is replaceable by N. The following result contains the details and the proof is omitted since it is similar to that of Theorem 2.4.

THEOREM 2.6. *Let* $x(n)$, $y(n)$ *satisfy the regression model* (1.1) *which* $x(n)$ *satisfies conditions* 2.C2 *and* $y(n)$ *satisfies* C1 *and* C2. *Also assume that* $\delta_0 \in \mathcal{D}$ *which is a closed and bounded subset of* R^{ts}. *If* $(\bar{\theta}_N, \bar{\delta}_N)$, $(\tilde{\theta}_N, \tilde{\delta}_N)$ *respectively minimise* $\bar{L}_N(\theta, \delta)$, $\tilde{L}_N(\theta, \delta)$ *over* $\Theta \times \mathcal{D}$ *then* $\bar{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$, $\tilde{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$, $\bar{\delta}_N \xrightarrow{\text{a.s.}} \delta_0$ *and* $\tilde{\delta}_N \xrightarrow{\text{a.s.}} \delta_0$.

Note that when $\bar{\delta}_N$ is restricted to $\mathcal{D}$ the concentration of $\bar{L}_N(\theta, \delta)$ with respect to δ for θ fixed is not necessarily the correct way of minimising $\bar{L}_N(\theta, \delta)$ over $(\theta, \delta) \in \Theta \times \mathcal{D}$. Thus some changes are required in the previous proof of Theorem 2.4. The same remark applies, of course, to $\tilde{L}_N$. Theorem 2.6 is not applicable to $\hat{L}_N$ in its present form since the additional conditions defining Θ^{**} in Corollary 2.5 are also needed for $\hat{\theta}_N \to \theta_0$ and not just $\hat{\delta}_N \to \delta_0$.

To close this section interpretation of the results in Theorem 2.4 and Corollary 2.5 will be made. When the $d_j(N)^2$ grow at a sufficiently fast rate the following holds.

COROLLARY 2.7. *Let* $y(n)$ *in Theorem* 2.4 *or Corollary* 2.5 *further satisfy, for* $j = 1, 2, \ldots, t$,

$$\lim_{N\to\infty} N^{-\alpha_j} d_j(N)^2 > 0 \tag{2.23}$$

where $\alpha_j \geq 1$. *Then* $\bar{\delta}_N \xrightarrow{\text{a.s.}} \delta_0$, $\tilde{\delta}_N \xrightarrow{\text{a.s.}} \delta_0$ *and* $\hat{\delta}_N \xrightarrow{\text{a.s.}} \delta_0$.

Proof. Since, for example, $\left\| N^{-\frac{1}{2}} D_N\left(\hat{\delta}_N - \delta_0\right)\right\| \xrightarrow{\text{a.s.}} 0$ and $\lim_{N\to\infty} N^{-\alpha_j} d_j(N)^2 > 0$, $j = 1, \ldots, t$, then $\lim_{N\to\infty} N^{-1} d_j(N)^2 > 0$ so that $\overline{\lim}_{N\to\infty} N d_j(N)^{-2} \leq b^2 < \infty$ for each j . But then

$$\begin{aligned}\|\hat{\delta}_N - \delta_0\| &= \left\| \left[N^{\frac{1}{2}} D_N^{-1}\right]\left[N^{-\frac{1}{2}} D_N\left(\hat{\delta}_N - \delta_0\right)\right]\right\| \\ &\leq b \,.\, N^{-\frac{1}{2}} D_N\left(\hat{\delta}_N - \delta_0\right) \xrightarrow{\text{a.s.}} 0 \,. \qquad \square\end{aligned}$$

For $\alpha_j < 1$ in (2.23) it is not possible to conclude directly that $\tilde{\delta}_N \to \delta_0$, for example. In relation to the CLT of §5 below (i.e. the asymptotic normality of $D_N\left(\tilde{\delta}_N - \delta_0\right)$, for example) the convergence $N^{-\frac{1}{2}} D_N\left(\tilde{\delta}_N - \delta_0\right) \xrightarrow{\text{a.s.}} 0$ is sufficient. The case $\alpha_j < 1$ will be further discussed in §4.

3. The strong law of large numbers for ARMA residuals

In this section the strong convergence results of §2 will be discussed for the case when the residual vector, $x(n)$, in (1.1) is generated by an ARMA process (see Chapter 3). Problems of identification (as discussed in

§3.2) and parameterisation (see §3.3) are not essentially different from the regression model (1.1) than for the ARMA model. The purpose of this section is to give a brief extension of the results of §2 to the case when $x(n)$ is ARMA along similar lines to the extension of the results of Chapter 2 given in Chapter 3.

To keep the following discussion to a reasonable length we will only consider in detail the case when it is assumed that θ_0 (defined in (3.1.2)) belongs to Θ_1 the space of simply identified ARMA (p, q) structures (see Chapter 3). We will therefore take $(\theta_0, \delta_0) \in \Theta_1 \times R^{ts}$ and convergence of sequences (θ_n, δ_n) to (θ_0, δ_0) in the Euclidean topology. The following result is not completely satisfactory in that it excludes certain regression sequences $y(n)$. A simple example is given after the proof of the theorem which illustrates this with a proof that the condition in C3.1 to be introduced below can be avoided in that example.

THEOREM 3.1. *In the regression model* (1.1) *let* $x(n)$ *be generated by the* ARMA *model* (3.1.1) *and let* Θ_1, Θ_{01} *be as described in Chapter* 3. *Let* $y(n)$ *be independent of* $x(n)$ *and satisfy* C1.1 *and* C1.3. *In addition assume that for* $V = p(s-1) + q$ *:*

C3.1 $\quad \Gamma_{V+1}^{(Y)} = \left[\Gamma^{(Y)}(m-n)\right]_{m,n=0,\ldots,V}$ *is nonsingular.*

Then $\bar{\theta}_N, \tilde{\theta}_N, \hat{\theta}_N$ *converges a.s. to* θ_0 *in the Euclidean topology and* $N^{-\frac{1}{2}}D_N(\bar{\delta}_N-\delta_0)$, $N^{-\frac{1}{2}}D_N(\tilde{\delta}_N-\delta_0)$, $N^{-\frac{1}{2}}D_N(\hat{\delta}_N-\delta_0)$ *converge a.s. to the null vector. Furthermore, when* $y(n)$ *additionally satisfies* (2.23) *for all* $\alpha_j \geq 1$ *then* $\bar{\delta}_N, \tilde{\delta}_N, \hat{\delta}_N$ *converge a.s. to* δ_0.

Proof. Only the proof for $\hat{\theta}_N, \hat{\delta}_N$ will be given since it is slightly more difficult than for the other cases. Constant reference will be made throughout this proof to the proofs of Theorems 3.4.1 and 2.4. Part (i) of the proof of Theorem 2.4 applies here without change (see also the proof of Theorem 3.4.1). For part (ii) of the proof of Theorem 2.4 we proceed as follows to show, just as it was shown in the proof of Theorem 3.4.1, that eventually $\hat{\theta}_N$ may be obtained by minimising over that subset of Θ_1 for which $\lambda_1(K_\theta) \geq b_1 > 0$, $\lambda_s(K(\theta)) \leq b_2 < \infty$ and for which $P_\theta(z) = [\text{adjoint } g(z;\ \theta)].h(z;\ \theta)$ has coefficient matrices with bounded norms (this set was called $\Theta_{K,P}$ in the proof of Theorem 3.4.1 - the notation of that proof will also be used in this proof). Now for $(\theta,\ \delta) \in \Theta_1 \times R^{ts}$,

$$\hat{L}_N(\theta,\ \delta) = N^{-1} \log \det \Gamma_N(\theta) + N^{-1}\left(z_N - \left(I_s \otimes \Delta\right)y_N\right)'\Gamma_N^{-1}(\theta)\left(z_N - \left(I_s \otimes \Delta\right)y_N\right)$$

$$\geq \log \det K_\theta + N^{-1}\left(z_N - \left(I_s \otimes \Delta\right)y_N\right)'\Gamma_N^{-1}(\theta)\left(z_N - \left(I_s \otimes \Delta\right)y_N\right)$$

(by Lemma 2.2.6)

$$\geq \log \det K_\theta + \gamma^{-2} \,.\, \mathrm{tr}\left[K_\theta^{-1} \sum_{j,k=0}^{V}\sum p_\theta(j)\left\{N^{-1} \sum_{n=V+1}^{N} \left(x(n-j)\right.\right.\right.$$

$$\left.\left.\left. -\left(\Delta-\Delta_0\right)y(n-j)\right)\left(x(n-k)-\left(\Delta-\Delta_0\right)y(n-k)\right)'\right\}p_\theta(k)\right]$$

(by the proof for $\hat{L}_N$ given in Theorem 3.4.1). (3.1)

But

$$\sum_0^V\sum_0^V p_\theta(j)\left\{N^{-1} \sum_{n=V+1}^{N} \left(x(n-j)-\left(\Delta-\Delta_0\right)y(n-j)\right)\left(x(n-k)-\left(\Delta-\Delta_0\right)y(n-k)\right)'\right\}p_\theta(k)'$$

$$= P_N(\theta)G_N P_N(\theta) \quad (3.2)$$

where

$$P_N(\theta) = \left[p_\theta(0) \vdots \cdots \vdots p_\theta(V) \vdots -p_\theta(0)\left(\Delta-\Delta_0\right)d(N).N^{-\frac{1}{2}} \vdots \cdots\right.$$

$$\left. \cdots \vdots -p_\theta(V)\left(\Delta-\Delta_0\right)d(N).N^{-\frac{1}{2}}\right]$$

and

$G_N =$

$$= \left[\begin{array}{c:c} \left[\frac{1}{N}\sum_{V+1}^{N} x(n-j)x(n-k)'\right]_{\substack{j,k=\\0,\dots,V}} & \left[N^{-\frac{1}{2}}\sum_{V+1}^{N} x(n-j)y(n-k)'d(N)^{-1}\right]_{\substack{j,k=\\0,\dots,V}} \\ \hdashline \left[N^{-\frac{1}{2}}\sum_{V+1}^{N} d(N)^{-1}y(n-j)x(n-k)'\right]_{\substack{j,k=\\0,\dots,V}} & \left[\sum_{V+1}^{N} d(N)^{-1}y(n-j)y(n-k)'d(N)^{-1}\right]_{\substack{j,k=\\0,\dots,V}} \end{array}\right] .$$

But

$$N^{-1}\sum_{V+1}^{N} x(n-j)x(n-k)' \xrightarrow{\text{a.s.}} \Gamma_0(j-k) ,$$

$$N^{-\frac{1}{2}}\sum_{V+1}^{N} x(n-j)y(n-k)'d(N)^{-1} \xrightarrow{\text{a.s.}} 0$$

by Lemma 2.3 and

$$\sum_{V+1}^{N} d(N)^{-1}y(n-j)y(n-k)'d(N)^{-1} \xrightarrow{\text{a.s.}} \Gamma^{(Y)}(j-k)$$

as in the proof of Lemma 2.1. Hence

$$G_N \xrightarrow{\text{a.s.}} \begin{bmatrix} \Gamma_{V+1}(\theta_0) & 0 \\ 0 & \Gamma_{V+1}^{(Y)} \end{bmatrix} \geq 2\varepsilon I_{(V+1)(t+s)} , \quad \varepsilon > 0$$

since the smallest eigenvalue of $\Gamma_{V+1}(\theta_0)$ is positive (see Lemma 3.4.3) and the smallest eigenvalue of $\Gamma_{V+1}^{(Y)}$ is positive by assumption C3.1. Hence if the N is sufficiently large so that

$$\left\| G_N - \begin{pmatrix} \Gamma_{V+1}(\theta_0) & 0 \\ 0 & \Gamma_{V+1}^{(Y)} \end{pmatrix} \right\| \leq \varepsilon$$

the smallest eigenvalue of G_N is then not less than $\varepsilon > 0$. Hence, using (3.2) in (3.1) and the facts concerning the eigenvalue of G_N just given we have

$$\hat{L}_N(\theta, \delta) \geq s \log \lambda_1(K_\theta) + \lambda_1^{-1}(K_\theta) \cdot \gamma^{-2} \cdot \varepsilon \circ \lambda_1(P_N(\theta)P_N(\theta)') .$$

But

$$\lambda_1(P_N(\theta)P_N(\theta)') = \lambda_1\left[\sum_0^V p_\theta(j)p_\theta(j)' + \sum_0^V p_\theta(j)(\Delta-\Delta_0)(N^{-1}d(N)^{-2})(\Delta-\Delta_0)p_\theta(j)'\right]$$

$$\geq \lambda_1\left[\sum_0^V p_\theta(j)p_\theta(j)'\right] \geq 1$$

since $p_\theta(0) \equiv I_s$. Hence

$$\hat{L}_N(\theta, \delta) \geq s \log \lambda_1(K_\theta) + \lambda_1^{-1}(K_\theta) \cdot b_3 \text{ , where } b_3 = \gamma^{-2} \cdot \varepsilon > 0 .$$

Hence, as in the proof of Theorem 3.4.1, $\lambda_1(K_\theta)$ remains bounded away from zero for N sufficiently large. A similar argument to the above also shows that $\sum_0^V \|p_\theta(j)\|^2 = \operatorname{tr} \sum_0^V p_\theta(j)p_\theta(j)'$ also remains uniformly bounded and thus attention may be restricted to $\Theta_{K,P}$ as in the proof of Theorem 3.4.1. Now since

$$\underline{\lim}_{N\to\infty} \hat{L}_N(\hat{\theta}_N, \hat{\delta}_N) \geq \sup_{\eta>0} \underline{\lim}_{N\to\infty} \left\{\log \det K_N + \operatorname{tr}\left[\int \frac{P_N^* K_N^{-1} P_N I_X(\omega)}{\left(|\det g_N|^2+\eta\right)} d\omega\right] - \hat{R}_{N,\eta}\right\}$$

where $\hat{R}_{N,\eta}$ is defined as in (1.5a) with $f^{-1}(\omega; \theta)$ replaced by $2\pi\left(P_N^* K_N^{-1} P_N\right)/\left(|\det g_N|^2+\eta\right)$. But $\hat{R}_{N,\eta}$ converges to zero a.s. by Lemmas 2.1 and 2.2 and the proof of Theorem 3.4.1 may be used to obtain $\hat{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$. Now that this is established it remains to show that $N^{-\frac{1}{2}} D_N(\hat{\delta}_N - \delta_0) \xrightarrow{\text{a.s.}} 0$. From the above proof for $\hat{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$ it follows that

$$s = \underline{\lim}_{N\to\infty} \hat{S}_N(\hat{\theta}_N, \hat{\delta}_N)$$

$$= \underline{\lim}_{N\to\infty} N^{-1}(x_N - (I_s \otimes (\hat{\Delta}_N - \Delta_0))y_N)'\Gamma_N^{-1}(\hat{\theta}_N)(x_N - (I_s \otimes (\hat{\Delta}_N - \Delta_0)y_N))$$

$$\geq \sup_{\eta>0} \underline{\lim}_{N\to\infty} \left\{N^{-1}(x_N - (I_s \otimes (\hat{\Delta}_N - \Delta_0)y_N))'\Gamma_N^{-1}\left[\Phi_\eta(\hat{\theta}_N)^{-1}\right] \cdot \right.$$

$$\left. \cdot N^{-1}(x_N - (I_s \otimes (\hat{\Delta}_N - \Delta_0))y_N)\right\} \quad (3.3)$$

where

$$\Phi(\omega;\ \theta) = (2\pi)h^*(e^{i\omega};\ \theta)\left[g(e^{i\omega};\ \theta)K(\theta)g(e^{i\omega};\ \theta)+\eta I_s\right]^{-1}h(e^{i\omega};\ \theta)$$

(see Chapter 2). But since $\hat{\theta}_N$ is eventually in a small neighbourhood of θ_0 the function Φ_η is a uniformly continuous function of $\omega \in [-\pi, \pi]$ and θ in this neighbourhood. Then in a very similar way to the proof of Lemma 2.2.5b the last line in (3.3) may be replaced by

$$\sup_{\eta>0} \varliminf_{N\to\infty} \left\{N^{-1}\left(x_N-\left(I_s \otimes \left(\hat{\Delta}_N-\Delta_0\right)\right)y_N\right)\left[\Gamma_N\left(\Phi_\eta\left(\hat{\theta}_N\right)\right)+\varepsilon_N^{(1)}I_{Ns}\right] \cdot \right.$$
$$\left. \cdot \left(x_N-\left(I_s \otimes \left(\hat{\Delta}_N-\Delta_0\right)y_N\right)\right)\right\} \quad (3.4)$$

where $\varepsilon_N^{(1)} \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$ and note that $\Gamma_N^{-1}\left(\Phi_\eta^{-1}\right)$ has been replaced by $\Gamma_N\left(\Phi_\eta\right)$. Now the expression in braces of (3.4) may be re-written as

$$\frac{1}{2\pi}\int \operatorname{tr}\left\{\left[\left(\hat{\Delta}_N-\Delta_0\right)I_Y\left(\hat{\Delta}_N-\Delta_0\right)'-\left(\Delta_N-\Delta_0\right)I_{YX}-I_{XY}\left(\Delta_N-\Delta_0\right)+I_X\right]\Phi_\eta\left(\omega;\ \hat{\theta}_N\right)\right\}d\omega$$

$$= \left[N^{-\frac{1}{2}}\left(D_N\left(\hat{\delta}_N-\delta_0\right)\right)' \ \vdots\ \left(\text{Vec } I_s\right)'\right] \cdot$$

$$\cdot \left\{\begin{bmatrix} \frac{1}{2\pi}\int \bar{F}_Y(d\omega) \otimes \Phi_\eta\left(\omega;\ \theta_0\right) & 0 \\ 0 & \frac{1}{2\pi}\int \bar{f}_0 \otimes \Phi_\eta\left(\theta_0\right)d\omega \end{bmatrix} - \varepsilon_N^{(2)}I_{(t+s)s}\right\} \cdot$$

$$\cdot \begin{bmatrix} N^{-\frac{1}{2}}D_N\left(\hat{\delta}_N-\delta_0\right) \\ \cdots\cdots\cdots\cdots \\ \left(\text{Vec } I_s\right) \end{bmatrix}$$

(see part (iv) of the proof of Theorem 2.4). Hence (3.4) may be replaced by

$$\sup_{\eta>0} \varlimsup_{N\to\infty} \left[N^{-\frac{1}{2}}\left(D_N\left(\hat{\delta}_N-\delta_0\right)\right)' \vdots \left(\text{Vec } I_s\right)'\right] .$$

$$\cdot \left\{\begin{bmatrix} \frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes \Phi_\eta\left(\omega;\ \theta_0\right) & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & \frac{1}{2\pi}\int f\left(\theta_0\right) \otimes \Phi_\eta\left(\theta_0\right) d\omega \end{bmatrix} - \varepsilon_N^{(3)} I_{(t+s)s}\right\} .$$

$$\cdot \begin{bmatrix} N^{-\frac{1}{2}} D_N\left(\hat{\delta}_N-\delta_0\right) \\ \cdots\cdots \\ \left(\text{Vec } I_s\right) \end{bmatrix} \qquad (3.5)$$

(where $\varepsilon_N^{(3)} = \varepsilon_N^{(1)} + \varepsilon_N^{(2)}$ and hence $\varepsilon_N^{(3)} \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$). But for η sufficiently small $\frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes \Phi_\eta\left(\omega;\ \theta_0\right) d\omega$ is positive definite (say with smallest eigenvalue $2\delta > 0$) since $\det h\left(\theta_0\right)$ has no zeros on the unit circle. Hence (3.5) is not less than, for N sufficiently large to make $\left|\varepsilon_N^{(3)}\right| \le \delta$,

$$\delta \cdot \varlimsup_{N\to\infty} \left\|N^{-\frac{1}{2}} D_N\left(\hat{\delta}_N-\delta_0\right)\right\|^2 + \sup_{\eta>0} \left\{\text{tr } \frac{1}{2\pi}\int f\left(\theta_0\right) \otimes \Phi_\eta\left(\theta_0\right) d\omega\right\} .$$

But the second term in this is equal to s so that (3.3) becomes

$$s \ge \varlimsup_{N\to\infty} \left\|N^{-\frac{1}{2}} D_N\left(\hat{\delta}_N-\delta_0\right)\right\|^2 \cdot \delta + s \ge s$$

from which it follows that $N^{-\frac{1}{2}} D_N\left(\hat{\delta}_N-\delta_0\right) \xrightarrow{\text{a.s.}} 0$ as required. The proofs for $\left(\tilde{\theta}_N, \tilde{\delta}_N\right)$, $\left(\overline{\theta}_N, \overline{\delta}_N\right)$ are simpler and similar to the above. □

The condition that $\Gamma_{V+1}^{(Y)}$ be nonsingular is satisfied when, for example, $F_Y(\omega)$ is absolutely continuous with density matrix $f_Y(\omega)$ nonsingular for all but a finite number of ω values. However $\Gamma_{V+1}^{(Y)} > 0$ is stronger than required for some examples of ARMA residuals as the following illustrates.

EXAMPLE 3.1. Let $z(n) = \delta y(n) + x(n)$ where all quantities are scalar (i.e. $s = t = 1$) and let $x(n)$ be generated by the first order autoregression

$$x(n) - \beta x(n-1) = \varepsilon(n) \ , \quad E\varepsilon(n) = 0 \ , \quad E\varepsilon(n)^2 = \sigma^2 \ .$$

The condition C3.1 means that here

$$\Gamma_2^{(Y)} = \begin{bmatrix} \int dF_Y & \int e^{i\omega} dF_Y \\ \int e^{i\omega} dF_Y & \int dF_Y \end{bmatrix}$$

is positive definite. But if $y(n)$ is such that dF_Y concentrates all its mass at a single point ω_0 say then $\Gamma_2^{(Y)}$ is singular since then it has determinant equal to $\left(1-e^{i\omega_0}e^{-i\omega_0}\right) = 0$. In particular if $y(n) \equiv 1$ (i.e. the regression is a mean) then our Theorem 3.1 above will not apply. However for the case under discussion the strong consistency of $\tilde{\beta}_N$, $\tilde{\delta}_N$ (for example) may still be established as follows. We take $\omega_0 = 0$ for simplicity. (The only other case of concern in the following is when $\omega_0 = \pm\pi$. When this is the case the following argument also applies.)

The difficulty in applying the proof of Theorem 3.1 to Example 3.1 arises in showing that $\tilde{R}_N(\theta) \xrightarrow{a.s.} 0$ uniformly in $\theta \in \overline{\Theta}_1$ (here note that $\overline{\Theta}_1 = [-1, 1]$) and this difficulty arises from the fact that when C3.1 is not satisfied the matrix $\tilde{M}_N(\theta)$ in $\tilde{R}_N(\theta)$ may become singular at certain θ values. (When $\omega_0 = 0$, this occurs at $\beta = 1$.) Thus we will now show that, for the example, $\tilde{R}_N(\beta) \xrightarrow{a.s.} 0$ uniformly in β belonging to $[-1, 1]$. Now $\tilde{R}_N(\beta) = \left|\left(\sqrt{\tilde{M}_N(\beta)}\right)^{-1}\tilde{m}_N(\beta)\right|^2$ where $\tilde{M}_N(\beta)$ and $\tilde{m}_N(\beta)$ are as follows:

$$\tilde{M}_N(\beta) =$$

$$= N^{-1} \sum_u I_Y |1-\beta e^{i\omega_u}|^2$$

$$= (1-\beta)^2 N^{-1} \sum_1^N y(n)^2 + \beta N^{-1} \sum_1^{N-1} \big(y(n+1)-y(n)\big)^2 + \beta N^{-1}\big(y(1)-y(N)\big)^2 , \quad (3.6)$$

$$\tilde{m}_N(\beta) = N^{-1} \sum_u I_{YX} |1-\beta e^{i\omega_u}|^2$$

$$= (1-\beta)^2 N^{-1} \sum_1^N y(n)x(n) - \beta N^{-1} \sum_1^{N-1} \big(y(n+1)-y(n)\big)x(n)$$

$$+ \beta N^{-1} \sum_1^{N-1} \big(y(n+1)-y(n)\big)x(n+1) + \beta N^{-1} x(1)\big(y(1)-y(N)\big) +$$

$$+ \beta N^{-1} x(N)\big(y(N)-y(1)\big) . \quad (3.7)$$

Letting $d_y(N)^2 = \sum_1^n y(n)^2$ it follows for $-1 \le \beta \le 1-\varepsilon$, $\varepsilon > 0$, that $Nd_y(N)^{-2}\tilde{M}_N(\beta) \xrightarrow{\text{a.s.}} \int |1-\beta e^{i\omega}|^2 F_Y(d\omega) = (1-\beta)^2 \ge \varepsilon^2 > 0$ and $N^{\frac{1}{2}} d_y(N)^{-1}\tilde{m}_N(\beta) \xrightarrow{\text{a.s.}} 0$ (see Lemmas 2.1 and 2.2). For $1-\varepsilon < \beta \le 1$ we proceed as follows. Any of the three terms in (3.6) may be used as a lower bound for $\tilde{M}_N(\beta)$ since each is positive. (If $y(n) \equiv 1$ then $\tilde{M}_N(\beta) = (1-\beta)^2$ and $\tilde{m}_N(\beta) = (1-\beta)^2\bar{x}$ so that $\tilde{M}_N(\beta)^{-\frac{1}{2}}\tilde{m}_N(\beta) \xrightarrow{\text{a.s.}} 0$ for all $1-\varepsilon < \beta \le 1$.) Then

$$\left|\tilde{M}_N^{-\frac{1}{2}}(\beta)\tilde{m}_N(\beta)\right| \le |1-\beta| . N^{-\frac{1}{2}} d_y(N)^{-1} \sum_1^N y(m)x(m)$$

$$+ \sqrt{\beta} \left| N^{-\frac{1}{2}} \left(\sum_1^{N-1} \big(y(n+1)-y(n)\big)^2\right)^{-\frac{1}{2}} \left(\sum_1^{N-1} \big(y(n+1)-y(n)\big)x(n)\right)\right|$$

$$+ \sqrt{\beta} \left| N^{-\frac{1}{2}} \left(\sum_1^{N-1} \big(y(n+1)-y(n)\big)^2\right)^{-\frac{1}{2}} \left(\sum_1^{N-1} \big(y(n+1)-y(n)\big)x(n+1)\right)\right|$$

$$+ \sqrt{\beta} \left|N^{-\frac{1}{2}} x(1)\right| + \sqrt{\beta} \left|N^{-\frac{1}{2}} x(N)\right| . \quad (3.8)$$

Now the first term on the r.h.s. of (3.8) converges a.s. to zero under condition C1.3 (see Lemma 2.3). The fourth converges a.s. to zero since $|x(1)| < \infty$ a.s. and the fifth converges a.s. to zero by ergodicity. Consider the second term (the third is dealt with in a similar way). Letting $u(m) = y(m+1) - y(m)$, $m = 1, 2, \ldots, N-1$ we need to show that

$$N^{-\frac{1}{2}}d_u(N)^{-1}\sum_1^{N-1} u(n)x(n) \xrightarrow{\text{a.s.}} 0 \qquad (3.9)$$

where $d_u^2(N) = \sum_1^{N-1} u(n)^2$. But (3.9) will follow by Lemma 2.3 provided $u(n)$ satisfies C1.3; that this is so for $u(n)$ follows from the inequality

$$\sum_1^{N-1} u(m)^2 \leq 4\sum_1^{N} y(m)^2$$

and the fact that $y(n)$ satisfies C1.3. Hence $\tilde{R}_N(\beta) \xrightarrow{\text{a.s.}} 0$ uniformly in $|\beta| \leq 1$. The remainder of the proof of Theorem 3.1 may be used to establish that $\tilde{\beta}_N \xrightarrow{\text{a.s.}} \beta_0$ where we assume that $|\beta_0| < 1$. Now, since $\int \left|1-\beta_0 e^{i\omega}\right|^2 dF_Y > 0$, because $\beta_0 \neq \pm 1$, the proof of Theorem 3.1 also gives $N^{-\frac{1}{2}}d_y(N)^2(\tilde{\delta}_N-\delta_0) \xrightarrow{\text{a.s.}} 0$.

In Hannan (1973a) regression of the type (1.1) is considered when $x(n)$ is scalar (i.e. $s = 1$) and $y(n)$ is a vector. The treatment of the strong consistency of $(\hat{\theta}_N, \hat{\delta}_N)$, $(\tilde{\theta}_N, \tilde{\delta}_N)$, $(\overline{\theta}_N, \overline{\delta}_N)$ (see Theorem 4, Hannan (1973a)) is incomplete since cases such as Example 3.1 are excluded by implicit assumptions in the proof given. It would be nice if a complete treatment of the regression problem with ARMA residuals could be given so that assumptions, such as C3.1, are not required. Alternatively the condition C3.1 may be dispensed with in Theorem 3.1 but at the expense (if the given proof is to be *easily* modified) of imposing extra conditions on the residual process $x(n)$ such as bounding the zeros of $(\det h)$ and $(\det g)$ away from the unit circle. We will not pursue the details of this here but suffice it to say an examination of the proof of Theorems 3.1 and 2.4 indicate how such alternative conditions may be framed.

The case where there are jumps present in $F_Y(\omega)$ which make $\Gamma_{V+1}^{(Y)}$ singular may, to some degree, be avoided by the following considerations. In certain regression vectors, $y(n)$, there may be a subset of variables which taken together constitute an "efficient regressor set" as follows: The $y_j(n)$ are of the form $n^{\nu} \cos \theta_j n$, $n^{\nu} \sin \theta_j n$, $0 \leq \theta_j \leq \pi$, $\nu = 0, 1, \ldots, r_j-1$, $j = 1, \ldots, t$ where *both* sine and cosine terms occur if one does so (see Hannan, 1970, Chapter VII). The estimation of the regression coefficients when the elements of $y_j(m)$ constitute an efficient regressor set may be efficiently done via ordinary least squares regression. Also such sequences have a corresponding discrete $F_Y(\omega)$. Thus if the only source of the discrete part of $F_Y(\omega)$ is from such an efficient regressor set we could proceed as follows. Let E be the transformation of $y(n)$ such that

$$Ey(n) = \begin{bmatrix} y^{(1)}(n) \\ \cdots\cdots\cdots \\ y^{(2)}(n) \end{bmatrix}$$

where $y^{(1)}(n)$ constitute an efficient regressor set. Then if $\Delta E' = \left[\Delta_E^{(1)} \;\vdots\; \Delta_E^{(2)}\right]$ consider

$$z(n) = \left[\Delta_E^{(1)} \;\vdots\; \Delta_E^{(2)}\right] \begin{bmatrix} y^{(1)}(n) \\ \cdots\cdots\cdots \\ y^{(2)}(n) \end{bmatrix} + x(n) \;;$$

that is $\left(z(n)-\Delta_E^{(1)} y^{(1)}(n)\right) = \Delta_E^{(2)} y^{(2)}(n) + x(n)$. Now if $\Delta_E^{(1)}$ is estimated by ordinary least squares regression of $z(n)$ on $y^{(1)}(n)$ the estimator $\dot{\Delta}_E^{(1)}$ will be efficient and one may form the residuals $\dot{z}(n) = z(n) - \dot{\Delta}_E^{(1)} y^{(1)}(n)$ and consider estimating $\Delta_E^{(2)}$ and θ specifying $x(n)$ by the regression

$$\dot{z}(n) = \Delta_E^{(2)} y^{(2)}(n) + x(n)$$

and use of the efficient methods $\hat{L}_N$, $\tilde{L}_N$ or $\overline{L}_N$ above. Insofar as the discrete part of $F_Y(\omega)$ has been "removed" by the preliminary regression of $z(n)$ on $y^{(1)}(n)$ it might be not unreasonably assumed that the asymptotic spectrum corresponding to $y^{(2)}(n)$, $F_{Y^{(2)}}(\omega)$ say, is absolutely continuous and has a positive definite spectral density matrix associated with it. Then the condition C3.1 will hold for $F_{Y^{(2)}}(\omega)$. Of course the procedure outlined above will not account for all conceivable situations but may be expected to cover a wide range. Also it may not always be possible to recognise or arrange that $y(n)$ has an efficient regressor subset in that, for example, certain sine terms (for which there is a cosine term) may not naturally appear. Finally the removal of $y^{(1)}(n)$ by the above least squares procedure will still work when $y^{(1)}(n)$ is *not* an efficient regressor set. In this case however the estimator of $\Delta_E^{(1)}$ will not be efficient.

4. Strong consistency of the regression coefficient estimators when the residual spectrum is known

In this section some results will be given concerning regression coefficient estimators $\delta_N(\Phi)$ (of δ in (1.1)) where

$$\delta_N(\Phi) - \delta_0 = M_N^{-1}(\Phi) m_N(\Phi)$$

in which $(M_N(\Phi),\ m_N(\Phi))$ denotes any choice of $(\hat{M}_N(\Phi),\ \hat{m}_N(\Phi))$, $(\tilde{M}_N(\Phi),\ \tilde{m}_N(\Phi))$ or $(\overline{M}_N(\Phi),\ \overline{m}_N(\Phi))$ (see equations (1.4) and (1.5)) with f^{-1} replaced by a general non-negative definite Hermitian matrix function $\Phi(\omega)$. When the true spectral density matrix $f(\omega;\ \theta_0)$, corresponding to $x(n)$, is known (an unlikely situation) the optimal choice of $\Phi(\omega)$ is $f^{-1}(\omega;\ \theta_0)$. The notation $\hat{\delta}_N(\Phi)$, $\tilde{\delta}_N(\Phi)$ and $\overline{\delta}_N(\Phi)$ will be used for the three methods mentioned above. Below we will only discuss the cases $\overline{\delta}_N(\Phi)$ and $\tilde{\delta}_N(\Phi)$. The latter case was included in the discussion of Hannan (1973b). In that article the following conditions concerning the rate of growth of $y_j(n)$ were used.

C4.1. For $1 \le j \le t$ and $\frac{1}{2} < \alpha_j$,

(a) $\overline{\lim_{N\to\infty}}\ N^{-\alpha_j} d_j(N)^2 < \infty$,

(b) $\underline{\lim_{N\to\infty}}\ N^{-\alpha_j} d_j(N)^2 > 0$.

The proof that $\tilde{\delta}_N(\Phi) - \delta_0$ converges a.s. to the null vector presented in Hannan (1973b) is incorrect when $\frac{1}{2} < \alpha_j < 1$, $1 \le j \le t$, as pointed out in a subsequent correction. A new proof of this result will be given below for $\overline{\delta}_N(\Phi)$ and $\tilde{\delta}_N(\Phi)$ when $\frac{1}{2} < \alpha_j < 1$, $1 \le j \le t$. The main motivation for the discussion of this section is an attempt to obtain the strong consistency of $\overline{\delta}_N(\overline{\theta}_N)$ and $\tilde{\delta}_N(\tilde{\theta}_N)$ (i.e. for $\Phi = f^{-1}(\omega;\ \overline{\theta}_N)$ or $f^{-1}(\omega;\ \tilde{\theta}_N)$) when $\frac{1}{2} < \alpha_j < 1$. This case was not covered by the results of §2 or §3, and only when condition C4.1(b) holds for $1 \le \alpha_j$, $1 \le j \le t$ was it possible to obtain the strong consistency of $\overline{\delta}_N(\overline{\theta}_N)$, $\tilde{\delta}_N(\tilde{\theta}_N)$ (see Corollary 2.7). The corresponding result for the case $\frac{1}{2} < \alpha_j < 1$ has only been established when $x(n)$ is generated by a vector autoregressive process (see below).

The estimator $\overline{\delta}_N(\Phi)$ will be considered first. Now

$$\overline{\delta}_N(\Phi) - \delta_0 = D_N^{-1}\left(ND_N^{-1}\overline{M}_N(\Phi)D_N^{-1}\right)^{-1}\left(ND_N^{-1}\overline{m}_N(\Phi)\right) . \tag{4.1}$$

Lemma 2.1 may be applied to show that $ND_N^{-1}\overline{M}_N(\Phi)D_N^{-1}$ converges a.s. to $\frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes \Phi(\omega)$ and for this C4.1 is not required. We will assume that $\frac{1}{2\pi}\int \overline{F}_Y(d\omega) \otimes \Phi(\omega)$ is positive definite (which is assured if $\Phi(\omega)$ is positive definite for all $\omega \in [-\pi, \pi]$) so that its inverse has finite norm. Then in order that $\overline{\delta}_N(\Phi) - \delta_0$ converges a.s. to the null vector it is sufficient to show that, for $1 \le j \le t$,

$$\overline{C}_{N,j}(\Phi) = Nd_j(N)^{-1}D_N^{-1}\overline{m}_N(\Phi) \tag{4.2}$$

converges a.s. to the null vector. (Recall that

$$\overline{m}_N(\Phi) = (2\pi)^{-1}\int \left(I_t \otimes \Phi\right)\text{Vec}\left(I_{XY}\right)d\omega \text{ .)}$$

The norming of $\overline{m}_N(\Phi)$ in (4.2) is stronger than that in Lemma 2.2 by the factor $N^{\left(1-\alpha_j\right)/2}$, $\alpha_j < 1$. The following result is quite general and in particular it is only required that $x(n)$ be weakly stationary.

THEOREM 4.1. *Let $y(n)$ satisfy* C1.1 *and* C4.1 *with $\frac{1}{2} < \alpha_j < 1$, $1 \le j \le t$, and let $x(n)$ be a second order stationary process with absolutely continuous spectrum having corresponding spectral density matrix $f(\omega)$ where $\|f(\omega)\| \le b < \infty$. Then for $1 \le j \le t$, $\overline{C}_{N,j}(\Phi)$ defined in* (4.2) *converges a.s. to the null vector provided Φ has an absolutely convergent Fourier series.*

The proof of this result is quite involved. Furthermore Professor E. J. Hannan has recently indicated to me that Theorem 4.1 may be considerably strengthened by use of Menchoff's inequality (see Stout (1974, p. 18) for example). It would appear that only C4.1(b) is needed and $\alpha_j > 0$ in Theorem 4.1. The use of this inequality also appears to simplify the proof. For these reasons we will give only a brief outline of the proof of Theorem 4.1. This proof is due to Professor Hannan in the scalar case and is unpublished. The proof for the vector case, given below, is a relatively straightforward modification of his proof.

Proof. Let $\tilde{y}(m) = d(N)^{-1}y(m)$. The variance of $\|\overline{C}_{N,j}(\Phi)\|$ is

$$E\left(\overline{C}_{N,j}(\Phi)^*\overline{C}_{N,j}(\Phi)\right) = d_j(N)^{-2} \operatorname{tr}\left[\sum_{l=1}^{t} \tilde{y}'_{N,l}A_N\Gamma_N A_N\tilde{y}_{N,l}\right] \tag{4.3}$$

where

$$\tilde{y}'_{N,l} = \left(\tilde{y}_l(1)I_s \,\vdots\, \cdots \,\vdots\, \tilde{y}_l(N)I_s\right) ,$$

$$\Gamma_N = \left[\int f(\omega)e^{i(m-n)\omega}d\omega\right]_{m,n=1,\ldots,N} ,$$

$$A_N = \left[\frac{1}{2\pi}\int \Phi(\omega)e^{i(m-n)\omega}d\omega\right]_{m,n=1,\ldots,N} .$$

But $\Gamma_N \le \left(\sup_\omega \|f(\omega)\|\right)I_{Ns} \le bI_{Ns}$, $b < \infty$, and

$$A_N A_N \le \left(\sup_\omega \|\Phi(\omega)\|\right)^2 I_{Ns} \le c^2 I_{Ns} , \quad c < \infty ,$$

so that

$$E\left(\overline{C}_{N,j}(\Phi)^*\overline{C}_{N,j}(\Phi)\right) \le b.c^2.s.t.d_j(N)^{-2} \quad \left(\text{since } \tilde{y}'_{N,l}\tilde{y}_{N,l} = I_s\right)$$

$$= O\left(N^{-\alpha_j}\right) \tag{4.4}$$

by use of C4.1(b). Thus if $\beta\alpha_j > 1$ then $\overline{C}_{M^\beta,j}(\Phi) \xrightarrow{\text{a.s.}} 0$ as $M \to \infty$ by the Borel-Cantelli lemma. The typical element of $\overline{C}_{N,j}(\Phi)$ is composed of at most s terms of the form

$$e(N) = d_j(N)^{-1}d_k(N)^{-1} \cdot \frac{1}{2\pi}\int \Phi_{ab}(\omega) \sum_{m=1}^{N} x_b(m)e^{im\omega} \sum_{n=1}^{N} y_k(n)e^{-in\omega}d\omega . \tag{4.5}$$

From now on we will drop the sbuscripts from $\Phi_{ab}(\omega)$ and $x_b(m)$ but *not* $y_k(n)$ and will let $N(M) = M^\beta$. Now writing $\sup_N$ for the supremum over $N(M) \le N \le N(M+1)$ then

$$\sup_N |e(N)-e(N(M))| \le \sup_N \left| e(N) - \frac{d_j(N(M))}{d_j(N)} \cdot \frac{d_k(N(M))}{d_k(N)} \cdot e(N(M)) \right|$$

$$+ \sup_N \left| \frac{d_j(N(M))}{d_j(N)} \cdot \frac{d_k(N(M))}{d_k(N)} - 1 \right| \cdot e(N(M)) \; . \tag{4.6}$$

The second term on the r.h.s. of this converges a.s. to zero since $e(N(M))$ does so and the factor multiplying $e(N(M))$ is dominated by unity. Consider, in relation to the first term on the r.h.s. of (4.6),

$$\left| e(N) - \frac{d_j(N(M))}{d_j(N)} \cdot \frac{d_k(N(M))}{d_k(N)} \cdot e(N(M)) \right|$$

$$\le \left| \sum_{m,n=N(M)+1}^{N} x(m)y_k(n)\Phi(m-n)/(d_j(N)d_k(N)) \right|$$

$$+ \left| \sum_{m=N(M)+1}^{N} \sum_{n=1}^{N} x(m)y_k(n)\Phi(m-n)/(d_j(N)d_k(N)) \right|$$

$$+ \left| \sum_{m=1}^{N(M)} \sum_{n=N(M)+1}^{N} x(m)y_k(n)\Phi(m-n)/(d_j(N)d_k(N)) \right| \; . \tag{4.7}$$

(i) Consider the first term on r.h.s. of (4.7). Now

$$E\left\{\sup_N \left| \sum_{m,n=N(M)+1}^{N} x(m)y_k(n)\Phi(m-n)/(d_j(N)d_k(N)) \right|^2\right\}$$

$$\le \left(\sup_\omega |\Phi(\omega)|\right)^2 . E\left\{\sup_N \left\{\left(\sum_{N(M)+1}^{N} x(m)^2\right)\left(\sum_{N(M)+1}^{N} y_k(m)^2\right) / \left(d_j(N)^2 d_k(N)^2\right)\right\}\right\}$$

$$\le c^2 . \{(M+1)^\beta - M^\beta\} . \left\{d_k((M+1)^\beta)^2 - d_k(M^\beta)^2\right\} / \left(d_j(M^\beta)^2 d_k(M^\beta)^2\right) \; , \tag{4.8}$$

and by precisely the same argument given in the proof of Theorem 1, Hannan (1973b) this last line, when summed over M from 1 to ∞ is finite. Hence by the Borel-Cantelli lemma the first term in (4.7) converges a.s. to zero.

(ii) Consider the second term on the r.h.s. of (4.7). Now this term has squared modulus equal to

$$d_j(N)^{-2}.d_k(N)^{-2}.\left|\sum_{l=1}^{N-1}\Phi(l)\sum_m{}' x(m)y_k(m-l)\right|^2$$

(where $\sum_m{}'$ is a sum from $\max(1+l,\ N(M+1))$ to $\min(N,\ N(M)+l)$)

$$\leq d_j(N)^{-2}.d_k(N)^{-2}\left(\sum_{l=1}^{N-1}|\Phi(l)|\right)\left[\sum_{l=1}^{N-1}|\Phi(l)|\left(\sum_m{}' x(m)^2\right)\left(\sum_m{}' y_k(m-l)^2\right)\right] \quad (4.9)$$

using the Cauchy-Schwarz inequality twice. Now the expected value of the supremum over N on the r.h.s. of (4.9) is no greater than

$$c_1d_j\big(N(M)\big)^{-2}.d_k\big(N(M)\big)^{-2}.\big(N(M+1)-N(M)\big).$$

$$.\left(\sum_{l=1}^{N(M+1)}|\Phi(l)|\right)\left(\sum_{l=1}^{N(M+1)}|\Phi(l)|\sum_m{}' y_k(m-l)^2\right)$$

$$\leq c_1.c_2.d_j\big(M^\beta\big)^{-2}d_k\big(M^\beta\big)^{-2}.M^{\beta-1}.\left(\sum_{l=1}^{(M+1)^\beta}|\Phi(l)|\sum_m{}' y_k(m-l)^2\right)$$

where $\sum_{l=0}^{\infty}|\Phi(l)| \leq c_2$, $\sup_\omega |f_{bb}(\omega)| \leq c_1$ and

$$N(M+1) - N(M) = (M+1)^\beta - M^\beta = O\big(M^{\beta-1}\big)$$

for M large. Now the r.h.s. of (4.10) may be re-written, apart from the constant factor $\{c_1.c_2\}$, as

$$d_j\big(M^\beta\big)^{-2}.d_k\big(M^\beta\big)^{-2}.M^{\beta-1}.\left\{\sum_{l=1}^{(M+1)^\beta-M^\beta}|\Phi(l)|\sum_{m=M^\beta+1}^{(M+1)^\beta}y_k(m-l)^2\right.$$

$$\left.+\sum_{l=(M+1)^\beta-M^\beta+1}^{M^\beta}|\Phi(l)|\sum_{m=M^\beta+1}^{(M+1)^\beta}y_k(m-l)^2+\sum_{l=M^\beta+1}^{(M+1)^\beta}|\Phi(l)|\sum_{m=l}^{(M+1)^\beta}y_k(n-l)^2\right\}. \quad (4.11)$$

Now the first term in (4.11) is dominated by

$$d_j(M^\beta)^{-2} d_k(M^\beta)^{-2} . M^{\beta-1} \left(\sum_{l=1}^{(M+1)^\beta - M^\beta} |\Phi(l)| \right) \left(\sum_{2M^\beta-(M+1)^\beta}^{M^\beta} y_k(m)^2 \right)$$

$$\leq c_2 . c_3 M^{-\beta(\alpha_j+\alpha_k)} . M^{\beta-1} \left| \sum_{2M^\beta-(M+1)^\beta}^{M^\beta} y_k(m)^2 \right| \quad (4.12)$$

where c_2 is as above and c_3 is a finite constant from C4.1. But the last factor on the r.h.s. of (4.12) is

$$\left[d_k(M^\beta)^2 - d_k((M-1)^\beta)^2\right] + \left[d_k((M-1)^\beta)^2 - d_k(2M^\beta-(M+1)^\beta)\right]$$

and for M large $2M^\beta - (M+1)^\beta - (M-2)^\beta$ is positive so that there are less terms in $d_k((M-2)^\beta)^2$ than in $d_k(2M^\beta-(M+1)^\beta)^2$. Hence

$$\sum_{2M^\beta-(M+1)^\beta}^{M^\beta} y_k(m)^2 \leq \left[d_k(M^\beta)^2 - d_k((M-1)^\beta)^2\right] + \left[d_k((M-1)^\beta)^2 - d_k((M-2)^\beta)^2\right] . \quad (4.13)$$

Using this in the r.h.s. of (4.12) yields two terms each of which may be shown to be finite when summed over $1 \leq M < \infty$ by the argument of Theorem 1, Hannan (1973b). The second term on the r.h.s. of (4.11) is no greater than

$$d_j(M^\beta)^{-2} d_k(M^\beta)^{-2} . M^{\beta-1} . \sum_{l=1}^{\infty} |\Phi(l)| \left\{ d_k((M+1)^\beta - l)^2 - d_k(M^\beta - l)^2 \right\}$$

which when summed over M becomes

$$\sum_{l=1}^{\infty} |\Phi(l)| \sum_{M=1}^{\infty} M^{-\beta(\alpha_j+\alpha_k)} M^{\beta-1} \left[\sum_{M^\beta+1-l}^{(M+1)^\beta - l} y_k(m)^2 \right] . \quad (4.14)$$

Now β may be chosen so that $1 - \alpha_j - \beta^{-1} < 0$ since $\alpha_j > \frac{1}{2}$. Then as in the proof of Theorem 1, Hannan (1973b), for $\gamma = -\left(1 - \alpha_j - \beta^{-1}\right)$, we have (4.14) dominated by

$$\sum_{l=1}^{\infty} |\Phi(l)| \left\{ \sum_{n=1}^{\infty} y_k(n)^2 n^{-\alpha_k - \gamma} \right\}$$

which is finite by C4.1(a) (see Hannan 1971b). Finally, in (4.11), the third term is dominated by

$$d_j(M^\beta)^{-2} \cdot d_k(M^\beta)^{-2} \cdot M^{\beta-1} \sum_{l=M^\beta+1}^{(M+1)^\beta} |\Phi(l)| \sum_{M^\beta+1}^{(M+1)^\beta} y_k(m-l)^2$$

which may be handled in the same way as was the second term in (4.11). Hence the second term of (4.7) converges a.s. to zero.

(iii) Consider the third term on the r.h.s. of (4.7). This may be shown to converge to zero by the same argument as just given for the second term in the r.h.s. of (4.7). □

To establish the analogous result for $\tilde{\delta}_N(\Phi) - \delta_0$ requires extra smoothness conditions on the elements of $\Phi(\omega)$.

COROLLARY 4.2. *If in addition to the conditions of Theorem* 4.1 *it is required that the elements of* Φ *belong to* Λ_γ *(the Lipschitz class of degree* γ *) where* $\gamma > (1-\alpha_j)/(2\alpha_j)$ *and these elements are of bounded variation then*

$$\tilde{C}_{N,j}(\Phi) = N d_j(N)^{-1} D_N^{-1} \tilde{m}_N(\Phi) \xrightarrow{\text{a.s.}} 0 , \quad 1 \le j \le t . \tag{4.15}$$

Proof. Let $\bar{C}_{N,j}(\Phi)$ be as in (4.2). We will show that

$$\tilde{C}_{N,j}(\Phi) - \bar{C}_{N,j}(\Phi) \xrightarrow{\text{a.s.}} 0 , \quad 1 \le j \le t . \tag{4.16}$$

Then Theorem 4.1 yields (4.15) since the conditions on Φ ensure those of Theorem 4.1 on Φ . Now let S_{N^β} be the ordinary partial sum of the Fourier series for Φ to $[N^\beta]$ terms where $[N^\beta]$ is the greatest integer not greater than N^β and $2\beta > (1-\alpha_j)/(2\gamma)$. Then the typical element of $\tilde{C}_{N,j}(\Phi) - \tilde{C}_{N,j}(S_{N^\beta})$ is composed of at most s terms of the form

$$(N')^{-1} \sum_u \left(\Phi(\omega_u) - S_{N^\beta}(\omega_u)\right) d_j(N)^{-1} d_k(N)^{-1} \sum_{m,n=1}^{N} x(m) y_k(n) e^{i(m-n)\omega_u} \quad (4.17)$$

in which the subscript (a, b) has been dropped from Φ and S_{N^β} and the subscript (b) has been dropped from $x(m)$. The modulus of (4.17) is dominated by

$$d_j(N)^{-1} d_k(N)^{-1} \left[\sum_1^N x(m)^2\right]^{\frac{1}{2}} \left[\sum_1^N y_k(m)^2\right]^{\frac{1}{2}} \cdot \sup_\omega \left|\Phi(\omega_u) - S_{N^\beta}(\omega_u)\right|$$

$$= d_j(N)^{-1} \left(N^{\frac{1}{2}} \sup_\omega \left|\Phi(\omega_u) - S_{N^\beta}(\omega_u)\right|\right) \left(N^{-1} \sum_1^N x(m)^2\right)^{\frac{1}{2}} . \quad (4.18)$$

But $\left(N^{-1} \sum_1^N x(m)^2\right) \xrightarrow{\text{a.s.}} \Gamma_{bb}(0) < \infty$ (by ergodicity) and by Zygmund (1959, p. 64), $\sup_\omega \left|\Phi(\omega) - S_{N^\beta}(\omega)\right| = O\left(N^{-\beta\gamma} \log N^\beta\right)$. Thus (4.18) is

$$O\left(N^{-\alpha_j/2} . N^{\frac{1}{2}} . N^{-\beta\gamma} \log N^\beta\right) = O\left(N^{\left(1-\alpha_j - 2\beta\gamma + 2\beta\varepsilon\right)/2}\right) , \quad \varepsilon > 0 .$$

But $\beta > \left(1-\alpha_j\right)/(2\gamma)$ so that there exists $\varepsilon > 0$ such that $2\beta > \left(1-\alpha_j\right)/(\gamma-\varepsilon)$ and thus $1 - \alpha_j - 2\beta(\gamma-\varepsilon) < 1 - \alpha_j - \left(1-\alpha_j\right) = 0$.

Hence $\tilde{C}_{N,j}(\Phi) - \tilde{C}_{N,j}\left(S_{N^\beta}\right) \xrightarrow{\text{a.s.}} 0$ and using a similar proof it may be shown that $\bar{C}_{N,j}(\Phi) - \bar{C}_{N,j}\left(S_{N^\beta}\right) \xrightarrow{\text{a.s.}} 0$, for $\beta > \left(1-\alpha_j /(2\gamma)\right.$. Thus the corollary will be established if it is shown that

$$\tilde{C}_{N,j}\left(S_{N^\beta}\right) - \bar{C}_{N,j}\left(S_{N^\beta}\right) \xrightarrow{\text{a.s.}} 0 . \quad (4.19)$$

Now $\tilde{C}_{N,j}\left(S_{N^\beta}\right) - \bar{C}_{N,j}\left(S_{N^\beta}\right)$ has typical element composed of a sum of s terms each of which is of the form (again dropping subscripts from S_{N^β} and $x(m)$)

$$(N')^{-1} \sum_u S_{N^\beta}(\omega_u) d_j(N)^{-1} d_k(N)^{-1} \sum_{m,n=1}^{N} x(m) y_k(n) e^{i(m-n)\omega_u}$$

$$- \frac{1}{2\pi} \int S_{N^\beta}(\omega) d_j(N)^{-1} d_k(N)^{-1} \sum_{m,n=1}^{N} x(m) y_k(n) e^{i(m-n)\omega} d\omega$$

$$= d_j(N)^{-1} d_k(N)^{-1} \Big\{ \sum_{N'-N^\beta \le m-n \le N-1} x(m) y_k(n) \Phi(m-n-N')$$

$$+ \sum_{-N+1 \le m-n \le -N'+N^\beta} x(m) y_k(n) \Phi(m-n+N') \Big\} \quad (4.20)$$

where $\Phi(l)$ is the lth Fourier coefficient of $\Phi_{ab}(\omega)$ and the summations over $(m-n)$ are zero if $N' - N^\beta > N - 1$. Now the contribution to (4.20) from the second term in braces is dominated by

$$d_j(N)^{-1} d_k(N)^{-1} \sum_{l=1}^{N^\beta} \left\{ |\Phi(l)| \left(l^{-1} \sum_{m=1}^{(N-N'+l)} x(m)^2 \right)^{\frac{1}{2}} \left(l \sum_{m=1}^{N-N'+l} y_k(m+N'-l)^2 \right)^{\frac{1}{2}} \right\}$$

$$\le c.d_j(N)^{-1} \sum_{l=1}^{N^\beta} \left(|\Phi(l)|.l^{\frac{1}{2}} \right), \quad c < \infty \quad (4.21)$$

since $l^{-1} \sum_{m=1}^{N-N'+l} x(m)^2 \le l^{-1} \sum_{m=1}^{l} x(m)^2$ which is uniformly bounded and similarly $d_k(N)^{-1} \left(\sum_{1}^{N-N'+l} y_k(m+N'-l)^2 \right)^{\frac{1}{2}}$ is uniformly bounded. But the r.h.s. of (4.21) is dominated by $d_j(N)^{-1} N^{\beta/2} \sum_{l=0}^{\infty} |\Phi(l)|$ which is $O\left(N^{-(\alpha_j-\beta)/2}\right)$ since $\sum_{-\infty}^{\infty} |\Phi(l)| < \infty$ (because when $\Phi \in \Lambda_\gamma$, $\gamma > 0$ and Φ is of bounded variation it has an absolutely convergent Fourier series - see Zygmund (1959), Theorem 3.6, p. 241). Now we may choose β such that $(1-\alpha_j)/(2\gamma) < \beta < \alpha_j$ since $\gamma > (1-\alpha_j)/(2\alpha_j)$ and thus $(1-\alpha_j)/(2\gamma) < \alpha_j$. Hence (4.21) converges a.s. to zero and hence so does the second term on the r.h.s. of (4.20). The contribution to (4.20) from the first term in braces is dominated by

$$d_j(N)^{-1}d_k(N)^{-1} \sum_{l=-N^\beta}^{-1} |\Phi(l)| \left\{\sum_{m=1}^{N'-N+l} x(l+m+N')^2\right\}^{\frac{1}{2}} \left\{\sum_{m=1}^{N'-N+l} y_k(m)^2\right\}^{\frac{1}{2}} .$$

which may be treated in the same way as the first term in (4.20) was. Hence (4.19) is satisfied and the proof is complete. □

If the conditions in Corollary 4.2 that Φ have elements belonging to Λ_γ , $\gamma > (1-\alpha_j)/(2\alpha_j)$, $\frac{1}{2} < \alpha_j < 1$, and these elements are of bounded variation are replaced by the condition that Φ have elements in Λ_γ , $\gamma > \frac{1}{2}$ then the above corollary continues to hold since with the latter specification Φ has an absolutely convergent Fourier series (see Bernstein's theorem, Zygmund, 1959, p. 240). In this case β in the above proof needs to be chosen, if $\varepsilon > 0$, so that $1 - \alpha_j - \beta + \varepsilon < 0$ and $\beta < \alpha_j$. That is $\alpha_j > \beta > 1 - \alpha_j - \varepsilon$. But since $\alpha_j > \frac{1}{2}$ such a β may be chosen.

The case where $\delta_N(\Phi)$ is $\hat{\delta}_N(\Phi)$ has not been investigated by us. No doubt a similar corollary to the above could be established for this case although stronger smoothness conditions on Φ may be required.

The application of the above results to the joint estimation problem discussed in §2, i.e. to $\tilde{\delta}_N = \tilde{\delta}_N(\tilde{\theta}_N)$, $\bar{\delta}_N = \bar{\delta}_N(\bar{\theta}_N)$, is not at all obvious except in special cases. One such special case occurs when $x(n)$ is generated by a vector autoregression and then, in fact, the proofs of Theorem 4.1 and Corollary 4.2 are very much simpler. That is, when

$$(2\pi f(\omega;\, \bar{\theta}_N))^{-1} = h\left(e^{i\omega};\, \bar{\theta}_N\right)^* . K^{-1}(\bar{\theta}_N) . h\left(e^{i\omega};\, \bar{\theta}_N\right) ,$$

$$h\left(e^{i\omega};\, \bar{\theta}_N\right) = \sum_{j=0}^{q} B(j;\, \bar{\theta}_N) e^{ij\omega}$$

we have the typical element of $\bar{C}_{N,j}\left(f^{-1}(\omega;\, \bar{\theta}_N)\right)$ composed of at most s terms of the form (see (4.5))

$$e(N) = \sum_{u=0}^{q} \sum_{v=0}^{q} \left[B\left(u;\ \overline{\theta}_N\right)^* K^{-1}\left(\overline{\theta}_N\right) B\left(v;\ \overline{\theta}_N\right)\right]_{ab} \cdot$$

$$\left\{d_j(N)^{-1} d_k(N)^{-1} \cdot \frac{1}{2\pi} \int e^{i(v-u)\omega} \sum_{m=1}^{N} x_b(m) e^{im\omega} \sum_{n=1}^{N} y_k(n) e^{-in\omega} d\omega\right\} . \quad (4.22)$$

But, here, the term in braces converges a.s. to zero by the use of Theorem 4.1 (but in a much simpler way) and

$$\left[B\left(u;\ \overline{\theta}_N\right)^* K^{-1}\left(\overline{\theta}_N\right) B\left(v;\ \overline{\theta}_N\right)\right]_{ab} \xrightarrow{\text{a.s.}} \left[B\left(u;\ \theta_0\right)^* K^{-1}\left(\theta_0\right) B\left(v;\ \theta_0\right)\right]_{ab}$$

by Theorem 2.4. Hence $e(N)$ in (4.22) converges a.s. to zero. The difficulty in applying Theorem 4.1 to the case where $f\left(\omega;\ \overline{\theta}_N\right)$ is a more general spectral density matrix arises because in such cases the Fourier coefficients of $f^{-1}\left(\omega;\ \overline{\theta}_N\right)$ may be nonnull at all lags. Then, in the proof of Theorem 4.1 it is no longer possible to take variances of the various quantities and consequently the strong consistency may not obviously be established by use of the Borel-Cantelli lemma. It may be that alternative methods of establishing Theorem 4.1 (such as those mentioned just before that proof) will yield a proof in which Φ in the definition of (4.2) may also depend upon $\overline{\theta}_N$.

5. The central limit theorem for the regression model

Below, the asymptotic distribution for the estimators obtained from $\overline{L}_N(\theta, \Delta)$ will first be considered. The other two procedures $\tilde{L}_N$, $\hat{L}_N$ will be discussed towards the end. For this section we will require that $x(n)$ satisfy the conditions of Chapter 2 (see 2.C2 and 2.C3). Recall that in §2.3, k was required to be parameterised by θ and K by μ where θ and μ varied independently of each other. This will be assumed here also. It will also be assumed that $\overline{\theta}_N \xrightarrow{\text{a.s.}} \theta_0$, $\overline{\mu}_N \xrightarrow{\text{a.s.}} \mu_0$ so that the spaces M_θ and M_μ of Chapter 2 may need to be further restricted by the extra conditions required for Theorem 2.4 to hold. It will be convenient to use the notation $X_N \xrightarrow{\mathcal{D}} N(0, A)$ to mean the random vectors X_N tend in distribution to the random vector having multivariate normal distribution with null mean vector and covariance matrix A.

THEOREM 5.1. *Let* $x(n)$ *satisfy the conditions of Theorem* 2.3.1 *and Theorem* 2.4. *Let* $y(n)$ *also satisfy the conditions of Theorem* 2.4. *Then*

(a) $$\begin{bmatrix} N^{\frac{1}{2}}\left(\overline{\theta}_N-\theta_0\right) \\ \\ D_N\left(\overline{\delta}_N-\delta_0\right)\end{bmatrix} \xrightarrow{\mathcal{D}} N\left(0, \begin{bmatrix}\Omega^{-1} & 0 \\ \\ 0 & \Sigma^{-1}\end{bmatrix}\right)$$ *where* Ω *is defined in Theorem* 2.3.1 *and*

$$\Sigma = \left[\int \overline{F}_Y(d\omega) \otimes f^{-1}\left(\omega;\ \theta_0,\ \mu_0\right)\right];$$

(b) $N^{\frac{1}{2}}\left(\overline{K}_N-\dot{K}_N\right) \xrightarrow{P} 0$ *where* $\dot{K}_N$ *is as in* §2.3 *and*

$$\overline{K}_N = (2\pi)^{-1} \int k\left(\overline{\theta}_N\right)^{-1}\left\{I_Z-I_{YZ}^*\overline{\Delta}_N^*-\overline{\Delta}_N I_{YZ}+\overline{\Delta}_N I_Y \overline{\Delta}_N^*\right\} k\left(\overline{\theta}_N\right)^{-1*} d\omega\ ;$$

(c) if it is further assumed that the $\varepsilon(m)$ *have finite fourth moment then*

$$N^{\frac{1}{2}}\left(\overline{\mu}_N-\mu_0\right) \xrightarrow{\mathcal{D}} N(0,\ \Pi)$$

where Π *is defined in Theorem* 2.3.1.

Proof of (a). For this proof we will use the notation

$k_N\left(e^{i\omega}\right) = k\left(e^{i\omega};\ \overline{\theta}_N\right)$, $K_N = K\left(\overline{\mu}_N\right)$, $f_N(\omega) = (2\pi)^{-1} k_N\left(e^{i\omega}\right) K_N k_N\left(e^{i\omega}\right)^*$, $\Delta_N = \overline{\Delta}_N$. Similarly k_0, K_0, f_0 and Δ_0 (or δ_0) will denote these quantities evaluated at $\left(\theta_0,\ \mu_0,\ \delta_0\right)$. We will also drop the overbar from expressions such as $\overline{R}_N$, $\overline{M}_N$ etc. To obtain the joint CLT for $\left(N^{\frac{1}{2}}\left(\theta_N-\theta_0\right)',\ \left(D_N\left(\delta_N-\delta_0\right)\right)'\right)'$ it will be convenient to consider the CLT for

$N^{\frac{1}{2}}(\theta_N-\theta_0)$ first as follows. For (θ, μ) fixed concentrate $\overline{L}_N$ with respect to δ. Thus rewriting equation (1.5c) to account for the partition of θ as (θ, μ) we have

$$\overline{L}_N(\theta, \mu, \delta_N(\theta, \mu)) = \log \det K(\mu) + Q_N(\theta, \mu) - R_N(\theta, \mu)$$

from which it follows (see the proof of Theorem 2.3.1) that

$$0 = N^{\frac{1}{2}} \cdot \frac{\partial Q_N(\theta_0, \mu_N)}{\partial \theta} - N^{\frac{1}{2}} \cdot \frac{\partial R_N(\theta_0, \mu_N)}{\partial \theta} + \left[\frac{\partial^2 Q_N(\dot{\theta}_N, \mu_N)}{\partial \theta^2} - \frac{\partial^2 R_N(\dot{\theta}_N, \mu_N)}{\partial \theta^2}\right] \cdot \left[N^{\frac{1}{2}}(\theta_N-\theta_0)\right] \quad (5.1)$$

where $\|\dot{\theta}_N-\theta_0\| \le \|\theta_N-\theta_0\|$. Consider each term on the r.h.s. of (5.1) as follows.

(i) By the proof of Theorem 2.3.1,

$$N^{\frac{1}{2}} \frac{\partial Q_N(\theta_0, \mu_N)}{\partial \theta} \xrightarrow{\mathcal{D}} N(0, \Omega) \quad (5.2)$$

where Ω is as in the statement of that theorem.

(ii) Also, by the proof of Theorem 2.3.1,

$$\frac{\partial^2 Q_N(\dot{\theta}_N, \mu_N)}{\partial \theta^2} \xrightarrow{\text{a.s.}} \Omega \,. \quad (5.3)$$

(iii) Next

$$N^{\frac{1}{2}} \frac{\partial R_N(\theta_0, \mu_N)}{\partial \theta_j} = 2\left(N^{\frac{1}{2}} D_N^{-1} \frac{\partial m_N(\theta_0, \mu_N)}{\partial \theta_j}\right)^* \left(ND_N^{-1} M_N(\theta_0, \mu_N) D_N^{-1}\right)^{-1} \left(ND_N^{-1} m_N(\theta_0, \mu_N)\right)$$
$$+ \left(N^{\frac{1}{2}} D_N^{-1} m_N(\theta_0, \mu_N)\right)^* \frac{\partial}{\partial \theta_j} \left(ND_N^{-1} M_N(\theta_0, \mu_N) D_N^{-1}\right)^{-1} \left(ND_N^{-1} m_N(\theta_0, \mu_N)\right) . \quad (5.4)$$

Now consider the first term on the r.h.s. of (5.4). The matrix in this bilinear form converges a.s. to Σ^{-1} (by a similar argument as used to establish Lemma 2.1) and this is positive definite (since $f^{-1}(\omega; \theta_0, \mu_0) > 0$, $\omega \in [-\pi, \pi]$ so that $\|\Sigma^{-1}\|$ is finite. In the vector $ND_N^{-1}m_N(\theta_0, \mu_N)$ in the above bilinear form we may replace K_N by K_0 as follows:

$$ND_N^{-1}m_N(\theta_0, \mu_N) - ND_N^{-1}m_N(\theta_0, \mu_0)$$

$$= N \text{ Vec}\left\{\int k_0^{-1*}\left(K_N^{-1}-K_0^{-1}\right)k_0^{-1}I_{XY}d(N)^{-1}d\omega\right\} \tag{5.5}$$

in which the typical element is

$$\sum_{\substack{b\ c\ d\\=1}}^{s} \left(K_N^{-1}-K_0^{-1}\right)_{bc}\left\{\frac{1}{2\pi}\int \left(k_0^{-1*}\right)_{ab}\left(k_0^{-1}\right)_{cd} \sum_{m=1}^{N} x_d(m)e^{im\omega} \sum_{n=1}^{N} \tilde{y}_e(n)e^{-in\omega}d\omega\right\} \tag{5.6}$$

where $\tilde{y}_e(n) = y_e(n)/d_e(N)$. Taking the variance of the quantity in braces in (5.6) yields, where we have put $\Phi(\omega) = \left(k_0^{-1*}\right)_{ab}\left(k_0^{-1}\right)_{cd}$,

$$(2\pi)^{-2} \iint \Phi(\omega)\Phi(\mu) \sum_{\substack{m\ n\\=1}}^{N}\sum \left(\gamma_0(m-n)\right)_{dd}e^{im\omega}e^{-in\mu}\left(\sum_{\substack{p\ q\\=1}}^{N}\sum \tilde{y}_e(p)e^{-ip\omega}\tilde{y}_e(q)e^{iq\mu}\right)d\omega d\mu$$

$$= \int \left|\frac{1}{2\pi}\int \Phi(\omega) \sum_{p=1}^{N} \tilde{y}_e(p)e^{-ip\omega} \sum_{m=1}^{N} e^{im(\omega-\lambda)}d\omega\right|^2 . \left(f_0(\lambda)\right)_{dd}d\lambda$$

$$\le b_1 \int \left|\sum_{m=1}^{N} \left(\frac{1}{2\pi}\int \Phi(\omega) \sum_{p=1}^{N} \tilde{y}_e(p)e^{-ip\omega}e^{im\omega}d\omega\right)e^{-im\lambda}\right|^2 d\lambda$$

$$\left(\text{where } \sup_{\omega}\left|\left(f_0(\omega)\right)_{dd}\right| \le b_1 < \infty\right)$$

$$= b_1 \sum_{m=1}^{N} \left|\frac{1}{2\pi}\int \Phi(\omega) \sum_{p=1}^{N} \tilde{y}_e(p)e^{i(m+p)\omega}d\omega\right|^2 . \tag{5.7}$$

Now using the convention that $\dot{y}_e(p) = \tilde{y}_e(p)$ if $1 \le p \le N$ and zero otherwise and letting $\Phi(j)$ be the jth Fourier coefficient of $\Phi(\omega)$ we may re-write this last line as

$$b_1 \cdot \sum_{m=1}^{N} \left| \frac{1}{2\pi} \int \sum_{j=-\infty}^{\infty} \Phi(j) e^{-ij\omega} \sum_{-\infty}^{\infty} \dot{y}_e(p) e^{-ip\omega} . e^{im\omega} d\omega \right|^2$$

$$= b_1 \cdot \sum_{m=1}^{N} \left| \sum_{j=-\infty}^{\infty} \Phi(j) \dot{y}_e(m-j) \right|^2$$

$$\leq b_1 \sum_{m=-\infty}^{\infty} \left| \sum_{j=-\infty}^{\infty} \Phi(j) \dot{y}_e(m-j) \right|^2 .$$

But $\sum_{j=-\infty}^{\infty} \Phi(j) \dot{y}_e(m-j)$ is the mth Fourier coefficient of the function

$\Phi(\omega) \sum_{-\infty}^{\infty} \dot{y}_e(k) e^{-ik\omega}$ so that this last line is equal to

$$b_1 \int |\Phi(\omega)|^2 \left| \sum_{k=-\infty}^{\infty} \dot{y}_e(k) e^{-ik\omega} \right|^2 d\omega \leq b_1 \cdot \sup_{\omega} |\Phi(\omega)|^2 \cdot \sum_{1}^{N} \tilde{y}_e(k)^2$$

which is almost surely bounded since the last factor is just $G_{ee}^{(Y)}(0)$.

But $K_N \xrightarrow{\text{a.s.}} K_0$, which is nonsingular, so that (5.6) converges in probability to zero and thus so does (5.5). We will later show (see Theorem 5.2 below) that $ND_N^{-1} m_N(\theta_0, \mu_0)$ is asymptotically normally distributed with covariance matrix $\not\Sigma$ which has finite norm. Thus provided

$$N^{\frac{1}{2}} D_N^{-1} \frac{\partial m_N(\theta_0, \mu_N)}{\partial \theta} \xrightarrow{P} 0 \tag{5.8}$$

the first term on the r.h.s. of (5.4) will converge in probability to zero. To establish (5.8) we first replace μ_N by μ_0 as follows. Now

$$N^{\frac{1}{2}} D_N^{-1} \left\{ \frac{\partial m_N(\theta_0, \mu_N)}{\partial \theta_j} - \frac{\partial m_N(\theta_0, \mu_0)}{\partial \theta_j} \right\} \tag{5.9}$$

has typical element

$$- \sum_{b,c,d=1}^{S}\sum\sum \left(K_N^{-1} - K_0^{-1}\right)_{de} \frac{N^{-\frac{1}{2}}}{2\pi} \int \left\{ \left[\left(k_0^{-1*}\right)_{da} \left(\frac{\partial k_0^{-1}}{\partial \theta_j}\right)_{eb} + \left(\frac{\partial k_0^{-1}}{\partial \theta_j}\right)_{da} \left(k_0^{-1}\right)_{eb} \right] \right.$$

$$\left. \cdot \sum_{m=1}^{N} x_b(m) e^{im\omega} \cdot \sum_{n=1}^{N} \tilde{y}_c(n) e^{-in\omega} \right\} d\omega . \tag{5.10}$$

But the function in square brackets appearing in the integrand has modulus bounded by $b_2 < \infty$ so that (5.10) has modulus dominated by

$b_2 \cdot G_{bb}^{(X)}(0)^{\frac{1}{2}} \cdot G_{cc}^{(Y)}(0)^{\frac{1}{2}}$ which is a.s. finite. Thus, since

$K_N \xrightarrow{a.s.} K_0$, (5.10) converges a.s. to zero and hence (5.9) converges a.s. to the null vector. Now consider

$$N^{\frac{1}{2}} D_N^{-1} \frac{\partial m_N(\theta_0,\mu_0)}{\partial \theta_j} = N^{\frac{1}{2}} D_N^{-1} \frac{1}{2\pi} \int \left(I_t \otimes \frac{\partial f_0^{-1}}{\partial \theta_j} \right) \text{Vec } I_{XY} d\omega . \tag{5.11}$$

As in the proof of Lemma 2.2, $\partial f_0^{-1}/\partial \theta_j$ may be replaced by its Cesaro sum to M terms and consequently (5.11) will converge a.s. to the null vector. Hence (5.8) holds. Hence the first term on the r.h.s. of (5.4) converges in probability to zero. The second term in (5.4) may also be shown to converge a.s. to zero by the same type of arguments as just used for the first term and noting that

$$\frac{\partial}{\partial \theta_j} \left(N D_N^{-1} M_N(\theta_0,\ \mu_N) D_N^{-1} \right)^{-1} = -\left[N D_N^{-1} M_N(\theta_0,\ \mu_N) D_N^{-1} \right]^{-1} \left[N D_N^{-1} \left(\frac{\partial}{\partial \theta_j} M_N(\theta_0,\ \mu_N) \right) D_N^{-1} \right] \cdot \left[N D_N^{-1} M_N(\theta_0,\ \mu_N) D_N^{-1} \right]^{-1} .$$

By similar methods as used to establish Lemma 2.1 the middle matrix in this may be shown to converge to $\frac{1}{2\pi} \int \overline{F}_Y(d\omega) \otimes \frac{\partial f_0^{-1}(\omega)}{\partial \theta_j}$ which has finite norm. Hence (5.4) converges in probability to zero.

(iv) Finally, in the consideration of (5.1), it may be shown by similar methods as used in (iii) above and the proof of Lemma 2.1 that

$$\frac{\partial^2 R_N(\dot{\theta}_N,\mu_N)}{\partial \theta^2} \xrightarrow{a.s.} 0 . \tag{5.12}$$

Combining (i)-(iv) it follows that

$$N^{\frac{1}{2}}(\theta_N - \theta_0) \xrightarrow{\mathcal{D}} N(0,\ \Omega^{-1}) . \tag{5.13}$$

Now consider the CLT for $D_N(\delta_N - \delta_0)$. We have

$$D_N(\delta_N-\delta_0) = \left\{ND_N^{-1}M_N(\theta_N,\ \mu_N)D_N^{-1}\right\}^{-1}.\left\{ND_N^{-1}m_N(\theta_N,\ \mu_N)\right\}\ . \tag{5.14}$$

But $\left\{ND_N^{-1}M_N(\theta_N,\ \mu_N)D_N^{-1}\right\} \xrightarrow{\text{a.s.}} \Sigma^{-1}$ as in (ii) above. Also

$$ND_N^{-1}m_N(\theta_N,\ \mu_N) = ND_N^{-1}m_N(\theta_0,\ \mu_0) + ND_N^{-1}\frac{1}{2\pi}\int I_t \otimes \left\{f_N^{-1}-f_0^{-1}\right\}(\text{Vec } I_{XY})d\omega\ . \tag{5.15}$$

But

$$ND_N^{-1}\frac{1}{2\pi}\int I_t \otimes \left\{f_N^{-1}-f_0^{-1}\right\}(\text{Vec } I_{XY})d\omega$$

$$= ND_N^{-1}\frac{1}{2\pi}\int I_t \otimes \left\{f^{-1}(\theta_N,\ \mu_N)-f^{-1}(\theta_0,\ \mu_N)\right\}(\text{Vec } I_{XY})d\omega$$

$$+ ND_N^{-1}\frac{1}{2\pi}\int I_t \otimes \left\{f^{-1}(\theta_0,\ \mu_N)-f^{-1}(\theta_0,\ \mu_0)\right\}(\text{Vec } I_{XY})d\omega\ . \tag{5.16}$$

In (iii) above it was shown that the second term converges a.s. to the null vector. Consider, therefore, the first term on the r.h.s. of (5.16). This may be rewritten as

$$\sum_{j=1}^{u}\left\{\left[N^{\frac{1}{2}}D_N^{-1}\frac{1}{2\pi}\int I_t \otimes \left(\frac{\partial f^{-1}(\dot{\theta}_N,\mu_N)}{\partial\theta_j}\right)(\text{Vec } I_{XY})d\omega\right].N^{\frac{1}{2}}(\theta_{N,j}-\theta_{0,j})\right\} \tag{5.17}$$

(where $\|\dot{\theta}_N-\theta_0\| \le \|\theta_N-\theta_0\|$). But $N^{\frac{1}{2}}(\theta_N-\theta_0) \xrightarrow{\mathcal{D}} N(0,\ \Omega^{-1})$ and so each component, $N^{\frac{1}{2}}(\theta_{N,j}-\theta_{0,j})$, has bounded variance. Again, using the methods outlined in (iii) above (see equation (5.8)), it may be shown that each term in square brackets in (5.17) converges a.s. to the null vector. Therefore (5.17), and hence (5.16), converges in probability to the null vector. This means that the CLT for $ND_N^{-1}m_N(\theta_N,\ \mu_N)$ may be obtained as the CLT for $ND_N^{-1}m_N(\theta_0,\ \mu_0)$ (see (5.15)). This latter CLT was needed in the argument leading to (5.13) and will now be established. (We will return to the remainder of the proof of Theorem 5.1 after the proof of the next theorem.)

The following theorem is just the extension of Theorem A, Hannan (1973a, Appendix) to the vector case and the procedure $\bar{L}_N$. As in that theorem it is possible to be more general and allow the $y(n)$ sequence to depend on N (so that we write $y^{(N)}(n)$ as before).

THEOREM 5.2. *Let* $y^{(N)}(n)$ *be a sequence of* t*-vectors satisfying* C1.1(a) *and* (c) *with* C1.1(b) *strengthened to*

$$\text{C5.1} \qquad \lim_{N\to\infty} \max_{1\le n\le N} \frac{\left|y_j^{(N)}(n)\right|}{d_j^{(N)}} = 0 .$$

Let $x(n)$ *be independent of* $y^{(N)}(n)$ *and satisfy*

$$x(n) = \sum_{j=0}^{\infty} C(j)\varepsilon(n-j) , \quad \sum_{j=0}^{\infty} \|C(j)\|^2 < \infty$$

where the $\varepsilon(n)$ *satisfy* 2.C3.3, *and the spectral density matrix corresponding to* $x(n)$, $f(\omega) = \frac{1}{2\pi} k\left(e^{i\omega}\right)Kk\left(e^{i\omega}\right)^*$,

$k\left(e^{i\omega}\right) = \sum_{j=0}^{\infty} C(j)e^{ij\omega}$, *is nonsingular for all* $\omega \in [-\pi, \pi]$. *Then*

$$ND_N^{-1}m_N\left(f^{-1}\right) \xrightarrow{\mathcal{D}} N\left(0, \Phi\right) ,$$

where

$$m_N\left(f^{-1}\right) = \frac{1}{2\pi}\int \left[I_t \otimes f^{-1}(\omega)\right]\left(\text{Vec } I_{XY}\right)d\omega .$$

Proof of Theorem 5.2. Let P_M be the Cesaro sum to M terms of the Fourier series for $f^{-1} + \varepsilon I_s$ and let $\Phi = P_M - f^{-1}$. Then since f^{-1} is continuous, f being continuous and nonsingular, M may be chosen so that

$$\sup_{\omega} \|\Phi(\omega)\| \le \varepsilon \quad \text{and} \quad \Phi(\omega) \ge 0 , \quad \omega \in [-\pi, \pi] . \tag{5.18}$$

Now, by a similar argument as used to establish Theorem 4.1, it (see equations (4.4)) may be shown that the variance of $\left\|ND_N^{-1}m_N(\Phi)\right\|$ is dominated by $b_1.\varepsilon^2.s.t$, where $\sup_{\omega} \|f(u)\| \le b_1$ and ε is as in (5.18). (Note that here, in contrast to the proof of Theorem 4.1, we need to consider $ND_N^{-1}m_N(f^{-1})$ and *not* $Nd_j(N)^{-1}D_N^{-1}m_N(f^{-1})$.) Hence the variance of $\|ND_N^{-1}m_N(\Phi)\|$ may be made arbitrarily small by choosing M large. This means that the CLT for $ND_N^{-1}m_N(f^{-1})$ may be obtained as the CLT for

$$ND_N^{-1}\left\{\frac{1}{2\pi}\int (I_t \otimes P_M)(\text{Vec } I_{XY})d\omega\right\}$$

$$= \text{Vec}\left\{\sum_{l=-M}^{M} P(l) \sum_{m=\max(1,1-l)}^{\min(N,N-l)} x(m+l)y^{(N)}(m)'d(N)^{-1}\right\} \quad (5.19)$$

$\left(\text{where } P_M(\omega) = \sum_{-M}^{M} P(l)e^{il\omega}\right)$.

It is sufficient to establish the CLT for any linear combination of the elements of the vector in (5.19). Letting A be the $t \times s$ matrix with a_{jk} as (j, k)th element then $(\text{Vec } A')'$ is such a linear combination so that we may consider $\left(\text{where } \tilde{y}_j^{(N)}(m) = y_j^{(N)}(m)/d_j(N)\right)$

$$\sum_{j=1}^{t}\sum_{k=1}^{s} a_{jk} \sum_{b=1}^{s}\sum_{l=-M}^{M} (P(l))_{kb} \sum_{m=\max(1,1-l)}^{\min(N,N-l)} \tilde{y}_j^{(N)}(m)x_b(m+l)$$

$$= \sum_{j=1}^{t}\sum_{k=1}^{s} a_{jk} \sum_{b=1}^{s}\sum_{l=0}^{M} (Q(l))_{kb} \sum_{m=1}^{N} \tilde{y}_j^{(N)}(m-l)x_b(m) + o(1) \quad (5.20)$$

where $Q(0) = P(0)$, $Q(l) = P(l) + P(-l)$ if $l > 0$, $\tilde{y}_j(u) = 0$ if $u \le 0$ and the $o(1)$ arises from terms which converge in probability to zero because of C5.1. Now the r.h.s. in (5.20) (apart from the $o(1)$ term) is

$$X_N = \sum_{u=-\infty}^{N} A_N(u)\varepsilon(u) \quad (5.21)$$

where

$$A_N(u) = \sum_{j=1}^{t}\sum_{k=1}^{s} a_{jk} \sum_{b=1}^{s}\sum_{l=0}^{M} (Q(l))_{kb} \sum_{m=1}^{N} \tilde{y}_j(m-l)C_b(m-u)$$

and $C_b(l)$ is the bth row of $C(l)$ (and is thus null if $l < 0$).

As in the proof of Theorem A, Hannan (1973a), X_N in (5.21) may be replaced by

$$Y_N = \sum_{u=-j_N}^{N} \eta_N(u)$$

$\left(\text{in which } \eta_N(u) = A_N(u)\varepsilon(u)\right)$ where $j_N \to \infty$ (as $N \to \infty$) in such a way as to make $E\left(X_N - Y_N\right)^2 \to 0$. Again as in Hannan (1973a) we will call $\zeta_N(u) = \eta_N\left(u-j_N\right)$ so that

$$Y_N = \sum_{u=0}^{N+j_N} \zeta_N(u) \ .$$

Now, for each fixed N , $\left\{\sum_{u=0}^{n} \zeta_N(u),\ F_{n-j_N-1},\ 0 \le n \le N+j_N\right\}$ is a martingale $\left(\text{where } F_{N-1} \text{ is the } \sigma\text{-algebra generated by the } \varepsilon(m) ,\ m \le N-1\right)$ so that Theorem 2 of Scott (1973) may be applied in a similar way as indicated in Hannan (1973a). The proof that Lindeberg's condition:

$$\lim_{N\to\infty} \sum_{u=0}^{N+j_N} \int_{|z|^2>\delta} z^2 F_{N,u}(dz) = 0 \ , \quad \delta > 0 \qquad (5.22)$$

where $F_{N,u}(z)$ is the distribution function of $\zeta_N(u)$ may be accomplished in almost the same way as in Hannan (1970, pp. 232-233). The proof that

$$\sum_{u=0}^{N+j_N} E\left(\zeta_N(u)^2 \mid F_{u-j_N-1}\right) \xrightarrow{P} c \ , \quad 0 < c < \infty \qquad (5.23)$$

may be accomplished by noting (as in Hannan (1973a)) that the quantity on the l.h.s. of (5.23) is a.s. equal to

$$\sum_{u=0}^{N+j_N} E\left(\zeta_N(u)^2\right) = \mathrm{Var}\left(Y_N\right)$$

by 2.C3.3. But the variance of Y_N converges to a nonzero and finite limit since the covariance matrix of (5.19) converges to a non-singular matrix with finite norm. The results (5.22) and (5.23) ensure Scott's

(1973) conditions. Hence Y_N is asymptotically normally distributed so that (5.19) has an asymptotic multivariate normal distribution. The asymptotic covariance matrix of (5.19) is easily established and this may be made arbitrarily close to Σ by taking M large. □

The application of this theorem to $ND_N^{-1}m_N(\theta_0, \mu_0)$ is obvious and we now return to the proof of Theorem 5.1.

CONTINUATION OF THE PROOF OF THEOREM 5.1. All that remains concerning the asymptotic distribution of $N^{\frac{1}{2}}(\theta_N-\theta_0)$, $D_N(\delta_N-\delta_0)$ is to show that they are *jointly* asymptotically normal with null asymptotic covariance between the vector $N^{\frac{1}{2}}(\theta_N-\theta_0)$ and the vector $D_N(\delta_N-\delta_0)$. Again it is sufficient to show that

$$\alpha'\left[N^{\frac{1}{2}}(\theta_N-\theta_0)\right] + \beta'[D_N(\delta_N-\delta_0)] \tag{5.24}$$

is asymptotically normally distributed for all non-null vectors α $(u \times 1)$ and β $(st \times 1)$. But by (i)-(iv) above the CLT for $\alpha'N^{\frac{1}{2}}(\theta_N-\theta_0)$ reduces to that for $\alpha'N^{\frac{1}{2}} \frac{\partial Q_N(\theta_0,\mu_N)}{\partial\theta}$ which in turn, by part (iv) of the proof of Theorem 2.3.1, reduces to that for a finite linear combination (R finite)

$$\sum_{r=1}^{R} N^{-\frac{1}{2}} \sum_{u=1}^{N} \varepsilon(u)'H(r)\varepsilon(u+r) \tag{5.25}$$

{where each $H(r)$ is an $s \times s$ constant matrix}. Equivalently (see Hannan and Heyde (1972, p. 2062)) we may consider instead

$$\sum_{r=1}^{R} N^{-\frac{1}{2}} \sum_{u=1}^{N} \varepsilon(u-r)'H(r)\varepsilon(u) = \sum_{u=1}^{N} \left\{N^{-\frac{1}{2}} \sum_{r=1}^{R} \varepsilon(u-r)'H(r)\right\}\varepsilon(u) \tag{5.26}$$

where the difference between (5. 25) and (5.26) converges in probability to zero. Calling

$$\chi_N(u) = \begin{cases} \left(N^{-\frac{1}{2}} \sum_{r=1}^{R} \varepsilon(u-j_N-r)'H(r)\right)\varepsilon(u-j_N) & , \quad j_N+1 \le u \le j_N+N , \\ 0 & , \quad u < j_N+1 \end{cases}$$

the r.h.s. of equation (5.26) may be re-written as $Z_N = \sum_{u=0}^{N+j_N} \chi_N(u)$.

Also the CLT for $\beta'\left(D_N(\delta_N-\delta_0)\right)$ reduces to that for Y_N (see the proof of Theorem 5.2) so that (5.24) will be asymptotically normally distributed if

$$W_N = aZ_N + bY_N \tag{5.27}$$

is, where a, b are non-zero constants. Now

$$\left\{\sum_{u=0}^{n} \left(a\chi_N(u)+b\zeta_N(u)\right), \ F_{n-j_N-1}, \ 0 \le n \le N+j_N\right\}$$

is a martingale for each N fixed so that Theorem 2 of Scott (1973) may again be applied provided the analogues of (5.22) and (5.23) hold here. Now letting $F_{N,u}^{(\chi,\zeta)}$ be the joint distribution of (χ, ζ) by the Cauchy-Schwarz inequality,

$$\iint_{|az+by|^2 \ge \delta} (az+by)^2 F_{N,u}^{(\chi,\zeta)}(dz, dy)$$

$$\le 2a^2 \int_{|z|^2 \ge \delta/2|a|} z^2 F_{N,u}^{(\chi)}(dz) + 2b^2 \int_{|y|^2 \ge \delta/2|b|} y^2 F_{N,u}^{(\zeta)}(dy) \tag{5.28}$$

the first term of which converges to zero by the proof of the CLT for (5.26) (see Hannan and Heyde (1972)) and the second term converges to zero by the proof of Theorem 5.2. Hence the l.h.s. of (5.28) converges to zero. The second thing that must be demonstrated is that

$$\sum_{u=0}^{N+j_N} E\left\{\left(a\chi_N(u)+b.\zeta_N(u)\right)^2 \mid F_{u-j_N-1}\right\} \xrightarrow{P} c , \quad 0 < c < \infty . \tag{5.29}$$

But the l.h.s. equals

$$a^2 \sum_{u=0}^{N+j_N} E\left(\chi_N(u)^2 \mid F_{u-j_N-1}\right) + b^2 \sum_{u=0}^{N+j_N} E\left(\zeta_N(u)^2 \mid F_{u-j_N-1}\right)$$

$$+ 2ab \sum_{u=0}^{N+j_N} E\left(\chi_N(u)\zeta_N(u) \mid F_{u-j_N-1}\right) . \quad (5.30)$$

The first term of (5.30) converges in probability to a constant by the proof of the CLT for (5.20). The second term converges in probability to a constant also, by the proof of Theorem 5.2. The third term can be shown to converge in probability to zero since the summation over r in the definition of $\chi_N(u)$ is from 1 to R and using 2.C3.3(b). Hence (5.29) holds also. Thus (5.24) is asymptotically normally distributed. That the asymptotic covariance between $N^{\frac{1}{2}}(\theta_N-\theta_0)$ and $D_N(\delta_N-\delta_0)$ is null may be seen by considering the covariance between (5.26) and Y_N. This is zero since only terms of the form $\varepsilon(n-r)'H(r)\varepsilon(n)$, *for* $r > 0$, appear in (5.26). This completes the proof of part (a) of Theorem 5.1.

Proof of (b). As in §2.3 we will put

$$\overline{K}_0 = \int k_0^{-1} I_X k_0^{-1*} d\omega .$$

Then

$$N^{\frac{1}{2}}\left(\overline{K}_N-\overline{K}_0\right)$$

$$= N^{\frac{1}{2}} . \int k_N^{-1}\Big\{-\left(d(N)^{-1}I_{YX}\right)^*\left[\left(\Delta_N-\Delta_0\right)d(N)\right]^* - \left[\left(\Delta_N-\Delta_0\right)d(N)\right].\left[d(N)^{-1}I_{YX}\right]$$

$$+ \left[\left(\Delta_N-\Delta_0\right)d(N)\right]\left[d(N)^{-1}I_Y d(N)^{-1}\right]\left[\left(\Delta_N-\Delta_0\right)d(N)\right]^*\Big\}k_N^{-1*} d\omega$$

$$+ N^{\frac{1}{2}} . \int \left(k_N^{-1} I_X k_N^{-1*} - k_0^{-1} I_X k_0^{-1*}\right) d\omega . \quad (5.31)$$

As in the proof of (b), Theorem 2.3.1, the second term on the r.h.s. of (5.31) converges in probability to the null matrix under the conditions of that theorem. The first term on the r.h.s. of (5.31) converges in probability to the null matrix by the asymptotic normality of

$D_N(\delta_N-\delta_0) = \text{Vec}[(\Delta_N-\Delta_0)d(N)]$ and the fact that $N^{-\frac{1}{2}} \int k_N^{-1}d(N)^{-1}I_{YX}k_N^{-1*}d\omega$ converges a.s. to the null vector. Now the rest of the proof of (b), Theorem 2.3.1, may be used to show that $N^{\frac{1}{2}}(\overline{K}_N-\dot{K}_N) \xrightarrow{P} 0$.

Proof of (c). We finally consider the CLT for $N^{\frac{1}{2}}(\mu_N-\mu_0)$ under the additional condition that the elements of $\varepsilon(m)$ have finite fourth moments. As in the above treatment for $N^{\frac{1}{2}}(\theta_N-\theta_0)$ consider

$$N^{\frac{1}{2}}(\mu_N-\mu_0) = -\Pi_N^{-1} \cdot N^{\frac{1}{2}} \frac{\partial \overline{L}_N(\theta_N,\theta_0)}{\partial\mu}$$

where

$$\Pi_N = \left[\frac{\partial^2}{\partial\mu_j\partial\mu_k} \overline{L}_N(\theta_N, \dot{\mu}_N)\right] , \quad \|\dot{\mu}_N-\mu_0\| \le \|\mu_N-\mu_0\|$$

and $\overline{L}_N(\theta, \mu)$ is just $\overline{L}_N(\theta, \mu, \delta)$ concentrated with respect to δ for (θ, μ) fixed. Hence

$$N^{\frac{1}{2}} \frac{\partial \overline{L}_N(\theta_N,\mu_0)}{\partial\mu} = N^{\frac{1}{2}}\left\{\frac{\partial}{\partial\mu} \log\det K(\mu)\Big|_{\mu_0} + \frac{\partial}{\partial\mu} Q_N(\theta_N, \mu)\Big|_{\mu_0}\right\} - N^{\frac{1}{2}} \frac{\partial}{\partial\mu} R_N(\theta_N, \mu)\Big|_{\mu_0} . \quad (5.32)$$

But the first term on the r.h.s. of this expression as an asymptotic normal distribution by the proof of (c), Theorem 2.3.1. Consider, therefore,

$$N^{\frac{1}{2}} \frac{\partial}{\partial\mu_j} R_N(\theta_N, \mu)\Big|_{\mu_0} = N^{\frac{1}{2}}\left\{2\left(\frac{\partial m_N(\theta_N,\mu_0)}{\partial\mu_j}\right)^* M_N^{-1}(\theta_N, \mu_0)m_N(\theta_N, \mu_0)\right.$$
$$\left. -m_N(\theta_N, \mu_0)^*M_N^{-1}(\theta_N, \mu_0)\left(\frac{\partial M_N(\theta_N,\mu_0)}{\partial\mu_j}\right)M_N^{-1}(\theta_N, \mu_0)m_N(\theta_N, \mu_0)\right\} . \quad (5.33)$$

By similar methods as those used above to demonstrate that (5.4) converged in probability to the null vector, the r.h.s. of (5.33) may also be shown to converge to the null vector. Likewise Π_N may be shown to converge a.s. Π

of Theorem 2.3.1. □

The CLT for the estimators obtained by using the $\tilde{L}_N$ procedure will now be discussed. The smoothness conditions on the elements of k and $\partial k/\partial\theta_j$ are exactly the same as those required in Corollary 2.3.2.

COROLLARY 5.3. *If, in addition to the conditions of Theorem* 5.1, *it is assumed that the elements of* $k\left(e^{i\omega};\ \theta_0\right)$ *and* $\partial k\left(e^{i\omega};\ \theta_0\right)/\partial\theta_j$, $1 \le j \le u$, *belong to* Λ_α , $\alpha > \frac{1}{2}$ *then the results of Theorem* 5.1 *apply to* $\tilde{\theta}_N$, $\tilde{\delta}_N$, $\tilde{\mu}_N$ *and* $\tilde{K}_N$ *where*

$$\tilde{K}_N = (2\pi)(N')^{-1} \sum_u k(\tilde{\theta}_N)^{-1}\{I_Z - I^*_{YZ}\tilde{\Delta}^*_N - \tilde{\Delta}_N I_{YZ} + \tilde{\Delta}_N I_Y \tilde{\Delta}^*_N\} k(\tilde{\theta}_N)^{-1*} .$$

Proof. The proof of this corollary is not markedly different from that of Theorem 5.1 and therefore the following discussion will be brief. To establish the CLT for $N^{\frac{1}{2}}(\tilde{\theta}_N-\theta_0)$ we may proceed as in the proof of Theorem 5.1, part (a) and thus consider the analogue of equation (5.1) to the present case. As in the previous proof the CLT for $N^{\frac{1}{2}}(\tilde{\theta}_N-\theta_0)$ will be that given in Corollary 2.3.2 provided it is shown that

$$N^{\frac{1}{2}} \frac{\partial \tilde{R}_N(\theta_0,\tilde{\mu}_N)}{\partial\theta} \xrightarrow{P} 0 \tag{5.34}$$

and

$$\frac{\partial^2 \tilde{R}_N(\dot{\theta}_N,\tilde{\mu}_N)}{\partial\theta^2} \xrightarrow{P} 0 \quad \text{where} \quad \|\dot{\theta}_N-\theta_0\| \le \|\tilde{\theta}_N-\theta_0\| . \tag{5.35}$$

The result (5.35) follows in a straightforward way from Lemma 2.1 and Lemma 2.2 (see parts (iii) and (iv) of the proof of Theorem 5.1). Consider (5.34). Exactly as in the proof that (5.4) converges in probability to zero the proof that (5.34) holds requires that

$$ND_N^{-1}\tilde{m}_N(\theta_0,\ \mu_0) \xrightarrow{\mathcal{D}} N(0,\ \Phi) \tag{5.36}$$

where $\sum\!\!\!\!\!\!\!\!\top$ is as in Theorem 5.1. That (5.36) is true follows in much the same way as the proof of Theorem 5.2 (see also Hannan (1973a, proof of Theorem A), and Robinson (1972, proof of Lemma 3). The remaining results of the corollary are proved by very similar arguments as those used to establish Theorem 5.1. □

The final result of this section concerns estimation via $\hat{L}_N$.

COROLLARY 5.4. *In addition to the conditions of Theorem* 5.1 *assume that the elements of* $k(e^{i\omega};\theta)$ *and* $\partial k(e^{i\omega};\theta)/\partial\theta_j$, $1 \le j \le u$, *are differentiable with respect to* ω *with these derivatives belonging to* Λ_γ , $\gamma > \frac{1}{2}$. *Assume also that* $y(n)$ *satisfies* C4.1 *for* $\alpha_j > \frac{1}{2}$, $1 \le j \le t$. *Then the results* (a) *and* (c) *of Theorem* 5.1 *apply to* $\hat{\theta}_N, \hat{\delta}_N, \hat{\mu}_N$.

Note that here, in contrast to Corollary 2.3.3, it is required that $\gamma > \frac{1}{2}$. (Also we have used γ here for what was called α in Corollary 2.3.3 to avoid confusion with the α_j .) The condition that $y(n)$ satisfy C4.1 was not required in Theorem 5.1 or Corollary 5.3. These extra conditions will be further discussed after the proof. Note also that no result concerning "$\hat{K}_N$" has been given since, as indicated before Corollary 2.3.3, it is difficult to obtain such a quantity when $\hat{L}_N$ is used.

Proof. The main difficulty here is to demonstrate that

$$ND_N^{-1}\hat{m}_N(\theta_0, \hat{\mu}_N) - ND_N^{-1}\hat{m}_N(\theta_0, \mu_0) \xrightarrow{P} 0 \qquad (5.37)$$

(compare equation (5.5)). The remainder of the proof for this corollary follows along similar lines to the proof of Theorem 5.1 above by use of Lemmas 2.1 and 2.2. Also Theorem 5.2 may be demonstrated to hold for $ND_N^{-1}\hat{m}_N(f^{-1})$ by noting that the variance of the norm of $ND_N^{-1}\hat{m}_N(\Phi)$ is

$$\operatorname{tr} E\left\{\left[d(N)^{-1}Y_N' \otimes I_s\right]\Gamma_N^{-1}(\Phi)x_Nx_N'\Gamma_N^{-1}(\Phi)\left[Y_Nd(N)^{-1} \otimes I_s\right]\right\}$$

$$= \operatorname{tr}\left\{\left[d(N)^{-1}Y_N' \otimes I_s\right]\Gamma_N^{-1}(\Phi)\Gamma_N(f)\Gamma_N^{-1}(\Phi)\left[Y_Nd(N)^{-1} \otimes I_s\right]\right\}$$

$$\leq b_1 \cdot \left(\sup_\omega \|\Phi\|\right)^2, \; b_1 < \infty$$

provided $\Phi(\omega) \gtrapprox 0$, $\omega \in [-\pi, \pi]$. This fact enables f^{-1} to be replaced by its Cesaro sum to a finite number of terms and the use of results of §2.2 and the proof of Theorem 5.2 enables Theorem 5.2 to be also established for $ND_N^{-1}\hat{m}_N(f^{-1})$. Now consider the proof of (5.37). Let σ_M, Φ_M, $\psi_N(e^{i\omega}; \theta, \mu)$ be as defined in the proof of Corollary 2.3.3 where M is chosen as N^β , $\beta < \frac{1}{2}$, $\beta(1+\gamma) = \frac{1}{2} + \varepsilon$, $\varepsilon > 0$. Then, letting

$$f_{0,N}(\omega) = \frac{1}{2\pi} k\left(e^{i\omega}; \theta_0\right)K_Nk\left(e^{i\omega}; \theta_0\right)^* \quad \text{where} \quad K_N = K(\hat{\mu}_N) ,$$

$$\sup_\omega \left\|f_{0,N}(\omega)-\psi_N\left(e^{i\omega}; \theta_0, \hat{\mu}_N\right)\right\| \leq \frac{1}{2\pi} \cdot \sup_\omega \left\{\left\|k_0-\Phi_{N^\beta}^{-1}\right\|\cdot\|K_Nk_0^*\|+\left\|\Phi_{N^\beta}^{-1}K_N\right\|\cdot\left\|k_0^*-\Phi_{N^\beta}^{-1*}\right\|\right\}$$

$$\leq b_1 \cdot N^{-(\frac{1}{2}+\varepsilon)} \text{ a.s.}, \; b_1 < \infty . \tag{5.38}$$

Letting $\chi_N = \psi_N - b_1\left(N^{-(1+\varepsilon)/2}\right)I_s$ it follows that $f_{0,N}(\omega) - \chi_N(\omega) \geq 0$ for all $\omega \in [-\pi, \pi]$ a.s. for N sufficiently large. Hence

$$\left\|-\left[d(N)^{-1}Y_N' \otimes I_s\right]\Gamma_N^{-1}(f_{0,N})x_N+\left[d(N)^{-1}Y_N' \otimes I_s\right]\Gamma_N^{-1}(\chi_N)x_N\right\|^2$$

$$\leq \left\{\left\|\left[d(N)^{-1}Y_N' \otimes I_s\right]\Gamma_N^{-1}(f_{0,N})\cdot\Gamma_N(f_{0,N}-\chi_N)\cdot\Gamma_N^{-1}(f_{0,N})\cdot\left[Y_Nd(N)^{-1} \otimes I_s\right]\right\|\right\}$$

$$\cdot \left\{x_N'\Gamma_N^{-1}(\chi_N)\cdot\Gamma_N(f_{0,N}-\chi_N)\cdot\Gamma_N^{-1}(\chi_N)x_N\right\}$$

$$\leq \left[b_1N^{-(1+\varepsilon)/2}\cdot b_2\cdot b_3\right]^2\cdot\left\|\left[d(N)^{-1}Y_N'Y_Nd(N)^{-1} \otimes I_s\right]\right\|\cdot x_N'x_N \tag{5.39}$$

where $\lambda_{\max}\left[\Gamma_N^{-1}(f_{0,N})\right] \leq b_2 < \infty$ a.s. and $\lambda_{\max}\left[\Gamma_N^{-1}(\chi_N)\right] \leq b_3 < \infty$ a.s.

But $\left\|\left[d(N)^{-1}Y_N'Y_Nd(N)^{-1} \otimes I_s\right]\right\|$ is bounded and $N^{-(1+\varepsilon)}x_N'x_N \xrightarrow{\text{a.s.}} 0$.

Hence (5.38) converges a.s. to zero. Also, for N sufficiently large,

$\psi_N\left(e^{i\omega};\ \theta_0,\ \hat{\mu}_N\right) > 0$ since $f_0(\omega) > 0$. Hence we may repeat the above argument to replace χ_N by ψ_N since $\psi_N - \chi_N$ is non-negative definite for all ω . This means that $f_{0,N}$ in $\hat{m}_N$ may be replaced by $\psi_N\left(e^{i\omega};\ \theta_0,\ \hat{\mu}_N\right)$. Now the relationship established in the proof of Lemma 2.2.5 for the inverse of $\Gamma_N\left(P^{-1}\right)$ (where P^{-1} is the spectral density matrix of an autoregressive process) may be applied to obtain $\Gamma_N^{-1}\left(\psi_N\left(\theta_0,\ \hat{\mu}_N\right)\right)$ since $\psi_N = (2\pi)^{-1}\Phi^{-1}_{N^\beta}K_N\Phi^{-1*}_{N^\beta}$ and Φ_{N^β} is a matrix of polynomials in $e^{i\omega}$ of degree N^β at most. Hence

$$ND_N^{-1}\hat{m}_N\left(\psi_N\left(\theta_0,\ \hat{\mu}_N\right)\right)$$

$$= \left(d(N)^{-1}Y_N' \otimes I_s\right)\begin{bmatrix}\Gamma^{-1}_{N^\beta}\left(\psi_N\left(\theta_0,\ \hat{\mu}_N\right)\right) & 0 \\ 0 & 0\end{bmatrix}x_N + \sum_{n=N^\beta+1}^{N} e_Y(n)'K_N^{-1}e_X(n) \tag{5.40}$$

where $e_X(n)$ is as in the proof of Lemma 2.2.5 and

$$e_Y(n)' = \sum_{j=0}^{N^\beta}\left(d(N)^{-1}y(n-j) \otimes I_s\right)\Phi_{N^\beta}(j)' \ ,$$

with $\Phi_{N^\beta}(j)$ being the coefficient matrix of $e^{ij\omega}$ in Φ_{N^β} . Now the typical element of the first st-vector on the r.h.s. of (5.40) has squared modulus (where E_k is the kth row of I_s) as

$$\left|d_j(N)^{-1}\left[y_j(1)E_k,\ \ldots,\ y_j\left(N^\beta\right)E_k\right]\Gamma^{-1}_{N^\beta}\left(\psi_{N^\beta}\left(\theta_0,\ \hat{\mu}_N\right)\right)x_{N^\beta}\right|^2$$

$$\leq b_1 \cdot d_j(N)^{-2} \cdot d_j\left(N^\beta\right)^{-2} \cdot x'_{N^\beta}x_{N^\beta}\ , \quad \left(b_1 < \infty \text{ a.s.}\right)$$

$$= O\left(N^{-\alpha_j}N^{\beta\alpha_j}N^{\beta}\right)\ . \tag{5.41}$$

But we may choose β such that $\alpha_j - \beta\left(1+\alpha_j\right) > 0$, i.e. such that

$\beta < \alpha_j/(1+\alpha_j)$ since $\alpha_j/(1+\alpha_j) > 1/3$, if $\alpha_j > \frac{1}{2}$. Also since $(\frac{1}{2}+\varepsilon)/(1+\gamma) < 1/3 + 2/3\,\varepsilon$, we may choose ε sufficiently small so that $1/3 + 2/3\,\varepsilon < \beta < \alpha_j/(1+\alpha_j)$. Then (5.41) converges a.s. to zero. Now consider the second vector on the r.h.s. of (5.40). This has typical element

$$\sum_{m=N^\beta+1}^{N} \left\{\sum_{l=0}^{N^\beta} d_j(N)^{-1} y_j(m-l) E_k \Phi_{N^\beta}(l)'\right\} K_N^{-1} \left\{\sum_{k=0}^{N^\beta} \Phi_{N^\beta}(k) x(m-k)\right\}$$

which is composed of a finite number of terms of the form

$$\left(K_N^{-1}\right)_{bc} \sum_{l=0}^{N^\beta} \sum_{k=0}^{N^\beta} \left(\Phi_{N^\beta}(l)'\right)_{ab} \left(\Phi_{N^\beta}(k)\right)_{cd} \left\{\sum_{N^\beta+1}^{N} x_d(m-k) y_j(m-l)/d_j(N)\right\} . \qquad (5.42)$$

The factor multiplying $\left(K_N^{-1}\right)_{bc}$ has variance dominated by a similar quantity as that in the last line of (5.7) except now $\Phi(j)$ is replaced by $\left(\Phi_{N^\beta}(j)'\right)_{ab}$. But $\sup_\omega \left\|\left(\Phi_{N^\beta}(\omega)\right)_{ab}\right\|$ is bounded, and $K_N \xrightarrow{a.s.} K_0$ so that $\left(K_N^{-1}\right)_{bc}$ in (5.42) may be replaced by $\left(K_0^{-1}\right)_{bc}$. Reversing the steps leading from (5.40) to (5.42) with $\left(K_0^{-1}\right)_{bc}$ replacing $\left(K_N^{-1}\right)_{bc}$ establishes (5.37). □

To close this section we note that if $\alpha_j \geq 1$ for $1 \leq j \leq t$ then the condition that $1 \geq \gamma > \frac{1}{2}$ may be replaced by $1 \geq \gamma > 0$ which is the same as in Corollary 2.3.3. The reason for this is that when $\alpha_j \geq 1$, $\alpha_j/(1+\alpha_j) > \frac{1}{2}$ and then (see below (5.41)) $\alpha_j - \beta(1+\alpha_j) > 0$ is automatically satisfied since $\beta < \frac{1}{2}$. That condition C4.1 is needed at all in the above proof arises because it is not clear whether or not the l.h.s. of (5.37) can be written in the form (5.6) directly. A different argument, to that given to establish (5.37), may not require C4.1.

REFERENCES

Auslander, L. and MacKenzie, R.E. (1963), *Introduction to differentiable manifolds*, McGraw-Hill, New York.

Berk, K.N. (1974), "Consistent autoregressive spectral estimates", *Ann. Statist.* 2, 489-502.

Billingsley, P. (1961), "The Lindeberg-Lévy theorem for martingales", *Proc. Amer. Math. Soc.* 12, 788-792.

Bloomfield, P. (1973), "An exponential model for the spectrum of a scalar time series", *Biometrika* 60, 217-226.

Caines, P.E. and Rissanen, J. (1974a), "Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes", *IEEE Transactions on Information Theory* IT-20, 102-104.

Caines, P.E. and Rissanen, J. (1974b), "Consistency of maximum likelihood estimators for ARMA processes", Control Systems Report No. 7424, Control Science and Engineering, University of Toronto.

Davies, R.B. (1973), "Asymptotic inference in stationary Gaussian time series", *Adv. Appl. Prob.* 5, 469-497.

Dunsmuir, W. and Hannan, E.J. (1976), "Vector linear time series models", *Adv. Appl. Prob.* 8.

Dzhapharidze, K.O. and Yaglom, A.M. (1974), "Asymptotically efficient estimation of the spectrum parameters of stationary stochastic processes" preprint.

Grenander, U. and Szego, G. (1958), *Toeplitz forms and their applications*, Univ. of California Press, Berkeley.

Hannan, E.J. (1969), "The identification of vector mixed autoregressive-moving average systems", *Biometrika* 56, 223-225.

Hannan, E.J. (1970), *Multiple time series*, Wiley, New York.

Hannan, E.J. (1971a), "Nonlinear time series regression", *J. Appl. Prob.* 8, 767-780.

Hannan, E.J. (1971b), "The identification problem for multiple equation systems with moving average errors", *Econometrica* 39, 751-765.

Hannan, E.J. (1973a), "The asymptotic theory of linear time series models", *J. Appl. Prob.* 10, 130-145.

Hannan, E.J. (1973b), "Multivariate time series analysis", *J. Multivariate Anal.* 3, 395-407.

Hannan, E.J. (1973c), "Central limit theorems for time series regression", *Z. Wahrscheinlichkeitstheorie verw. Geb.* 16, 157-170.

Hannan, E.J. (1974), "Time series analysis", *IEEE Transactions on Automatic Control* AC-19, 706-715.

Hannan, E.J. (1975a), "Multivariate ARMA theory", in *Essays in honour of Professor A.W. Phillips*, ed. M. Peston, Wiley, New York.

Hannan, E.J. (1975b), "The estimation of ARMA models", *Ann. Statist.* 3, 975-981.

Hannan, E.J. (1976a), "The identification and parameterisation of ARMA and state space forms", *Econometrica* 44.

Hannan, E.J. (1976b), "The asymptotic distribution of serial covariances", *Ann. Statist.* 4.

Hannan, E.J. and Heyde, C.C. (1972), "On limit theorems for quadratic functions of discrete time series", *Ann. Math. Statist.* 43, 2058-2066.

Kang, K.M. (1973), "A comparison of least squares and maximum likelihood estimation for moving average processes", paper presented at the Australian Statistical Conference, Perth.

MacDuffee, C.C. (1956), *The theory of matrices*, Chelsea, New York.

Mackey, G.W. (1974), "Ergodic theory and its significance for statistical mechanics and probability"theory", *Advances in Math.* 12, 178-268.

Marden, M. (1949), *The geometry of the zeros of a polynomial in a complex variable*, American Mathematical Society, New York.

Matsushima, Y. (1972), *Differentiable manifolds*, Marcel Dekker, New York.

Neudecker, H. (1969), "Some theorems of matrix differentiation with special reference to Kronecker matrix products", *J. Amer. Stat. Assoc.* 64, 953-963.

Neveu, J. (1965), *The mathematical foundations of the calculus of probability*, Holden-Day, San Fransisco.

Pagano, M. (1974), "Estimation of models of autoregressive signal plus white noise", *Ann. Statist.* 2, 99-108.

Rao, C.R. (1965), *Linear statistical inference and its applications*, Wiley, New York.

Robinson, P.M. (1972), "Non-linear regression for multiple time series", *J. Appl. Prob.* 9, 758-768.

Rozanov, Yu.A. (1967), *Stationary random processes*, Holden-Day, San Fransisco.

Scott, D.J. (1973), "Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach", *Adv. Appl. Prob.* 5, 119-137.

Stout, W.F. (1974), *Almost sure convergence*, Academic Press, New York.

Walker, A.M. (1964), "Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary non-deterministic time-series", *J. Austral. Math. Soc.* 4, 363-384.

Ward, L.E. (1972), *Topology*, Marcel Dekker, New York.

Whittle, P. (1951), *Hypothesis testing in time series analysis*, Almqvist and Wiksell, Uppsala.

Whittle, P. (1953), "The analysis of multiple stationary time series", *J. Roy. Statist. Soc. (B)* 15, 125-139.

Whittle, P. (1954), *A study in the analysis of stationary time series* (Wold) Appendix 2, Almqvist and Wiksell, Uppsala.

Whittle, P. (1961), "Gaussian estimation in stationary time series", *Bull. I.S.I.* 33, 1-26.

Zellner, A. and Palm, F. (1974), "Time series analysis in simultaneous equation econometric models", *J. of Econometrics* 2, 17-54.

Zygmund, A. (1959), *Trigonometric series*, Volume I, Cambridge University Press, Cambridge.