ESTIMATION AND IDENTIFICATION FOR VECTOR LINEAR TIME SERIES MODELS

by

ROBERT KOHN

A thesis submitted to the
Australian National University
for the degree of
Doctor of Philosophy,
September 1977

DECLARATION

The results of this thesis are my own, except where otherwise acknowledged. The results of sections 10 and 11 of Chapter 2, Chapter 3 and section 5 of Chapter 5 will appear in Kohn [1977c, 1977b, 1977a]. Other results in the thesis are also being prepared, or have been prepared, for publication.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor R.D. Terrell for introducing me to time series and for his faith and ecnouragement throughout the writing of this thesis. I am also indebted to Professor E.J. Hannan for his advice on several issues that arose in the course of the thesis.

I am grateful for having had the opportunity to study at the Australian National University and for being supported by an Australian Government Research Award, supplemented by the Australian National University, during the course of this study.

I would like to thank my parents for supporting me and encouraging my study through the formative high school and university years. I would like to dedicate this thesis to them.

I would also like to thank Zac for her love and help during the past hectic months.

My final thanks go to Mrs I. Kewley for her excellent typing of the manuscript.

SUMMARY

This thesis is primarily concerned with obtaining asymptotic identification and estimation results for vector linear time series models which can be parameterized by a finite and fixed number of parameters and for which a sequence of discrete, equally spaced observations is available. We note, however, that Chapter 3 presents results of a more general nature.

The main results are given in Chapters 3, 4, 7 and 8. In Chapter 3 we consider a general nonlinear multivariate time series model which can be rewritten, if necessary, in a form such that the disturbances are stationary martingale differences. We prove the strong consistency and asymptotic Normality of the Gaussian Estimators of the parameters.

Chapter 4 applies the results of Chapter 3 to a general vector linear time series model, a special case of which is the vector ARMAX model.

Chapters 7 and 8 consider the estimation of an unknown vector parameter in a multivariate regression model with stationary residuals. Applications of the general theory of Chapters 3, 4, 7 and 8 are given.

Chapter 2 gives the main identification results used in the thesis. Chapter 5 discusses some of the properties of the Gauss-Newton algorithm from a computational point of view. It also relates the algorithm to some of the estimators proposed in the literature.

In Chapter 6 we give a general method for obtaining

recursive estimators for time series models by using a Gauss-Newton linearization.

Chapter 9 obtains conditions for which the methods of estimation in Chapter 4 and Chapters 7 and 8 are equally as efficient (asymptotically) when applied to dynamic simultaneous equations models with autoregressive or moving average residuals.

CONTENTS

CHAPTER 1

INTRODUCTION

1. THE THESIS IN PERSPECTIVE

This thesis is primarily concerned with obtaining asymptotic identification and estimation results for vector linear time series models which can be parameterized by a finite and fixed number of parameters and for which a sequence of discrete, equally spaced, observations is available. We note, however, that Chapter 3 presents results of a more general nature.

The models discussed in the thesis have useful empirical applications (prediction, control, structural analysis) in Engineering and Econometrics, although it may sometimes be necessary to transform the data, for example by trend or seasonal adjustment, before analysis can take place. Data transformation is carried out to ensure that the variables we finally work with satisfy (approximately) the assumptions of the theoretical analysis.

A general procedure we use throughout the thesis is to prove, subject to regularity conditions, the consistency and asymptotic Normality of parameter estimates obtained by maximizing a Gaussian Likelihood, or an approximation to it; the proofs, however, will not require the observations to be Normally distributed.

Following Whittle [1962] we call estimates thus obtained *Gaussian Estimates* to distinguish them from Maximum Likelihood Estimates. Subject to regularity conditions, the Gaussian Estimates are *efficient* in the sense that they have the same asymptotic distribution as the Maximum Likelihood Estimates for Normally distributed observations. This definition of efficiency follows Hannan [1970, p.377].

An important feature of the thesis is the use of the Gauss-Newton algorithm to provide, subject to appropriate regularity conditions, a sequence of iterates that are strongly consistent and asymptotically efficient, and that converge for N sufficiently large, almost surely, to the Gaussian Estimate. Thus we can obtain asymptotically efficient estimators of the parameters in a fixed and finite number of steps. The Gauss-Newton algorithm has often been used in the past as a computational algorithm, which, when iterated to convergence, gives us a stationary point of the Normal Equations. The statistical properties of the individual iterates have not been discussed before, although as we will see in Chapter 5, the Gauss-Newton iterates have been used to provide efficient two-step estimators for particular models without explicit recognition being made of this fact.

Chapter 2 gives the basic identification results that will be used throughout the thesis, these results being based mainly on Hannan's [1969a, 1971a] work, with Hannan [1969a] being a special case of Hannan [1971a]. Unfortunately there are small errors in Hannan [1971a]. We amend these and also extend Wegge's [1965] classical identification result to ARMAX structures. We will also show how the ARMAX identification results can be used to identify ARMAX structures when identities are present, as well as to identify structures with autoregressive or stationary residuals.

The latter part of Chapter 2 discusses and extends some recent work on identification which has appeared in the Econometric literature. In particular, we amend and extend an identification result of Hatanaka [1976] and discuss Deistler's [1976] work on vector linear models, where he obtains identification results under conditions which appear to be weaker than those in Hannan [1971a]; we also show the link between identification and estimation, especially for the time series context, by using Sargan's [1975] concept of asymptotic identification.

In Chapter 3 we consider a general nonlinear multivariate time series model which can be parameterized by a finite and fixed number of parameters and which can be rewritten, if necessary, in a form such that the disturbances are stationary martingale differences. Subject to regularity assumptions, we prove the strong consistency and asymptotic Normality of the Gaussian Estimators of the parameters, the parameters possibly being subject to nonlinear constraints. Because the Normal Equations are usually highly nonlinear it may be difficult to obtain explicit expressions for the Gaussian Estimates. To overcome this problem we use the Gauss-Newton algorithm to obtain a sequence of iterates which converge to, and have the same asymptotic properties as, the Gaussian Estimates.

Because the approach in Chapter 3 is so general, we can expect that in any specific application we may be able to prove strong consistency under weaker assumptions on the admissible parameter set. I feel, however, that any modification of the proof, necessary to weaken the assumptions in the specific case, will be minor. The other results of the chapter only require local (in a neighbourhood of the true parameter point) conditions on the parameter set and the assumptions we make are quite weak.

The results of Chapter 3 are a natural generalization of the results in Jennrich [1969] who considered the scalar model

$$y(n) = f(n;\theta_o) = u(n) \qquad n = 1,2,\ldots$$

where the $u(n)$ are independent and identically distributed residuals with a zero mean and a finite variance, and the $f(n;\cdot)$ are known, continuous, nonstochastic functions on a compact set.

Hannan [1971b] generalized Jennrich in a different direction from us by permitting $u(n)$ to be a general stationary residual and, for a frequency domain criterion function, obtained the asymptotic properties of the Gaussian Estimator of θ_o.

Chapter 4 applies the results of Chapter 3 to the vector linear time series model

$$\sum_{j=o}^{r} B_j(\theta_o)y(n-j) = \sum_{j=o}^{\infty} C_j(\theta_o)x(n-j) + \sum_{j=o}^{\infty} A_j(\theta_o)e(n-j) \qquad n = 0,\pm 1,\ldots \qquad (1.10)$$

where the $\{y(n)\}$ are endogenous variables, the $\{x(n)\}$ are exogenous variables and the $\{e(n)\}$ are stationary martingale differences. The assumption that the $e(n)$ are martingale differences is a very natural one to make for the linear system (1.10), as then the best linear predictor of $y(n)$ given the past is the best predictor (see Remark 2.1(ii) of Chapter 4). We use the vector ARMAX model

$$\sum_{j=o}^{r} B_j y(n-j) = \sum_{j=o}^{s} C_j x(n-j) + \sum_{j=o}^{t} A_j e(n-j)$$

and the dynamic simultaneous model with autoregressive residuals to illustrate the theory.

Chapter 5 discusses some of the properties of the Gauss-Newton algorithm from a computational point of view. It also relates the Gauss-Newton algorithm to some of the estimators proposed in the literature.

In Chapter 6 we show how to obtain recursive estimators for multivariate time series models by using the Gauss-Newton linearization. This gives us a general method of obtaining recursive estimators whose nonrecursive counterparts have desirable asymptotic properties.

In Chapters 7 and 8 we consider the estimation of the unknown vector parameter θ_o appearing in the multivariate regression model

$$y(n) = \sum_{j=o}^{\infty} D_j(\theta_o)x(n-j)+v(n) \tag{1.11}$$

with $v(n)$ a stationary residual. No specific parametric structure is assumed for $v(n)$. We again show that the Gaussian Estimates are strongly consistent and asymptotically Normal and obtain the asymptotic properties of the constrained Gauss-Newton estimators. Chapter 8 also deals in detail with the application of the general theory to the linear model

$$\sum_{j=o}^{r} B_j y(n-j) = \sum_{j=o}^{s} C_j x(n-j) + u(n) \tag{1.12}$$

where $u(n)$ is a stationary residual.

The model (1.11) is a particular case of a more general continuous time model considered by Robinson [1972b, 1975]. Using a criterion function which is similar to ours, Robinson also obtains consistency and asymptotic Normality results for the parameter estimates. Because our model is more specific than Robinson's we have been able to weaken his conditions in some respects.

In Chapter 9 we take (1.12) to be an identified, dynamic simultaneous equations model with $u(n)$ either a stationary autoregression or a stationary moving average. Given that the parametric treatment of the residuals is correct, we show that an asymptotically efficient estimator of the B_j and C_k matrices which takes account of the parametric nature of the residuals will be at least as efficient (asymptotically) as, and generally more than, the corresponding estimator discussed in Chapters 7 and 8, the latter using only the stationarity property of the residuals. A necessary and sufficient condition for the two approaches to be equally as efficient is given. These results generalize previous results obtained by Espasa [1975] and Espasa and Sargan [1975].

2. BASIC DEFINITIONS

In this section we define some technical terms that are used later in the thesis.

We take $(\Omega, \mathcal{F}, P)$ as our basic probability space, with Ω a set, $\mathcal{F}$ a σ-field of subsets of Ω and P a probability measure defined on $\mathcal{F}$.

A. Martingales

Let J be an interval of the form $[m,n]$ of the set $(\ldots,-1,0,1,\ldots)$ and let $\{\mathcal{F}_n, n \in J\}$ be an increasing sequence of σ-fields such that $\mathcal{F}_n \subseteq \mathcal{F}$. If $\{X_n, n \in J\}$ is a sequence of random variables defined on Ω which satisfy: (i) X_n is $\mathcal{F}_n$

measurable; (ii) $E(\|X_n\|) < \infty$; (iii) $E(X_n | F_m) = X_m$ a.s. for all $m < n$ with $m,n \in J$,

then the sequence $\{X_n, n \in J\}$ is a *martingale* (with respect to the σ-fields $\{F_n, n \in J\}$).

If X_n is a vector, then each component of X_n is F_n measurable. Let $Y_n = X_n - X_{n-1}$ $(n \in J)$. Then $\{Y_n, F_n, n \in J\}$ is a sequence of *martingale differences*.

B. Regular Stationary Process

We follow Hannan [1973b]. Let $\{X_n, n = 0,\pm1,\ldots\}$ be a strictly stationary process. Let F_a be the σ-algebra generated by $\{X_n, n \leq a\}$ and put $\bigcap_a F_a = F_{-\infty}$. If $F_{-\infty}$ is the trivial σ-field, then X_n is a regular process.

C. Process with Orthogonal Increments

We follow Ibragimov and Linnik [1971, p.292]. A random process $Y(\lambda)$ is said to have orthogonal increments if for any values $\lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4$,

$$E[\{Y(\lambda_4)-Y(\lambda_3)\}\{Y(\lambda_2) - Y(\lambda_1)\}^*] = 0.$$

D. Miscellaneous

The *Finite Fourier Transform* (FFT) of $\{X_n, n = 1\ldots,N\}$ at the frequency λ is

$$(2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} X_n e^{in\lambda}.$$

Let A be an arbitrary $m\times n$ matrix with columns a_1 to a_n. Then, vec(A) is the vector $(a_1', \ldots, a_n')'$. Let B be an arbitrary $p\times q$ matrix. Then, $(A\otimes B)$ is the $mp\times nq$ block matrix with $(a_{ij}B)$ in the $(i,j)^{th}$ block, where a_{ij} is the $(i,j)^{th}$ element of A; $i = 1$ to m, $j = 1$ to n.

Unless stated otherwise, $\|A\|$ will stand for any proper matrix norm of A. We can make the selection of the norm arbitrary because any two matrix norms are equivalent in the following sense. If $\|\cdot\|_1$ and $\|\cdot\|_2$ are two matrix norms, then there exist positive finite constants c_1 and c_2 such that $\|\cdot\|_1 \le c_1\|\cdot\|_2 \le c_2\|\cdot\|_1$. Similarly, if b is a vector, then $\|b\|$ will stand for any proper vector norm of b.

3. NOTATION

We now list some of the symbols and abbreviations used in the text. Those symbols that are not standard are either explained now, or will be explained when they first appear in the text.

(i) Internal Referencing

§4	Section 4 of current chapter
2.§4	Section 4 of Chapter 2
(5.10)	equation number (5.10), which is equation 10 in Section 5 of current chapter
(2.5.10)	equation (5.10) of Chapter 2
Theorem 5	Theorem 5 of current chapter
Theorem 2.5	Theorem 5 of Chapter 2
	The same notation as applies to theorems, applies to lemmas and corollaries. We note that lemmas, theorems and corollaries are numbered consecutively within chapters.
Remark 6.2	Remark 2 in §6 of current chapter

(ii) Abbreviations

AR	autoregressive
ARMA	autoregressive moving average
ARMAX	autoregressive moving average with exogenous variables
a.s.	almost surely
CLT	central limit theorem
FFT	finite Fourier transform

g.c.d.	greatest common divisor
g.c.l.d.	greatest common left divisor
G.N.	Gauss-Newton
iff	if and only if
i.i.d.	independent and identically distributed
MA	moving average
ML	maximum likelihood
NSC	necessary and sufficient condition
p.d.	positive definite
p.s.d.	positive semi-definite
r.v.	random variable
s.t.	such that
w.r.t.	with respect to

(iii) Mathematical Notation

(a) Limits

$\rightarrow$	converges to
$\uparrow$	monotonically increasing convergence
$\overset{P}{\rightarrow}$	convergence in probability
$\overset{a.s.}{\rightarrow}$	convergence a.s.
$\overset{\mathcal{D}}{\rightarrow}$	convergence in distribution
$\lim_n$	limit as $n \rightarrow \infty$
$a_n = O(1)$	$\{a_n\}$ is a bounded sequence
$a_n = o(1)$	$a_n \rightarrow 0$ as $n \rightarrow \infty$

(b) Probability

E	expectation operator
$E(X\|F)$	conditional expectation of the r.v. X w.r.t. the σ-field F
$\overset{\mathcal{D}}{=}$	has the same distribution as
$N(a,\Phi)$	multivariate Normal distribution with mean a and covariance matrix Φ.
$a_n=O_p(1)$	the sequence $\{a_n\}$ is bounded in probability
$a_n=o_p(1)$	the sequence $\{a_n\}$ converges to zero in probability

(c) Vectors and Matrices

Let A be an arbitrary matrix, B a square matrix, C a Hermitian matrix and b an arbitrary vector.

$I_{(p)}$	the identity matrix of dimension p
$\lambda_{max}(C)$	maximum eigenvalue of the matrix C
$\lambda_{min}(C)$	minimum eigenvalue of the matrix C
$\mathrm{tr}\ B$	trace of the matrix B
$\det B$	determinant of the matrix B
A'	transpose of the matrix A
A^*	complex transpose of the matrix A
A^+	Moore-Penrose generalized inverse of the matrix A
$\Vert A\Vert$	any proper matrix norm of A
$\Vert b\Vert$	any proper vector norm of b
$\Vert A\Vert_S$	spectral norm of $A = \{\lambda_{max}(A'A)\}^{\frac{1}{2}}$

(d) Miscellaneous

$\square$	end of proof; end of statement
$\in$	belongs to
$[x]$	largest integer less than or equal to x
$\approx$	approximately
$c\Theta$	complement of the set Θ
$\overline{\Theta}$	closure of the set Θ
$\Longleftrightarrow$	iff

CHAPTER 2

IDENTIFICATION IN ARMAX AND RELATED STRUCTURES

1. INTRODUCTION

The first part of the chapter gives identification conditions for the simultaneous ARMAX structure

$$\sum_{j=o}^{r} B_j \; y(n-j) = \sum_{j=o}^{s} C_j \; x(n-j) + u(n), \; n = 0, \pm 1, \ldots \quad (1.10)$$

with

$$u(n) = \sum_{j=o}^{t} A_j \; e(n-j) \quad (1.11)$$

where we allow identities to be present in (1.10). As far as Econometric applications are concerned, the most important ARMAX identification results have been given by Hannan in a series of papers, Hannan [1969a, 1971a, 1976a, 1976b]. Of these, Hannan [1969a, 1971a] are the seminal, and perhaps the most important, ones with Hannan [1969a] being a special case of Hannan [1971a]. Unfortunately there are small errors in some of Hannan's [1971a] theorems; we give corrected versions of these theorems. We also generalize to the ARMAX case the classical identification result of Wegge [1965] and Rothenberg [1971, pp.587-588].

We will also discuss briefly how to obtain analogous identification results when the residuals u(n) are autoregressive, or are taken to be stationary with no specific parametric structure assumed.

The second part of the chapter looks at some other recent contributions to identification that have appeared in the Econometric literature. More specifically, we amend and extend an identification result of Hatanaka [1976]; we show

how some of the derivations in Deistler [1976] may be simplified and the realism of one of Deistler's assumptions is discussed; the implications of Sargan's [1975] concept of asymptotic identification are explored, especially for the time series case.

We have not discussed Akaike's [1974a, 1974b and 1975] (see also Hannan [1976a]) and Deistler, Dunsmuir and Hannan [1977]) important results because we mainly direct our attention to those identification conditions that preserve any prior information we may have about the structure. Akaike's just identified canonical form for reduced form ARMA models would, in general, make prior information impossible to apply.

In later chapters we will often refer to the results of this chapter.

The chapter is structured as follows. §2 states the assumptions we make for the ARMAX case; §3 defines identification terms such as model and structure, while §4 gives some results for matrix polynomials. §§5 and 6 give the main ARMAX results; §§7 and 8 discuss identification when the residuals are stationary and autoregressive respectively; §9 considers identification when identities are present; §10 extends Hatanaka's [1976] result; §11 looks at Sargan's concept of asymptotic identification while §12 discusses Deistler's work; proofs are placed in §13.

We adopt the following notation in the chapter: (1) For a function of the complex variable z, $g(z)$ say, we will write $g(e^{i\lambda})$ as $g(\lambda)$; (2) We will use z both as a complex variable and as a lag operator, e.g. $zy(n) = y(n-1)$; its usage will be clear from the context; (3) $\rho(A)$ denotes the rank of the matrix A.

2. ASSUMPTIONS FOR THE ARMAX MODEL

Put $B(z) = \sum_{j=o}^{r} B_j z^j$, $C(z) = \sum_{j=o}^{s} C_j z^j$, $A(z) = \sum_{j=o}^{t} A_j z^j$

A1. $\{e(n),\ n = 0,\pm 1,\ldots\}$ is a sequence of p dimensional i.i.d. r.v.'s having zero mean and covariance matrix Ω, where Ω is a nonsingular matrix.

A2. The q dimensional exogenous variables $\{x(n),\ n = 0,\pm 1,\ldots\}$ are stationary, ergodic, have finite second moments and are independent of the $e(n)$ sequence. For $j = 0,\pm 1,\ldots$ put

$$\Gamma_x(j) = E\{x(n)x(n+j)'\}.$$

Furthermore, we assume that the $\{x(n)\}$ have a spectral density matrix, $f_x(\lambda)$ say, which is continuous on $[-\pi,\pi]$ and nonsingular on an open subset of $[-\pi,\pi]$.

A3. The $y(n)$ are stationary.

A4.(i) The matrices B_j, C_k and A_ℓ are $p\times p$, $p\times q$ and $p\times p$ respectively, where $0 \le j \le r$, $0 \le k \le s$ and $0 \le \ell \le t$;

(ii) $\det B(z) \neq 0$ for $|z| \le 1$;

(iii) $\det A(z) \neq 0$ for $|z| < 1$. □

To simplify the discussion we have imposed the simplest, and therefore also the strongest, assumptions possible on the $e(n)$, $x(n)$ and $y(n)$. As we shall see below, as long as we can obtain $B(\lambda)^{-1}C(\lambda)$ and $B(\lambda)^{-1}A(\lambda)\Omega A(\lambda)^*B(\lambda)^{-1*}$ uniquely from one complete realization, any other assumptions (see for example 4.§2) will do just as well.

3. IDENTIFICATION DEFINITIONS

For $j = 0,\pm1,\ldots$ we put $\Gamma_y(j) = E\{y(n)y(n+j)'\}$, $\Gamma_{yx}(j) = E\{y(n)x(n+j)'\}$. For most of the chapter, our identification results are based on a knowledge of the second order moments and cross-moments of the observed variates, i.e. on a knowledge of $\{\Gamma_y(j), \Gamma_{yx}(j), \Gamma_x(j);\ j = 0,\pm1,\ldots\}$ or equivalently on $f_x(\lambda)$, $f_{yx}(\lambda)$, $f_y(\lambda), \lambda \in (-\pi,\pi)$, where $f_y(\lambda)$ is the spectral density of $y(n)$ and $f_{yx}(\lambda)$ is the cross-spectral density of y and x.

It seems useful to define what we mean by terms such as structure, model, identification and observational equivalence, for the following reasons. (i) In much of the Econometric literature identification is defined through the likelihood, e.g. Hurwicz [1950], Koopmans *et al.* [1950], Rothenberg [1971]; (ii) in the time series context integer valued, as well as continuously varying, parameters are present; they describe the lag lengths and do matter.

The approach below, with a structure defined first and then a model, is in the spirit of Hurwicz [1950]. Some of our definitions are quite similar to those in Deistler [1976, p.32, §3].

We construct an ARMAX structure by choosing non-negative integer values for r, s and t and then alloting numerical values to elements of the matrices B_j, C_k, A_ℓ and Ω, where $0 \leq j \leq r$, $0 \leq k \leq s$, $0 \leq \ell \leq t$. *Unless stated otherwise, we will always assume that the above choice is made in such a way that* Ω *is p.d. and assumption A4. is satisfied.*

For our purposes the properties of the $x(n)$, $y(n)$ and $e(n)$ are taken as given and so do not enter our definitions of structure or model.

By a *model* we mean a class of structures satisfying certain prescribed conditions. We denote a structure by S, a model by M and let G be the class of all ARMAX structures. Let F be the class of all second order moments $\{\Gamma_{yx}(j), \Gamma_y(j), j = 0,\pm1,\ldots\}$ generated by members of G with F a typical member of F. It is easy to see that a given S determines F uniquely. We take ϕ to be the mapping of G onto F.

Two structures S_1 and S_2 will be called *observationally equivalent* if $\phi(S_1) = \phi(S_2)$. The *equivalence classes* of a model M are defined as $\{\phi^{-1}(F) \cap M : F \in F\}$, where $\phi^{-1}(F) \cap M = \{S \in M : \phi(S) = F\}$ Some of these model equivalence classes may be empty.

A model is *globally* - (i) *identified*, if each of its equivalence classes has at most one member in it; (ii) *over-identified*, if it is globally identified and there exists at least one empty equivalence class; (iii) *just identified*, if it is identified, but not overidentified; (iv) *under-identified*, if it is not identified.

A structure S, belonging to the model M, is *globally identified* if it is the sole member of its model equivalence class.

We can look upon an ARMAX structure as a point in some suitably dimensioned Euclidean space, i.e., given r, s and t, we put $\theta = \text{vec}\{B_0,\ldots,B_r, C_0,\ldots,C_s, A_0,\ldots,A_t, \Omega\}$, and regard θ as representing the structure. This leads to the following definition. A structure S, belonging to the model M, is locally identified if there is a neighbourhood of S such that no other structure belonging to M and observationally equivalent to S lies in the neighbourhood.

4. RESULTS FOR MATRIX POLYNOMIALS

The following results, which are well known and not our own, will be used in later sections.

$D(z)$ is called a $p\times q$ *matrix polynomial of degree* r if it can be written as

$$D(z) = \sum_{j=o}^{r} D_j z^j$$

where the D_j are constant $p\times q$ dimensional matrices.

For the rest of this section we suppose that $H(z)$, $H_1(z)$, $H_2(z)$ are $p\times p$ matrix polynomials having determinants that are not identically zero and $K(z)$, $K_1(z)$ and $K_2(z)$ $p\times q$ matrix polynomials.

A $p\times p$ matrix polynomial $U(z)$ is called *unimodular* if there exists another $p\times p$ matrix polynomial, $V(z)$ say, such that $U(z)V(z) = I_{(p)}$.

The following result (e.g. MacDuffee [1956, p.30, Theorem 20.1]) gives NSC's for a matrix to be unimodular.

<u>*Lemma 1*</u> Each of the following is a NSC for $U(z)$ to be a unimodular matrix.

(i) det $U(z)$ is a nonzero constant.

(ii) $U(z)^{-1}$ exists for any z and is a matrix polynomial. For a proof see MacDuffee. □

An example of a unimodular matrix is

$$U(z) = \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} \quad \text{with} \quad U(z)^{-1} = \begin{pmatrix} 1 & -z \\ 0 & 1 \end{pmatrix}, \ \det U(z)=1.$$

The $p\times p$ matrix polynomial $L(z)$ is called a *common left divisor* (c.l.d.) of $H(z)$ and $K(z)$ if there exist matrix polynomials $H_1(z)$ and $K_1(z)$ such that

$$\{H(z), K(z)\} = L(z)\{H_1(z), K_1(z)\}$$

$L(z)$ is called a *greatest common left divisor* (g.c.l.d.) of $H(z)$ and $K(z)$ if for any other c.l.d. of $H(z)$ and $K(z)$, $L_1(z)$ say, there exists a matrix polynomial $P(z)$ such that $L(z) = L_1(z)P(z)$; i.e. $L_1(z)$ is a left divisor of $L(z)$.

By MacDuffee [1956, p.35, Corollary 23.12] if $H(z)$ and $K(z)$ have a g.c.l.d. $L(z)$, then every g.c.l.d. of $H(z)$ and $K(z)$ is of the form $L(z)U(z)$ where U is unimodular.

We will say that $H(z)$ and $K(z)$ are *relatively left prime (or left coprime)* if all their g.c.l.d.'s are unimodular.

<u>*Lemma 2*</u> $H(z)$ and $K(z)$ are left coprime iff the rank of $[H(z), K(z)]$ is p for any z in the complex plane. For a proof of this result see Rosenbrock [1970, p.71, Theorem 6.1(i)]. □

The next lemma is due to Hannan [1969a]. Its proof is given implicitly in the proof of Theorem 1 of that paper. See also Deistler [1976, Theorem 1, pp.35-36].

<u>*Lemma 3*</u> Suppose that $[H_2(z), K_2(z)] = U(z)[H_1(z), K_1(z)]$ ($\Rightarrow U(z) = H_2(z)H_1(z)^{-1}$). Then: (i) If $H_1(z)$ and $K_1(z)$ are left coprime, then $U(z)$ is a matrix polynomial;

(ii) if, in addition, $H_2(z)$ and $K_2(z)$ are also left coprime, then $U(z)$ is unimodular. □

Lemma 4 If $H(0)$ is nonsingular, then there exists a unimodular matrix $U(z)$ such that $U(z)H(z)$ is lower triangular with the degree of the (j,j) element of $U(z)H(z)$ greater than the degree of the $(i,j)^{th}$ element $(i > j)$ and with $U(0)H(0) = I_{(p)}$.

The proof follows from MacDuffee [1956, p.32, Theorem 22.1]. □

Lemma 5 Suppose that $k_1(z)$ and $k_2(z)$ are two $p \times p$ matrices with elements that are rational functions of z and have real coefficients. The determinants of $k_1(z)$ and $k_2(z)$ have no poles in or on the unit circle, and no zeros inside the unit circle. If Ω_1, Ω_2 are two $p \times p$ positive definite matrices, $k_1(0) = k_2(0) = I_{(p)}$ and $k_2(\lambda)\Omega_2 k_2(\lambda)^* = k_1(\lambda)\Omega_1 k_1(\lambda)^*$ for $\lambda \in [-\pi, \pi]$, then $k_1(z) = k_2(z)$ $\forall$ z and $\Omega_1 = \Omega_2$.

For a proof of this lemma see Hannan [1970, p.163, Theorem 1"]. □

5. IDENTIFICATION FOR ARMAX STRUCTURES

For convenience we denote Hannan [1971a] by H. In this section we will give amended versions of Theorems 3, 4 and 5 of H and modify Theorem 1 of H to make it more useful for estimation. All identification results will be global and we will not mention this again. Throughout we assume that Assumption A4. holds and Ω is p.d. For convenience we denote the ARMAX structures S_1 and S_2 as

$$S_j := [B_j(z), C_j(z), A_j(z), \Omega_j], \; j \quad j = 1,2.$$

The following lemma gives NSC for two ARMAX structures to be observationally equivalent.

Lemma 6 (i) and (ii) are equivalent NSC's for the two ARMAX structures S_1 and S_2 to be observationally equivalent.

(i) $B_1(\lambda)^{-1}C_1(\lambda) = B_2(\lambda)^{-1}C_2(\lambda)$ and

$$B_1^{-1}(\lambda)A_1(\lambda)\Omega_1 A_1(\lambda)^* B_1^{-1}(\lambda)^* = B_2^{-1}(\lambda)A_2(\lambda)\Omega_2 A_2(\lambda)^* B_2(\lambda)^{-1*}$$

for $\lambda \in [-\pi,\pi]$.

(ii) There exists a $p\times p$ matrix $U(z)$, with elements that are rational functions of z, and a $p\times p$ constant non-singular matrix F such that $[B_2(z),C_2(z),A_2(z)] = U(z)[B_1(z),C_1(z),A_1(z)F]\ \forall\ z$ and $\Omega_1 = F\Omega_2 F'$.

(i) is due to H (eqns (6) and (7) on p.754); (ii) is given in Deistler [1976, Proposition 2, p.35 and Lemma 2, p.41] but his assumptions on the exogenous variables, and hence his method of proof, are quite different from ours. □

The next theorem is a minor modification of Theorem 1 of H. We replace Hannan's requirement that $B_o = A_o$ by $A_o = I_{(p)}$. The practical advantage of this is that there is now no longer any need to impose the restriction $B_o = A_o$ when estimating, in the case that $B_o \neq I_{(p)}$.

Theorem 1 Suppose that the ARMAX model is defined by the following conditions: (i) $B(z)$ is lower triangular with diagonal elements $b_{jj}(z)$ satisfying $b_{jj}(0) = 1$ and the off-diagonal elements satisfying either:

(a) $\deg\{b_{ij}(z)\} < \deg\{b_{jj}(z)\}$ $(i > j)$, or:

(b) $\deg\{b_{ij}(z)\} \leq \deg\{b_{jj}(z)\}$ $(i > j)$ and $B(0) = I_{(p)}$;

(ii) $A(0) = I_{(p)}$; (iii) $A(z)$, $B(z)$ and $C(z)$ are left coprime.

Then the model is just identified. □

The next example shows that the conditions of Theorems 3 and 4 of H are insufficient to ensure identification.

Example 1 Consider the two ARMAX structures $S_1 = [I_{(p)}, C, I_{(p)}, I_{(p)}]$ and $S_2 = [I_{(p)}, C, P, I_{(p)}]$, where C is an arbitrary constant matrix and P is an orthogonal matrix. Both S_1 and S_2 satisfy the conditions of Theorems 3 and 4 of H and their observational equivalence can be deduced from Lemma 6(ii). □

It is interesting to investigate whether the imposition of the extra condition $A(0) = I_{(p)}$ is sufficient to ensure identification. Although the example above is no longer relevant, the following example shows that there are still insufficient conditions for identification.

Example 2 Let $P = (p_1,p_2,p_3)$, where $p_1' = \frac{1}{2}(\sqrt{2}, 1, 1)$, $p_2' = \frac{1}{2}(\sqrt{2},-1,-1)$ and $p_3' = 2^{-\frac{1}{2}}(0,1,-1)$; let $Q = (q_1,q_2,q_3)$ where $q_1 = p_1 + (\sqrt{2}-1)p_2 =$ $= \{1, (2-\sqrt{2})/2,(2-\sqrt{2})/2\}'$, $q_2 = p_2 + 3/\sqrt{2}\, p_3 = (\sqrt{2}/2,1,-2)'$ and $q_3 = \sqrt{2}(\sqrt{2}+1)p_1 + p_3 = (\sqrt{2} + 1, \sqrt{2} + 1,1)'$. It can be easily verified that P is an orthogonal matrix and

$$P'Q = \begin{pmatrix} 1 & 0 & 2+\sqrt{2} \\ \sqrt{2}-1 & 1 & 0 \\ 0 & 3/\sqrt{2} & 1 \end{pmatrix};$$

det $P'Q = 4$ so that $P'Q$ is nonsingular. We define the two ARMAX structures S_1 and S_2 as $S_1 = (Q,C,I_{(p)},I_{(p)})$ and $S_2 = (P'Q, P'C, I_{(p)}, I_{(p)})$ where C is an arbitrary constant matrix. S_1 and S_2 certainly satisfy the conditions of Theorems 3 and 4 of H and their observational equivalence can be deduced from Lemma 6(ii). □

The following theorem provides two possible amendments of Theorem 3 of H.

Theorem 2A Put $T(z) = [B(z), C(z), A(z)]$ and define the model by the following conditions: (i) In each row of $T(z)$ we prescribe at least $(p-1)$ zero elements; (ii) 1 is the greatest common divisor of the elements of each row of $T(z)$; (iii) the diagonal elements of $B(0)$ are unity; (iv) $B(0)=A(0)$.

A NSC that a structure belonging to this model is identified is: C1. for each row, the matrix consisting of the columns of $T(z)$ having prescribed null elements in that row is of rank $(p-1)$.

Theorem 2B Let $T(z)$ be redefined as $[B(z), C(z)]$. We define the model by: (i), (ii) and (iii) are the same as in (A); (iv) $A(0) = I_{(p)}$. A NSC for a structure belonging to the model to be globally identified is C1. of Theorem 2A. There are ARMAX structures that are not observationally equivalent to any structure of the form just described. □

In the following theorems, and Theorems 5 and 6 of §6, we will assume that the degrees of $B(z)$, $C(z)$ and $A(z)$ are prescribed as r, s and t respectively.

The next theorem gives two possible amendments of Theorem 4 of H.

Theorem 3A Put $N = (B_o, \ldots, B_r, C_o, \ldots, C_s, A_o, \ldots, A_t)$. We define the model by: (i) at least $(p-1)$ zero elements are prescribed in each row of N; (ii) the diagonal elements of B_o are unity; (iii) $B_o = A_o$.

Sufficient conditions for a structure belonging to the model to be identified are: C1. $A(z)$, $B(z)$ and $C(z)$ are left coprime; C2. $[B_r, C_s, A_t]$ has rank p; C3. the rank of each submatrix of N obtained by taking the columns of N with prescribed zeros in a certain row is $(p-1)$.

Let M' be the class of all structures belonging to the model and satisfying C1. and C2.; C3. is then also a necessary condition for a structure belonging to M' to be identified.

<u>*Theorem 3B*</u> We define N as $(B_o,\ldots,B_r, C_o,\ldots,C_s)$ and let the model be defined by: (i) and (ii) are the same as in part A; (iii) $A_o = I_{(p)}$.

Sufficient conditions for a structure belonging to the model to be identified are the same as in part A. □

Theorems 2 and 3 indicate that if $B_o \neq I_{(p)}$ and we want to use restrictions on the A_j's for identification, then, in general, we must impose the constraint $B_o = A_o$ as well.

It will often be the case in the Econometric context that the matrix $[B_r, C_s, A_t]$ will be very sparse and hence not of full row rank. Therefore, condition C2. of Theorems 3A and 3B will not be satisfied. To cope with this situation we give an alternative version of Theorem 3, based on the results of Theorem 2 of H.

<u>*Theorem 3'*</u> Put $T(z) = [B(z), C(z), A(z)]$. In Theorems 3A and 3B, instead of prescribing the degrees of $B(z)$, $C(z)$ and $A(z)$ we will have instead: (0') At least p components are chosen from the vectors $y(n)$, $x(n)$ and $e(n)$, and for each the maximum lag is prescribed; e.g. if $y_1(n)$

is chosen, then the degree of $(b_{11}(z),\ldots,b_{p1}(z))'$ is prescribed.

Let N be the matrix of coefficients of these maximum lags; e.g. if $y_j(n)$ is specified to have a maximum lag ℓ, then the j^{th} column of B_ℓ is included in N.

We also replace condition C2. in Theorems 3A and 3B by:
C2.' rank $(N) = p$.

If the rest of the conditions of Theorems 3A and 3B remain as they are, then the results of both parts of Theorem 3 still hold. □

A similar result may be obtained by using condition $(iii)_4$ on p.758 of H instead of prescribing the degrees of $B(z)$, $C(z)$ and $A(z)$.

Put $K(z) = B(z)^{-1}A(z)$, $D(z) = B(z)^{-1}C(z)$. By assumption A4. we can write $K(z)$ and $D(z)$ as

$$K(z) = \sum_{j=o}^{\infty} K_j z^j, \quad D(z) = \sum_{j=o}^{\infty} D_j z^j \tag{5.10}$$

for $|z| \leq 1$.

Theorem 5 of H is motivated by rewriting (1.10) and (1.11) as

$$y(n) = \sum_{j=o}^{\infty} D_j \, x(n-j) + \sum_{j=o}^{\infty} K_j \, e(n-j).$$

The next example shows that the conditions of Theorem 5 of H are insufficient to ensure identification.

Example 3 Put $p = 3$, $r = s = t = 0$ and let the matrices P and Q be defined as in Example 2. We define the two ARMAX structures S_1 and S_2 by $S_1 = (Q'^{-1}, Q'^{-1}, I_{(3)}, I_{(3)})$, $S_2 = (P'Q'^{-1}, P'Q'^{-1}, I_{(3)}, I_{(3)})$. Hence $K_1(z) = Q'$,

$D_1(z) = I_{(3)} = D_2(z)$, $K_2(z) = Q'P$ so that $K_1(0)$ and $K_2(0)$ have 1's on the diagonal. If $D(0)$ is prescribed as diagonal, then $D_1(0)$ and $D_2(0)$ certainly satisfy this condition and thus there are two zero elements in each row of the matrix A which is given in the statement of Theorem 5 of H. By Lemma 6(ii) S_1 and S_2 are observationally equivalent and it is evident that they both satisfy the conditions of Theorem 5 of H. □

The following is an amended version of Theorem 5 of H.

Theorem 4 Put $N = (K_o', K_1', K_2', \ldots)$, where the K_j, $j \geq 0$, are defined by (5.10). We define the model by: (i) at least $(p-1)$ zero elements are prescribed in each row of N; (ii) one element in each row of K_o' is prescribed as unity; (iii) $A_o = I_{(p)}$; (iv) $B(z)$, $C(z)$ and $A(z)$ are left coprime.

The NSC's for a structure belonging to the model to be globally identified are: C1. $[B_r, C_s, A_t]$ has rank p; C2. the rank of each submatrix of N obtained by taking the columns of N with prescribed zeros in a certain row is $(p-1)$. Unlike Theorem 5 of H, our conditions (i) - (iv) do not involve the D_j. □

6. SOME FURTHER ARMAX IDENTIFICATION RESULTS

The vector ARMAX model will usually contain a large number of moving average parameters. For certain data sets a more parsimonious representation may be adequate; e.g. $B(z)$ or $A(z)$ may be taken as diagonal. Identification with $B(z)$ diagonal is an easy consequence of Theorem 1; for the special case where all the diagonal elements of $B(z)$ are equal (usually called scalar identification), identification has been discussed by Hannan [1976a, p.717]. See also Zellner and Palm [1974] for an application. A drawback of taking $B(z)$ diagonal is that if we have some information on the structural coefficients B_j, $0 \leq j \leq r$, then we will almost certainly not be able to use it if $B(z)$ is chosen canonically as diagonal.

The next theorem gives sufficient conditions for the identification of a structure when $A(z)$ is diagonal.

Theorem 5 Put $N = (B_o, \ldots, B_r, C_o, \ldots, C_s)$. We define the model by: (i) for each row of the matrix N there is prescribed at least one zero element; (ii) the diagonal elements of B_o are unity; (iii) $A_o = I_{(p)}$.

Sufficient conditions for a structure belonging to the model to be globally identified are: C1. $A(z)$, $B(z)$ and $C(z)$ are left coprime; C2. (B_r, C_s, A_t) has rank p; C3. $a_i(z) \neq a_j(z)$, $i,j = 1,\ldots,p$, where $a_i(z)$ is the i^{th} diagonal element of $A(z)$; C4. In each row of N there is at least one prescribed zero element such that there is no other zero element in the column to which this zero element belongs.□

We now give an example of an identified structure which satisfies the conditions of Theorem 5.

Example 4 Let $p = 3$, $r = 1 = t$, $s = 0$ and put

$$A(z) = \begin{pmatrix} 1+\frac{1}{2}z & 0 & 0 \\ 0 & 1+\frac{1}{3}z & 0 \\ 0 & 0 & 1+\frac{1}{4}z \end{pmatrix}, \quad B(z) = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} z,$$

$$C(z) = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \quad \Omega = I_{(3)}.$$ The structure certainly

satisfies (i) - (iii) and conditions C3. and C4. Because $\det C(z) \not\equiv 0$ it follows that C2. is satisfied. By Lemma 2, C1. is also satisfied. □

The next theorem extends the Wegge [1965, Theorem II, p.71] and Rothenberg [1971, Theorem 9, p.588] results to the ARMAX case. Because nonlinear restrictions are allowed the identification will be local.

Theorem 6 Put $\Delta = (B_o,\ldots,B_r, C_o,\ldots,C_s)$, $\delta = \text{vec}\,\Delta$, $\alpha = \text{vec}(A_1,\ldots,A_t)$, $\omega = \text{vec}\,\Omega$ and $\theta = \text{vec}(\delta,\alpha,\omega)$. We define the model by the following conditions: (i) $A_o = I_{(p)}$; (ii) θ satisfies the restrictions $\psi(\theta) = 0$, where ψ is an $R\times 1$ vector function which is symmetric with respect to ω; i.e. $\psi(\delta,\alpha,\omega) = \psi(\delta,\alpha,L\omega)$ with L the permutation matrix defined by $\text{vec}(F') = L\,\text{vec}\,F$ for any $p\times p$ matrix F.

(a) Suppose that ψ is continuously differentiable in a neighbourhood of the structure S_o, where S_o belongs to the model. Put $\Psi_\delta = \partial\psi/\partial\delta'$, $\Psi_\alpha = \partial\psi/\partial\alpha'$, $\Psi_\omega = \partial\psi/\partial\omega'$,

$$J'_\delta = (\Delta \otimes I_{(p)}),\quad J'_\alpha = (A_1 \otimes I_{(p)},\ldots, A_t \otimes I_{(p)}) - (I_{(p)} \otimes A'_1,\ldots,I_{(p)} \otimes A'_t).$$

The following conditions are sufficient for the structure S_o to be locally identified: C1. $A(z)$, $B(z)$ and $C(z)$ are left coprime; C2. (B_r,C_s,A_t) has rank p; C3. the rank of the matrix $G(\theta) = \Psi_\delta J_\delta + \Psi_\alpha J_\alpha + 2\,\Psi_\omega(\Omega \otimes I_{(p)})$ is p^2 at S_o.

(b) Let M' be the subclass of the model containing all those structures satisfying conditions C1., C2. and $\det A(z) \not\equiv 0$ for $|z| \leq 1$. If $S_o \in M'$ and is a regular point of the matrix $G(\cdot)$, i.e. $G(\cdot)$ has constant rank in a neighbourhood of S_o (see the comment made at the top of p.77), then C3. is also a necessary condition for S_o to be locally identified.

(c) If Ψ_δ is a constant matrix and ψ is independent of α and ω, then we can replace the words 'local identification' by 'global identification' in (a) and (b) above. □

The next lemma is used in the proof of Theorem 6. Because of its independent interest we give a statement of it here rather than in §13.

Lemma 7 For r and s positive integers define the permutation matrix $L(r,s)$ by $\text{vec } R' = L(r,s) \text{ vec } R$ where R is an arbitrary $r\times s$ matrix. Obviously $L(r,s)'L(r,s) = I_{(rs)}$ and $L(r,s)' = L(s,r)$. Suppose that A is an $m\times n$ matrix with columns a_i, $i = 1,\dots,n$, and rows α_i', $i = 1,\dots,m$ and B is a $p\times q$ matrix with columns b_i, $i = 1,\dots,q$, and rows β_i', $i = 1,\dots,p$.

$$\text{Then, } L(p,m)(A\otimes B) = (b_1\otimes a_1,\dots,b_q\otimes a_1,\dots,b_1\otimes a_n,\dots,b_q\otimes a_n)$$
$$= (\alpha_1\otimes\beta_1,\dots,\alpha_m\otimes\beta_1,\dots,\alpha_1\otimes\beta_p,\dots,\alpha_m\otimes\beta_p)',$$

and
$$L(p,m)(A\otimes B)L(n,q) = B\otimes A. \quad \square$$

Remark 6.1 (i) I wish to thank Dr Tony Hall for pointing out an error in the original proof of Lemma 7.

(ii) We now give an application of Lemma 7. Put $f = \text{vec}(AB'CBD)$, $\beta = \text{vec } B$ where A is $m\times n$, B is $r\times n$, C is $r\times r$ and D is $n\times s$.

Then,

$$f = \{(CBD)'\otimes A\}L(r,n)\beta = (D\otimes AB'C)\beta = L(s,m)\{A\otimes (CBD)'\}\beta$$

by Lemma 7. Therefore

$$\partial f/\partial\beta' = L(s,m)\{A\otimes (CBD)'\} + \{D\otimes AB'C\} \tag{13.11}$$

By using Lemma 7 we can obtain an explicit expression for the first term on the right in (13.11). □

7. IDENTIFICATION WITH STATIONARY RESIDUALS

In this section we show how the ARMAX identification results of the previous two sections can be specialized to obtain identification conditions for (1.10) when $u(n)$ is a general stationary process with no particular parameterization of it assumed. We assume that assumptions A2., A3. and A4. (i) and (ii) hold; A4.(iii) becomes irrelevant. A1. is replaced by: A1.' $\{u(n),\ n = 0,\pm 1,\ldots\}$ is a second order stationary process with a spectral density matrix $f_u(\lambda)$. We assume that $\{x(n)\}$ is uncorrelated with $\{u(m)\}$.

The definition of a structure will be similar to the ARMAX case, except that it will now involve $B(z)$, $C(z)$ and f_u, rather than $B(z)$, $C(z)$, $A(z)$ and Ω. All other identification definitions given in §3 remain relevant. We denote the structures S_1 and S_2 by

$$S_j := [B_j(z),\ C_j(z),\ f_u^{(j)}],\ j = 1,2.$$

Because $B(\lambda)f_{yx}(\lambda) = C(\lambda)f_x(\lambda)$ and $B(\lambda)f_y(\lambda)B(\lambda)^* = C(\lambda)f_x(\lambda)C(\lambda)^* + f_u(\lambda)$, it follows readily that $B(\lambda)^{-1}C(\lambda)$ and $B(\lambda)^{-1}f_u(\lambda)B(\lambda)^{-1*}$ will be uniquely determined from a complete realization. Therefore the analogue of Lemma 6(ii) is (Deistler [1975, p.20, Lemma 3]):

Lemma 8 The following is a NSC for the two structures S_1 and S_2 to be observationally equivalent: There exists a $p\times p$ nonsingular matrix $U(z)$ with elements that are rational functions of z so that

$[B_2(z), C_2(z)] = U(z)[B_1(z), C_1(z)] \; \forall \; z$ and

$f_u^{(2)}(\lambda) = U(\lambda) f_u^{(1)}(\lambda) U(\lambda)^*, \; \lambda \in [-\pi, \pi]$.

The proof of the lemma is immediate and will not be given. □

If we are willing to only use information on the $B(z)$ and $C(z)$ matrices, then an obvious specialization of the ARMAX results yields identification conditions for the stationary case. Since we will rarely have more additional information on f_u than that contained in assumption A1.', using only $B(z)$ and $C(z)$ for identification involves little loss of generality.

We illustrate the correspondence between the ARMAX and stationary cases by obtaining an analogue of Theorem 1.

Theorem 7 We define the model by the following conditions: (i) $B(z)$ is lower triangular with diagonal elements $b_{jj}(z)$ satisfying $b_{jj}(0) = 1$ and the off diagonal elements satisfying either: (a) $\deg\{b_{ij}(z)\} < \deg\{b_{jj}(z)\}$ $(i > j)$, or: (b) $\deg\{b_{ij}(z)\} \leq \deg\{b_{jj}(z)\} (i > j)$ and $B(0) = I_{(p)}$; (ii) $B(z)$ and $C(z)$ are left coprime.

Then the model is just identified. Theorem 7 follows immediately from Theorem 1 and a proof will not be given. □

Hatanaka [1975] obtains an analogue of Theorem 2B for the stationary case using a completely different method of proof to that given for Theorem 2B. Given the above remarks, Hatanaka's result is just a simple corollary of Theorem 2B.

8. IDENTIFICATION WITH AUTOREGRESSIVE RESIDUALS

We now briefly discuss how the ARMAX identification results may be specialized to give identification conditions for the dynamic model (1.10) when the residuals $u(n)$ are autoregressive, i.e.

$$\sum_{j=o}^{t} A_j\ u(n-j) = e(n) \tag{8.10}$$

For convenience we call (1.10) coupled with (1.11) an ARMAX (r,s,t) structure and (1.10) coupled with (8.10) an ARARX (r,s,t) structure. Assumptions A1. - A4. stay as they are except that A4.(iii) is strengthened to: A4.(iii') $\det A(z) \neq 0$ for $|z| \leq 1$. The definitions of model, structure, etc. given in §3 will apply here as well. Because an ARARX (r,s,t) structure is a special type of ARMAX (r+t,s+t,0) structure, the following lemma is an immediate consequence of Lemma 6.

Lemma 9 (i) and (ii) are equivalent NSC's for the ARARX structures S_1 and S_2 to be observationally equivalent.

(i) $B_1(\lambda)^{-1}\ C_1(\lambda) = B_2^{-1}(\lambda)C_2(\lambda)$ and

$\{A_1(\lambda)B_1(\lambda)\}^{-1}\Omega_1\{A_1(\lambda)B_1(\lambda)\}^{-1*}=\{A_2(\lambda)B_2(\lambda)\}^{-1}\Omega_2\{A_2(\lambda)B_2(\lambda)\}^{-1*}$

for $\lambda \in [-\pi,\pi]$.

(ii) There exists a $p\times p$ matrix $U(z)$, with elements that are rational functions of z, and a constant nonsingular matrix F such that $[B_2(z),C_2(z)]=U(z)[B_1(z),C_1(z)]$, $A_2(z)U(z) = F\ A_1(z)$ and $\Omega_2 = F\Omega_1F'$. □

(i) can be deduced from Lemma 6(ii) while (ii) follows easily from Lemma 5. (ii) has also been obtained by

Deistler [1976] but his assumptions, and therefore his method of proof, are completely different from ours.

As in the stationary case, if we are willing to only use information on the $B(z)$ and $C(z)$ matrices, and also to impose $A_o = I_{(p)}$, then an obvious specialization of the ARMAX results yields ARARX identification conditions. Using the notation of Lemma 9(ii), this is because once $B(z)$ and $C(z)$ are identified we have $U(z) = I_{(p)}$; because $A_1(0) = A_2(0)$ we also have $F = I_{(p)}$ so that $A_1(z) = A_2(z)$ and $\Omega_1 = \Omega_2$.

We note that to eliminate redundancy in ARARX structures we can always assume that $B(z)$ and $C(z)$ are left coprime. This is not so, in general, for ARMAX structures. For suppose we have an ARARX structure $S := [B(z),C(z),A(z),\Omega]$ such that the g.c.l.d. of $B(z)$ and $C(z)$ is $U(z)$, with $U(z)$ not unimodular. If we define $[B_1(z),C_1(z)]=U(z)^{-1}[B(z),C(z)]$, $A_1(z) = A(z)U(z)$ and $\Omega_1 = \Omega$, then the ARARX structure $S_1 := [B_1(z),C_1(z),A_1(z),\Omega_1]$ is observationally equivalent to S by Lemma 9 (ii) and $B_1(z)$, $C_1(z)$ are left coprime by construction.

Given the just-mentioned conditions, we illustrate the correspondence between ARMAX and ARARX identification by obtaining an ARARX analogue of Theorem 1.

Theorem 8 We define the ARARX model by the conditions of Theorem 7 together with $A_o = I_{(p)}$. Then the model is just identified. □

If identification conditions (besides $A_o = I_{(p)}$) are also imposed on $A(z)$, then, in general, the analysis becomes much messier and usually the only useful identification results obtainable are local ones. See §10 for example.

9. IDENTIFICATION WITH IDENTITIES PRESENT

We now discuss identification for the ARMAX model (1.10), (1.11) when identities are present. Bearing in mind the remarks made in §§7 and 8, it is easy to see that analogous results may be obtained if the residuals are stationary or autoregressive.

Supposing that the last $(p-p_1)$ equations $(0 < p_1 < p)$ in (1.10) are identities, we can write

$$B(z)=\begin{bmatrix} B_{11}(z) & \vdots & B_{12}(z) \\ \cdots & \vdots & \cdots \\ B_{21}(z) & \vdots & B_{22}(z)\end{bmatrix},\quad C(z)=\begin{bmatrix} C_1(z) \\ \cdots \\ C_2(z)\end{bmatrix},\quad A(z)=\begin{bmatrix} A_{11}(z) & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 0\end{bmatrix},$$

$$\Omega=\begin{bmatrix} \Omega_{11} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 0\end{bmatrix},\quad u(n)'=\{u_1(n)',0'\},\quad e(n)'=\{e_1(n)',0'\},$$

$y(n)'=\{y_1(n)',y_2(n)'\}$, where $B_{11}(z)$ is $p_1\times p_1$, $B_{12}(z)$ is $p_1\times(p-p_1)$, $C_1(z)$ is $p_1\times q$, $A_{11}(z)$ is $p_1\times p_1$, Ω_{11} is $p_1\times p_1$, $u_1(n)$ is $p_1\times 1$, $e_1(n)$ is $p_1\times 1$ and $y_1(n)$ is $p_1\times 1$; $B_{21}(z)$, $B_{22}(z)$ and $C_2(z)$ are known.

We make the following changes to assumptions A1. to A4. In assumption A1., Ω_{11} rather than Ω is nonsingular now; in A4.(ii) we will also assume that $\det\{B_{22}(z)\} \neq 0$ for $|z| \leq 1$; instead of A4.(iii) we will now assume that $\det\{A_{11}(z)\} \neq 0$ for $|z| < 1$.

Assumptions A2. and A3. remain as they are. Given these assumptions we can rewrite (1.10) and (1.11) as

$$\tilde{B}(z)y_1(n) = \tilde{C}(z)x(n) + u_1(n), u_1(n) = A_{11}(z)e_1(n) \quad (9.10)$$

where $\tilde{B}(z) = B_{11}(z)-B_{12}(z)B_{22}(z)^{-1}B_{21}(z)$ and $\tilde{C}(z) = C_1(z)-$ $- B_{12}(z)B_{22}(z)^{-1}C_2(z)$, with $\det \tilde{B}(z) \neq 0$ for $|z| \leq 1$. Although identities have been eliminated in (9.10), (9.10) is not usually a suitable form for identification because it will be difficult to state prior knowledge on $B(z)$ and $C(z)$ in terms of $\tilde{B}(z)$ and $\tilde{C}(z)$.

We will denote the two ARMAX structures S_1 and S_2 by $S_j := [B_j(z),\ C_j(z),\ A_{11,j}(z),\ \Omega_{11,j}]$, $j = 1,2$.

The following lemma gives NSC's for two structures to be observationally equivalent.

Lemma 10 Put $\tilde{A}(z)' = \{A_{11}(z)',0'\}$. The two structures S_1 and S_2 are observationally equivalent if and only if there exists a nonsingular matrix $U(z)$ with elements that are rational functions of z, and a $p_1 \times p_1$ nonsingular constant matrix F, such that

$[B_2(z),\ C_2(z), \tilde{A}_2(z)] = U(z)[B_1(z), C_1(z),\ \tilde{A}_1(z)F] \forall\ z$ and

$\Omega_{11,1} = F\Omega_{11,2}F'$. □

Because $B_{21}(z)$, $B_{22}(z)$ are known and $U(z)=B_2(z)B_1(z)^{-1}$, $U(z)$ must be of the form

$$\begin{bmatrix} U_{11}(z) & \vdots & U_{12}(z) \\ \cdots & \cdots & \cdots \\ 0_{(p-p_1),p_1} & \vdots & I_{(p-p_1)} \end{bmatrix}.$$

Using Lemma 10 we can obtain identification results similar to those of §§5 and 6. We illustrate by obtaining an analogue of Theorem 3B.

Theorem 9 Put $N = (B_o,\ldots,B_r,\ C_o,\ldots,C_s)$. We define the model by the following conditions: (i) at least $(p-1)$ zero elements are prescribed in each of the first p_1 rows of N; (ii) the first p_1 diagonal elements of B_o are unity; (iii) $A_{11}(0) = I_{(p_1)}$.

Sufficient conditions for a structure belonging to the model to be globally identified are: C1. $B(z)$, $C(z)$ and $\tilde{A}(z)$ are left coprime;
C2. $(B_r, C_s, \tilde{A}_t)$ has rank p; C3. the rank of each submatrix of N obtained by taking the columns of N with prescribed zeros in a certain row is $(p-1)$.

The proof of the theorem follows directly from the proof of Theorem 3A and will not be given. □

Hannan [1971a] also allows for the presence of identities by permitting Ω to be singular; in the statements of his theorems, however, he does not seem to take account of the fact that $B_{21}(z)$, $B_{22}(z)$ and $C_2(z)$ are known matrices. In addition we feel that for the type of structures we are considering, our derivation (Lemma 10) of the NSC's for two structures to be observationally equivalent is simpler than that given in Hannan [1971a, p.763].

Deistler [1975, p.21] also mentions the problem of identification when identities are present. His analysis seems to be deficient, however, because he assumes that $U_{12}(z) \equiv 0$ and this is not necessarily the case.

10. AN EXTENSION OF A THEOREM OF HATANAKA

Hatanaka [1976, pp.200-203] gives a local identification result for the ARARX (1,0,1) model

$$B_o y(n) + B_1 y(n-1) = C_o x(n)+u(n), u(n)+A_1 u(n-1)=e(n) \qquad (10.09).$$

We note, however, that condition (II') of Hatanaka's theorem (p.201) is not necessary. For example, consider the ARARX (1,0,1) model with $p=q=2$ and with zeros prescribed on B_o, B_1 and C_o as in the following structure

$$B(z) = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} + \begin{pmatrix} 0 & \frac{1}{4} \\ \frac{1}{4} & 0 \end{pmatrix} z, \quad C_o(z) = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}, A(z)=I_{(2)} + \tfrac{1}{2} I_{(2)} z,$$

$$\Omega = I_{(2)} \qquad (10.10)$$

Below we use Hatanaka's (p.191) definition of the integers m_1 and s_1. Under the usual normalization convention, the structure given in (10.10) is certainly identified. However $m_1 = 1$ and $s_1 = 2$ implying that $m_1 + s_1 = 3$, while the matrix in (II') (Hatanaka, p.201) can be of rank two at most. In addition Hatanaka's condition (I') is satisfied because C_o is of full rank. Hence we have demonstrated that (II') can be violated for an identified structure.

We also note that Hatanaka's result is not directly applicable to models having lagged exogenous variables. To see this suppose we had $C_o x(n) + C_1 x(n-1)$ instead of $C_o x(n)$ in (10.09). The former expression could be rewritten as $\tilde{C}_o \tilde{x}(n)$ where $\tilde{C}_o = \{C_o, C_1\}$ and $\tilde{x}(n)' = \{x(n)', x(n-1)'\}$. Unfortunately, $\tilde{x}(n)$ does not satisfy Hatanaka's (p.200) condition (III). Thus we cannot reduce a situation where there are lags on the exogenous variables to the unlagged one of (10.09).

We now show how the issues raised above may be overcome. The role of condition (I') (Hatanaka, p.200), whose function is a little unclear in the Hatanaka analysis, is shown, and a more straightforward proof than Hatanaka's is given. The degrees of $B(z)$, $C(z)$ and $A(z)$ are allowed to be arbitrarily prescribed, and conditions under which global identification is achieved are given. Analogous results are obtained for ARMAX structures. For simplicity we assume that there are no identities and the conditions of §3 and §8 hold for ARMAX and ARARX structures respectively. We denote ARMAX and ARARX structures as in previous sections.

To show the simplicity of our approach we first deal with Hatanaka's case.

Theorem 10 We define the ARARX model by: (i) the degrees of $B(z)$, $C(z)$ and $A(z)$ are prescribed as 1,0 and 1 respectively.

We put $N = (B_o, B_1, C_o)$.

(ii) In each row of N there are prescribed at least (p-1) zero elements; (iii) $A_o = I_{(p)}$; (iv) the diagonal elements of B_o are prescribed as unity.

Denote the structure S_o as $[B_o(z), C_o(z), A_o(z), \Omega_o]$, where we write $B_o(z)$, etc. as $B_{o,o}+B_{1,o}z+\ldots+B_{r,o}z^r$, etc. Put $\alpha = \text{vec } A_1$, $\alpha_o = \text{vec } A_{1,o}$, $J_1(\alpha) = (A_1 \otimes I_{(p)})$, $J_2 = (I_{(p)} \otimes B_{o,o}^{-1'} B_{1,o}')$, $J_3 = (I_{(p)} \otimes C_{o,o}')$, $H(\alpha)' = [\{J_2 - J_1(\alpha)\}', J_3']$.

If the structure S_o belongs to the model and satisfies

$$\rho[H(\alpha_o)] = p^2, \qquad (10.15)$$

then it is locally identified if and only if the following condition holds: C1. The rank of each submatrix of N obtained by taking the columns of N with prescribed zeros in a certain row is (p-1).

Proof Suppose $S := [B(z), C(z), A(z), \Omega]$ is another structure belonging to the model and observationally equivalent to S_o. By Lemma 9(ii),

$$[B(z), C(z)] = U(z)[B_o(z), C_o(z)],\ A(z)U(z) = FA_o(z), \Omega = F\Omega_o F', \tag{10.18}$$

where $U(z) = B(z)B_o(z)^{-1}$ and $F = U(0)$.

By assumption A4.(ii) we can write $U(z)$ as

$$U(z) = \sum_{j=o}^{\infty} U_j z^j \quad \text{for} \quad |z| \leq 1 \tag{10.19}$$

By (i)

$$U_1 B_{1,o} + U_2 B_{o,o} = 0,\ U_1 C_{o,o} = 0,\ A_1 U_1 + U_2 = 0 \tag{10.20}$$

implying that $A_1 U_1 - U_1 B_{1,o} B_{o,o}^{-1} = 0$ and $U_1 C_{o,o} = 0$. Hence $H(\alpha)u = 0$, where $u = \text{vec}\{U_1'\}$. For S close to S_o α will be close to α_o, so that $H(\alpha)$ will be of full column rank by (10.15). It follows that $U_1 = 0$. Because $A_1 U_j + U_{j+1} = 0\ \forall\ j \geq 1$, it follows recursively that $U_j = 0$ for $j \geq 1$. Hence $U(z)$ must be a constant matrix for S close to S_o. Having reduced the identification problem to the classical case, it is clear that C1. is a NSC for S_o to be locally identified. □

Corollary 1 Subject to conditions (i) - (iv) of Theorem 10, if

$$\rho[B_{1,o} B_{o,o}^{-1} C_{o,o}] = p \tag{10.22}$$

then C1. of Theorem 10 is a NSC for the *global* identification of S_o.

Proof Because $U_2 C_{o,o} = 0$ it follows from (10.20) that $U_1 B_{1,o} B_{o,o}^{-1} C_{o,o} = 0$. If (10.22) holds, then $U_1 = 0$ and therefore $U_j = 0$ for $j \geq 1$ as in the proof of Theorem 10.

The proof is now completed as for Theorem 10. □

Remark 10.1 (i) It is now clear that condition (10.15) above, (i.e. condition (I') of Hatanaka) ensures that the matrix $U(z)$ is constant for structures close to S_o.

(ii) We note that by imposing the stronger condition (10.22), instead of (10.15), we obtain a global identification result, rather than the local identification one of Theorem 10. □

Define the permutation matrix P_d by

$$\text{vec}(U_1 \ldots, U_d) = P_d \, \text{vec}\{(U_1', \ldots, U_d')'\}$$

where the U_j, $1 \le j \le d$, are arbitrary $p \times p$ matrices. It is easy to verify that $P_d = (\tilde{P}_d \otimes I_{(p)})$ where $\tilde{P}_d$ is the $pd \times pd$ permutation matrix mapping the $(j-1)d + \ell$ element into the $(\ell-1)p+j$ position, $1 \le j \le p$, $1 \le \ell \le d$. For example, if $p = 2$, then

$$\tilde{P}_3 = \begin{pmatrix} 1 & 0 & 0 & . & 0 & 0 & 0 \\ 0 & 0 & 0 & . & 1 & 0 & 0 \\ & . & . & . & . & . & \\ 0 & 1 & 0 & . & 0 & 0 & 0 \\ & & & . & & & \\ 0 & 0 & 0 & . & 0 & 1 & 0 \\ & . & . & . & . & . & \\ 0 & 0 & 1 & . & 0 & 0 & 0 \\ 0 & 0 & 0 & . & 0 & 0 & 1 \end{pmatrix}$$

For the rest of this section we suppose that the degrees of $B(z)$, $C(z)$ and $A(z)$ are prescribed as r,s and t respectively. The next theorem is a generalization of Theorem 10.

Theorem 11 Put $N = (B_o, \ldots, B_r, C_o, \ldots, C_s)$. We define the model by: (i) In each row of N there are prescribed at least $(p-1)$ zero elements; (ii) $A_o = I_{(p)}$; (iii) the diagonal elements of B_o are prescribed as unity. Denote the structure S_o by $[B_o(z),C_o(z),A_o(z),\Omega_o]$ and put $d = \max(r,s,t)$, $\alpha = \mathrm{vec}(A_1,\ldots,A_t)$,

$$K_1(\alpha) = \begin{bmatrix} A_d & A_{d-1} & \cdots & A_1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & A_{d-1} \\ 0 & & & A_d \end{bmatrix}, \quad K_2(\alpha) = \begin{bmatrix} I_{(p)} & & & 0 \\ A_1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ A_{d-1} & \cdots & A_1 & I_{(p)} \end{bmatrix},$$

$$K_3 = \begin{bmatrix} B_{d,o} & & & 0 \\ B_{d-1,o} & \ddots & & \\ \vdots & \ddots & \ddots & \\ B_{1,o} & \cdots & B_{d-1,o} & B_{d,o} \end{bmatrix},$$

$$K_4 = \begin{bmatrix} B_{o,o} & B_{1,o} & \cdots & B_{d-1,o} \\ & \ddots & \ddots & \vdots \\ & & \ddots & B_{1,o} \\ 0 & & & B_{o,o} \end{bmatrix}$$

and let K_5 and K_6 be defined with respect to the $C_{j,o}$ in the same way as K_3 and K_4 were defined with respect to the $B_{j,o}$; put $J_1(\alpha) = [I_{(p)} \otimes K_2^{-1}(\alpha) K_1(\alpha)]$, $J_2 = (K_4'^{-1} K_3' \otimes I_{(p)})$, $\tilde{J}(\alpha) = P_d J_1(\alpha) P_d'$ and

$$H(\alpha) = \begin{bmatrix} \tilde{J}(\alpha) - J_2 \\ \cdots \\ (K_5' \otimes I_{(p)}) - (K_6' \otimes I_{(p)}) J_2 \end{bmatrix}.$$

Let α_o be the value of α at S_o. If the structure S_o belongs to the model and

$$\rho[H(\alpha_o)] = p^2 d,$$

then S_o is locally identified if and only if C1. of Theorem 10 holds. □

The following lemma may prove useful in obtaining a simple expression for $\tilde{J}(\alpha)$.

Lemma 11 Let P_d be as described above. Suppose A is an arbitrary $p\times p$ matrix, and D is an arbitrary $d\times d$ block matrix with the $p\times p$ submatrices D_{ij} $(i,j = 1,\ldots,d)$. Then:

(i) $P_d'P_d = I_{(p^2d)}$.

(ii) $P_d(A\otimes D)P_d'$ is a $d\times d$ block matrix with the submatrix $(A\otimes D_{ij})$ in the $(i,j)^{th}$ position.

(i) follows because P_d is a permutation matrix, while the proof of (ii) is easy but messy and hence is omitted. □

Corollary 2 (i) Subject to conditions (i) - (iii) of Theorem 11, if

$$\rho(K_5 - K_3K_4^{-1}K_6) = pd, \tag{10.24}$$

then C1. of Theorem 10 is a NSC for the *global* identification of S_o.

If (10.24) holds, then: (ii) $B_o(z)$ and $C_o(z)$ are left coprime and the rank of $(B_{r,o}, C_{s,o})$ is p. □

Corollary 3 Put $\underline{d} = \min\,(r,t)$,

$$T(\alpha) = \begin{bmatrix} H(\alpha) \\ \cdots\cdots\cdots \\ I_{(p^2\underline{d})} \;\vdots\; 0_{p^2\underline{d},(d-\underline{d})p^2} \end{bmatrix}.$$

Subject to the conditions of Theorem 11, if α_o is a regular point of $T(\alpha)$ and $H(\alpha)$ and if

$$\rho\{T(\alpha_o)\} = \rho\{H(\alpha_o)\}, \tag{10.25}$$

then C1. of Theorem 10 is a NSC for the local identification of S_o. □

Remark 10.2 (i) Because $K_2(\alpha_o)$, K_4 are block triangular matrices, $K_2^{-1}(\alpha_o)K_1(\alpha_o)$ and $K_3K_4^{-1}$ may be computed recursively without actually evaluating $K_2^{-1}(\alpha_o)$ and K_4^{-1} explicitly. In addition, because K_1 to K_6 are block triangular, we can speed up the calculation of $H(\alpha_o)$ by taking explicit account of blocks of zeros.

(ii) Using Lemma 12, we can obtain an explicit expression for $\tilde{J}(\alpha)$.

(iii) Identification using Theorem 11 just involves checking two rank conditions. They will be easier to check than the usual criteria for determining whether left coprimeness holds.

(iv) Corollary 2 shows that (10.24) will in general not be a necessary condition for S_o to be identified. This is because (see §8) the conditions stated in Corollary 2(ii) together with C1. of Theorem 10 suffice to identify S_o.

(v) By the obvious rewording of condition C1. in Theorems 10 and 11 and in Corollary 2, we may obtain local identification results (global in the case of Corollary 2) for a single equation or a subset of equations.

(vi) For a result comparable to Corollary 3 see Wald [1950, p.244, Theorem 3.3]. □

We now obtain identification results for ARMAX structures analogous to those obtained above. The identification will now be global rather than local.

Theorem 12 Put $N = (B_o,..,B_r, C_o,...,C_s)$. We define the ARMAX model by the following conditions: (i) In each row of N there are prescribed at least $(p-1)$ zero elements; (ii) $A_o = I_{(p)}$; (iii) the diagonal elements of B_o are prescribed as unity.

Denote the ARMAX structure S_o by $[B_o(z), C_o(z), A_o(z), \Omega_o]$ and put $d = \max(r,s,t)$ and $H = (K_1 - K_3K_4^{-1}K_2, K_5 - K_3K_4^{-1}K_6)$, where K_3, K_4, K_5 and K_6 are defined in the same way as in Theorem 11, while K_1 and K_2 are defined with respect to the $A_{j,o}$ in the same way as K_3 and K_4 are defined with respect to the $B_{j,o}$.

If S_o belongs to the model and

$$\rho(H) = pd, \tag{10.30}$$

then S_o is globally identified if and only if C1. of Theorem 10 holds. □

Corollary 4 If (10.30) holds, then $B_o(z)$, $C_o(z)$ and $A_o(z)$ are left coprime and $\rho(B_{r,o}, C_{s,o}, A_{t,o}) = p$. □

Corollary 5 Put $\underline{d} = \min(r,t)$. Subject to conditions (i) - (iii) of Theorem 12, if

$$\rho \begin{bmatrix} H & \vdots & I_{(p\underline{d})} \\ & \vdots & \cdots \\ & \vdots & 0 \end{bmatrix} = \rho(H),$$

then C1. of Theorem 10 is a NSC for the global identification of S_o. □

Remark 10.3 (i) Remarks 10.2 (i), (iii) to (vi) apply equally well to the ARMAX results just obtained. □

11. IDENTIFICATION WITH RESPECT TO A SEQUENCE OF CRITERION FUNCTIONS

For observations generated by a probability distribution which is known up to the specification of a parameter vector, and which possesses a probability density function, Rothenberg [1971, pp.581-582] obtained conditions for the local identification of the true parameter θ_o in the presence of nonlinear restrictions.

Using an information inequality, Bowden [1973] generalized Rothenberg's results by obtaining a global identification condition and dispensed with the requirement that the probability density function of the observations should exist.

Although constituting a significant contribution to identification theory, the Rothenberg - Bowden results have the following shortcomings: (i) the results depend on a knowledge of the distribution function of the observations. Because this information is usually not available, it would be preferable to have a distribution free method of identifying parameters of interest; (ii) there is little attempt to link the identification theory with common estimation procedures to demonstrate how the lack of identification affects estimation; (iii) there is no link between the results stated above and the time series identification results we have discussed in previous sections.

In an unpublished paper Sargan [1975] overcomes some of the problems raised in (i) to (iii). Subject to regularity conditions analogous to those given below, Sargan defines the identifiability of a parameter (calling it asymptotic identifiability) with respect to the uniform limit of a sequence of criterion functions. It is the purpose of this section to examine the implications of Sargan's definition. In so doing we will obtain an intuitive understanding of results obtained more rigorously in later chapters. In particular Theorem 13, Application 1 and Applications 3 and 4 introduce informally important ideas and techniques.

To estimate θ_o, the true value of the unknown parameter θ, we usually choose a sequence $\{L_N(\theta), N \geq 1\}$ of stochastic criterion functions indexed by the time parameter N and select as an estimate of θ_o that value of θ which minimizes $L_N(\theta)$ over an appropriate parameter space, Θ say. For example, in Applications 3 and 4 we select $L_N(\theta)$ as $-2/N \times \log$ of a Gaussian likelihood. If the estimation of θ_o is to be meaningful, then, at least for large samples, the criterion function should distinguish between θ_o and other values of θ.

Let m be the dimension of θ and denote the permissible parameter space by Θ; $\theta_o \in \Theta$. We suppose that there exists a nonstochastic function of θ, $L(\theta)$ say, such that $L_N(\theta) \overset{a.s.}{\to} L(\theta)$, uniformly in Θ. We now make the following assumptions:

B1. $L(\theta_o) \leq L(\theta)\ \forall\ \theta \in \Theta$.

B2. $h(\theta)$ is an f dimensional, $0 \leq f < m$, continuous vector function of $\theta \in \Theta$. Furthermore $h(\theta_o) = 0$.

Definition 1 θ is observationally equivalent to θ_o if $L(\theta) = L(\theta_o)$.

Definition 2 Subject to the restrictions $h(\theta) = 0$, we will say that θ_o is globally identified with respect to the sequence of criterion functions $\{L_N(\theta)\}$ if there is no other $\theta \in \Theta$ satisfying $h(\theta) = 0$ and observationally equivalent to θ_o.

Definition 2 involves exactly the same idea as the definition of asymptotic identification given in Sargan [1975, p.8]. We prefer our own terminology as it suggests the link between the sequence of criterion functions that are minimized to obtain an estimate of θ, and the identification process.

Definition 3 Subject to the restrictions $h(\theta) = 0$, we will say that θ_o is locally identified with respect to the sequence of criterion functions $\{L_N(\theta)\}$ if there exists a neighbourhood of θ_o such that there is no other $\theta \in \Theta$ lying in the neighbourhood, which is observationally equivalent to θ_o and satisfies the constraints.

To obtain an analogue of Theorem 2 of Rothenberg [1971, p.581] we need the following additional assumption.

B3.(i) θ_o is an interior point of Θ.

(ii) There exists a neighbourhood of θ, N say, such that $h(\theta)$ is continuously differentiable for $\theta \in N$. Furthermore $\partial^2 L_N(\theta)/\partial\theta\partial\theta'$ and $\partial^2 L(\theta)/\partial\theta\partial\theta'$ exist and are continuous for $\theta \in N$.

(iii) $\partial^2 L_N(\theta)/\partial\theta\partial\theta' \xrightarrow{p} \partial^2 L(\theta)/\partial\theta\partial\theta'$, uniformly in $\theta \in N$. Put $H(\theta) = \partial h(\theta)/\partial\theta'$, $H = H(\theta_o)$, $\Delta(\theta) = \partial^2 L(\theta)/\partial\theta\partial\theta'$. $\Delta = \Delta(\theta_o)$.

Theorem 13 Suppose θ_o is a regular point of the matrix $[\Delta(\theta)', H(\theta)']$. Given assumptions B1.-B3., (i), (ii) and (iii) are equivalent NSC's for θ_o to be locally identified in the sense of Definition 3.

(i) $\rho(\Delta',H') = m$

(ii) $$\rho\left\{\begin{bmatrix}\Delta & . & H' \\ \cdots & \cdots & \cdots \\ H & . & 0\end{bmatrix}\right\} = m + \tilde{f}, \text{ where } \tilde{f} = \rho(H)$$

Put $E = I_{(m)} - H'(HH')^{+}H.$

(iii) $\rho(E\Delta E) = m - \tilde{f}.$ □

Application 1 This example illustrates the connection between estimation and local identification. The constrained minimization of $L_N(\theta)$ yields

$$\partial L_N(\theta)/\partial\theta + H'(\theta)\lambda = 0,\ h(\theta) = 0 \tag{11.10}$$

where λ is a vector of Lagrange multipliers. Because we cannot usually solve (11.10) for θ and λ we linearize about $\hat{\theta}$, a consistent estimate of θ_o in the manner of Aitchison and Silvey [1958], to obtain

$$\partial L_N(\hat{\theta})/\partial\theta + \Delta_N(\hat{\theta})(\tilde{\theta}-\hat{\theta}) + H'(\hat{\theta})\tilde{\lambda} = 0 \tag{11.11}$$

$$h(\hat{\theta}) + H(\hat{\theta})(\tilde{\theta}-\hat{\theta}) = 0 \tag{11.12}$$

where $\tilde{\theta}$, $\tilde{\lambda}$ are the linearized approximate solutions to (11.10). We can rewrite (11.11) and (11.12) as

$$\begin{bmatrix}\Delta_N(\hat{\theta}) & \vdots & H'(\hat{\theta}) \\ \cdots & \vdots & \cdots \\ H(\hat{\theta}) & \vdots & 0\end{bmatrix}\begin{bmatrix}\tilde{\theta}-\hat{\theta} \\ \cdots \\ \tilde{\lambda}\end{bmatrix} = -\begin{bmatrix}\partial L_N(\hat{\theta})/\partial\theta \\ \cdots \\ h(\hat{\theta})\end{bmatrix} \tag{11.13}$$

Without loss of generality we can assume that $H(\theta_o)$ is of full row rank. If θ_o is locally identified, then Theorem 13(ii) implies that with a high probability the matrix in (11.13) will be nonsingular for N large so we can solve (11.13) for $(\tilde{\theta}-\hat{\theta})$ and $\tilde{\lambda}$. Conversely, if θ_o is not locally identified then, the matrix on the left side of

(11.13) will be close to singularity for N large and it will be numerically unstable to solve (11.13) for $(\tilde{\theta}-\hat{\theta})$ and $\tilde{\lambda}$. □

We need the following lemma in the next application.

Lemma 12 Let E be a $p\times p$ symmetric, idempotent matrix of rank f $(0 < f < p)$ and A an arbitrary $p\times p$ matrix. Then:

(i) $E(EAE)^{+} = (EAE)^{+} = (EAE)^{+}E$.

For the next two results we assume that $\rho(E) = \rho(EAE)$.

(ii) $E\{I_{(p)} - (EAE)^{+}EAE\} = 0$

(iii) Suppose that $(EAE)\,x = y$ is a consistent equation for x. Then $x^{\dagger} = (EAE)^{+}y$ is the only solution of the equation which satisfies $Ex = x$. □

Application 2 We again illustrate the connection between local identification and estimation. We assume that $H(\theta)$ is constant, i.e. the restrictions are linear. Rather than use Lagrange multipliers, as was done in the previous application, we impose the restrictions using a symmetric idempotent matrix. This method of handling linear restrictions is used in Hannan and Terrell [1972, p.190 and 1973, p.306].

Assuming that the $\hat{\theta}$ defined above satisfies the constraints $h(\theta) = 0$, (11.12) is equivalent to $E(\tilde{\theta}-\hat{\theta})=(\tilde{\theta}-\hat{\theta})$, where E is defined in Theorem 13. (11.11) and (11.12) imply that

$$\{E\Delta_N(\hat{\theta})E\}(\tilde{\theta}-\hat{\theta}) = -E\{\partial L_N(\hat{\theta})/\partial\theta\} \qquad (11.15)$$

Put $Q_N = E\Delta_N(\hat{\theta})E$. Assuming that H is of full row rank and θ_o is identified, it follows from Theorem 13(iii) that for N large $\rho(Q_N) = \rho(E)$ with a high probability. It therefore follows from Lemma 12(iii) that, with probability tending to 1 as N increases, $\tilde{\theta} = \hat{\theta} - Q_N^+ \, \partial L_N(\hat{\theta})/\partial\theta$ is the unique solution to (11.15) that satisfies the linear restrictions. □

Using the approach taken in this section we will re-derive the NSC's for two ARMAX, or ARARX, structures to be observationally equivalent and thereby show the close link between the identification results of previous sections and common estimation practice as depicted by Application 1. Although the criterion functions we will work with correspond approximately to Gaussian likelihoods (or more precisely to $-2/N \times \log$ of a Gaussian Likelihood) they are often used to derive estimators (see Chapter 4) irrespective of whether the data is Gaussian or not. The results below also do not depend on Normality.

It will be convenient to adopt the convention that a zero subscript indicates the value of a function at the true parameter point.

Application 3 We now consider ARMAX structures as described by (1.10) and (1.11). Let r,s and t be prescribed and take R_1 as an arbitrarily large positive number and ε_1 as an arbitrarily small positive number. Putting $\theta = \mathrm{vec}[B_o,\ldots,B_r, C_o,\ldots,C_s, A_o,\ldots,A_t, \Omega]$, we take the parameter space Θ as the closure of the set Θ_o, where Θ_o is defined as follows:

C1. Ω is p.d. and $|\det \Omega| \geq \varepsilon_1$

C2. (i) $\|\theta\| \leq R_1$; (ii) $|\det B(0)| \geq \varepsilon_1$, $|\det B(z)| > 0$ for $|z| \leq 1$; $|\det A(z)| \geq \varepsilon_1$ for $|z| \leq 1$.

We assume that $\theta_o \in \Theta_o$.

We put $\tilde{\Omega} = B(0)^{-1}A(0)\ \Omega A(0)'B(0)^{-1'}$,

$\Phi(z;\theta) = B(0)^{-1}A(0)A(z)^{-1}\{C(z)-B(z)B_o(z)^{-1}C_o(z)\}$

$\Psi(z;\theta) = B(0)^{-1}A(0)A(z)^{-1}B(z)B_o(z)^{-1}A_o(z)A_o(0)^{-1}B_o(0)$

and define the sequence $\{v(n;\theta),\ n \geq 1\}$ constructively by

$$A(z)A(0)^{-1}B(0)v(n;\theta) = B(z)y(n) - C(z)x(n)$$

with $v(n;\theta) = 0$ for $n \leq 0$. We define the criterion function $L_N(\theta)$, $N \geq 1$, as

$$N^{-1}\ \mathrm{tr}\ \tilde{\Omega}^{-1} \sum_{n=1}^{N} v(n;\theta)v(n;\theta)' + \log\det\tilde{\Omega} \qquad (11.20)$$

It follows from Lemmas 4.1 and 4.3 that

$$L_N(\theta) \overset{a.s.}{\to} L(\theta) = \mathrm{tr}\ \tilde{\Omega}^{-1}\int_{-\pi}^{\pi}\Phi(\lambda;\theta)f_x(\lambda)\Phi(\lambda;\theta)^*d\lambda +$$

$$+ (2\pi)^{-1}\ \mathrm{tr}\ \tilde{\Omega}^{-1}\int_{-\pi}^{\pi}\Psi(\lambda;\theta)\tilde{\Omega}_o\Psi(\lambda;\theta)^*d\lambda + \log\det\tilde{\Omega}$$

uniformly in $\theta \in \Theta$. We will now show that $L(\theta) \geq L(\theta_o)$ with equality holding if and only if

$C(z) = B(z)B_o(z)^{-1}C_o(z)$, $A(z) = B(z)B_o(z)^{-1}A_o(z)F$, $\Omega_o = F\Omega F'$, where $F = A_o(0)^{-1}B_o(0)B(0)^{-1}A(0)$.

This result is precisely Lemma 6(ii). We can write $\Phi(z;\theta)$ and $\Psi(z;\theta)$ respectively as

$$\sum_{j=o}^{\infty}\Phi_j(\theta)z^j \quad \text{and} \quad \sum_{j=o}^{\infty}\Psi_j(\theta)z^j,$$

the series converging for $|z| \leq 1$; $\Psi_o(\theta) = I_{(p)}$. For fixed

$\theta \in \Theta$,

$$L(\theta) \geq \operatorname{tr} \tilde{\Omega}^{-1} \sum_{j=o}^{\infty} \Psi_j(\theta)\tilde{\Omega}_o\Psi_j(\theta)' + \log\det\tilde{\Omega} \geq$$

$$\geq \operatorname{tr} \tilde{\Omega}^{-1}\tilde{\Omega}_o + \log\det\tilde{\Omega} =$$

$$= \operatorname{tr} \tilde{\Omega}^{-\frac{1}{2}}\tilde{\Omega}_o\tilde{\Omega}^{-\frac{1}{2}} - \log\det\tilde{\Omega}^{-\frac{1}{2}}\tilde{\Omega}_o\tilde{\Omega}^{-\frac{1}{2}} + \log\det\tilde{\Omega}_o =$$

$$= \sum_{j=1}^{p} (\lambda_j - \log\lambda_j) + \log\det\tilde{\Omega}_o \geq p + \log\det\tilde{\Omega}_o = L(\theta_o),$$

because $(\lambda - \log\lambda) \geq 1 \ \forall\ \lambda > 0$, with equality only at $\lambda = 1$; the λ_j $(1 \leq j \leq p)$ are the eigenvalues of $\tilde{\Omega}^{-\frac{1}{2}}\tilde{\Omega}_o\tilde{\Omega}^{-\frac{1}{2}}$. We can therefore deduce that $L(\theta) \geq L(\theta_o)$ with equality if and only if

$$\operatorname{tr} \tilde{\Omega}^{-1} \int_{-\pi}^{\pi} \Phi(\lambda;\theta)f_x(\lambda)\Phi(\lambda;\theta)^*d\lambda = 0,\ \tilde{\Omega} = \tilde{\Omega}_o \text{ and}$$

$\Psi(z;\theta) = I_{(p)}$. Because $\operatorname{tr} \tilde{\Omega}^{-1} \int_{-\pi}^{\pi} \Phi(\lambda;\theta)f_x(\lambda)\Phi(\lambda;\theta)^*d\lambda = 0$ implies that $\Phi(z;\theta) = 0$, the required result follows. □

Application 4 We now consider the ARARX structures described by (1.10) combined with (8.10). Suppose the degrees of $B(z)$, $C(z)$ and $A(z)$ are prescribed as r, s and t respectively and let R_1 and ε_1 be two positive numbers, the first arbitrarily large and the second arbitrarily small. Defining θ as in Application 3, we take the permissible parameter space Θ to be the closure of the set Θ_o which is defined by:

D1. Ω is p.d. and $|\det\Omega| \geq \varepsilon_1$;

D2. (i) $\|\theta\| \leq R_1$; $|\det B(0)| \geq \varepsilon_1$; $|\det B(z)| > 0$ and $|\det A(z)| > 0$ for $|z| \leq 1$.

We assume that $\theta_o \in \Theta_o$.

Put $\tilde{\Omega} = B(0)^{-1}A(0)^{-1}\Omega A(0)^{-1'}B(0)^{-1'}$ and define the sequence $v(n;\theta)$ constructively by

$$v(n;\theta) = B(0)^{-1}A(0)^{-1}A(z)[B(z)y(n)-C(z)x(n)], \; n \geq 1.$$

We now define the criterion function $L_N(\theta)$, $N \geq 1$, as in (11.20) above. Because an ARARX (r,s,t) structure is a special case of an ARMAX (r+t,s+t,o) structure we can deduce from Application 3 that θ is observationally equivalent to θ_o (in the sense of Definition 1 of this section) if and only if

$$C(z) = B(z)B_o(z)^{-1}C_o(z), \; A(z)B(z)=FA_o(z)B_o(z), \; \Omega = F\Omega_oF'$$

with $F = A(0)B(0)B_o(0)^{-1}A_o(0)^{-1}$.

This is precisely the result of Lemma 9(ii). □

12. A COMMENT ON DEISTLER [1976]

Deistler [1976] gives identification results for linear time series models under conditions that are in some respects weaker than the corresponding conditions which were stated in previous sections. We now make three observations on Deistler's paper. The first two simplify some of Deistler's derivations; hopefully such simplification may make Deistler's paper more accessible. The third observation concerns the realism of one of Deistler's assumptions. Throughout this section we use Deistler's notation and all references are to Deistler unless stated otherwise.

(i) Defining $D(z)$ and $T(z)$ as in Deistler (pp.29 and 30 respectively), $D(z)$ and $T(z)$ can be expressed as

$$D(z) = \sum_{j=o}^{\infty} D_j z^j, \; T(z) = \sum_{j=o}^{\infty} T_j z^j,$$

the expansions being valid for small z because of the non-singularity of A_o.

Because of the zero initial conditions (II), we can rewrite eqn (1) (p.27) as

$$y(n) = \sum_{j=o}^{n} T_j x(n-j) + \sum_{j=o}^{n} D_j u(n-j) \tag{12.10}$$

Defining $K_y(n,m)$, $K_x(n,m)$, $K_u(n,m)$ $(n,m \geq 0)$ as in Deistler we deduce from (12.10) that

$$K_{yx}(n,m) = \sum_{j=o}^{m} T_j K_x(n-j,m) \tag{12.11}$$

and

$$K_y(n,m) - \sum_{j=o}^{n} \sum_{k=o}^{m} T_j K_x(n-j,m-k)T_k' = \sum_{j=o}^{n} \sum_{k=o}^{m} D_j K_u(n-j,m-k)D_k' \tag{12.12}$$

From assumption (IV), $x(j) = 0$ a.s. for $0 \leq j < t$ and $K_x(t,t)$ is nonsingular. It follows from (12.11) above that

$$K_{yx}(j,t) = T_o K_x(j,t)+T_1 K_x(j-1,t)+...+T_{j-t} K_x(t,t) \text{ for}$$

$j \geq t$. We can therefore compute the T_j, $j \geq 0$, recursively. Since the T_j, as well as the second moments of the observed variates, are now known, the left side of (12.12) above is known and hence also the right side. We have now effectively obtained equations (8) to (9c) (Deistler, p.34) without having to use isomorphisms between formal matrix power series and infinite block triangular matrices as is done in Deistler.

(ii) We now give a substantially shorter proof of Deistler's Theorem 3 (p.42). Unlike Deistler's proof, the proof does not require Deistler's Lemmas 3-5. Consider equation 18(a) (Deistler, p.41). From Theorem 1 (p.35), $U(z)$ is a polynomial matrix. Suppose $\deg\{U(z)\} = f > 0$. Then ,

$$z^f U(z^{-1})\{z^p G^*(z^{-1})\}^{-1}\{z^q H^*(z^{-1})\}S^{-1} = z^f\{z^p G(z^{-1})\}^{-1}\{z^q H(z^{-1})\}.$$

Letting $z \to 0$ in the above equation we obtain $U_f G_p^{*-1} H_q^* = 0$. Because G_p^*, H_q^* are nonsingular by assumption (XIII), $U_f = 0$. Hence $U(z)$ must be a constant matrix.

(iii) As shown in the previous section, identification is closely related to estimation. With this consideration in mind we look at Deistler's assumptions on the second moments of the observed variates. He assumes that the exogenous variables are stochastic and that $E(y_t x_s')$ and $E(x_t x_s')$ are known $\forall$ s,t ≥ 0.

Now, even if the $\{x_t\}$ and $\{y_t\}$ sequences tended to stationarity for large t, we could still not evaluate the abovementioned moments from a single, complete realization of the underlying process; x_t and y_t cannot be stationary because of the zero initial conditions (II). This must cast some doubt on the usefulness of Deistler's results for pre-estimation identification.

13. PROOFS

Proof of Lemma 6 (i) follows from Hannan [1971a, equations (6) and (7)].

(ii) Sufficiency - given the conditions of (ii) we have that $B_2^{-1}(\lambda)C_2(\lambda) = B_1^{-1}(\lambda)C_1(\lambda)$, $B_2^{-1}(\lambda)A_2(\lambda) = B_1^{-1}(\lambda)A_1(\lambda)F$ and $\Omega_1 = F\Omega_2 F'$. It is now easy to check that the conditions of (i) hold.

Necessity - suppose the structures S_1 and S_2 are observationally equivalent. Put $U(z) = B_2(z)B_1(z)^{-1}$, $K_j(z) = B_j(z)^{-1}A_j(z)$, j = 1,2. Using Lemma 5 we can deduce from (i) that $K_2(z)K_2(0)^{-1} = K_1(z)K_1(0)^{-1}$ and

$K_2(0)\Omega_2K_2(0)' = K_1(0)\Omega_1K_1(0)'$ so that $A_2(z)=U(z)A_1(z)F$ and $\Omega_1 = F\Omega_2F'$, with $F = K_1(0)^{-1}K_2(0)$. Furthermore, it follows from (i) that $[B_2(z),C_2(z)] = U(z)[B_1(z),C_1(z)]$. This completes the proof of (ii). □

Proof of Theorem 1 We first show that the model is identified. Suppose S_1 and S_2 are two observationally equivalent structures belonging to the model. From Lemma 6(ii)

$$[B_2(z),C_2(z),A_2(z)] = U(z)[B_1(z),C_1(z),A_1(z)F].$$

It follows from Lemma 3 and (iii) that $U(z)$ is unimodular and hence $\det U(z)$ is a nonzero constant. Because $U(z)$ is also lower triangular $(U(z) = B_2(z)B_1(z)^{-1})$ its diagonal elements must be nonzero constants.

To deal with conditions (i)(a) or (i)(b) we take $p=2$ for convenience, but the proof applies to a general p. Put

$$B_1(z) = \begin{pmatrix} b_{11}(z) & 0 \\ b_{21}(z) & b_{22}(z) \end{pmatrix}, \quad B_2(z) = \begin{pmatrix} \alpha(z) & 0 \\ \beta(z) & \gamma(z) \end{pmatrix},$$

$$U(z) = \begin{pmatrix} u_1 & 0 \\ u_{21}(z) & u_2 \end{pmatrix}.$$

We have shown that u_1 and u_2 are nonzero constants. We have $\alpha(z) = u_1b_{11}(z)$, $\beta(z) = u_{21}(z)b_{11}(z) + u_2b_{21}(z)$. Because $\deg \beta(z) \leq \deg \alpha(z)$ and $\deg b_{21}(z) \leq \deg b_{11}(z)$ it follows that $u_{21}(z)$ is a constant. If (i)(a) applies, then u_{21} is zero so that U is a constant diagonal matrix. It follows that $U = I_{(p)}$ because the diagonal elements of $B(0)$ are unity. If (i)(b) applies, then $U = I_{(p)}$ because $B_1(0) = B_2(0) = I_{(p)}$. Therefore (i) implies that $U = I_{(p)}$.

F now equals $I_{(p)}$ by (ii).

We now show that the model is just identified. Suppose $S_1 := [B_1(z), C_1(z), A_1(z), \Omega_1]$ is an arbitrary ARMAX structure; we construct an observationally equivalent structure belonging to the model. Let $L(z)$ be a g.c.l.d. of $B_1(z), C_1(z)$ and $A_1(z)$. By Lemma 4 there exists a unimodular matrix $U(z)$ such that $U(z)L(z)^{-1}B_1(z)$ is lower triangular with the degree of its $(j,j)^{th}$ element greater than the degree of the $(i,j)^{th}$ element $(i > j)$, and with $U(0)L(0)^{-1}B_1(0) = I_{(p)}$. Put $F = \{U(0)L(0)^{-1}A_1(0)\}^{-1}$, $G(z) = U(z)L(z)^{-1}$. Then the structure $S_2 := [G(z)\{B_1(z), C_1(z), A_1(z)F\}, F^{-1}\Omega_1 F^{-1'}]$ is observationally equivalent to S_1 by Lemma 6(ii) and certainly belongs to the model. □

Proof of Theorem 2A *Sufficiency:* Suppose S_1 and S_2 are two observationally equivalent structures belonging to the model and furthermore S_1 satisfies C1.. By (iv) and Lemma 6(ii), $[B_2(z), C_2(z), A_2(z)] = U(z)[B_1(z), C_1(z), A_1(z)]$ and $\Omega_2 = \Omega_1$. By (i) and C1, $U(z)$ must be diagonal. We can write the diagonal elements of $U(z)$, $u_j(z)$ say $(1 \leq j \leq p)$, as $m_j(z)/n_j(z)$ where, for each j, $m_j(z)$ and $n_j(z)$ are polynomials with no common roots. By (ii), each of the $m_j(z)$ and $n_j(z)$ must be constant and it now follows from (iii) that $u_j(z) = 1$ for $1 \leq j \leq p$. Hence $U = I_{(p)}$.

Necessity: Suppose the structure S_1 belongs to the model but does not satisfy C1.. Without loss of generality we suppose that the rank of the matrix having prescribed null elements in the first row is of rank less than $(p-1)$.

There exists a $p\times 1$ vector polynomial $m(z)$, $m(z)' = (1,\alpha(z)')$ with $\alpha(z) \not\equiv 0$, $\gamma = m(0)'b(0) \not\equiv 0$ and $m(z)$ orthogonal to the abovementioned matrix; $b(z)$ is the first column of $B_1(z)$. Put $T_1(z)=[B_1(z),C_1(z),A_1(z)]$ and let $\delta(z)$ be the g.c.d. of the elements of the row vector $m'(z)T_1(z)$ with $\delta(0) = \gamma$. Put

$$R(z) = \begin{pmatrix} \delta(z)^{-1}m'(z) \\ \cdots\cdots\cdots \\ 0 \quad \vdots\; I_{(p-1)} \end{pmatrix}$$

Then the structure $[R(z)\{B_1(z),C_1(z),A_1(z)\},\Omega_1]$ is observationally equivalent to S_1 and belongs to the model. Because $R(z) \not\equiv I_{(p)}$ by construction, S_1 cannot be identified. $\square$

Proof of Theorem 2B The proof that C1. is a NSC for a structure to be identified is very similar to the proof given for Theorem 2A and is omitted. We now show that the last statement in Theorem 2B holds. For $p=1$ we consider the ARMAX structure S_1, $S_1 := (1+\alpha z, 1+\alpha z, 1, 1)$, $\alpha \not\equiv 0$. Clearly S_1 does not satisfy (ii) of the theorem, i.e. the g.c.d. of $(1+\alpha z,\ 1+\alpha z)$ is not unity, and no ARMAX structure observationally equivalent to S_1 can satisfy (ii). $\square$

Proof of Theorem 3A Let S_1 and S_2 be two observationally equivalent structures belonging to the model with S_1 satisfying conditions C1., C2. and C3. Then, by (iii) and Lemma 6(ii),

$$[B_2(z),C_2(z),A_2(z)] = U(z)[B_1(z),C_1(z),A_1(z)],\ \Omega_1 = \Omega_2.$$

By Lemma 3 and C1. $U(z)$ must be a polynomial matrix and because the degrees of $B_1(z),B_2(z),C_1(z),C_2(z),A_1(z)$ and $A_2(z)$

are prescribed it follows from C2. that $U(z)$ must be a constant matrix. We have reduced the proof of the theorem to the classical case. (i), (ii) and C3. now imply that $U = I_{(p)}$. Hence S_1 is identified.

The proof that C3. is a NSC for a structure belonging to the model to be identified is straightforward and we omit it. □

The proofs of Theorems 3B and 3' are very similar to that of Theorem 3A and will be omitted.

Proof of Theorem 4 *Sufficiency:* Suppose that S_1 and S_2 are two observationally equivalent structures belonging to the model with S_1 satisfying C1. and C2.. It follows from (iv), C1. and Lemma 6(ii) that

$$[B_2(z),C_2(z),A_2(z)]=U[B_1(z),C_1(z),A_1(z)F]$$

with U a constant matrix. From (iii), $F = U^{-1}$. Hence $K_2(z) = K_1(z)U^{-1}$ so that $K'_{j,2} = VK'_{j,1}$, $j \geq 0$, where $V = U^{-1'}$. (i), (ii) and C2. now imply that $V = I_{(p)}$. The required result follows.

The proof of necessity is easy and is omitted. □

Proof of Theorem 5 Suppose S_1, S_2 are observationally equivalent structures belonging to the model with S_1 satisfying C1. to C4.. As in the proof of Theorem 4 we can show that

$$[B_2(z),C_2(z),A_2(z)] = U[B_1(z),C_1(z),A_1(z)U^{-1}]$$

with U a constant matrix. Hence $A_2(z)U = UA_1(z)$. We now show that one, and only one, element in each row of U is non-zero; because U is nonsingular, at least one non-zero element exists in each row of U. Let the j^{th} diagonal elements of $A_1(z)$ and $A_2(z)$ be $a_j^{(1)}(z)$ and $a_j^{(2)}(z)$ respectively. The first row of $A_2(z)U$ equals the first row of $UA_1(z)$ so that

$$\{a_1^{(2)}(z)u_{11},\ldots,a_1^{(2)}(z)u_{1p}\}=\{u_{11}a_1^{(1)}(z),\ldots,u_{1p}a_p^{(1)}(z)\}.$$

Suppose two (or more) elements in the first row of U are non-zero; we take them to be u_{11} and u_{12} for convenience. Then $a_1^{(2)}(z) = a_1^{(1)}(z)$ and $a_1^{(2)}(z) = a_2^{(1)}(z)$ so that $a_1^{(1)}(z) = a_2^{(1)}(z)$, contradicting C3. Hence there can only be one non-zero element in the first row of U and we can similarly show that this is true of all other rows as well. (i), (ii) and C4. now imply that $U = I_{(p)}$. □

Proof of Lemma 7

$$A\otimes B = (a_1\otimes b_1,\ldots,a_1\otimes b_q,\ldots,a_n\otimes b_1,\ldots,a_n\otimes b_q)$$

and

$$(B\otimes A)' = (\beta_1\otimes\alpha_1,\ldots,\beta_1\otimes\alpha_m,\ldots,\beta_p\otimes\alpha_1,\ldots,\beta_p\otimes\alpha_m). \qquad (13.10)$$

Hence,

$$\begin{aligned} L(p,m)(A\otimes B) &= (b_1\otimes a_1,\ldots,b_q\otimes a_1,\ldots,b_1\otimes a_n,\ldots,b_q\otimes a_n) \\ &= (\alpha_1\otimes\beta_1,\ldots,\alpha_m\otimes\beta_1,\ldots,\alpha_1\otimes\beta_p,\ldots,\alpha_m\otimes\beta_p)' \end{aligned}$$

Hence

$$\begin{aligned} L(q,n)\{L(p,m)(A\otimes B)\}' &= (\beta_1\otimes\alpha_1,\ldots,\beta_1\otimes\alpha_m,\ldots,\beta_p\otimes\alpha_1,\ldots,\beta_p\otimes\alpha_m) \\ &= (B\otimes A)' \text{ by } (13.10) \end{aligned}$$

The result follows. □

Proof of Theorem 6 (a) Suppose the structure S_o belongs to the model, satisfies conditions C1. to C3., but is not locally identified. Then there exists a sequence of structures, $\{S_n := [B_n(z),C_n(z),A_n(z),\Omega_n],\ n \geq 1;\ \ S_n \neq S_o\ \forall\ n \geq 1\}$, that are observationally equivalent to S_o, belong to the model, and converge to S_o. For $n \geq 0$ let Δ, δ, α, ω, θ take the values Δ_n, δ_n, α_n, ω_n, θ_n at $S = S_n$. As in the proof of Theorem 4, we can show that (i), C1. and C2 imply that

$$[B_n(z), C_n(z), A_n(z)] = U_n[B_o(z), C_o(z), A_o(z)U_n^{-1}], \ n \geq 1$$

where $U_n = B_n(0)B_o(0)^{-1}$. Therefore $\Delta_n - \Delta_o = (U_n - I_{(p)})\Delta_o$, $A_{j,n} - A_{j,o} = (U_n - I_{(p)})A_{j,o}U_n^{-1} - A_{j,o}U_n^{-1}(U_n - I_{(p)})$, $j = 1,\ldots,t$, $\Omega_n - \Omega_o = (U_n - I_{(p)})\Omega_o U_n' + \Omega_o(U_n' - I_{(p)})$.

Put $u_n = \mathrm{vec}(U_n - I_{(p)})$, $P_2^n = (U_n\Omega_o \otimes I_{(p)}) + (I_{(p)} \otimes \Omega_o)L$,

$$P_1^n = \begin{pmatrix} U_n^{-1'}A_{1,o}' \otimes I_{(p)} \\ \vdots \\ U_n^{-1'}A_{t,o}' \otimes I_{(p)} \end{pmatrix} - \begin{pmatrix} I_{(p)} \otimes A_{1,o}U_n^{-1} \\ \vdots \\ I_{(p)} \otimes A_{t,o}U_n^{-1} \end{pmatrix}$$

where the permutation matrix L was defined in the statement of the theorem.

Because $S_n \neq S_o \ \forall \ n \geq 1$, $u_n \neq 0 \ \forall \ n \geq 1$. We can obtain from the above that

$$\delta_n - \delta_o = J_\delta u_n, \ \alpha_n - \alpha_o = P_1^n u_n, \omega_n - \omega_o = P_2^n u_n, \ n \geq 1 \quad (13.15)$$

with J_δ evaluated at S_o. Because S_o and S_n satisfy the constraints $\psi(\theta) = 0$, we have that

$$0 = \psi(\theta_n) - \psi(\theta_o) = \Psi_\delta^\dagger(\delta_n - \delta_o) + \Psi_\alpha^\dagger(\alpha_n - \alpha_o) + \Psi_\omega^\dagger(\omega_n - \omega_o)$$

where $\Psi_\delta^\dagger$, $\Psi_\alpha^\dagger$ and $\Psi_\omega^\dagger$ are symbolic derivatives evaluated at $\theta_N^\dagger$, with each row of $[\Psi_\delta^\dagger, \Psi_\alpha^\dagger, \Psi_\omega^\dagger]$ evaluated at a possibly different value of $\theta_N^\dagger$ and with $\|\theta_n^\dagger - \theta_o\| \leq \|\theta_n - \theta_o\|$. Substituting in from (13.15) and putting $v_n = u_n/\|u_n\|$, $n \geq 1$, we obtain

$$[\Psi_\delta^\dagger J_\delta + \Psi_\alpha^\dagger P_1^n + \Psi_\omega^\dagger P_2^n]v_n = 0$$

Because the sequence $(v_n, n \geq 1)$ lies on a compact set, it it has a convergent subsequence $(v_{m(n)}, n \geq 1)$. Put $v = \lim_n v_{m(n)}$ $(\|v\| = 1)$. Then,

$$0 = \lim_n [\Psi_\delta^\dagger J_\delta + \Psi_\alpha^\dagger P_1^{m(n)} + \Psi_\omega^\dagger P_2^{m(n)}] v_{m(n)}$$

$$= [\Psi_\delta J_\delta + \Psi_\alpha J_\alpha + 2\Psi_\omega (\Omega_o \otimes I_{(p)})] v,$$

where we have used

$\Psi_\omega (I_{(p)} \otimes \Omega_o) L = \Psi_\omega L (I_{(p)} \otimes \Omega_o) L$ (by the symmetry of Ψ with respect to ω)

$$= \Psi_\omega (\Omega_o \otimes I_{(p)})$$

by Lemma 7. Because $v \neq 0$, $G(\theta_o)$ cannot be of rank p^2.

(b) The proof uses a technique of Rothenberg [1971,p.580]. If $G(\cdot)$ is regular at S_o and has less than full column rank there, then there exists a neighbourhood of S_o, N say, and a vector function $c(\theta)$ defined on N, so that

$$G(\theta) c(\theta) = 0, \qquad \theta \in N \tag{13.20}$$

If N is sufficiently small, then $\forall\ S \in N$, $\det A(z) \neq 0, |z| \leq 1$, and $\det B(z) \neq 0$ for $|z| \leq 1$; it also follows (Deistler, Dunsmuir and Hannan [1977, Theorem 3]) that C1. and C2. are satisfied.

Choosing N small we put $\beta_j = \text{vec } B_j$, $\gamma_k = \text{vec } C_k$ and $\alpha_\ell = \text{vec } A_\ell$; $0 \leq j \leq r$, $0 \leq k \leq s$ and $1 \leq \ell \leq t$. We now define the curve $\{\theta(x), x \geq 0\}$ in N by:

$$d\beta_j(x)/dx = \{B_j'(x) \otimes I_{(p)}\} c(\theta;x), \quad 0 \leq j \leq r$$

$$d\gamma_j(x)/dx = \{C_j'(x) \otimes I_{(p)}\} c(\theta;x), \quad 0 \leq j \leq s$$

$$d\alpha_j(x)/dx = [\{A_j'(x) \otimes I_{(p)}\} - \{I_{(p)} \otimes A_j(x)\}] c(\theta;x),$$
$$1 \leq j \leq t; \quad A_o(x) = I_{(p)}$$

$$d\omega(x)/dx = [\{\Omega(x) \otimes I_{(p)}\} + \{I_{(p)} \otimes \Omega(x)\} L] c(\theta;x)$$

with $\theta(0) = \theta_o$. Denote S by S_x for $\theta = \theta(x)$. We now show that the structures S_x ($x > 0$ and small) are observationally equivalent to S_o and satisfy the restrictions $\psi(\theta) = 0$. First we deal with observational equivalence. For $x > 0$

$$dB_o(x)^{-1}B_j(x)/dx = B_o(x)^{-1}dB_j(x)/dx - B_o(x)^{-1}dB_o(x)/dxB_o(x)^{-1}B_j(x).$$

Hence

$$\text{vec } dB_o(x)^{-1}B_j(x)/dx = \{I_{(p)} \otimes B_o(x)^{-1}\}\{B_j(x)' \otimes I_{(p)}\}c\{\theta(x),x\} -$$
$$- [B_j(x)'B_o(x)^{-1'} \otimes B_o(x)^{-1}]\{B_o(x)' \otimes I_{(p)}\}c\{\theta(x);x\} = 0$$

so that

$$B_o(x)^{-1}B_j(x) = B_o(0)^{-1}B_j(0), \text{ i.e. } B_j(x) = U_xB_j(0), \; j=0,\ldots,r;$$

$U_x = B_o(x)B_o(0)^{-1}$. We can similarly show that $C_j(x) = U_xC_j(0)$, $A_k(x) = U_xA_k(0)U_x^{-1}$ and $\Omega_x = U_x\Omega_oU_x'$ for $j = 0,\ldots,s$ and $k = 1,\ldots,t$. By Lemma 6(ii), the structures S_x ($x > 0$) are observationally equivalent to S_o.

Using Lemma 7 it is easy to check that for $x \geq 0$

$$d\psi\{\theta(x)\}/dx = G\{\theta(x)\}c\{\theta(x),x\}$$

so that $d\psi\{\theta(x)\}/dx = 0$ by (13.20). Hence ψ is constant along the curve. Because the curve passes through S_o we have thus shown that for x small and positive the structures S_x belong to M' and are observationally equivalent to S_o. It follows that S_o cannot be locally identified.

(c) If the conditions in (c) hold, then $G(\theta)$ depends only on S_o and the desired result follows readily from the proofs of (a) and (b). □

Proof of Lemma 10 Put $\Phi(z) = -B_{22}(z)^{-1}B_{21}(z)$, $\Psi(z) = -B_{22}(z)^{-1}C_2(z)$ and let $f_{y_1}(\lambda)$, $f_{y_1x}(\lambda)$ be respectively the spectral density of y_1 and the cross-spectral density of y_1 and x. We immediately obtain that

$$\begin{bmatrix} \tilde{B}(z) \\ \cdots \\ 0 \end{bmatrix} = B(z) \begin{bmatrix} I_{(p)} \\ \cdots \\ \Phi(z) \end{bmatrix}, \quad \begin{bmatrix} \tilde{C}(z) \\ \cdots \\ 0 \end{bmatrix} = C(z) + B(z) \begin{bmatrix} 0 \\ \cdots \\ \Psi(z) \end{bmatrix} \tag{13.25}$$

Putting $\tilde{D}(z) = \tilde{B}(z)^{-1}\tilde{C}(z)$ and $\tilde{G}(z) = \tilde{B}(z)^{-1}A_{11}(z)$ it follows from (9.10) that

$$f_{y_1x}(\lambda) = \tilde{D}(\lambda)f_x(\lambda), \quad f_{y_1}(\lambda) = \tilde{D}(\lambda)f_x\tilde{D}(\lambda)^*+\tilde{G}(\lambda)\Omega_{11}\tilde{G}(\lambda)^*.$$

Therefore, given the second moments of the observed variates, we can obtain $\tilde{D}(\lambda)$ and $\tilde{G}(\lambda)\Omega_{11}\tilde{G}(\lambda)^*$ $(\lambda\in[-\pi,\pi])$ uniquely.

Let $\{\tilde{B}_j(z),\tilde{C}_j(z)\}$, $j = 1,2$, denote $\{\tilde{B}(z),\tilde{C}(z)\}$ at the structures S_1 and S_2 respectively. We now give two sets of NSC's for the structures S_1 and S_2 to be observationally equivalent. The first set of conditions (NSC1.) follows directly from above, while the second set (NSC2.) follows from the first by an application of Lemma 5; for a closely related result see how Lemma 6(i) is shown to be equivalent to Lemma 6(ii).

NSC1. $\tilde{B}_2^{-1}(\lambda)\tilde{C}_2(\lambda) = \tilde{B}_1^{-1}(\lambda)\tilde{C}_1(\lambda)$ and

$$\tilde{B}_2^{-1}A_{11,2}(\lambda)\Omega_{11,2}A_{11,2}(\lambda)^*\tilde{B}_2^{-1}(\lambda)^*=\tilde{B}_1^{-1}A_{11,1}(\lambda)\Omega_{11,1}A_{11,1}(\lambda)^*\tilde{B}_1^{-1}(\lambda)^*$$

for $\lambda \in[-\pi,\pi]$.

NSC2. $[\tilde{B}_2(z),\tilde{C}_2(z),A_{11,2}(z)] =$

$$= \tilde{U}(z)[\tilde{B}_1(z),\tilde{C}_1(z),\tilde{A}_{11,1}(z)F] \ \forall\ z$$

and $\Omega_{11,1} = F\,\Omega_{11,2}F'$, where $\tilde{U}(z) = \tilde{B}_2(z)\tilde{B}_1(z)^{-1}$ and F is a constant $(p_1\times p_1)$ matrix.

Put $U(z) = B_2(z)B_1(z)^{-1}$. From (13.25)

$$B_2(z)^{-1}\begin{pmatrix}\tilde{B}_2(z)\\ \cdots\\ 0\end{pmatrix} = B_1(z)^{-1}\begin{pmatrix}\tilde{B}_1(z)\\ \cdots\\ 0\end{pmatrix}$$

so that

$$U(z)\begin{pmatrix}I_{(p_1)}\\ \cdots\\ 0\end{pmatrix} = \begin{pmatrix}\tilde{U}(z)\\ \cdots\\ 0\end{pmatrix} \qquad (13.26)$$

Suppose that S_1 and S_2 are observationally equivalent. Then, by NSC2., $[\tilde{C}_2(z), A_{11,2}(z)] = \tilde{U}(z)[\tilde{C}_1(z), A_{11,1}(z)F]$, so that

$$\begin{pmatrix}\tilde{C}_2\\ \cdot\cdot\\ 0\end{pmatrix} = C_2 + B_2\begin{pmatrix}0\\ \cdot\cdot\\ \Psi\end{pmatrix} = \begin{pmatrix}\tilde{U}\ \tilde{C}_1\\ \cdots\\ 0\end{pmatrix} = U\begin{pmatrix}\tilde{C}_1\\ \cdot\cdot\\ 0\end{pmatrix} = U\left\{C_1 + B_1\begin{pmatrix}0\\ \cdot\cdot\\ \Psi\end{pmatrix}\right\}$$

using (13.25) and (13.26). Therefore $C_2 = UC_1$ because $B_2 = UB_1$. We can similarly show that

$$\begin{pmatrix}A_{11,2}\\ \cdots\\ 0\end{pmatrix} = U\begin{pmatrix}A_{11,1}\\ \cdots\\ 0\end{pmatrix}F$$

Therefore the conditions of Lemma 10 hold. Conversely, if the conditions of Lemma 10 hold, then by reversing the arguments above we can show that NSC2. holds. □

<u>*Proof of Theorem 11*</u> As in the proof of Theorem 10, if $S := [B(z), C(z), A(z), \Omega]$ is an ARARX structure which is observationally equivalent to S_o, then (10.18) holds and $U(z)$ can be expanded as in (10.19). If S belongs to the model, then

$$K_1(\alpha)G_1 + K_2(\alpha)G_2 = 0,\ F_1K_3 + F_2K_4 = 0,\ F_1K_5 + F_2K_6 = 0$$

where $G_1' = (U_1', \ldots, U_d'), G_2' = (U_{d+1}', \ldots, U_{2d}'), F_j = (U_{(j-1)d+1}, \ldots, U_{jd})$ for $j \geq 1$. Hence,

$\{I_{(p)} \otimes K_2^{-1}(\alpha)K_1(\alpha)\}g_1+g_2 = 0, \{K_4^{-1'} K_3' \otimes I_{(p)}\}f_1 + f_2 = 0$

and $\{K_5' \otimes I_{(p)}\}f_1 + \{K_6' \otimes I_{(p)}\}f_2 = 0,$

where $f_j = \text{vec } F_j$ $(j \geq 1)$ and $g_j = \text{vec } G_j$ $(j=1,2)$. Eliminating g_1 and g_2 by using $g_1 = P_d'f_1$, $g_2 = P_d'f_2$ we obtain that $H(\alpha)f_1 = 0$. As in the proof of Theorem 10, $f_1 = 0$ for S close to S_o and we can then recursively show that $f_j = 0$ for $j > 1$. This again reduces us to the classical identification situation. □

Proof of Corollary 2. We use the notation of the proof of Theorem 11. (i) We know from the proof of Theorem 11 that $F_1(K_5-K_3K_4^{-1}K_6) = 0$, so if (10.24) holds, then $F_1 = 0$ for all structures S belonging to the model and observationally equivalent to S_o. The rest of the proof proceeds as for Theorem 11. Global, rather than local, identification is obtained because the matrix $(K_5-K_3K_4^{-1}K_6)$ depends on S_o only.

(ii) Let $U(z)$ be a g.c.l.d. of $B_o(z)$ and $C_o(z)$ and put

$$V(z) = U(z)^{-1} = \sum_{j=o}^{\infty} V_j z^j , \quad |z| \leq 1.$$

Let $[B(z), C(z)] = V(z)[B_o(z),C_o(z)]$; $B(z)$ and $C(z)$ are matrix polynomials. Let g be the maximum degree of $B_o(z)$, $C_o(z)$, $B(z)$ and $C(z)$. Then

$$V_{j+1} B_{o,o}+\ldots+V_{j+1-d} B_{d,o} = 0, \; j \geq g$$

and

$$V_{j+1} C_{o,o}+\ldots+V_{j+1-d} C_{d,o} = 0, \; j \geq g.$$

As in the proof of Theorem 11 we can now show that $V_j = 0$ for $j \geq g+1$ if (10.24) holds. Hence $V(z)$ is a matrix polynomial so that $U(z)$ is a unimodular matrix;

i.e. $B_o(z)$ and $C_o(z)$ are left coprime. We now show that the rank of $(B_{r,o}, C_{s,o})$ is p. Suppose $U(B_{r,o},C_{s,o}) = 0$ where U is a constant matrix. Put $[B(z),C(z)] = (I_{(p)}+Uz)[B_o(z),C_o(z)]$. Then, because $\deg B(z) \leq \deg B_o(z)$ and $\deg C(z) \leq \deg C_o(z)$, it can be shown as in (i) that $U = 0$ if (10.24) holds. Hence $(B_{r,o},C_{s,o})$ is of rank p. □

Proof of Corollary 3 We use the same notation as in the proof of Theorem 11. To be specific we suppose that $\underline{d} = t$. From the proof of Theorem 11 we know that $H(\alpha)f_1 = 0$. Hence, if (10.25) holds, then in a small neighbourhood of S_o, $T(\alpha)f_1 = 0$ so that $U_j = 0$ for $j=1\ldots,r$. Because

$$A_rU_j + A_{r-1}U_{j+1}+\ldots+ U_{j+r} = 0 \qquad (j \geq 1)$$

we can show that $U_j = 0$ for $j \geq 1$. The rest of the proof follows that of Theorem 10. □

Proof of Theorem 13 It follows from Lemma 3.9 and its proof that each of (i) and (ii) are NSC's for θ_o to be locally identified. We now show that (i) $\Longleftrightarrow$(iii). For any vector x

$$Ex = x \Longleftrightarrow Hx = 0 \tag{13.30}$$

If (iii) does not hold, then there exists an $x \neq 0$ such that $Ex = x$ and $x'E\Delta Ex = 0$ so that $x'\Delta x = 0$ and hence $\Delta x = 0$. From (13.30), (i) cannot hold. Conversely, if (i) does not hold, then there exists an $x \neq 0$ such that $\Delta x = 0$ and $Hx = 0$. Hence $Ex = x$ (by (13.30)) and $x'E\Delta Ex = 0$ so that (iii) does not hold. □

Proof of Lemma 12 (i) There exists an orthogonal $(p\times p)$ matrix M such that (Rao[1973, p.20 and p.28])

$$MEM' = \begin{bmatrix} I_{(f)} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix}$$

Put $B = MAM'$ and partition B as

$$\begin{bmatrix} B_{11} & \vdots & B_{12} \\ \cdots & \vdots & \cdots \\ B_{21} & \vdots & B_{22} \end{bmatrix}$$

with B_{11} $f \times f$. We note that for an arbitrary matrix A and an orthogonal M, $(MAM')^+ = MA^+M'$

$$ME(EAE)^+M' = MEM'(MEM'MAM'MEM')^+ = \begin{bmatrix} B_{11}^+ & 0 \\ 0 & 0 \end{bmatrix};$$

similarly

$$M(EAE)^+M' = \begin{bmatrix} B_{11}^+ & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore $E(EAE)^+ = (EAE)^+$. We can similarly show that $(EAE)^+E = (EAE)^+$.

(ii) If $\rho(E) = \rho(EAE)$, then B_{11} is nonsingular and (ii) follows easily.

(iii) Put $Q = EAE$. A general solution to the equation is (Rao [1973, p.25, (ii)(d)])

$x = x^\dagger + (I_{(p)} - Q^+Q)\eta$ where η is arbitrary. If $Ex = x$, then we deduce that $E(I_{(p)} - Q^+Q)\eta = (I_{(p)} - Q^+Q)\eta$. Hence $(I_{(p)} - Q^+Q)\eta = 0$ by (ii). □

CHAPTER 3

ASYMPTOTIC PROPERTIES OF TIME DOMAIN GAUSSIAN ESTIMATORS

1. INTRODUCTION

In this chapter we obtain parameter estimates by maximizing a particular approximation to a Gaussian Likelihood. In obtaining the asymptotic properties of the estimates, however, we will not require the observations to be Normally distributed. Following Whittle [1962] we call estimates so obtained Gaussian Estimates in order to distinguish them from Maximum Likelihood Estimates. The two coincide if the observations are Normally distributed. We note that our use of the term Gaussian Estimates slightly extends Whittle's usage of it.

Let θ_o be the true value of the parameter θ. We consider models of the form

$$v(n;\theta_o) = e(n), \quad n = 1,2,\ldots \tag{1.10}$$

where $(e(n), F_n, n \geq 1)$ is a sequence of stationary, square-integrable, martingale differences with $E\{e(n)e(n)'|F_{n-1}\} = \Omega_o$. Ω_o is assumed to be a constant nonsingular matrix and $v(n;\theta)$ is a nonlinear vector function of the observations and the parameter θ. We obtain estimates of θ_o and Ω_o by minimizing the criterion function

$$L_N(\theta,\Omega) = \tfrac{1}{2} \log \det \Omega + \tfrac{1}{2} \mathrm{tr}\, \Omega^{-1}\{N^{-1} \sum_{n=1}^{N} v(n;\theta)v(n;\theta)'\} \tag{1.11}$$

over all θ lying in a compact set and satisfying nonlinear restrictions of the form $h(\theta) = 0$, and over all positive definite matrices Ω. If the $e(n)$ are independent and Normally distributed, then $\exp\{-N \times L_N(\theta_o,\Omega)\}$ is the likelihood for $\text{vec}\{e(1),\ldots,e(N)\}$. This observation, together with (1.10), provides the motivation for using (1.11). The way in which (1.11) is an approximation to a Gaussian Likelihood and the reason for using this approximation are given below.

Subject to the assumptions stated below, we prove the strong consistency and asymptotic Normality of the Gaussian Estimator of θ_o, as well as the strong consistency of the Gaussian Estimator of Ω_o.

The Normal Equations will usually be highly nonlinear so that their closed form solution is unlikely to exist. Because of this, we give a Gauss-Newton type algorithm and show that, with probability tending to 1 as the sample size increases, the iterates converge to the Gaussian Estimator if the iteration is initialized by a consistent estimator. In addition, we show that the iterates are consistent and have the same asymptotic distribution as the Gaussian Estimator.

In the next chapter our results are applied to the vector linear time series model

$$\sum_{j=o}^{r} B_j(\theta_o)y(n-j) = \sum_{j=o}^{\infty} C_j(\theta_o)x(n-j) + \sum_{j=o}^{\infty} A_j(\theta_o)e(n-j) \quad n=0,\pm 1,\ldots \tag{1.12}$$

$B_o(\theta) \equiv I_{(p)}$, $A_o(\theta) \equiv I_{(p)}$. The exogenous variables $\{x(n),\ n = 0,\pm 1,\ldots\}$ are assumed to be independent of the $e(n)$. If, for $|z| \leq 1$, $\sum_{j=o}^{\infty} A_j(\theta)z^j$ is a nonsingular and analytic function of z and we put

$$\left(\sum_{j=o}^{\infty} A_j(\theta)z^j\right)^{-1} = \sum_{j=o}^{\infty} D_j(\theta)z^j,$$

then (1.12) can be rewritten as (1.10) with

$$v(n;\theta) = \sum_{k=o}^{\infty} D_k(\theta)\left\{\sum_{j=o}^{r} B_j(\theta)y(n-j-k) - \sum_{j=o}^{\infty} C_j(\theta)x(n-j-k)\right\}$$

Our results have a much wider application than (1.12), however, and apply to general nonlinear models.[1]

We now briefly discuss the way in which $\{-N\times L_N(\theta,\Omega)\}$ is an approximation to the log of a Gaussian Likelihood. Consider the scalar first order AR process

$$y(n) = \theta\, y(n-1) + e(n) \qquad n = 1,\dots,N \qquad (1.15)$$

where $|\theta| < 1$. Put $y_N' = \{y(1),\dots,y(N)\}$, $\Gamma_N = E(y_N y_N')$. If $\{y(n)\}$ is stationary and Gaussian, then $-N^{-1}\times$log-likelihood of y_N is

$$(2N)^{-1} \log \det \Gamma_N + (2N)^{-1} y_N' \Gamma_N^{-1} y_N \qquad (1.16)$$

If we put $v(n;\theta) = y(n) - \theta y(n-1)$ and if $y(o)$ is known, then $L_N(\theta,\Omega)$ is given by (1.11). It is easy to see that (1.11) and (1.16) are not the same. The reason for using (1.11) rather than the exact log-likelihood is that the former will usually be much easier to work with. For example, for the model (1.15), the presence of the matrix Γ_N makes (1.16) a more complicated function of θ than $L_N(\theta,\Omega)$.

Under certain circumstances $\exp\{-N\times L_N(\theta,\Omega)\}$ will be exactly a Gaussian Likelihood; e.g. for the model (1.15) with $y(o)$ a known constant and the $e(n)$ independent and Normal.

Our results may be considered as a generalization of the results given in Jennrich [1969] who considered the scalar model

$$y(n) = f(n;\theta_o) + u(n) \qquad n = 1,2,\dots$$

where the $u(n)$ are independent and identically distributed residuals with zero mean and finite variance $\sigma^2 > 0$. The $f(n;\theta)$ are known continuous nonstochastic functions on a

[1] See the comment on p.135 concerning the application of the theory in this chapter to the model (1.12).

compact subset Θ of Euclidean space. Jennrich proved the consistency and asymptotic Normality of the Gaussian estimator of θ_o and the strong consistency of the Gaussian Estimator of σ^2.

Hannan [1971b] generalized Jennrich, in a direction different from ours, by permitting $u(n)$ to be a stationary residual and, for a frequency domain criterion function, obtained the asymptotic properties of the Gaussian Estimator of θ_o. Robinson [1972a] generalized Hannan's analysis to the multivariate case.

The chapter is structured as follows: §2 contains the assumptions on which subsequent analysis is based, §3 contains the results, while all proofs are placed in §4.

2. ASSUMPTIONS

The underlying probability space is taken as $(\emptyset, F, P)$ with $\emptyset$ having typical element ω.

A1. $\{e(n), F_n, n \geq 0\}$ is a p dimensional sequence of stationary, square-integrable, martingale differences satisfying $E\{e(n)e(n)'|F_{n-1}\} = \Omega_o$ $n \geq 1$, where Ω_o is a constant nonsingular matrix.

A2. Θ, the space of permissible parameter points, is a compact subset of m dimensional Euclidean space. θ_o, the true parameter point, belongs to Θ.

A3. $\{v(n);\theta), n \geq 1\}$ is a p dimensional sequence of random variables depending on the parameter θ such that:

(i) $v(n;\theta_o) = e(n)$ $\forall n \geq 1$ and $\forall \theta, \theta_+ \in \Theta$, $\{v(n;\theta) - v(n;\theta_+)\}$ is F_{n-1} measurable.

Put $\quad \Omega_N(\theta) = N^{-1} \sum_{n=1}^{N} v(n;\theta)v(n;\theta)'$. (2.10)

(ii) a.s., and uniformly in $\theta \in \Theta$, $\Omega_N(\theta)$ converges to a nonstochastic finite limit, $\Omega(\theta)$ say.

(iii) a.s., and uniformly in $\theta \in \Theta$, $N^{-1} \sum_{n=1}^{N} v(n;\theta)e(n)'$ converges to a nonstochastic, finite, limit.

(iv) $\forall\, n \geq 1$ and a.s. $v(n;\theta)$ is a continuous function of θ for $\theta \in \Theta$.

$\forall\, \alpha > 0$ put $U_\alpha = (\theta : \|\theta-\theta_o\| \leq \alpha)$.

A4.(i) θ_o is an interior point of Θ. $\exists\, \alpha_1 > 0$ such that $U_{\alpha_1} \subseteq \Theta$ and $\forall\, n \geq 1$, and almost surely, $v(n;\theta)$ is twice differentiable with respect to θ, with continuous first and second partial derivatives for $\theta \in U_{\alpha_1}$.

Put $V(n;\theta) = (\partial/\partial\theta')v(n;\theta)$, $W^{(j)}(n;\theta) = (\partial/\partial\theta_j)V(n;\theta)$, $V(n) = V(n;\theta_o)$ and $W^{(j)}(n) = W^{(j)}(n;\theta_o)$ $(j = 1,\ldots,m)$.

(ii) $N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial\theta_i} e(n)'$ and $N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial\theta_i} \frac{\partial v(n;\theta)'}{\partial\theta_j}$

converge in probability, uniformly in $\theta \in U_{\alpha_1}$, to finite nonstochastic limits; $(i,j=1,\ldots,m)$.

Put $\lim_N N^{-1} \sum_{n=1}^{N} V(n)'\Omega_o^{-1} V(n) \overset{p}{=} \Delta$ (2.15)

(iii) $N^{-1} \sum_{n=1}^{N} \frac{\partial^2 v(n;\theta)}{\partial\theta_i\partial\theta_j} e(n)'$ and $N^{-1} \sum_{n=1}^{N} \frac{\partial^2 v(n;\theta)}{\partial\theta_i\partial\theta_j} \frac{\partial^2 v(n;\theta)'}{\partial\theta_k\partial\theta_\ell}$

converge in probability, uniformly in $\theta \in U_{\alpha_1}$, to finite nonstochastic limits; $(i,j,k,\ell = 1,\ldots,m)$.

A5.(i) $\lim_N N^{-\frac{1}{2}} \sum_{n=1}^{N} V(n)'\Omega_o^{-1}e(n) \overset{\mathcal{D}}{=} N(0,\Delta)$

(ii) $\limsup_N N^{-1} \sum_{n=1}^{N} E(\|V(n)\|^2) < \infty$

A6.(i) For $\theta \in \Theta$, $h(\theta)$ is an f dimensional, $(0 \leq f < m)$ continuous vector function of θ. In addition $h(\theta_o) = 0$.

(ii) $h(\theta)$ has continuous first and second partial derivatives for $\theta \in U_{\alpha_1}$.

Put $H(\theta) = (\partial/\partial\theta')h(\theta)$, $H = H(\theta_o)$ and define Θ_h as $\Theta_h = \Theta \cap \{h(\theta) = 0\}$.

(iii) H has rank f

(iv) The matrix
$$\begin{bmatrix} \Delta & \vdots & H' \\ \cdots & & \cdots \\ H & \vdots & 0 \end{bmatrix}$$

is nonsingular.

(v) If $\theta \in \Theta_h$ and $\Omega(\theta) = \Omega(\theta_o)$, then $\theta = \theta_o$. □

<u>*Remark 2.1*</u> (i) It follows from A3.(i) and A4.(i) that $(\partial/\partial\theta_i)v(n;\theta)$ and $(\partial^2/\partial\theta_i\partial\theta_j)v(n;\theta)$ are F_{n-1} measurable for $\theta \in U_{\alpha_1}$.

(ii) In practice θ_o is unknown so that we will have to check assumptions A4. and A6. for each point in a subset of Θ, Θ' say, in which we think that θ_o is likely to lie.

(iii) For models of the type (1.12), the requirement that the $e(n)$ are martingale differences is quite a natural one as then the best predictor of $y(n)$ given the past, i.e. $E(y(n)|F_{n-1})$, will be the best linear predictor, both in a least squares sense. For a further discussion of this point see Hannan and Heyde [1972, p.2059].

3. RESULTS

Instead of working with the criterion function (1.11) we can concentrate Ω out to obtain the criterion function

$$\ell_N(\theta) = \frac{1}{2} \log \det \Omega_N(\theta)$$

where $\Omega_N(\theta)$ is defined in (2.10). Below we take $\ell_N(\theta)$ as the criterion function to be minimized. If $\hat{\theta}_N$ minimizes $\ell_N(\theta)$, then we take $\Omega_N(\hat{\theta}_N)$ as the Gaussian Estimate of Ω_o. This is justified by the following lemma.

Lemma 1 $L_N(\theta,\Omega) \geq \ell_N(\theta) + \frac{1}{2}p$ with equality if and only if $\Omega = \Omega_N(\theta)$. □

For N large, and almost surely, $\ell_N(\theta)$ will be well defined because, by Lemma 3, $\Omega_N(\theta)$ will then be positive definite.

We first obtain a strong consistency result.

Theorem 1 Suppose assumptions A1.-A3. and A6.(i), (v) hold. Then,

(i) a.s. $\exists$ a $\hat{\theta}_N$ such that $\hat{\theta}_N$ minimizes $\ell_N(\theta)$ for $\theta \in \Theta_h$. For each N, choosing one such $\hat{\theta}_N$, we have:

(ii) $\lim_N \hat{\theta}_N = \theta_o$ a.s. and $\lim_N \Omega_N(\hat{\theta}_N) = \Omega_o$ a.s. □

Let $\alpha_2 = \frac{1}{2}\alpha_1$ and for each $N \geq 1$ put

$$A_N = (\omega : \hat{\theta}_N \in U_{\alpha_2}) \tag{3.09}$$

We know from Theorem 1 that $\lim_N P(A_N) = 1$. Because $\hat{\theta}_N$ minimizes $\ell_N(\theta)$ we know that for $\omega \in A_N$ there exists an f dimensional vector of Lagrange Multipliers, $\hat{\lambda}_N(\omega)$, such that

$$\frac{\partial \ell_N(\hat{\theta}_N)}{\partial \theta} + H'(\hat{\theta}_N)\hat{\lambda}_N = 0 \tag{3.10}$$

$$h(\hat{\theta}_N) = 0 \tag{3.11}$$

We put $\hat{\lambda}_N(\omega)$ equal to zero for $\omega \notin A_N$. (3.10) and (3.11) are called the Normal Equations.

Put $\Delta_N(\theta) = N^{-1} \sum_{n=1}^{N} V(n;\theta)'\Omega_N^{-1}(\theta)V(n;\theta)$, $J_N(\theta) = \begin{bmatrix} \Delta_N(\theta) & . & H'(\theta) \\ \cdots & & \cdots \\ H(\theta) & . & 0 \end{bmatrix}$,

(3.12)

$$\ddagger_N(\theta) = J_N(\theta)^{-1}\begin{bmatrix} \Delta_N(\theta) & \cdot & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdot & 0 \end{bmatrix} J_N(\theta)^{-1}, \quad J = \begin{bmatrix} \Delta & \cdot & H' \\ \cdots & \cdots & \cdots \\ H & \cdot & 0 \end{bmatrix}$$

$$J^{-1} = \begin{bmatrix} J^{11} & \cdot & J^{12} \\ \cdots & \cdots & \cdots \\ J^{21} & \vdots & J^{22} \end{bmatrix}, \quad \ddagger = \begin{bmatrix} J^{11} & \cdot & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdot & -J^{22} \end{bmatrix} \text{ and } L = I_{(m)} - H'(HH')^{-1}H \quad (3.15)$$

Put $\hat{\tau}_N' = (\hat{\theta}_N', \hat{\lambda}_N')$ $(N \geq 1)$, $\tau_o' = (\theta_o', 0')$. We now come to the central limit theorem.

Theorem 2 Suppose assumptions A1.-A6. hold. Then, $\lim_N N^{\frac{1}{2}}(\hat{\tau}_N - \tau_o) \overset{\mathcal{D}}{=} \mathcal{N}(0, \ddagger)$ and $\ddagger_N(\hat{\theta}_N) \to \ddagger$ a.s. □

Remark 3.1 (i) By the proof of Theorem 2 $\ddagger_N(\hat{\theta}_N)$ is a strongly consistent estimate of $\ddagger$.

(ii) It follows from Theorem 2 and Lemma 5 of §4 that $N^{\frac{1}{2}}(\hat{\theta}_N - \theta_o) \overset{\mathcal{D}}{\to} \mathcal{N}[0, (L\Delta L)^+]$. □

A natural way of obtaining the Gaussian Estimate of θ_o would be to solve the Normal Equations (3.10) and (3.11). In practice, however, these equations are likely to be highly nonlinear so that their closed form solution may not exist. Because of this, it will often prove useful to re-express the problem - of solving the Normal Equations - in a fixed point framework and iterate to the solution, beginning with apriori given initial values of θ and λ. Put $k_N(\theta) = (\partial/\partial\theta')\ell_N(\theta)$ $(\theta \in U_{\alpha_1})$ and define the iteration scheme by

$$\tau_N^{(j+1)} = \tau_N^{(j)} - J_N^{-1}(\theta_N^{(j)})K_N(\tau_N^{(j)}) \quad (j \geq 0) \qquad (3.20)$$

$\tau_N^{(j)} = (\theta_N^{(j)'}, \lambda_N^{(j)'})$, $\tau' = (\theta', \lambda')$ and

$$K_N(\tau) = [\{k_N(\theta) + H'(\theta)\lambda\}', h(\theta)']'; \qquad (3.21)$$

$\tau_N^{(o)}$ is given prior to the commencement of the iteration. (3.20) may be written in the following, simpler, form

$$\tau_N^{(j+1)} = \begin{bmatrix} \theta_N^{(j)} \\ \cdots\cdots \\ 0 \end{bmatrix} - J_N(\theta_N^{(j)})^{-1} \begin{bmatrix} k_N(\theta) \\ \cdots\cdots \\ h(\theta) \end{bmatrix}_{\theta=\theta_N^{(j)}} \quad (j \geq 0)$$

Theorem 3 There exists an $\alpha' > 0$ ($\alpha' < \alpha_1$) such that:

(i) In the set $U_{\alpha'} \times (\lambda : \|\lambda\| \leq \alpha')$, $\hat{\tau}_N$ is, with probability tending to 1 as $N \to \infty$, the unique solution to the Normal Equations.

(ii) If $\tau_N^{(o)} \in U_{\alpha'} \times (\lambda : \|\lambda\| \leq \alpha')$, then, with probability tending to 1 as $N \to \infty$, the iterate $\tau_N^{(j)}$ converges to $\hat{\tau}_N$ as $j \to \infty$. □

Remark 3.2 (i) The algorithm (3.20) is motivated by the Gauss-Newton algorithm (see 5.§2) and the iteration scheme given in Aitchison and Silvey [1958, p.827]. It is clear that if there were no restrictions, then (3.20) would be exactly the Gauss-Newton iteration scheme.

(ii) By assumptions A4.(ii) and A6.(iv) and equation (2.15), $J_N(\theta)$ will be nonsingular, with probability tending to 1 as $N \to \infty$, for θ close to θ_o. Hence the algorithm given in (3.20) will be well defined if the $\theta_N^{(j)}$ are close to θ_o. □

The next theorem deals with the asymptotic properties of the iterates $\tau_N^{(j)}$, ($j \geq 1$), when the iteration scheme (3.20) is initialized by $\tau_N^{(o)}$, where $\tau_N^{(o)\prime} = (\theta_N^{(o)\prime}, 0')$ and $\theta_N^{(o)}$ is a consistent estimator of θ_o.

Theorem 4 Suppose that for the iteration scheme (3.20) we put $\tau_N^{(o)\prime} = (\theta_N^{(o)\prime}, 0')$.

(i) If $\theta_N^{(o)} \to \theta_o$ p, then, with probability tending to 1 as $N \to \infty$, $\lim_j \tau_N^{(j)} = \hat{\tau}_N$.

(ii) If, in addition, $N^{\frac{1}{2}}(\theta_N^{(o)}-\theta_o) = O_p(1)$, then $N^{\frac{1}{2}}(\tau_N^{(j)}-\tau_o) \xrightarrow{\mathcal{D}} N(0,\mathfrak{F})$ for $j \geq 1$. □

Corollary 1 If we replace convergence in probability by almost sure convergence in assumptions A4.(ii) and A4.(iii) and we also have that $\theta_N^{(o)} \to \theta_o$ a.s. in Theorem 4(i), then we can replace the words "with probability tending to 1 as $N \to \infty$" by "almost surely for N large" in Theorem 3 and Theorem 4(i). □

Remark 3.3 (i) Corollary 1 generalizes Theorem 8 in Jennrich [1969].

(ii) Subject to the assumptions in Theorem 4 and Corollary 1, the iterates $\theta_N^{(j)}$ are consistent and asymptotically as efficient as the Gaussian Estimate $\hat{\theta}_N$. Thus, to simplify the computation, we may prefer to choose as the estimate of θ_o, $\theta_N^{(j)}$, for some small positive j, rather than iterate to convergence. Whether we do so or not will depend on the size of N and the class of models we are considering.

(iii) We show in §§2 and 3 of Chapter 5 that the Gauss-Newton iterates are easy to obtain once we have subroutines to compute $v(n;\theta)$ and $V(n;\theta)$. This is especially so when θ is unrestricted or the restrictions can be substituted out. □

For some problems assumptions A5.(i) and A6.(iv) may be quite difficult to check from first principles. Because of this the following simple sufficient conditions may prove useful in applications.

Put

$$\ell(\theta) = \frac{1}{2}\log\det\Omega(\theta) \quad \forall\, \theta \in \Theta \tag{3.24}$$

It follows from the proof of Theorem 1 that $\ell_N(\theta) \to \ell(\theta)$ a.s. for $\theta \in \Theta$. We now make the following additional assumption:

A7. θ_o is a regular point of the matrix $[(\partial^2/\partial\theta\partial\theta')\ell(\theta),\ H(\theta)']'$, i.e. the matrix has constant rank in some neighbourhood of θ_o.

This assumption involves little loss in generality because (Fisher [1966, p.166, Theorem 5.A.2]) the set of nonregular points has Lebesgue measure zero.

<u>*Theorem 5*</u> Suppose that assumptions A1.-A4.,A6.(i), A6.(ii), A6.(iii) and A7. hold. If, in addition,

$$N^{-1}\sum_{n=1}^{N}\frac{\partial v(n;\theta)}{\partial\theta_i}\frac{\partial v(n;\theta)'}{\partial\theta_j} \quad \text{and} \quad N^{-1}\sum_{n=1}^{N}\frac{\partial^2 v(n;\theta)}{\partial\theta_i\partial\theta_j}v(n;\theta)'$$

$(i,j = 1,\ldots,m)$, converge almost surely, and uniformly in $\theta \in U_{\alpha_1}$, to finite nonstochastic limits, then assumption A6.(v) implies assumption A6.(iv). □

For an application of the next theorem see the proof of Theorem 4.2. We will not discuss its meaning here.

<u>*Theorem 6*</u> Suppose that for each $n \geq 1$ we can write

$$V(n) = V_r(n) + \tilde{V}_r(n), \quad r = 1,2,\ldots \tag{3.25}$$

and

$$V_r(n) = V_{1r}(n) + V_{2r}(n) \tag{3.26}$$

where the decompositions (3.25) and (3.26) have the following properties:

(i) $V_{1r}(n)$ is stationary, F_{n-1} measurable, and square-integrable; $V_{2r}(n)$ is nonstochastic.

(ii) $\lim_N N^{-1}\sum_{n=1}^{N} V_{2r}(n)V_{2r}(n)'$ exists and is finite.

(iii)(a) $\lim_N N^{-1}\sum_{n=1}^{N} V_r(n)'\Omega_o^{-1}V_r(n) \overset{p}{=} \Delta_r$ $(r \geq 1)$, where Δ_r is a nonstochastic matrix.

(b) $\lim_r \Delta_r = \Delta$.

(iv) $\forall\ \varepsilon > 0\ \exists\, R = R(\varepsilon)$ such that

$$\limsup_N N^{-1} \sum_{n=1}^{N} E(\|\tilde{V}_r(n)\|^2) < \varepsilon \quad \text{if} \quad r \geq R.$$

Put $u_N = N^{-\frac{1}{2}} \sum_{n=1}^{N} V(n)'\Omega_o^{-1} e(n)$. Then $u_N \xrightarrow{\mathcal{D}} N(0,\Delta)$.

4. PROOFS

Throughout this section c is a universal constant, not always the same.

Proof of Lemma 1 Let $\lambda_i (i = 1,\ldots,p)$ be the eigenvalues of $\Omega^{-\frac{1}{2}}\Omega_N(\theta)\Omega^{-\frac{1}{2}}$. $L_N(\theta,\Omega)-\ell_N(\theta) = \frac{1}{2} \sum_{i=1}^{p} (\lambda_i - \log \lambda_i) \geq p/2$ with equality if and only if $\lambda_i = 1$ for $i=1,\ldots,p$. This is because $\lambda - \log \lambda \geq 1$ for $\lambda > 0$ with equality if and only if $\lambda = 1$. □

To prove Theorem 1 we need the following two lemmas.

Lemma 2 Suppose that $(a_n, G_n, n \geq 0)$ is a scalar sequence of square - integrable martingale differences, b_n is G_{n-1} measurable for $n \geq 1$, and there exists a non-stochastic positive constant c such that $E(a_n^2 | G_{n-1}) \leq c$ a.s. $\forall\ n \geq 1$.

(i) If $\limsup_N N^{-1} \sum_{n=1}^{N} b_n^2 \leq c < \infty$(a.s.), then

$\lim_N N^{-1} \sum_{n=1}^{N} a_n b_n = 0$ a.s..

(ii) If $\lim_N P(N^{-1} \sum_{n=1}^{N} b_n^2 \leq c) = 1$ for some finite positive constant c, then

$$\lim_N N^{-1} \sum_{n=1}^{N} a_n b_n = 0\ p.$$

Proof (i) Put $B_n = 1 + \sum_{j=1}^{n} b_j^2$ and $Z_n = \sum_{j=1}^{n} B_j^{-1} b_j a_j$; $(Z_n, G_n, n \geq 1)$ is a square-integrable martingale.

$$E(Z_n^2) \leq c\, E\left\{\sum_{j=1}^{\infty} B_j^{-2} b_j^2\right\} \leq c \int_0^{\infty} (1+t)^{-2} dt < \infty$$

the analysis proceeding as in Stout [1974, p.157, Theorem 3.3.10]. It follows from the Martingale Convergence Theorem (Stout [1974, p.42, Theorem 2.7.2]) that there exists a random variable Z such that $E(|Z|) < \infty$ and $Z_n \to Z$ a.s.. Hence

$$N^{-1} \sum_{n=1}^{N} b_n a_n = N^{-1} \sum_{n=1}^{N-1} (B_n - B_{n+1})(Z_n - Z) + N^{-1} B_N (Z_N - Z) + N^{-1} B_1 Z \to 0 \text{ a.s.,}$$

the proof proceeding in the same way as the proof of the Kronecker Lemma (e.g. Neveu [1975, Lemma VII - 2-5, p.152]). The proof of (ii) is almost the same as that of (i) and is omitted. □

Lemma 3 $$\Omega(\theta_o) = \lim_N N^{-1} \sum_{n=1}^{N} e(n)e(n)' \overset{a.s.}{=} \Omega_o. \qquad (4.10)$$

and

$$\Omega(\theta) \geq \Omega(\theta_o) \forall\ \theta \in \Theta \qquad (4.11).$$

($\Omega(\theta)$ is defined in assumption A3.(ii)).

Proof (4.10) follows immediately from assumption A1. and Heyde [1973, Lemma p.150].

$$\Omega_N(\theta) - \Omega_N(\theta_o) \geq N^{-1} \sum_{n=1}^{N} [e(n)\{v(n;\theta) - e(n)\}' + \{v(n;\theta) - e(n)\}e(n)'] \qquad (4.12)$$

For fixed $\theta \in \Theta$, the right side of (4.12) converges to zero almost surely by Lemma 2(i) and assumptions A1. and A3. (4.11) now follows from (4.10). □

Proof of Theorem 1 (i) Θ_h is compact because Θ is compact and $h(\theta)$ is continuous. The required result now follows by noting that $\ell_N(\theta)$ is a continuous function of θ

by A3.(iv).

(ii) Put $\ell(\theta) = \frac{1}{2} \log \det \Omega(\theta)$. By (4.11) of Lemma 3 and assumption A3.(ii), $\lim_N \ell_N(\theta) = \ell(\theta)$ a.s. $\geq \ell(\theta_o)$, the convergence being uniform in $\theta \in \Theta$. We note also that $\ell(\theta)$ is continuous in $\theta \in \Theta$. The proof that $\hat{\theta}_N \to \theta_o$ now proceeds by contradiction as in Jennrich [1969, p.638, Theorem 6].

If $\hat{\theta}_N \to \theta_o$, then the proof is complete. If it does not, then there exists a $\dot{\theta} \in \Theta_h - \theta_o$ and a subsequence $\{\hat{\theta}_{M(N)}, N \geq 1\}$ of $\{\hat{\theta}_N\}$ such that $\lim_N \hat{\theta}_{M(N)} = \dot{\theta}$. It follows that

$$\ell(\dot{\theta}) = \lim_N \ell_{M(N)}(\hat{\theta}_{M(N)}) \leq \lim_N \ell_{M(N)}(\theta_o) = \ell(\theta_o).$$

But, from above, $\ell(\dot{\theta}) \geq \ell(\theta_o)$. Hence $\ell(\dot{\theta}) = \ell(\theta_o)$ implying that $\Omega(\dot{\theta}) = \Omega(\theta_o)$. It follows from assumption A6.(v) that $\dot{\theta} = \theta_o$, giving us a contradiction. Therefore $\hat{\theta}_N \to \theta_o$ a.s. The proof that $\Omega_N(\hat{\theta}_N) \to \Omega_o$ a.s. is now immediate. □

To prove Theorem 2 we need the following lemmas.

Lemma 4 Let X be a subset of Euclidean space.

(a) Suppose that: (i) $(f_n(x), n \geq 1)$ is a sequence of stochastic functions that are continuous in $x \in X$ a.s..

(ii) $\lim_n f_n(x) = f(x)$ a.s., the convergence being uniform in $x \in X$.

(iii) $f(x)$ is finite a.s. $\forall\ x \in X$.

If $(x_n, n \geq 1)$ is a sequence of random variables, $\bar{x}$ an interior point of X and $x_n \to \bar{x}$ a.s., then $f_n(x_n) \to f(\bar{x})$ a.s.

(b) Suppose that: (i) There exists a sequence of subsets of $\emptyset$, $(B_n, n \geq 1)$ say, such that $\lim_n P(B_n) = 1$.

(ii) $(f_n(x), n \geq 1)$ is a sequence of random functions of

x such that $f_n(x)$ is continuous in $x \in X$ for $\omega \in B_n$.

(iii) $f_n(x) \to f(x)$ p, the convergence being uniform in $x \in X$, and $f(x)$ is finite a.s. $\forall\, x \in X$.

If $(x_n,\ n \geq 1)$ is a sequence of random variables, $\bar{x}$ an interior point of X and $x_n \to \bar{x}$ p, then $f_n(x_n) \to f(\bar{x})$ p.

Proof (a) By Rudin [1964, p.136, Theorem 7.12] $f(x)$ is continuous at $\bar{x}$. $x_n \in X$ for n sufficiently large because $\bar{x}$ is an interior point of X. Hence

$$|f_n(x_n) - f(\bar{x})| \leq |f_n(x_n) - f(x_n)| + |f(x_n) - f(\bar{x})| \to 0 \text{ a.s.}$$

by the uniform convergence of the $\{f_n(x),\ n \geq 1\}$ sequence and the continuity of $f(x)$ at $\bar{x}$.

(b) We first show that $f(x)$ is continuous at $\bar{x}$ a.s. $\forall\ \delta > 0$ let $A_n(\delta) = (\omega: \sup_{y \in X} |f_n(y) - f(y)| \leq \delta) \cap B_n$. For h sufficiently small $(\bar{x}+h) \in X$. For $\omega \in A_n(\delta)$,

$$|f(\bar{x}+h) - f(\bar{x})| \leq |f(\bar{x}+h) - f_n(\bar{x}+h)| + |f_n(\bar{x}+h) - f_n(\bar{x})| +$$
$$+ |f_n(\bar{x}) - f(\bar{x})| \leq 2\delta + |f_n(\bar{x}+h) - f_n(\bar{x})| \leq 3\delta$$

if we fix n and take h sufficiently small. The desired result follows by noting that $P\{A_n(\delta)\} \to 1$ as $n \to \infty$, δ is arbitrary and $|f(\bar{x}+h) - f(\bar{x})|$ is independent of n and δ. Hence

$$|f_n(x_n) - f(\bar{x})| \leq |f_n(x_n) - f(x_n)| + |f(x_n) - f(\bar{x})| \to 0 \text{ p}$$

because of the uniform convergence of the $\{f_n(x),\ n \geq 1\}$ sequence and the continuity of $f(x)$. $\square$

Lemma 5 We take J^{-1}, $\ddagger$ and L to be defined as in equation (3.15). Then

$$J^{-1}\begin{pmatrix} \Delta & \cdot & 0 \\ \cdot & \cdot & \cdot \\ 0 & \cdot & 0 \end{pmatrix} J^{-1} = \ddagger \quad \text{and} \quad J^{11} = (L\Delta L)^{+}$$

Proof It can be readily verified that

$$J^{-1}\begin{pmatrix} \Delta & . & 0 \\ \cdots & \cdots & \cdots \\ 0 & . & 0 \end{pmatrix} J^{-1} = \begin{pmatrix} J^{11}\Delta J^{11} & . & J^{11}\Delta J^{12} \\ \cdots & \cdots & \cdots \\ J^{21}\Delta J^{11} & . & J^{21}\Delta J^{12} \end{pmatrix},$$

$$\Delta J^{11} + H'J^{21} = I_{(m)} \tag{4.20}$$

$$HJ^{11} = 0 \tag{4.21}$$

$$HJ^{12} = I_{(f)} \tag{4.22}$$

$$\Delta J^{12} + H'J^{22} = 0 \tag{4.23}$$

From (4.21) $LJ^{11} = J^{11}$. It follows that $(L\Delta L)J^{11} = L$. Put $M = L\Delta L$. By Rao [1973, (ii)(d) p.25]

$$J^{11} = M^{+}L + (I_{(m)} - M^{+}M)\eta$$

for some matrix η. By Theorem 2.13 rank (L) = rank$(L\Delta L)$. It follows from Lemma 12 of 2.§11 that $LM^{+} = M^{+}$ so that $(I_{(m)} - M^{+}M)\eta = 0$, i.e. $J^{11} = (L\Delta L)^{+}$. From (4.20) $J^{11}\Delta J^{12} = J^{12} - J^{12}HJ^{12} = 0$ using (4.22). From (4.23) $J^{21}\Delta J^{12} = -(HJ^{12})'J^{22} = -J^{22}$ using (4.22). □

Lemma 6 Suppose assumptions A1.-A6. hold. Then.

$$N^{\frac{1}{2}}k_N(\theta_o) \xrightarrow{\mathcal{D}} N(0,\Delta). \tag{4.24}$$

Proof

$$N^{\frac{1}{2}}k_N(\theta_o) = N^{-\frac{1}{2}}\sum_{n=1}^{N} V(n)'\{\Omega_N^{-1}(\theta_o) - \Omega_o^{-1}\}e(n) + N^{-\frac{1}{2}}\sum_{n=1}^{N} V(n)'\Omega_o^{-1}e(n) \tag{4.25}$$

By (4.10) $\Omega_N^{-1}(\theta_o) \to \Omega_o^{-1}$ a.s.. By Chebyshev's inequality and assumption A5.(ii), $\forall\delta > 0$

$$P(|N^{-\frac{1}{2}}\sum_{n=1}^{N} V_{ij}(n)e_k(n)| > K) \le c\,N^{-1}\sum_{n=1}^{N} E\{V_{ij}(n)^2\}/K < \delta$$

for K sufficiently large. It now follows that the first term on the right in (4.25) tends to zero in probability, the proof proceeding as in Rao [1973, p.122 (x)(a)]. (4.24) now follows from assumption A5.(i) and Rao [1973, p.122(x)(b)].□

Lemma 7 Let $\Delta^{(j)}$ be the j^{th} column of Δ, θ_j the j^{th} element of θ, $1 \le j \le m$, and suppose that assumptions A1. - A6. hold. If $\tilde{\theta}_N \to \theta_o$ p, then

$$(\partial/\partial\theta_j)k_N(\tilde{\theta}_N) \overset{p}{\to} \Delta^{(j)} \tag{4.26}$$

Proof

$$\frac{\partial k_N(\theta)}{\partial\theta_j} = N^{-1}\sum_{n=1}^{N}(\partial/\partial\theta_j)\{V(n;\theta)'\Omega_N(\theta)^{-1}v(n;\theta)\} =$$

$$= N^{-1}\sum_{n=1}^{N} W^{(j)}(n;\theta)'\Omega_N^{-1}(\theta)v(n;\theta)+N^{-1}\sum_{n=1}^{N} V(n;\theta)'\frac{\partial}{\partial\theta_j}\{\Omega_N^{-1}(\theta)\}v(n;\theta)+$$

$$+ N^{-1}\sum_{n=1}^{N} V(n;\theta)'\Omega_N^{-1}(\theta)\frac{\partial v(n;\theta)}{\partial\theta_j} \tag{4.27}$$

$$= I_N(\theta) + II_N(\theta) + III_N(\theta), \text{ say.}$$

$$I_N(\theta) = N^{-1}\sum_{n=1}^{N} W^{(j)}(n;\theta)'\Omega_N(\theta)^{-1}\{v(n;\theta)-e(n)\}+$$

$$+ N^{-1}\sum_{n=1}^{N} W^{(j)}(n;\theta)'\Omega_N^{-1}(\theta)e(n) \tag{4.28}$$

By assumption A3., $N^{-1}\Sigma_n\|v(n;\theta)-e(n)\|^2$ converges a.s., and uniformly in θ, to a finite continuous, nonstochastic limit. Hence

$$N^{-1}\sum_{n=1}^{N}\|v(n;\tilde{\theta}_N) - e(n)\|^2 \overset{p}{\to} 0 \tag{4.29}$$

By an application of the Cauchy-Schwarz inequality and (4.29) it follows that the first term on the right in (4.28), when evaluated at $\tilde{\theta}_N$, tends to zero in probability. By assumption A4.(iii), Lemma 4b and Lemma 2(ii), the second term on the right side of (4.28), when evaluated at $\tilde{\theta}_N$, tends to zero in probability. Thus, $I_N(\tilde{\theta}_N) \overset{p}{\to} 0$ and similarly $II_N(\tilde{\theta}_N) \overset{p}{\to} 0$.

Finally,

$$\lim_N N^{-1} \sum_{n=1}^{N} V(n;\tilde{\theta}_N)'\Omega_N^{-1}(\tilde{\theta}_N)\{(\partial/\partial\theta_j)v(n;\tilde{\theta}_N)\} \overset{p}{=} \Delta^{(j)}$$

by Lemma 4(b), assumption A4.(ii), (2.15) and (4.10). Therefore (4.26) holds. □

Proof of Theorem 2 First we expand $k_N(\hat{\theta}_N)$ and $h(\hat{\theta}_N)$ about θ_o to obtain

$$k_N(\hat{\theta}_N) = k_N(\theta_o) + \{(\partial/\partial\theta')k_N^*(\theta_N^*)\}(\hat{\theta}_N - \theta_o) \tag{4.30}$$

$$0 = H^{**}(\theta_N^{**})(\hat{\theta}_N - \theta_o) \tag{4.31}$$

where the symbolism $\{(\partial/\partial\theta')k_N^*(\theta_N^*)\}$ is meant to indicate, here and in the sequel, that each row of $\{(\partial/\partial\theta')k_N^*(\partial_N^*)\}$ is to be evaluated at a possibly different θ_N^* such that $\|\theta_N^* - \theta_o\| \leq \|\hat{\theta}_N - \theta_o\|$.

A similar explanation holds for $H^{**}(\theta_N^{**})$.

From (3.10), (3.11), (4.30) and (4.31) it follows that

$$\begin{bmatrix} (\partial/\partial\theta')k_N^*(\theta_N^*) & \vdots & H'(\hat{\theta}_N) \\ \cdots\cdots & \vdots & \cdots\cdots \\ H^{**}(\theta_N^{**}) & \vdots & 0 \end{bmatrix} \cdot N^{\frac{1}{2}}(\hat{\tau}_N - \tau_o) = -\begin{bmatrix} N^{\frac{1}{2}}k_N(\theta_o) \\ \cdots\cdots \\ 0 \end{bmatrix}$$

for $\omega \in A_N$, where the set A_N is defined in (3.09). By (4.24), (4.26), Lemma 5, assumptions A6.(ii) and A6.(iv), and Rao [1973, p.122 (x)(d)] we obtain that

$$N^{\frac{1}{2}}(\hat{\tau}_N - \tau_o) \xrightarrow{\mathcal{D}} N(0,\ddagger). \quad \square$$

To prove Theorem 3 we need the results of the following lemma. Its proof is modelled on that of a contraction mapping theorem (e.g. Petrovski [1966, p.44, Theorem 1]).

Lemma 8 Suppose that there exists a point x_o having a neighbourhood $X_\alpha = (x : \|x - x_o\| \le \alpha)$ and $\{f_n(x), n \ge 1\}$ is a sequence of stochastic functions defined on X_α such that:

(a) X_α is a subset of s dimensional Euclidean space for some positive integer s and the $f_n(x)$ are s dimensional vector functions.

(b) $f_n(x_o) \to x_o$ p.

(c) There exists a sequence of subsets of $\emptyset$, $(B_n, n \ge 1)$, such that $\lim_n P(B_n) = 1$ and $\forall\ x \in X_\alpha$, $n \ge 1$ and $\omega \in B_n$, $f_n(x)$ is differentiable with respect to x and has a continuous first derivative.

Put $F_n(x) = (\partial/\partial x')f_n(x)$ whenever the derivative exists.

(d) $\lim_n P(\sup_{x \in X_\alpha} \|F_n(x)\| < {}^3/4) = 1$

For each $n \ge 1$ we define the sequence $(x_n^{(j)}, j \ge 1)$ by

$$x_n^{(j+1)} = f_n(x_n^{(j)}),\ j \ge 0,$$

with $x_n^{(o)}$ given apriori.

If $x_n^{(o)} \in X_\alpha$, then, with probability tending to 1 as $n \to \infty$, there exists an $\tilde{x}_n$ such that: (i) $\lim_j x_n^{(j)} = \tilde{x}_n$, (ii) $\tilde{x}_n \in X_\alpha$, and (iii) $\tilde{x}_n$ is the unique solution of the equation

$$x = f_n(x) \quad \ldots \text{ for } \quad x \in X_\alpha \quad . \tag{4.35}$$

Proof By (b), (c) and (d) there exists a sequence of subsets of $\emptyset$, $(C_n, n \ge 1)$ say, such that $\lim_n P(C_n) = 1$ and for $\omega \in C_n$

$$\sup_{x \in X_\alpha} \|F_n(x)\| < {}^3/_4 \quad \text{and} \quad \|f_n(x_o) - x_o\| < \tfrac{1}{4}\alpha.$$

For convenience we drop the subscript n from $x_n^{(k)}$. We first show that if $\omega \in C_n$ and $x^{(o)} \in X_\alpha$, then $x^{(k)} \in X_\alpha$ $\forall$ $k \geq 1$. Proceeding by induction, suppose that $x^{(j)} \in X_\alpha$ for $j \leq k$. Then

$$\|x^{(k+1)}-x_o\| \leq \|f_n(x^{(k)})-f_n(x_o)\|+ \|f_n(x_o)-x_o\| \leq {}^3/_4\|x^{(k)}-x_o\| + $$
$$+ {}^1\!/_4\alpha \leq \alpha.$$

Hence $x^{(k)}\in X_\alpha$ $\forall$ $k \geq 1$. Iterating backwards we have that $\|x^{(k)}-x^{(k-1)}\| \leq ({}^3/_4)^{k-1}\|x^{(1)}-x^{(o)}\|$ $\forall$ $k \geq 1$, so that

$$\sum_{k=1}^{\infty} \|x^{(k)}-x^{(k-1)}\| + \|x^{(o)}\| < \infty.$$

Thus for $\omega \in C_n$ and $x^{(o)} \in X_\alpha$, the series

$$\sum_{k=1}^{\infty} (x^{(k)}-x^{(k-1)}) + x^{(o)}$$

is absolutely convergent and hence convergent. (i) now follows. We now show that (ii) holds. From (i), $\tilde{x}_n$ is a limit point of X_α. Because X_α is compact, $\tilde{x}_n \in X_\alpha$.

$$\|\tilde{x}_n-f_n(\tilde{x}_n)\| \leq \|\tilde{x}_n-x_n^{(j+1)}\|+ \|f_n(x_n^{(j)})-f_n(\tilde{x}_n)\| \leq$$
$$\leq \|\tilde{x}_n-\tilde{x}_n^{(j+1)}\|+ {}^3/_4\|x_n^{(j)}-\tilde{x}_n\|$$

Since the right side of this last set of inequalities can be made arbitrarily small by increasing j, while the left side is independent of j, $\tilde{x}_n$ must be a solution of (4.35). To prove uniqueness, suppose that $\hat{x}_n$ is another solution of (4.35), with $\hat{x}_n \in X_\alpha$. Then $\|\hat{x}_n-\tilde{x}_n\| \leq {}^3/_4\|\hat{x}_n-\tilde{x}_n\| \Rightarrow \hat{x}_n=\tilde{x}_n$.□

Corollary 2 If in Lemma 8 conditions (b), (c) and (d) are changed to:

(b') $\lim_n f_n(x_o) = x_o$ a.s.,

(c') $\forall$ $x \in X_\alpha$ and $\forall$ $n \geq 1$ and a.s. $f_n(x)$ is differentiable with respect to x and has a continuous first derivative, (d') $\limsup_n \sup_{x \in X_\alpha} \|F_n(x)\| < 3/4$ a.s.,

then for $x_n^{(o)} \in X_\alpha$ (i), (ii) and (iii) of Lemma 8 hold a.s., for n sufficiently large.

Proof The proof is very similar to that of Lemma 8 and is omitted. □

Proof of Theorem 3 In the proof of this theorem α will stand for a positive constant, not always the same, such that $\alpha \leq \alpha_1$, where α_1 is defined in assumption A4.(i). Put $f_N(\tau) = \tau - J_N(\theta)^{-1} K_N(\tau)$, where $J_N(\theta)$ and $K_N(\tau)$ are defined in (3.12) and (3.21) respectively. The proof of the theorem consists of checking that conditions (b), (c) and (d) of Lemma 8 are satisfied. Once this is shown the result of the theorem follows because $\hat{\tau}_N \to \tau_o$ a.s. and with probability tending to 1 as $N \to \infty$, $\hat{\tau}_N$ is a solution of the Normal Equations.

By assumptions A3.(ii) and A4.(ii), $J_N(\theta)$ converges in probability, uniformly in $\theta \in U_{\alpha_1}$, to a finite nonstochastic limit which can be shown to be continuous. It follows from assumption A6.(iv) that $\lim_N J_N(\theta_o)$ is nonsingular. Hence there exists an $\alpha > 0$ and a sequence of sets $(B_N, N \geq 1)$ with $\lim_N P(B_N) = 1$ such that for $\omega \in B_N$ and $\theta \in U_\alpha$, $J_N(\theta)^{-1}$ is differentiable with respect to θ and has a continuous derivative. It follows from assumptions A4. and A6. that condition (c) of Lemma 8, when applied to $f_N(\tau)$, is satisfied.

Put $F_N(\tau) = (\partial/\partial\tau') f_N(\tau)$. Then,

$$F_N(\tau) = I_{(m+f)} - \{K_N(\tau)'\otimes I_{(m+f)}\}[(\partial/\partial\tau')\text{vec } J_N(\theta)^{-1}] -$$

$$- J_N(\theta)^{-1}\{(\partial/\partial\tau')K_N(\tau)\} \tag{4.40}$$

with

$$\frac{\partial K_N(\tau)}{\partial\tau'} = \begin{pmatrix} (\partial/\partial\theta')k_N(\theta)+(\lambda'\otimes I_{(m)})\{(\partial/\partial\theta')\text{vec } H'(\theta)\}. & H'(\theta) \\ & . \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots \\ H(\theta) & . \quad 0 \end{pmatrix} \tag{4.41}$$

Let θ_j be the j^{th} element of θ. $(\partial/\partial\theta_j)k_N(\theta)$ is given by (4.27). Let $T_{1j}(\theta)$ and $T_{2j}(\theta)$ be, respectively, the first and second terms in (4.28). An application of the Cauchy-Schwarz inequality, the discussion immediately following (4.28), together with assumptions A3.(ii) and A4.(iii) enable us to deduce that $\forall\ \varepsilon > 0\ \exists\ \alpha = \alpha(\varepsilon)$ such that

$$\lim_N P(\sup_{\theta\in U_\alpha} \|T_{1j}(\theta)\| \leq \varepsilon) = 1 \qquad j = 1,\ldots,m.$$

By assumption A4.(iii) and Lemma 2(ii) the second term in (4.28) tends to zero in probability, uniformly in $\theta \in U_{\alpha_1}$. Hence $\forall\ \varepsilon > 0\ \exists\ \alpha = \alpha(\varepsilon)$ such that

$$\lim_N P(\sup_{\theta\in U_\alpha} \|T_{1j}(\theta) + T_{2j}(\theta)\| \leq \varepsilon) = 1 \quad j = 1,\ldots,m.$$

A similar result may be obtained for the secondterm on the right in (4.27). For any $\alpha > 0$ put $Q_\alpha = U_\alpha \times (\lambda : \|\lambda\| \leq \alpha)$. It now follows from (4.27), (3.12), (4.41), assumption A6.(ii), and the argument above, that $\forall\ \varepsilon > 0\ \exists\ \alpha = \alpha(\varepsilon)$ such that

$$\lim_N P(\sup_{\tau\in Q_\alpha} \|(\partial/\partial\tau')K_N(\tau) - J_N(\theta)\| \leq \varepsilon) = 1 \tag{4.42}$$

We can show similarly that $\forall\ \varepsilon > 0,\ \exists\ \alpha = \alpha(\varepsilon)$ and a finite positive constant c such that:

$$\lim_N P(\sup_{\tau\in Q_\alpha} \|K_N(\tau)\| \leq \varepsilon) = 1 \tag{4.43}$$

$$\lim_N P(\sup_{\theta\in U_\alpha} \|(\partial/\partial\theta_j)J_N(\theta)\| \leq c) = 1 \qquad j = 1,\ldots,m \quad (4.44)$$

$$\lim_N P(\sup_{\theta\in U_\alpha} \|J_N(\theta)^{-1}\| \leq c) = 1 \qquad (4.45)$$

From (4.40) - (4.45) we can now deduce that $\exists\ \alpha' > 0$ such that

$$\lim_N P(\sup_{\tau\in Q_{\alpha'}} \|F_N(\tau)\| < {}^3/_4) = 1.$$

It follows that condition (d) of Lemma 8 is satisfied. Finally, we show that condition (b) is also satisfied. With $\tau_o' = (\theta_o', 0')$ we have that $f_N(\tau_o)=\tau_o-J_N(\theta_o)^{-1}\{k_N(\theta_o)',0'\}'$. By (4.24) $k_N(\theta_o) \to 0$ p, and it now follows from (4.45) that $f_n(\tau_o) \to \tau_o$ p. This completes the proof of the theorem. □

Proof of Theorem 4 (i) follows from Theorem 3(ii). (ii) is obtained using a standard argument (e.g. Dhrymes and Taylor [1976, p.370, Lemma 10]). We note, however, that the requirement in Dhrymes and Taylor that $N^{\frac{1}{2}}(\theta_N^{(o)} - \theta_o)$ converges in distribution is unnecessary, the condition given in Theorem 4(ii) being sufficient for the argument to go through.

We proceed by induction. Suppose that for $j = k$

$$N^{\frac{1}{2}}(\tau_N^{(k)} - \tau_o) = 0_p(1) \qquad (4.50)$$

For convenience we put $\tilde{\tau} = \tau_N^{(k+1)}$, $\hat{\tau} = \tau_N^{(k)}$. From (3.20)

$$\tilde{\tau} = \hat{\tau} - J_N^{-1}(\hat{\theta})\begin{pmatrix} k_N(\theta_o) \\ \cdots\cdots \\ 0 \end{pmatrix} - J_N^{-1}(\hat{\theta})\begin{bmatrix} (\partial/\partial\theta')k_N^*(\theta^*) & . & H'(\hat{\theta}) \\ \cdots\cdots & \cdots & \cdots\cdots \\ H^{**}(\theta^{**}) & . & 0 \end{bmatrix}\begin{pmatrix} \hat{\theta} - \theta_o \\ \cdots\cdots \\ \hat{\lambda} \end{pmatrix}$$

where the symbolic derivatives $(\partial/\partial\theta')k_N^*(\theta^*)$, $H^{**}(\theta^{**})$ have the same meaning as in the proof of Theorem 2 and $\|\theta^* - \theta_o\|$, $\|\theta^{**} - \theta_o\|$ are less than $\|\hat{\theta}-\theta_o\|$. It follows that

$$N^{\frac{1}{2}}(\tilde{\tau}-\tau_o) = \left\{ I_{(m+f)} - J_N^{-1}(\hat{\theta})\begin{bmatrix} (\partial/\partial\theta')k_N^*(\theta^*) & \vdots & H'(\hat{\theta}) \\ \cdots\cdots & \vdots & \cdots \\ H^{**}(\theta^{**}) & \vdots & 0 \end{bmatrix}\right\} . N^{\frac{1}{2}}(\hat{\tau}-\tau_o) -$$

$$- J_N(\hat{\theta})^{-1} \begin{pmatrix} N^{\frac{1}{2}} k_N(\theta_o) \\ \cdots\cdots \\ 0 \end{pmatrix}$$

The first term on the right in the above expression tends to zero in probability because $\hat{\tau}$ satisfies (4.50) and the term in $\{\cdot\}$ converges to zero in probability by an analysis similar to the one given in Theorem 2. It now follows, as in the proof of Theorem 2, that

$$\lim_N N^{\frac{1}{2}}(\tau_N^{(k+1)} - \tau_o) \overset{\mathcal{D}}{=} N(0,\ddagger).$$

From this we can deduce that $N^{\frac{1}{2}}(\tau_N^{(k+1)} - \tau_o) = 0_p(1)$. Because $N^{\frac{1}{2}}(\tau_N^{(o)} - \tau_o) = 0_p(1)$, the proof of the theorem is complete. □

Proof of Corollary 1 The proof of this corollary follows from Corollary 2 and a simple adaptation of the proofs of Theorems 3 and 4. □

We need the following lemma to prove Theorem 5.

Lemma 9 Suppose that: (i) $g(x)$ is a nonstochastic scalar function of an m dimensional vector argument and is twice differentiable, with continuous first and second derivatives, in a neighbourhood X_α of the point x_o; $X_\alpha = (x:\|x-x_o\| \leq \alpha)$.

(ii) The f dimensional, $(0 \leq f < m)$, vector function $h(x)$ is well defined, and has continuous first derivatives, in X_α.

Put $X_\alpha(h) = (x:h(x) = 0) \cap X_\alpha$,

$$H(x) = (\partial/\partial x')h(x),\ M(x) = (\partial^2/\partial x \partial x')g(x),$$

$$Q(x) = \begin{pmatrix} M(x) & . & H'(x) \\ \cdots & \cdots & \cdots \\ H(x) & . & 0 \end{pmatrix},\ T(x)' = \{M(x)', H(x)'\}.$$

(iii) $H(x_o)$ is of rank f.

(iv) $x_o \in X_\alpha(h)$ and $g(x) \geq g(x_o)\ \forall\ x \in X_\alpha$.

(v) x_o is a *regular* point of the matrix $T(x)$.

Then x_o is an isolated local minimum of $g(x)$ for $x \in X_\alpha(h)$ if and only if $Q(x_o)$ is nonsingular.

<u>*Proof*</u> We use an argument similar to that given in Rothenberg [1971, Theorem 1]. First we note that (Rao [1973, p.296, eqn (4i.1.21)]) $Q(x_o)$ is nonsingular if and only if $T(x_o)$ has rank m. If x_o is not an isolated local minimum of $g(x)$, then there exists a sequence of points $(x_n, n \geq 1)$ with $x_n \in X_\alpha(h) - x_o$, $g(x_n) = g(x_o)$ and $\lim_n x_n = x_o$. Put

$$\delta_n = (x_n - x_o)/\|x_n - x_o\| \;\forall\; n \geq 1 \;(\Rightarrow \|\delta_n\| = 1 \;\;\forall\; n \geq 1).$$

Then,

$$0 = \{h(x_n)-h(x_o)\}/\|x_n-x_o\| = H^{**}(x_n^{**})\delta_n \qquad (4.55)$$

and

$$0 = \{g(x_n)-g(x_o\}/\|x_n-x_o\|^2 = \{(\partial/\partial x')g(x_o)\}\delta_n/\|x_n-x_o\|+$$
$$+ \tfrac{1}{2}\delta_n' M^*(x_n^*)\delta_n = \tfrac{1}{2}\delta_n' M^*(x_n^*)\delta_n \qquad (4.56)$$

where $M^*(x_n^*)$, $H^{**}(x_n^{**})$ are symbolic derivatives whose meaning is explained in the proof of Theorem 2 and

$$\|x_n^*-x_o\|,\; \|x_n^{**}-x_o\| \leq \|x_n-x_o\| \quad \forall\; n \geq 1.$$

Because the sequence $(\delta_n, n \geq 1)$ lies on a compact set it has a convergent subsequence (Rudin [1964, p.33]), $(\zeta_n, n \geq 1)$ say, such that $\lim_n \zeta_n = \bar{\zeta}$ with $\|\bar{\zeta}\| = 1$. It follows from (4.55) and (4.56) that $\bar{\zeta}' M(x_o)\bar{\zeta} = 0$ and $H(x_o)\bar{\zeta} = 0$. Hence $T(x_o)\bar{\zeta} = 0$ because $M(x_o)$ is a positive semi-definite matrix. Hence the rank of $T(x_o)$ is less than m, implying that $Q(x_o)$ is singular.

Conversely, if $T(x)$ has rank less than m at x_o, then it has rank less than m in some neighbourhood of x_o

by the regularity assumption. Hence $\exists\ \alpha'$, $0 < \alpha' \leq \alpha$, and a vector function $c(x)$ defined on $X_{\alpha'}$, such that $T(x)c(x) = 0 \ \forall\ x \in X_{\alpha'}$.

We now define the curve $\{x(t),\ t \geq 0\}$ in $X_{\alpha'}$ by:

$$\frac{dx(t)}{dt} = c\{x(t)\},\ x(0) = x_o.$$

Then,

$$\frac{d}{dt}[\{(\partial/\partial x')g(x(t))\},\ h(x(t))'] = [T\{x(t)\}c\{x(t)\}]' = 0,$$

i.e. $[(\partial/\partial x)g(x), h(x)]$ is constant along the curve. Because $[(\partial/\partial x)g(x),\ h(x)] = 0$ at $x = x_o$ it follows that it is zero at every point on the curve. Furthermore

$$\frac{d}{dt}\{g(x(t))\} = \{(\partial/\partial x')g(x(t))\}c(x(t)) = 0$$

so that $g(x)$ is constant along the curve. Hence x_o cannot be an isolated local minimum of $g(x)$ for $x \in X_\alpha(h)$ because the curve passes through x_o by construction. This completes the proof of the lemma. □

Proof of Theorem 5 It follows from the assumptions of Theorem 5 that $(\partial^2/\partial\theta_i\partial\theta_j)\Omega_N(\theta)$, $(i,j = 1,\ldots,m)$, converges a.s., and uniformly in $\theta \in U_{\alpha_1}$, to a finite limit. Because $(\partial/\partial\theta_i)\Omega_N(\theta)$ and $\Omega_N(\theta)$ converge a.s. to finite limits for $\theta = \theta_o$, it follows (Rudin [1964, p.140, Theorem 7.17]) that

$$\lim_N \frac{\partial\Omega_N(\theta_o)}{\partial\theta_i} \overset{a.s.}{=} \frac{\partial\Omega(\theta_o)}{\partial\theta_i}$$

and

$$\lim_N \frac{\partial^2\Omega_N(\theta_o)}{\partial\theta_i\partial\theta_j} \overset{a.s.}{=} \frac{\partial^2\Omega(\theta_o)}{\partial\theta_i\partial\theta_j} \qquad i,j = 1,\ldots,m.$$

From assumption A1., Lemma 2(i), and the assumptions of Theorem 5 it follows that $(\partial/\partial\theta_i)\Omega_N(\theta_o) \to 0$ a.s. so that

$$(\partial/\partial\theta_i)\Omega(\theta_o) = 0 \qquad i = 1,\ldots,m \tag{4.60}$$

by above.

We can similarly show that

$$\mathrm{tr}\ \Omega_o^{-1}\{(\partial^2/\partial\theta_i\partial\theta_j)\Omega(\theta_o)\} = 2(\Delta)_{ij} \quad i,j=1,\ldots,m \tag{4.61}$$

Let $\ell(\theta)$ be defined by (3.24). If assumption A6.(v) holds, then, as we have already shown in the proof of Theorem 1, θ_o is an isolated minimum of $\ell(\theta)$ for $\theta \in U_{\alpha_1} \cap \Theta_h$. In addition $(\partial^2/\partial\theta\partial\theta')\ell(\theta)$ and $H(\theta)$ are continuous functions of θ and $H(\theta_o)$ has rank f. Hence the conditions of Lemma 9 are satisfied and it follows that

$$\left(\begin{array}{c:c} \dfrac{\partial^2\ell(\theta_o)}{\partial\theta\partial\theta'} & H'(\theta_o) \\ \hdashline H(\theta_o) & 0 \end{array}\right)$$

is nonsingular. But

$$2\,\frac{\partial^2\ell(\theta_o)}{\partial\theta_i\partial\theta_j} = \mathrm{tr}\ \Omega_o^{-1}\,\frac{\partial^2\Omega(\theta_o)}{\partial\theta_i\partial\theta_j} - \mathrm{tr}\ \Omega_o^{-1}\,\frac{\partial\Omega(\theta_o)}{\partial\theta_i}\,\Omega_o^{-1}\,\frac{\partial\Omega(\theta_o)}{\partial\theta_j} =$$

$$= 2(\Delta)_{ij} \quad (i,j = 1,\ldots,m)$$

by (4.60) and (4.61). This completes the proof of the theorem.□

Remark 4.1 Although the theorems quoted from Rudin assume that θ lies in a bounded interval of the real line, their conclusions apply also to our situation. □

To prove Theorem 6 we need the following lemma.

Lemma 10 Suppose that there exists a sequence of subsets of $\mathcal{A}$, $(B_n, n \geq 1)$ say, such that $\lim_n P(B_n) = 1$. If $X \geq 0$ a.s., $E(X) < \infty$, then $\lim_n \int_{B_n} X dP = 0$

Proof By Halmos [1950, p.85, Theorem B and p.112 Theorem B] there exists a sequence $(X_j, j \geq 1)$ of non-negative simple functions such that $X_j \uparrow X$ (a.s.) and $\forall\ \varepsilon > 0\ \exists J=J(\varepsilon)$ such that $\int_{\not\emptyset}(X-X_j)dP < \varepsilon$ for $j \geq J$.

Fixing j ($>J$) we have

$$\int_{B_n} X\ dP \leq \int (X-X_j)dP + \int_{B_n} X_j < \varepsilon + \int_{B_n} X_j\ dP.$$

Hence

$$\limsup_n \int_{B_n} X\ dP \leq \varepsilon$$

because X_j is a simple function. Since $\limsup_n \int_{B_n} X\ dP$ is independent of ε, the result follows. □

Proof of Theorem 6 By Rao [1973, p.123 (xi)] it is sufficient to show that for each $t \neq 0$

$$\lim_N t'u_N \overset{\mathcal{D}}{=} N(0,t'\Delta t) \tag{4.65}$$

By (iii), (iv) and Anderson [1971, p.425 Theorem 7.7.1] (4.65) holds if

$$\lim_N N^{-\frac{1}{2}} t' \left\{ \sum_{n=1}^{N} V_r(n)'\Omega_o^{-1} e(n) \right\} \overset{\mathcal{D}}{=} N(0,t'\Delta_r t) \tag{4.66}$$

Fix t and r and for $1 \leq n \leq N$ put

$Y_n = t'V_{1r}(n)'\Omega_o^{-1}e(n)$, $Z_n = t'V_{2r}(n)'\Omega_o^{-1}e(n)$,

$X_{Nn} = N^{-\frac{1}{2}}(Y_n+Z_n)$, $F_{Nn} = F_n$, $\sigma^2_{Nn} = E(X^2_{Nn}|F_{N,n-1})$,

$b_N = \max_{1\leq n \leq N} \sigma^2_{Nn}$, $T_N = \sum_{n=1}^{N} \sigma^2_{Nn}$.

Below we can, without loss of generality, take t to be normalized so that $\|t\|=1$.

Sufficient conditions for (4.66) to hold are (Brown and Eagleson [1971, Corollary 1]): (a) $b_N \to 0$ p,

(b) $T_N \overset{p}{\to} t' \Delta_r t$, and (c) $\forall\ \varepsilon > 0$

$$\lim_N \sum_{n=1}^{N} E\{X_{Nn}^2 I(|X_{Nn}| \geq \varepsilon) | F_{N,n-1}\} \overset{p}{=} 0.$$

By Chebyshev's inequality, a sufficient condition for (c) to hold is:

(c') $\quad \forall\ \varepsilon > 0 \quad \lim_N \sum_{n=1}^{N} E\{X_{Nn}^2 I(|X_{Nn}| \geq \varepsilon)\} = 0.$

By the ergodic theorem (Stout [1974, p.176, Theorem 3.5.6]) $\lim_N N^{-1} \| V_{1r}(N) \| = 0$ a.s. and it follows (Anderson [1971, p.25, Lemma 2.6.1]) that

$$\lim_N \max_{1 \leq n \leq N} N^{-1} \| V_{1r}(n) \|^2 = 0 \text{ a.s.} \tag{4.67}$$

From condition (ii) of Theorem 6 $N^{-1} \| V_{2r}(N) \|^2 \to 0$ a.s. and it again follows that

$$\lim_N N^{-1} \max_{1 \leq n \leq N} \| V_{2r}(n) \|^2 = 0 \tag{4.68}$$

By condition (ii) of Theorem 6

$$N^{-1} \sum_{n=1}^{N} \| V_{2r}(n) \|^2 = 0(1) \tag{4.69}$$

We will also have that

$$I(|X_{Nn}| \geq \varepsilon) \leq I(|Y_n| \geq \tfrac{1}{2} N^{\frac{1}{2}} \varepsilon) + I(|Z_n| \geq \tfrac{1}{2} N^{\frac{1}{2}} \varepsilon) \tag{4.70}$$

and

$$X_{Nn}^2 \leq 2\, N^{-1} (Y_n^2 + Z_n^2) \tag{4.71}$$

Now $\sigma_{Nn}^2 = N^{-1} t' V_r(n)' \Omega_o^{-1} V_r(n) t \leq c\ N^{-1} (\| V_{1r}(n) \|^2 + \| V_{2r}(n) \|^2)$.

(a) now follows from (4.67) and (4.68), while (b) follows from condition (iii)(a) of Theorem 6.

We now show that (c') holds.

$$\lim_N N^{-1} \sum_{n=1}^{N} E\{Y_n^2 I(|Y_n| \geq \tfrac{1}{2} N^{\frac{1}{2}} \varepsilon)\} = 0 \tag{4.72}$$

because Y_n is stationary and square-integrable. Put

$$\zeta(N) = \max_{1\leq n\leq N} N^{-\frac{1}{2}}\|V_{2r}(n)\|.$$

Then,

$$N^{-1}\sum_{n=1}^{N} E\{Z_n^2 I(|Z_n| \geq \tfrac{1}{2} N^{\frac{1}{2}}\varepsilon)\}$$

$$\leq c\{N^{-1}\sum_{n=1}^{N}\|V_{2r}(n)\|^2\}.E\{\|e(n)\|^2 I(\|e(n)\| \geq c\ \zeta(N)^{-1})\}=$$

$$= c\{N^{-1}\sum_{n=1}^{N}\|V_{2r}(n)\|^2\}.E\{\|e(1)\|^2 I(\|e(1)\| \geq c\ \zeta(N)^{-1})\} \quad (4.73)$$

because of the stationarity of the $e(n)$. It now follows from (4.68), (4.69), (4.73) and the square-integrability of $e(1)$, that

$$\lim_N N^{-1}\sum_{n=1}^{N} E\{Z_n^2 I(|Z_n| \geq \tfrac{1}{2} N^{\frac{1}{2}}\varepsilon)\} = 0 \quad (4.74)$$

$$N^{-1}\sum_{n=1}^{N} E\{Y_n^2 I(|Z_n| \geq \tfrac{1}{2}N^{\frac{1}{2}}\varepsilon)\}\leq N^{-1}\sum_{n=1}^{N} E\{Y_n^2 I(\|e(n)\| \geq c\ \zeta(N)^{-1})\} =$$

$$= E\{Y_1^2 I(\|e(1)\| \geq c\ \zeta(N)^{-1})\} \to 0 \text{ as } N \to \infty \quad (4.75)$$

using the stationarity of Y_n and $e(n)$, (4.68) and applying Lemma 10.

Finally,

$$N^{-1}\sum_{n=1}^{N} E\{Z_n^2 I(|Y_n|\geq \tfrac{1}{2} N^{\frac{1}{2}}\varepsilon)\} \leq$$

$$\leq c\{N^{-1}\sum_{n=1}^{N}\|V_{2r}(n)\|^2\}.E\{\|e(1)\|^2 I(|Y_1| \geq \tfrac{1}{2} N^{\frac{1}{2}}\varepsilon)\}\to 0 \text{ as } N\to\infty \quad (4.76)$$

using (4.69) and applying Lemma 10.

Condition (c') now follows from (4.70), (4.71), (4.72) and (4.74) - (4.76). This completes the proof of the theorem.□

CHAPTER 4

SOME ASYMPTOTIC RESULTS FOR VECTOR LINEAR TIME SERIES MODELS

1. INTRODUCTION

We apply the results of the previous chapter to the vector linear time series model

$$\sum_{j=o}^{r} B_j(\theta_o)y(n-j) = \sum_{j=o}^{\infty} C_j(\theta_o)x(n-j) + \sum_{j=o}^{\infty} A_j(\theta_o)e(n-j), n=0,\pm 1,.. \quad (1.10)$$

where the $y(n)$ and $x(n)$ are the endogenous and exogenous variables respectively; we observe the $y(n)$ and $x(n)$, but not the $e(n)$.

We also obtain analogous results for Gaussian Estimates minimizing an equivalent frequency domain criterion function.

The theory is then applied to some of the time series models appearing in the literature.

When we refer to (1.10) as a linear model we mean that it is linear in the variables $y(n)$, $x(n)$ and $e(n)$; the B_j, C_j and A_j matrices may be nonlinear functions of the unknown parameter vector θ_o.

The chapter is structured as follows: §2 states the main assumptions required for the results to hold; the time domain results are given in §3 and the frequency domain ones in §4; §§5 and 6 discuss applications of the theory; all proofs are placed in §7.

We will adopt the following notation in the chapter:

1. For a function of the complex variable z, $g(z)$ say, we will write $g(e^{i\lambda})$ as $g(\lambda)$. 2. $\sum_{j=a}^{b} \equiv 0$ if $a > b$.

2. ASSUMPTIONS

Put $B(z;\theta) = \sum_{j=o}^{r} B_j(\theta)z^j$, $C(z;\theta) = \sum_{j=o}^{\infty} C_j(\theta)z^j$ and $A(z;\theta) = \sum_{j=o}^{\infty} A_j(\theta)z^j$. We make the following assumptions:

A1. $(e(n), G_n, n = 0,\pm1,..)$ is a sequence of p dimensional, stationary martingale differences satisfying $E(e(n)e(n)'|G_{n-1}) = \Omega_o$, where Ω_o is a constant nonsingular matrix and G_n is the σ-field generated by the $e(m)$, $m \leq n$.

A2. The q dimensional exogenous variables $(x(n), n = 0,\pm1,...)$ are assumed to be either:

(i) a stationary, ergodic, sequence having finite second moments and independent of the $\{e(n)\}$ sequence. Put $\Gamma_x(j) = E\{x(n)x(n+j)'\}$ for $j = 0,\pm1,...$.

OR

(i') a nonstochastic sequence such that the following limits exist and are finite for all non-negative integers j.

$$\lim_N N^{-1} \sum_{n=1}^{N-j} x(n)x(n+j)' = \Gamma_x(j)$$

We will also assume that

$$\lim_N N^{-1}\|x(-N)\|^2 = 0.$$

In either (i) or (i') we put $\Gamma_x(-j) = \Gamma_x(j)'$ for $j \geq 0$.

For either (i) or (i') it follows (Doob [1953, p.474, Theorem 3.2]) that there exists a distribution matrix $F_x(\lambda)$ such that

$$\Gamma_x(j) = \int_{-\pi}^{\pi} e^{ij\lambda} dF_x(\lambda) \quad , \; j = 0,\pm1,...$$

We will further assume that:

(ii) For $\lambda \in [-\pi,\pi]$ there exists a continuous matrix function $f_x(\lambda)$ s.t. $F_x(\lambda) = \int_{-\pi}^{\lambda} f_x(\lambda)d\lambda$, with $f_x(\lambda)$ nonsingular on an open subset of $[-\pi,\pi]$ having Lebesgue measure greater than zero.

A3.(i) Θ, the permissible parameter space, is a compact subset of m dimensional Euclidean space. θ_o, the true parameter point, belongs to Θ.

(ii) The degree, r, of the polynomial $B(z;\theta)$ is chosen prior to estimation.

(iii) The matrices $B_j(\theta)$, $0 \leq j \leq r$, are $p\times p$; the $C_j(\theta)$, $j \geq 0$, are $p\times q$; the $A_j(\theta)$, $j \geq 0$, are $p\times p$.

(iv) $\forall\ \theta \in \Theta$, $B_o(\theta) \equiv A_o(\theta) \equiv I_{(p)}$.

(v) $\exists\, R > 1$ s.t. $\forall\ \theta \in \Theta$ and for $|z| \leq R$, $\det A(z;\theta) \neq 0$ and the elements of $A(z;\theta)$ and $C(z;\theta)$ are analytic functions of z, given θ, and continuous functions of (z,θ).

(vi) For $\theta \in \Theta$, the elements of $B_j(\theta)$ $(j=1,\ldots,r)$ are continuous functions of θ. $\det\{B(z;\theta_o)\} \neq 0$ for $|z| \leq 1$.

$\forall\ \alpha > 0$ put $U_\alpha = (\theta : \|\theta-\theta_o\| \leq \alpha)$

A4.(i) $\exists\, \alpha_1 > 0$ s.t. $U_{\alpha_1} \subseteq \Theta$

(ii) $\forall\ \theta \in U_{\alpha_1}$, the elements of $B_j(\theta)$ $(1 \leq j \leq r)$, $C_j(\theta)$ $(j \geq 0)$ and $A_j(\theta)$ $(j \geq 1)$ are twice differentiable functions of θ with continuous first and second derivatives.

(iii) Let R be defined by assumption A3.(v)

For $|z| \leq R$ and $\theta \in U_{\alpha_1}$,

$$\sum_{k=o}^{\infty} \frac{\partial A_k(\theta)}{\partial\theta_i} z^k, \quad \sum_{k=o}^{\infty} \frac{\partial^2 A_k(\theta)}{\partial\theta_i\partial\theta_j} z^k, \quad \sum_{k=o}^{\infty} \frac{\partial C_k(\theta)}{\partial\theta_j} z^k,$$

$$\sum_{k=o}^{\infty} \frac{\partial^2 C_k(\theta)}{\partial\theta_i\partial\theta_j} z^k, \quad i,j = 1,\ldots,m$$

are analytic functions of z, given θ, and continuous functions of (z,θ).

A5.(i) For $\theta \in \Theta$, $h(\theta)$ is an f dimensional, $(f < m)$, continuous, vector function of θ. $h(\theta_o) = 0$.

(ii) $h(\theta)$ has continuous first and second partial derivatives for $\theta \in U_{\alpha_1}$.

Put $H(\theta) = (\partial/\partial\theta')h(\theta)$, $H = H(\theta_o)$.

(iii) H has rank f.

Put

$$\Theta_h = \Theta \cap \{\theta: h(\theta) = 0\}, \quad \ell(\theta) = \tfrac{1}{2} \log \det \Omega(\theta)$$

where $\Omega(\theta)$ is defined in (2.10) below.

A6.(i) The matrix $[\{(\partial^2/\partial\theta\partial\theta')\ell(\theta)\},H'(\theta)]$ is regular at the point θ_o, i.e. the matrix has constant rank in some neighbourhood of θ_o.

(ii) If $\theta \in \Theta_h$ and

$$[C(z;\theta),A(z;\theta)] = B(z;\theta)B(z;\theta_o)^{-1}[C(z;\theta_o),A(z;\theta_o)]$$

for $|z| = 1$, then $\theta = \theta_o$. □

Remark 2.1 (i) As far as the results of this chapter are concerned, there is no loss of generality in assuming that $B_o(\theta) \equiv I_{(p)}$. If this was not so, then, assuming $B_o(\theta)$ is nonsingular $\forall\ \theta \in \Theta$, we could always premultiply (1.10) by $B_o(\theta_o)^{-1}$ and make it so by a simple redefinition of the $A_j(\theta_o)$ and $e(n)$. More specifically, we would put $A_j(\theta_o)$ equal to $B_o(\theta_o)^{-1}A_j(\theta_o)B_o(\theta_o)$ and $e(n)$ equal to $B_o(\theta_o)^{-1}\ e(n)$. For a further clarification of this point see the treatment of the simultaneous linear ARMAX model in §5. □

(ii) As we have already mentioned in Remark 2.1(iii) of 3.§2, the assumption that the $e(n)$ are martingale differences is quite a natural one for the model (1.10) and its derivatives, as it is equivalent to requiring that the best predictor of $y(n)$ given $y(j)$ $(j < n)$ and $x(k)$ $(k \leq n)$ is equal to the best linear predictor (both in a least squares sense).

Put $D(z) = B(z;\theta_o)^{-1}C(z;\theta_o) = \sum_{j=o}^{\infty} D_j z^j$,

$$E(z;\theta) = A(z;\theta)^{-1} = \sum_{j=o}^{\infty} E_j(\theta)z^j,$$

$$F(z;\theta) = A^{-1}(z;\theta)C(z;\theta) = \sum_{j=o}^{\infty} F_j(\theta)z^j, G(z;\theta)=A^{-1}(z;\theta)B(z;\theta)=$$

$$= \sum_{j=o}^{\infty} G_j(\theta)z^j,\ H(z) = B(z;\theta_o)^{-1} = \sum_{j=o}^{\infty} H_j z^j,$$

$$K(z) = G(z;\theta_o)^{-1} = \sum_{j=o}^{\infty} K_j z^j$$

$$\Phi(z;\theta) = G(z;\theta)D(z)-F(z;\theta) = \sum_{j=o}^{\infty} \Phi_j(\theta)z^j$$

$$\Psi(z;\theta) = G(z;\theta)K(z) = \sum_{j=o}^{\infty} \Psi_j(\theta)z^j$$

$$\Omega(\theta) = \int_{-\pi}^{\pi} \Phi(\lambda;\theta)dF_x(\lambda)\Phi(\lambda;\theta)^* + (2\pi)^{-1}\int_{-\pi}^{\pi} \Psi(\lambda;\theta)\Omega_o\Psi(\lambda;\theta)^* d\lambda \quad (2.10)$$

We can deduce from our assumptions that the above series expansions are valid for $\theta \in \Theta$ and for $|z| < R'$, for some $R' > 1$.

3. TIME DOMAIN RESULTS

We put

$$v(n;\theta) = \sum_{j=o}^{\infty} G_j(\theta)y(n-j) - \sum_{j=o}^{\infty} F_j(\theta)x(n-j) \qquad (3.10)$$

and

$$\Omega_N(\theta) = N^{-1} \sum_{n=1}^{N} v(n;\theta)v(n;\theta)' \qquad (3.11)$$

To obtain the Gaussian Estimate of θ_o, $\hat{\theta}_N$ say, we minimize

$$\ell_N(\theta) = \tfrac{1}{2} \log \det \Omega_N(\theta)$$

over $\theta \in \Theta_h$ and take $\Omega_N(\hat{\theta}_N)$ as the Gaussian Estimate of Ω_o. This last step is justified in 3.§3.

Theorem 1 Suppose assumptions A1. - A3., A5.(i) and A6.(ii) hold. Then: (i) a.s., for N sufficiently large, $\exists$ a $\hat{\theta}_N$ such that $\hat{\theta}_N$ minimizes $\ell_N(\theta)$ for $\theta \in \Theta_h$.

For each N, choosing one such $\hat{\theta}_N$, we have: (ii) $\lim_N \hat{\theta}_N = \theta_o$ a.s. and $\lim_N \Omega_N(\hat{\theta}_N) = \Omega_o$ a.s. □

We now suppose that assumptions A1.-A6. hold. Then we know from Theorem 1 that $\hat{\theta}_N$ lies, almost surely, in the interior of U_{α_1} for N sufficiently large and therefore there exists an f dimensional vector of Lagrange multipliers, $\hat{\lambda}_N$, such that

$$(\partial/\partial\theta)\ell_N(\hat{\theta}_N) + H'(\hat{\theta}_N)\hat{\lambda}_N = 0 \qquad (3.15)$$

$$h(\hat{\theta}_N) = 0 \qquad (3.16)$$

Put $V(n;\theta)=(\partial/\partial\theta')v(n;\theta)$,

$$\Delta_N(\theta)= N^{-1} \sum_{n=1}^{N} V(n;\theta)'\Omega_N(\theta)^{-1}V(n;\theta),$$

$$J_N(\theta) = \begin{bmatrix} \Delta_N(\theta) & \vdots & H'(\theta) \\ \cdots & \cdots & \cdots \\ H(\theta) & \vdots & 0 \end{bmatrix}, \quad \ddagger_N(\theta) = J_N(\theta)^{-1} \begin{bmatrix} \Delta_N(\theta) & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} J_N(\theta)^{-1}$$

$$\Delta = (\partial^2/\partial\theta\partial\theta')\ell(\theta_o), \quad J = \begin{bmatrix} \Delta & \vdots & H' \\ \cdots & \cdots & \cdots \\ H & \vdots & 0 \end{bmatrix}, \quad \ddagger = J^{-1} \begin{bmatrix} \Delta & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} J^{-1}$$

$$L = I_{(m)} - H'(HH')^{-1}H.$$

We now obtain an explicit representation for Δ. Let Δ_{ij} be the $(i,j)^{th}$ element of Δ $(i,j = 1,\ldots,m)$. From (2.10) it is fairly easy to obtain that

$$\Delta_{ij} = \mathrm{tr}\ \Omega_o^{-1}\left\{\int_{-\pi}^{\pi} \frac{\partial\Phi(\lambda;\theta_o)}{\partial\theta_i} f_x(\lambda) \frac{\partial\Phi(\lambda;\theta_o)^*}{\partial\theta_j}\, d\lambda + \right.$$

$$\left. + (2\pi)^{-1}\int_{-\pi}^{\pi} \frac{\partial\Psi(\lambda;\theta_o)}{\partial\theta_i}\, \Omega_o\, \frac{\partial\Psi(\lambda;\theta)^*}{\partial\theta_j}\, d\lambda \right\}$$

Put $\phi(\lambda) = (\partial/\partial\theta')\mathrm{vec}\ \Phi(\lambda;\theta_o)$, $\psi(\lambda) = (\partial/\partial\theta')\mathrm{vec}\ \Psi(\lambda;\theta_o)$ Using the identity $\mathrm{tr}\ AB \equiv [\mathrm{vec}\{A'\}]'\mathrm{vec}\ B$, we will have that

$$\Delta = \int_{-\pi}^{\pi} \phi(\lambda)^*\{f_x'(\lambda)\otimes\Omega_o^{-1}\}\phi(\lambda)\,d\lambda + (2\pi)^{-1}\int_{-\pi}^{\pi} \psi(\lambda)^*(\Omega_o \otimes \Omega_o^{-1})\psi(\lambda)\,d\lambda.$$

We now come to the central limit theorem. Let $\hat{\tau}_N' = (\hat{\theta}_N', \hat{\lambda}_N')$, $\tau_o' = (\theta_o', 0')$.

Theorem 2 Given assumptions A1. - A6.

$\lim_N N^{\frac{1}{2}}(\hat{\tau}_N - \tau_o) \overset{\mathcal{D}}{=} N(0,\ddagger)$ and $\lim_N N^{\frac{1}{2}}(\hat{\theta}_N - \theta_o) \overset{\mathcal{D}}{=} N(0,\ddagger_\theta)$, where $\ddagger_\theta = (L\Delta L)^+$. Furthermore $\ddagger_N(\hat{\theta}_N) \to \ddagger$ a.s., i.e. $\ddagger_N(\hat{\theta}_N)$ is a strongly consistent estimate of $\ddagger$. □

We will see in Chapter 5 that for most time series models we cannot obtain closed form solutions of the Normal equations (3.15, 3.16). Because of this we again set up the constrained G.N. iteration

$$\tau_N^{(j+1)} = \begin{pmatrix} \theta_N^{(j)} \\ \cdots \\ 0 \end{pmatrix} - J_N^{-1}(\theta_N^{(j)}) \begin{pmatrix} k_N(\theta_N^{(j)}) \\ \cdots\cdots \\ h(\theta_N^{(j)}) \end{pmatrix}, j \geq 0, \tag{3.20}$$

where

$$k_N(\theta) = (\partial/\partial\theta')\ell_N(\theta) = N^{-1}\sum_{n=1}^{N} V(n;\theta)'\Omega_N(\theta)v(n;\theta)$$

and $\tau_N^{(j)'} = (\theta_N^{(j)'}, \lambda_N^{(j)'})$; $\tau_N^{(o)}$ initiates the iteration. Subject to the conditions of the next theorem *the* $\theta_N^{(j)}$ $(j \geq 1)$ *will be consistent and have the same asymptotic distribution as* $\hat{\theta}_N$.

<u>*Theorem 3*</u> Suppose that for the iteration scheme (3.20) we put $\tau_N^{(o)'} = (\theta_N^{(o)'}, 0')$. Given that assumptions A1.-A6. hold we will have that:

(i) $\exists\, \alpha'$, $0 < \alpha' < \alpha_1$, s.t. in the set $U_{\alpha'} \times (\lambda: \|\lambda\| \leq \alpha')$, $\hat{\tau}_N$ is, a.s., the unique solution of the Normal Equations for N sufficiently large.

(ii) If $\theta_N^{(o)} \to \theta_o$ p, then, with probability tending to 1 as $N \to \infty$, $\lim_j \tau_N^{(j)} = \hat{\tau}_N$.

(iii) If $\theta_N^{(o)} \to \theta_o$ a.s., then, a.s. for N sufficiently large, $\lim_j \tau_N^{(j)} = \hat{\tau}_N$.

(iv) If, in addition, in both (ii) and (iii), $\theta_N^{(o)} - \theta_o = 0_p(N^{-\frac{1}{2}})$, then $\lim_N N^{\frac{1}{2}}(\tau_N^{(j)} - \tau_o) \overset{\mathcal{D}}{=} N(0, \ddagger)$ for $j \geq 1$.□

We note that $\ddagger_N(\theta_N^{(j)})$ is a consistent estimator of $\ddagger$.

It will often be more realistic to assume that (1.10) holds for $n \geq 1$ only, and that the pre-observation-period values of $y(n)$ and $e(n)$ are given constants. The next theorem shows that in this slightly modified situation the conclusions of Theorems 1 - 3 still hold.

We change assumption A1. to:

A1'.(i) (1.10) holds for $n \geq 1$ only

(ii) This is the same as assumption A1. but with n taking on positive values only. G_o is assumed to be the trivial σ-field.

(iii) For some non-negative integers r and t, known prior to estimation, $y(o),\ldots,y(1-r)$, $e(o),\ldots,e(1-t)$ are known constants. $y(n)$ and $e(n)$ are identically zero for $n < 1-r$ and $n < 1-t$ respectively.

Assumptions A2.-A6. stay as they are. □

We change $v(n;\theta)$ to

$$v(n;\theta) = -\sum_{k=o}^{n-1} E_k(\theta)\left\{\sum_{j=n-k}^{t} A_j(\theta)e(n-j-k)\right\} + \sum_{k=o}^{n-1} E_k(\theta)\times$$

$$\times\left\{\sum_{j=o}^{r} B_j(\theta)y(n-j-k) - \sum_{j=o}^{\infty} C_j(\theta)x(n-j-k)\right\}, n \geq 1 \quad (3.25)$$

The definition of $v(n;\theta)$ in (3.25) is motivated by the fact that $v(n;\theta_o) = e(n)$ (see 3.§1 for a discussion of the criterion function $L_N(\theta,\Omega)$ which is given there). Given $v(n;\theta)$, $\Omega_N(\theta)$ and $\ell_N(\theta)$ will still be defined by (3.10) and (3.11).

Theorem 4 Suppose assumptions A1'., A2. - A6. are given. Then Theorems 1 - 3 still hold. □

4. FREQUENCY DOMAIN ESTIMATION

Let $w_y(\lambda)$ and $w_x(\lambda)$ be the FFT's of the $y(n)$ and $x(n)$ respectively. Put $\lambda_n = 2\pi n/N$, $-\tfrac{1}{2}N < n \leq [\tfrac{1}{2}N]$,

$$v^{(F)}(n;\theta) = A(\lambda_n;\theta)^{-1}\{B(\lambda_n;\theta)w_y(\lambda_n) - C(\lambda_n;\theta)w_x(\lambda_n)\}$$

$$\Omega_N^{(F)}(\theta) = 2\pi N^{-1}\Sigma_n v^{(F)}(n;\theta)v^{(F)}(n;\theta)^*, \quad \ell_N^{(F)}(\theta) = \tfrac{1}{2}\log\det\Omega_N^{(F)}(\theta)$$

The superscript F stands for frequency. Let $w_e(\lambda)$ be the FFT of the $e(n)$. Then the definitions of $\Omega_N^{(F)}(\theta)$ and $\ell_N^{(F)}(\theta)$ are motivated by the fact that $v^{(F)}(n;\theta_o)$ is approximately $w_e(\lambda_n)$ and the easily proved identity

$$2\pi N^{-1} \sum_n w_e(\lambda_n) w_e(\lambda_n)^* = N^{-1} \sum_{n=1}^{N} e(n)e(n)'$$

See also Lemma 9 and Theorem 8.

To obtain the Gaussian Estimate of θ_o, $\hat{\theta}_N^{(F)}$ say, we minimize $\ell_N^{(F)}(\theta)$ over $\theta \in \Theta_h$. We then take $\Omega_N^{(F)}(\hat{\theta}_N^{(F)})$ as the Gaussian Estimate of Ω_o.

In order that the central limit theorem results of Theorems 2 and 3 remain valid for the frequency domain estimators we need the following additional assumption if the exogenous variables are nonstochastic.

A2.(i'') If the exogenous variables are nonstochastic, then there exists a finite positive constant c such that $\|x(n)\| \leq c$ for $n = 0,\pm 1,\ldots$ □

<u>*Theorem 5*</u> (i) If assumptions A1.-A3., A5.(i) and A6.(ii) hold, then $\hat{\theta}_N^{(F)} \to \theta_o$ a.s. and $\Omega_N^{(F)}(\hat{\theta}_N^{(F)}) \to \Omega_o$ a.s.

(ii) If assumptions A1.-A6. hold, then the analogues of Theorem 3(i), (ii) and (iii) also hold for the frequency domain estimators.

(iii) If assumption A2.(i'') holds as well, then the results of Theorem 2 and Theorem 3(iv) will also apply to the frequency domain estimators. □

5. APPLICATIONS

We now apply our results to some of the time series models proposed in the literature. Throughout this section we will make the following assumptions:

1. Where relevant, assumptions A1., A2.(i) and (i'), A5. and A6.(i) hold. As well, A2.(i'') holds for frequency domain estimation.

2. The degrees of all lag polynomials are non-negative integers which are specified prior to estimation.

3. R_1 and ε_1 are two positive numbers chosen before the start of the estimation. R_1 can be taken very large and ε_1 will be chosen close to zero.

MODEL 1

The first model we look at is the simultaneous equations model with AR residuals

$$\sum_{j=o}^{r'} \underline{B}_j\, y(n-j) = \sum_{j=o}^{s'} \underline{C}_j\, x(n-j)+u(n), \quad \sum_{j=o}^{t} \underline{A}_j u(n-j) = \underline{e}(n)$$

$$\underline{A}_o \equiv I_{(p)} \tag{5.10}$$

with $E(\underline{e}(n)\underline{e}(n)'|G_{n-1}) = \underline{\Omega}_o$, where $\underline{\Omega}_o$ is nonsingular; $\underline{B}_o$ is nonsingular and not necessarily the unit matrix. We put $\theta = \text{vec}\{\underline{B}_o,\ldots,\underline{B}_r,\ \underline{C}_o,\ldots,\underline{C}_s,\ \underline{A}_1,\ldots,\underline{A}_t\}$, $r = r'+t$, $s = s'+t$, $\Omega_o = \underline{B}_o^{-1}\underline{\Omega}_o\underline{B}_o^{-1'}$ and $e(n) = \underline{B}_o^{-1}\ \underline{e}(n)$. Let the matrices $B_j(\theta)$, $0 \le j \le r$, $C_k(\theta)$, $0 \le k \le s$, be defined by

$$\sum_{j=o}^{r} B_j(\theta) z^j = \underline{B}_o^{-1}\left(\sum_{j=o}^{t} \underline{A}_j z^j\right)\left(\sum_{j=o}^{r'} \underline{B}_j z^j\right),$$

$$\sum_{j=o}^{s} C_j(\theta) z^j = \underline{B}_o^{-1}\left(\sum_{j=o}^{t} \underline{A}_j z^j\right)\left(\sum_{j=o}^{s'} \underline{C}_j z^j\right)$$

We can now rewrite Model 1 in the standard form (1.10). We will make the following assumptions:

B1. Let $\chi_\gamma(z) = (1, z, \ldots, z^s)$. The matrix $\int_{-\pi}^{\pi} \chi_\gamma(\lambda)\chi_\gamma(\lambda)^* \otimes dF_x(\lambda)$ is nonsingular.

The parameter space Θ is the closure of the set Θ_o, where Θ_o is defined as follows:

B2.(i) $\|\theta\| \leq R_1$, $|\det \underline{B}_o| \geq \varepsilon_1$.

(ii) $|\det(\Sigma_j \underline{B}_j z^j)| > 0$, $|\det(\Sigma_j \underline{A}_j z^j)| > 0$ for $|z| \leq 1$.

(iii) A g.c.l.d. of $\Sigma_j \underline{B}_j z^j$ and $\Sigma_j \underline{C}_j z^j$ is $I_{(p)}$.

(iv) The matrix $[\underline{B}_{r'}, \underline{C}_{s'}]$ has rank p.

We will assume that $\theta_o \in \Theta_o$.

Put $\underline{B}(z;\theta) = \Sigma_j \underline{B}_j(\theta) z^j$, $\underline{C}(z;\theta) = \Sigma_j \underline{C}_j(\theta) z^j$, $\underline{A}(z;\theta) = \Sigma_j \underline{A}_j(\theta) z^j$. We have written $\underline{B}_j(\theta)$ etc. instead of $\underline{B}_j$ etc. in order that we may distinguish between arguments at two distinct values of θ.

B3. If $\theta \in \Theta_h$ and $[\underline{B}(z;\theta), \underline{C}(z;\theta)] = U[\underline{B}(z;\theta_o), \underline{C}(z;\theta_o)]$, $\underline{A}(z;\theta)U = U\underline{A}(z;\theta_o)$ where U is some p×p matrix, then $\theta = \theta_o$. □

Theorem 6 Subject to assumptions B1.-B3., the results of Theorems 1-5 hold for Model 1. □

Remark 5.1 (i) Expressions for $\ddagger$ and $\ddagger_\theta$ (see the statement of Theorem 2) are given in 9.§3.

(ii) Dhrymes and Erlat [1974] give a proof of the strong consistency and asymptotic Normality of a three-stage-least-squares type of estimator for Model 1 when $r' = t = 1$ and $s' = 0$ under the following conditions:

DE1. The e(n) are i.i.d. r.v.'s and have finite *sixth* moments.

DE.2 (i) The $x(n)$ sequence is nonstochastic and uniformly bounded.

(ii) A matrix labelled M (see Dhrymes and Erlat [1974, p.253]) exists and is nonsingular.

DE.3 We have initial consistent estimates of $(\underline{B}_o, \underline{B}_1, \underline{C}_o)$ prior to estimation. (see p.253 in Dhrymes and Erlat).

DE.4 $\det(\underline{B}_o + \underline{B}_1 z) \neq 0$ and $\det(I_{(p)} + \underline{A}_1 z) \neq 0$ for $|z| \leq 1$.

DE.5 There are simple exclusion and normalization restrictions on the elements of $(\underline{B}_o, \underline{B}_1, \underline{C}_o)$. □

It is not clear what condition DE.2(ii) means as it is a mixed condition involving the exogenous variables, endogenous variables and the parameters of the model. It does appear, moreover, that the classical identification conditions imposed by Dhrymes and Erlat will be insufficient to identify the model. For example, the two structures

(I) $y(n) + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} y(n-1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x(n) + u(n)$,

$u(n) + \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} u(n-1) = e(n)$ and

(II) $y(n) + \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} y(n-1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x(n) + u(n)$,

$u(n) + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} u(n-1) = e(n)$ will certainly satisfy assumptions DE.4, DE.5 and the classical identification conditions and yet they are observationally equivalent. It could be, of course, that condition DE.2(ii) does ensure identification but this remains to be shown. Finally, it seems that the limits in assumption A2.(i') must also exist if the analysis is to be valid.

Dhrymes and Taylor [1976] give a proof of the consistency and asymptotic Normality of an efficient two-step estimator (see 5.§4) for the model discussed in Dhrymes and Erlat. They make the following assumptions:

DT.1 The $(\underline{e}(n),\ n = 0,\pm 1,\ldots)$ are independent $N(0,\underline{\Omega}_o)$ r.v.'s with $\underline{\Omega}_o$ nonsingular.

DT.2(i) This is condition (iii) of Theorem 1 in Dhrymes and Taylor.

(ii) The ML estimators of $(\underline{B}_o,\ \underline{B}_1,\ \underline{C}_o,\ \underline{\Omega}_o)$ have well behaved limiting distributions.

DT.3 This is the same as DE.3 - DE.5.

Again it appears that there is a need for additional identification conditions besides the classical ones. Furthermore, a condition of the type DE.2(i) seems to be needed since Dhrymes and Taylor [1976, p.369 Lemma 9] use a result from Dhrymes and Erlat where this condition is assumed.

Finally, both the above papers do not, in general, allow lagged exogenous variables to be present. To demonstrate this we will, for convenience, depart from our own notation and use that in Dhrymes and Taylor and Dhrymes and Erlat. If w_n was of the form $w_n' = (t_n',\ t_{n-1}')$, then the $Q'Q$ matrix (Dhrymes and Erlat p.250, eqn (13); Dhrymes and Taylor p.363 eqn (5)) would always be singular, contrary to their assumptions.

(iii) Hatanaka [1974] proves the consistency and asymptotic Normality of a two-step estimator (see 5.§4) for the scalar version of (5.10). It is not clear whether our conditions on the exogenous variables are stronger than Hatanaka's or not. On the one hand, the Grenander conditions (A.2a) - (A.2c) imposed by Hatanaka [1974, p.212] are weaker than A2.(1'); on the other hand condition (A.2e) in Hatanaka is a mixed condition involving the exogenous variables and the parameters.

(iv) When $t = 0$ (5.10) becomes the classical simultaneous equations model (e.g. Schmidt [1976, p.119])

$$\sum_{j=o}^{r'} \underline{B}_j \, y(n-j) = \sum_{j=o}^{s'} \underline{C}_j \, x(n-j) + \underline{e}(n) \tag{5.15}$$

Even for (5.15) our results seem to be superior to those obtained previously. For example, Schmidt [1976, p.221, Proposition 5] gives a proof of the consistency and asymptotic Normality of the time domain Gaussian Estimates assuming that the $e(n)$ are independently, identically and Normally distributed, the exogenous variables are uniformly bounded, and the restrictions on θ are simple exclusion and normalization ones. It should be noted, however, that the result used in Schmidt's proof (Schmidt [1976, p.255, Theorem 19]) does not apply to (5.15) as it is for i.i.d. r.v's only.

(v) If $\underline{B}_o \equiv I_{(p)}$ in (5.15), i.e. we are dealing with a reduced form system, then we can redefine θ as $\theta = \text{vec}(\underline{B}_1, \ldots, \underline{B}_r, \underline{C}_o, \ldots, \underline{C}_s)$. When there are no restrictions on θ we can obtain a closed form solution to the Normal Equations and strong consistency can be proved from first principles without imposing any boundedness assumptions on θ.

For the reduced form case, Anderson [1971, Theorem 5.5.13, p.210] proves the weak consistency and asymptotic Normality of the Gaussian Estimates when the $\underline{e}(n)$ are *independent* , with $E(\underline{e}(n)) = 0$ and $E\{\underline{e}(n)\underline{e}(n)'\} = \underline{\Omega}_o$, and have *bounded fourth moments;* the $x(n)$ are assumed to be nonstochastic, uniformly bounded and satisfy assumption A2.(i').

Therefore our results are stronger than Anderson's.

Schönfeld [1971, p.125, Theorem 2] and Schmidt [1976, p.99, Theorem 3] also prove the consistency and asymptotic Normality of the Gaussian Estimates, for the case just considered, but under stronger assumptions than in Anderson.

MODEL 2

The second model we look at is the simultaneous ARMAX model

$$\sum_{j=o}^{r} \underline{B}_j\, y(n-j) = \sum_{j=o}^{s} \underline{C}_j\, x(n-j) + \sum_{j=o}^{t} \underline{A}_j \underline{e}(n-j),\quad A_o \equiv I_{(p)}$$

We put $\theta = \text{vec}(\underline{B}_o,\ldots,\underline{B}_r,\ \underline{C}_o,\ldots,\underline{C}_s,\ \underline{A}_1,\ldots,\underline{A}_t)$ and define $B_j(\theta)$, $C_k(\theta)$, $A_\ell(\theta)$, Ω_o and $e(n)$, respectively, as $B_j(\theta) = \underline{B}_o^{-1}\underline{B}_j$, $C_k(\theta) = \underline{B}_o^{-1}\underline{C}_k$, $A_\ell(\theta) = B_o^{-1}\underline{A}_\ell\underline{B}_o$, $\Omega_o = \underline{B}_o^{-1}\underline{\Omega}_o\underline{B}_o^{-1'}$ and $e(n) = \underline{B}_o^{-1}\underline{e}(n)$; $0 \le j \le r$, $0 \le k \le s$, $0 \le \ell \le t$. We can now rewrite Model 2 in the standard form (1.10). We make the following assumptions:

C1. This is the same as A2.(ii).

We take the permissible parameter space Θ to be the closure of the set Θ_o, where Θ_o is defined as follows:

C2.(i) $\|\theta\| \le R_1$, $|\det \underline{B}_o| \ge \varepsilon_1$, $|\det(\Sigma_j \underline{B}_j z^j)| > 0$ for $|z| \le 1$

(ii) $|\det(\Sigma_j \underline{A}_j z^j)| \ge \varepsilon_1$ for $|z| \le 1$.

(iii) A g.c.l.d. of $\Sigma_j \underline{B}_j z^j$, $\Sigma_j \underline{C}_j z^j$ and $\Sigma_j \underline{A}_j z^j$ is $I_{(p)}$.

(iv) The matrix $[\underline{B}_r, \underline{C}_s, \underline{A}_t]$ has rank p.

We will assume that $\theta_o \in \Theta_o$.

C3.(i) If $\theta \in \bar{\Theta}_h$ and $[B(z;\theta), C(z;\theta), A(z;\theta)] = [B(z;\theta_o), C(z;\theta_o), A(z;\theta_o)]$, then $\theta = \theta_o$. □

Theorem 7 Subject to assumptions C1.-C3., the results of Theorems 1-5 hold for Model 2. □

Remark 5.2 (i) Expressions for $\ddagger$ and $\ddagger_\theta$ (see the statement of Theorem 2) are given in 9.§3.

(ii) To the best of my knowledge, no complete treatment of the ARMAX case has yet been given in the literature. When the exogenous variables are missing and $\underline{B}_o \equiv I_{(p)}$ we obtain the ARMA model. The literature relating to Gaussian Estimates of the ARMA model is quite extensive, the most comprehensive treatment being given by Dunsmuir and Hannan [1976].

(iii) For the scalar version of Model 2, Hannan and Nicholls [1972, p.538, Theorem 1] obtain a two-step frequency domain estimator (see Remark (ii)in 5.§6) and show that it is strongly consistent and asymptotically efficient (in the sense of 1.§1). Using the notation in Hannan and Nicholls, it appears, however, that their proof of asymptotic Normality is incomplete in that the step of going from (Hannan and Nicholls [1972, top of p.540]) $|hw_y + \sum_{j=1}^{r} \delta(j)w_j|$ to I_w needs a result of the type given in Theorem 8 of §6 to make it valid. This step is also not justified in Nicholls [1971], to which Hannan and Nicholls refer for a complete proof.

(iv) The boundedness requirements on the $\underline{B}_j$ and $\underline{C}_k$ $(1 \leq j \leq r,\ 0 \leq k \leq s)$ matrices and the condition C2.(ii) are of no *practical* importance because R_1 can be chosen very large and ε_1 chosen very close to zero. For theoretical purposes, however, they are both unnecessary and undesirable. We can still obtain the strong consistency results even if the $\underline{B}_j$ and $\underline{C}_k$ $(1 \leq j \leq r,\ 0 \leq k \leq s)$ are not bounded and C2.(ii) is changed to: $\det(\Sigma_j \underline{A}_j z^j) \neq 0$ for $|z| \leq 1$, and θ_o can belong to $\Theta - \Theta_o$. The proof in the frequency domain would follow Hannan [1973c] and Dunsmuir and Hannan [1976] (see also Theorems 7.1 and 8.6 and their proofs); analogous techniques can be used in the time domain.

Unfortunately, I can see no way, at present, of relaxing the requirement that $\varepsilon_1 \leq |\det \underline{B}_o| \leq R' < \infty$, where R' is a constant.

6. SOME FURTHER APPLICATIONS

We now briefly discuss two other, potentially useful, applications of the theory.

Example 1 - Estimating the coefficients in a differential equation model.

Rather than give a general theory we illustrate the methodology using a simple example taken from Robinson [1976a, p.759-760]. We suppose that $y(t)$ is a scalar stationary process generated by

$$y(t) + \alpha \frac{dy(t)}{dt} = \beta x(t)+u(t), \; t \in (-\infty,\infty) \tag{6.10}$$

where $x(t)$ and $u(t)$ are stationary processes whose spectra are respectively $f_x^c(\lambda) = \sigma_1^2[(2\pi)(1+\mu^2\lambda^2)]^{-1}$ and $f_u^c(\lambda) = \sigma_2^2[2\pi(1+\nu^2\lambda^2)]^{-1}$, $\lambda \in (-\infty,\infty)$.

It follows from Robinson [1976a, p.760] that we can regard $y(n)$ $(n = 0,\pm1,\ldots)$ as being generated by

$$y(n) - ay(n-1) = bx(n) + cx(n-1) + \xi(n) \tag{6.11}$$

where $\xi(n)$ is an ARMA process

$$\xi(n)+\psi\xi(n-1)=e(n)+\rho e(n-1), E(e(n)e(m))=\delta_{nm}\sigma_e^2 \tag{6.12}$$

a, b, c, ψ and ρ are nonlinear functions of α, β, μ and ν. If we observe (6.10) for $t = 0, \pm 1, \ldots$ only, then we can obtain estimates of α, β, μ and ν by applying the theory of §§3 and 4 to (6.11) and (6.12). In principle the discussion generalizes to higher order, and vector, differential equations systems.

Example 2 - Nonrational $A(z;\theta)$ and $C(z;\theta)$

So far all our examples have dealt with $A(z;\theta)$ and $C(z;\theta)$ rational functions of z. Our theory, however, gives us more flexibility than this. For example, consider the simple model

$$y(n) = \sum_{j=0}^{\infty} \frac{\theta^j}{j!} x(n-j) + e(n)$$

Here $C(z;\theta) = e^{\theta z}$ and certainly satisfies our assumptions. Therefore we can estimate θ by the methods of §§3 and 4.

7. PROOFS

Throughout this section c is a universal constant, not always the same. Unless stated otherwise, it should be assumed that all convergence results below mean a.s. convergence; the mode of convergence will often not be indicated.

Let F_x be the sub-σ-field generated by the $x(n)$ sequences and put $F_n = F(G_n, F_x)$, i.e. F_n is the smallest σ-field containing the elements of G_n and F_x. When the $x(n)$ are nonstochastic F_x is the trivial σ-field.

Let $v(n;\theta)$ be defined by (3.10). It follows from Chapter 3 that:

1. Theorem 1 holds if A1.-A3., A5.(i) and A6.(ii) imply that

conditions S1.-S3. and S6.(i), (ii) hold. (The S. conditions are given below).

2. Theorems 2 and 3 hold if A1.-A6. imply that S1.-S6. hold.

S1. ($e(n)$, F_n, $n = 0,\pm1,\ldots$) is a sequence of p dimensional, stationary martingale differences with $E(e(n)e(n)'|F_{n-1}) = \Omega_o$.

S2. This is just assumption A3.(i).

S3.(i) $\forall\ n \geq 1$ and $\forall\ \theta, \theta_+ \in \Theta$, $v(n;\theta_o) = e(n)$ and $\{v(n;\theta) - v(n;\theta_+)\}$ is F_{n-1} measurable. (ii) a.s., and uniformly in $\theta \in \Theta$, $\Omega_N(\theta)$ converges to $\Omega(\theta)$. (iii) a.s., and uniformly in $\theta \in \Theta$,

$$N^{-1} \sum_{n=1}^{N} v(n;\theta)e(n)'$$

converges to a nonstochastic finite limit. (iv) $\forall\ n \geq 1$, and a.s., $v(n;\theta)$ is a continuous function of θ for $\theta \in \Theta$.

S4.(i) $\forall\ n \geq 1$, $\forall\ \theta \in U_{\alpha_1}$ and a.s., $v(n;\theta)$ is twice differentiable with respect to θ, with continuous first and second derivatives.

$$\text{(ii)}\ N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial \theta_i} e(n)',\ N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial \theta_i} \frac{\partial v(n;\theta)'}{\partial \theta_j},$$

$$N^{-1} \sum_{n=1}^{N} \frac{\partial^2 v(n;\theta)}{\partial \theta_i \partial \theta_j} v(n;\theta)',\ N^{-1} \sum_{n=1}^{N} \frac{\partial^2 v(n;\theta)}{\partial \theta_i \partial \theta_j}\ e(n)'$$

and $N^{-1} \sum_{n=1}^{N} \frac{\partial^2 v(n;\theta)}{\partial \theta_i \partial \theta_j} \frac{\partial^2 v(n;\theta)'}{\partial \theta_k \partial \theta_\ell}$ converge a.s., and uniformly in $\theta \in U_{\alpha_1}$, to finite nonstochastic limits; ($i,j,k,\ell=1,\ldots,m$).

Put $V(n) = (\partial/\partial\theta')v(n;\theta_o)$

S5.(i) $\limsup_N N^{-1} \sum_{n=1}^{N} E(\|V(n)\|^2) < \infty$

(ii) For each $n \geq 1$ we can write

$V(n) = V_r(n) + \tilde{V}_r(n)$, $r = 1,2,\ldots$ and :

$V_r(n) = V_{1r}(n) + V_{2r}(n)$ where: (a) $V_{1r}(n)$ is stationary, F_{n-1} measurable, and square-integrable, while $V_{2r}(n)$ is nonstochastic.

(b) $\lim_N N^{-1} \sum_{n=1}^{N} V_{2r}(n)V_{2r}(n)'$ exists and is finite.

(c) $\lim_N N^{-1} \sum_{n=1}^{N} V_r(n)'\Omega_o^{-1}V_r(n) \overset{p}{=} \Delta_r \;\; \forall\, r \geq 1$, where Δ_r is a nonstochastic matrix and $\lim_r \Delta_r = \Delta$.

(d) $\forall\, \varepsilon > 0 \;\; \exists R = R(\varepsilon)$ s.t.

$$\limsup_N N^{-1} \sum_{n=1}^{N} E\|\tilde{V}_r(n)\|^2 < \varepsilon \quad \text{for} \quad r \geq R.$$

S6.(i) This is A5.(i).

Let $\Omega(\theta)$ be defined by (2.10).

(ii) If $\theta \in \Theta_h$ and $\Omega(\theta) = \Omega(\theta_o)$, then $\theta = \theta_o$.

(iii) This is A5.(ii) and (iii) and A6.(i).□

To obtain the required results we need the following lemma:

Lemma 1 Suppose that: (i) $\underset{\sim}{X}$ is a subset of Euclidean space with typical element x; (ii) $(X_j(x), j \geq 0)$ and $(Y_j(x), j \geq 0)$ are two matrix sequences with elements that are functions of x for $x \in \underset{\sim}{X}$; (iii) $(z(n), n = 0,\pm1,\ldots)$ is a stochastic vector sequence satisfying

$$\lim_N N^{-1} \sum_{n=1}^{N} z(n-j)z(n-k)' \overset{a.s.}{=} \Gamma(j-k)$$

where $\Gamma(j)$ is a finite nonstochastic matrix $\forall\, j$, and $\limsup_N N^{-1} \|z(-N)\|^2 = 0$ a.s.

If $(m_j, j \geq 0)$ is a sequence of non-negative scalar constants such that:

$\sum_{j=o}^{\infty} jm_j < \infty$ and $\|X_j(x)\| \leq m_j$, $\|Y_j(x)\| \leq m_j \; \forall \; x \in \underset{\sim}{X}$ and

$\forall \; j \geq 0$, then: (a) For all $n \geq 1$

$$\sum_{j=o}^{\infty} X_j(x)z(n-j)$$

converges uniformly in $x \in \underset{\sim}{X}$.

(b) $\lim_N N^{-\frac{1}{2}} \sum_{j=0}^{\infty} X_j(x)z(-N-j) = 0$ a.s.

uniformly in $x \in \underset{\sim}{X}$.

(c) Put

$$\zeta_N(x,x_+) = N^{-1} \sum_{n=1}^{N} \left\{ \sum_{j=o}^{\infty} X_j(x)z(n-j) \right\} \left\{ \sum_{k=o}^{\infty} Y_k(x_+)z(n-k) \right\}'.$$

Then

$$\lim_N \zeta_N(x,x_+) = \zeta(x,x_+) = \sum_{j=o}^{\infty} \sum_{k=o}^{\infty} X_j(x)\Gamma(j-k)Y_k(x_+) \text{ a.s.,}$$

the convergence being uniform in x, $x_+ \in \underset{\sim}{X}$.

Proof (a) Fix $n \geq 1$ and choose $r > n$. It is sufficient to show that $\forall \; \varepsilon > 0$, $\exists \; J = J(\varepsilon)$, J possibly dependent on n and also on the realization, such that

$$\sup_{x \in \underset{\sim}{X}} \left\| \sum_{j=r+1}^{\infty} X_j(x)z(n-j) \right\| < \varepsilon \quad \text{for} \quad r \geq J.$$

Now,

$$\left\| \sum_{j=r+1}^{\infty} X_j(x)z(n-j) \right\| \leq \sum_{j=r+1}^{\infty} j^{\frac{1}{2}} m_j \times \{ j^{-\frac{1}{2}} \| z(n-j) \| \}$$

and the result follows.

(b) The proof of (b) is trivial.

(c) Put

$$C_N(j,k) = N^{-1} \sum_{n=1}^{N} z(n-j)z(n-k)' \quad \forall \; j,k.$$

For each $r > 0$, let

$\lambda_{rN}(x,x_+) = \sum_{j=o}^{r} \sum_{k=o}^{r} X_j(x)C_N(j,k)Y_k(x_+)$, $\lambda_r(x,x_+) =$

$= \sum_{j=o}^{r} \sum_{k=o}^{r} X_j(x)\Gamma(j-k)Y_k(x_+)$ and $\mu_{rN}(x,x_+) =$

$= \zeta_N(x,x_+) - \lambda_{rN}(x,x_+)$. For fixed r,

$$\|\lambda_{rN}(x,x_+)-\lambda_r(x,x_+)\| \leq \sum_{j=o}^{r} \sum_{k=o}^{r} m_j m_k \|C_N(j,k)-\Gamma(j-k)\| \to 0$$

as $N \to \infty$. Hence $\forall\ \varepsilon > 0\ \exists\ N_1 = N_1(\varepsilon,r)$, possibly realization dependent, s.t.

$$\sup_{x,x_+\in \underset{\sim}{X}} \|\lambda_{rN}(x,x_+)-\lambda_r(x,x_+)\| < \varepsilon \text{ for } N \geq N_1 \tag{7.09}$$

$$\|\lambda_r(x,x_+)-\zeta(x,x_+)\| \leq \sum_{j=r+1}^{\infty} \sum_{k=o}^{\infty} m_j m_k \|\Gamma(j-k)\| +$$

$$+ \sum_{j=o}^{r} \sum_{k=r+1}^{\infty} m_j m_k \|\Gamma(j-k)\| \to 0 \text{ as } r \to \infty$$

because $\|\Gamma(j)\| \leq \operatorname{tr} \Gamma(o)\ \forall\ j$.

Hence $\forall\ \varepsilon > 0\ \exists\ J_1 = J_1(\varepsilon)$ s.t.

$$\sup_{x,x_+\in \underset{\sim}{X}} \|\lambda_r(x,x_+)-\zeta(x,x_+)\| < \varepsilon \quad \text{for } r \geq J_1 \tag{7.10}$$

$$\sum_{j=r}^{\infty} m_j (N^{-1} \sum_{n=1}^{N} \|z(n-j)\|^2) \leq \sum_{j=r}^{\infty} j m_j (N^{-1} \sum_{n=1}^{j} j^{-1}\|z(n-j)\|^2) +$$

$$+ \sum_{j=r}^{\infty} m_j (N^{-1} \sum_{n=1}^{N} \|z(n)\|^2) \to 0 \text{ as } r \to \infty \tag{7.11}$$

$$\mu_{rN}(x,x_+) = \sum_{j=r+1}^{\infty} \sum_{k=o}^{\infty} X_j(x) C_N(j,k) Y_k(x_+) +$$

$$+ \sum_{j=o}^{r} \sum_{k=r+1}^{\infty} X_j(x) C_N(j,k) Y_k(x_+).$$

Hence,

$$\|\mu_{rN}(x,x_+)\| \leq \sum_{j=r+1}^{\infty} \sum_{k=o}^{\infty} m_j m_k (N^{-1} \sum_{n=1}^{N} \|z(n-j)\|^2 + N^{-1} \sum_{n=1}^{N} \|z(n-k)\|^2) +$$

$$+ \sum_{j=o}^{r} \sum_{k=r+1}^{\infty} m_j m_k (N^{-1} \sum_{n=1}^{N} \|z(n-j)\|^2 + N^{-1} \sum_{n=1}^{N} \|z(n-k)\|^2)$$

It follows from (7.11) that $\forall\ \varepsilon > 0\ \exists$ a.s. a $J_2 = J_2(\varepsilon)$ s.t.

$$\sup_{x,x_+\in \underset{\sim}{X}} \|\mu_{rN}(x,x_+)\| < \varepsilon \qquad \text{for} \quad r \geq J_2 \tag{7.12}$$

Given $\varepsilon > 0$, we choose $r \geq \max\{J_1(\varepsilon), J_2(\varepsilon)\}$ and then keep it fixed. We then choose $N \geq N_1(\varepsilon, r)$.

Because

$$\|\zeta_N(x,x_+)-\zeta(x,x_+)\| \leq \|\lambda_{rN}(x,x_+)-\lambda_r(x,x_+)\|+$$
$$+ \|\lambda_r(x,x_+)-\zeta(x,x_+)\|+\|\mu_{rN}(x,x_+)\|$$

it follows from (7.09), (7.10) and (7.12) that

$$\sup_{x,x_+\in \underset{\sim}{X}} \|\zeta_N(x,x_+)-\zeta(x,x_+)\| \leq 3\varepsilon \text{ (a.s.)} \quad \text{if} \quad N \geq N_1(\varepsilon, r).$$

This completes the proof because neither ζ_N nor ζ depends on r. □

Lemma 2 Let the matrix functions $G(z;\theta)$, $F(z;\theta)$, $D(z)$ and $K(z)$ be defined as in §2, and suppose assumption A3. holds. Then there exists a positive constant c and a number $R_1 > 1$, c and R_1 being independent of j and θ, s.t. for each $j \geq 0$ the norms of $G_j(\theta)$, $F_j(\theta)$, D_j and K_j are each less than c/R_1^j, uniformly in $\theta \in \Theta$.

Proof The lemma is an immediate consequence of assumptions A3.(v) and (vi). □

Lemma 3 Suppose that assumptions A1., A2.(i) or (i') and A3. hold. Then, for any integer j,

$$\lim_N N^{-1} \sum_{n=1}^{N} x(n)e(n+j)' = 0 \qquad \text{a.s.} \tag{7.15}$$

$$\lim_N N^{-1} \sum_{n=1}^{N} y(n)y(n+j)' = \int_{-\pi}^{\pi} e^{ij\lambda}D(\lambda)dF_x(\lambda)D(\lambda)^* +$$
$$+ (2\pi)^{-1} \int_{-\pi}^{\pi} e^{ij\lambda}K(\lambda)\Omega_o K(\lambda)^* d\lambda \qquad \text{a.s.} \tag{7.16}$$

$$\lim_N N^{-1} \sum_{n=1}^{N} y(n)x(n+j)' = \int_{-\pi}^{\pi} e^{ij\lambda} D(\lambda)\, dF_x(\lambda) \quad \text{a.s.} \tag{7.17}$$

$$\lim_N N^{-1} \| y(-N) \|^2 = 0 \qquad \text{a.s.} \tag{7.18}$$

Proof (5.15) follows from Lemma 3.2(i) and assumptions A1. and A2.(i) or A2.(i'). From (1.10)

$$y(n) = \sum_{j=o}^{\infty} D_j x(n-j) + \sum_{j=o}^{\infty} K_j e(n-j) \tag{7.20}$$

By Lemma 2 $\Sigma_j\, j\|D_j\| < \infty$, $\Sigma_j\, j\|K_j\| < \infty$ so that the right side of (5.20) converges for any n by Lemma 1(a). (7.16) and (7.17) follows from (7.15) and Lemma 1(c), while (7.18) follows from Lemma 1(b). □

Lemma 4 Suppose assumptions A1.-A3., A5.(i) and A6.(ii) hold. Then conditions S1.-S3. and S6.(i) and (ii) also hold.

Proof S1., S2., S3.(i) and S6.(i) hold trivially. $\Omega_N(\theta) \to \Omega(\theta)$ a.s. and uniformly in $\theta \in \Theta$ by Lemmas 1-3. Therefore S3.(ii) holds. S3.(iii) holds by a similar argument. Finally, we show that S6.(ii) holds.

Let $\Phi(z;\theta)$ and $\Psi(z;\theta)$ be as defined in §2. For fixed $\theta \in \Theta$, $\Omega(\theta) \geq \sum_{j=o}^{\infty} \Psi_j(\theta)\Omega_o\Psi_j(\theta)' \geq \Omega_o$ (because $\Psi_o \equiv I_{(p)}$) with equality if and only if $\Psi_j(\theta) = 0 \;\forall j \geq 1$ and

$$\int_{-\pi}^{\pi} \Phi(\lambda;\theta) f_x(\lambda) \Phi(\lambda;\theta)^* d\lambda = 0 \tag{7.25}$$

It follows from assumption A2.(ii) that (7.25) holds if and only if $\Phi(\lambda;\theta) = 0 \;\forall\, \lambda \in [-\pi,\pi]$. Hence $\Omega(\theta) = \Omega(\theta_o)$ if and only if

$$[C(\lambda;\theta), A(\lambda;\theta)] = B(\lambda;\theta)B(\lambda;\theta_o)^{-1}[C(\lambda;\theta_o), A(\lambda;\theta_o)] \;\forall \lambda \in [-\pi,\pi].$$

S6.(ii) now follows from A6.(ii). □

Having shown that assumptions A1.-A3., A5.(i) and A6.(ii) *imply that conditions* S1.-S3. *and* S6.(i) *and* (ii) *hold, it follows from Chapter 3 that Theorem 1 holds.*

Lemma 5 Suppose assumptions A3. and A4. hold. Then there exists a positive number c and a number $R_2 > 1$, c and R_2 being independent of k and θ, s.t. the norms of $(\partial/\partial\theta_i)G_k(\theta)$, $(\partial^2/\partial\theta_i\partial\theta_j)G_k(\theta)$, $(\partial/\partial\theta_i)F_k(\theta)$ and $(\partial^2/\partial\theta_i\partial\theta_j)F_k(\theta)$ are each less than c/R_1^k, uniformly in $\theta \in U_{\alpha_1}$; $(i,j = 1,\ldots,m)$.

Proof The lemma is an immediate consequence of assumptions A4.(i)-(iii). □

Lemma 6 If assumptions A1.-A4. hold, then conditions S4.(i) and (ii) hold.

Proof (i) By Lemmas 1-3 and 5

$$\sum_{k=o}^{\infty} \frac{\partial G_k(\theta)}{\partial\theta_i} y(n-k) - \sum_{k=o}^{\infty} \frac{\partial F_k(\theta)}{\partial\theta_i} x(n-k) \qquad i = 1,\ldots,m \tag{7.30}$$

and

$$\sum_{k=o}^{\infty} \frac{\partial^2 G_k(\theta)}{\partial\theta_i\partial\theta_j} y(n-k) - \sum_{k=o}^{\infty} \frac{\partial^2 F_k(\theta)}{\partial\theta_i\partial\theta_j} x(n-k) \quad i,j=1,\ldots,m \tag{7.31}$$

converge a.s. , and uniformly in $\theta \in U_{\alpha_1}$, for $n \geq 1$ and $i,j = 1,\ldots,m$.

By Lemmas 1-3 the right side of (3.10) also converges for $n \geq 1$ and for $\theta = \theta_o$. It follows (Rudin [1964,p.136, Theorem 7.12 and p.140, Theorem 7.17]) that $v(n;\theta)$ is twice differentiable with continuous first and second derivatives for $n \geq 1$ and for $\theta \in U_{\alpha_1}$; $(\partial/\partial\theta_i)v(n;\theta)$ and $(\partial^2/\partial\theta_i\partial\theta_j)v(n;\theta)$ are given by (7.30) and (7.31) respectively. Hence S4.(i) holds.

Using Lemmas 1-3 and 5, we can deduce from (3.10), (7.30) and (7.31) that S4.(ii) also holds. □

Lemma 7 If assumptions A1.-A4. hold, then S5. holds.

Proof For convenience we separate the case where assumption A2.(i) holds from the case where A2.(i') holds. By (3.10) and (7.20)

$$v(n;\theta) = \sum_{j=o}^{\infty} \Phi_j(\theta)x(n-j) + \sum_{j=o}^{\infty} \Psi_j(\theta)e(n-j)$$

For each $r \geq 1$ we put

$$\Phi_r(z;\theta) = \sum_{j=o}^{r} \Phi_j(\theta)z^j, \quad \Psi_r(z;\theta) = \sum_{j=o}^{r} \Psi_j(\theta)z^j$$

(a) Suppose A2.(i) holds. We put

$$v_r(n;\theta) = \sum_{j=o}^{r} \Phi_j(\theta)x(n-j) + \sum_{j=o}^{r} \Psi_j(\theta)e(n-j)$$

$V_{1r}(n) = (\partial/\partial\theta')v_r(n;\theta_o)$, $V_{2r}(n) \equiv 0$, $V_r(n) = V_{1r}(n)$,

$\tilde{V}_r(n) = V(n) - V_r(n)$ for $r \geq 1$.

For this decomposition of $V(n)$, S5.(ii)(a) and (b) are certainly satisfied.

$$E(\|\frac{\partial v(n;\theta_o)}{\partial\theta_i}\|^2) = \mathrm{tr}\left\{\sum_{j,k=o}^{\infty} \frac{\partial\Phi_j(\theta_o)}{\partial\theta_i}\Gamma_x(j-k)\frac{\partial\Phi_k(\theta_o)'}{\partial\theta_i} + \right.$$

$$\left. + \sum_{j=o}^{\infty} \frac{\partial\Psi_j(\theta_o)}{\partial\theta_i}\,\Omega_o\,\frac{\partial\Psi_j(\theta_o)'}{\partial\theta_i}\right\} < \infty, \quad i = 1,\ldots,m \quad (7.40)$$

Hence S5.(i) holds.

$$\lim_N \mathrm{tr}\,\Omega_o^{-1}\left\{N^{-1}\sum_{n=1}^{N} \frac{\partial v_r(n;\theta_o)}{\partial\theta_i}\,\frac{\partial v_r(n;\theta_o)'}{\partial\theta_j}\right\} =$$

$$= \mathrm{tr}\,\Omega_o^{-1}\left[\int_{-\pi}^{\pi} \frac{\partial\Phi_r(\lambda;\theta_o)}{\partial\theta_i}\,dF_x(\lambda)\,\frac{\partial\Phi_r(\lambda;\theta_o)^*}{\partial\theta_j} + \right.$$

$$+ (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{\partial \Psi_r(\lambda;\theta_o)}{\partial \theta_i} \Omega_o \frac{\partial \Psi_r(\lambda;\theta_o)^*}{\partial \theta_j} d\lambda \Bigg] \text{ a.s.} = (\Delta_r)_{ij}$$

say. Obviously $\lim_r \Delta_r = \Delta$ so that S5.(ii)(c) is satisfied. Finally, the i^{th} column of $\tilde{V}_r(n)$, $\tilde{V}_r(n,i)$ say, equals

$$\sum_{j=r+1}^{\infty} \frac{\partial \Phi_j(\theta_o)}{\partial \theta_i} x(n-j) + \sum_{j=r+1}^{\infty} \frac{\partial \Psi_j(\theta_o)}{\partial \theta_i} e(n-j)$$

so that

$$E(\|\tilde{V}_r(n,i)\|^2) \leq c \operatorname{tr} \int_{-\pi}^{\pi} \|(\partial/\partial\theta_i)\{\Phi(\lambda;\theta_o)-\Phi_r(\lambda;\theta_o)\}\|^2 dF_x(\lambda) +$$

$$+ c \int_{-\pi}^{\pi} \|(\partial/\partial\theta_i)\{\Psi(\lambda;\theta_o)-\Psi_r(\lambda;\theta_o)\}\|^2 d\lambda \to 0 \text{ as } r \to \infty$$

by dominated convergence. Because $V_r(n)$ is stationary, S5.(ii)(d) follows.

(b) We now briefly discuss the proof when A2.(i') holds. For each $r \geq 1$ we put

$$v_{1r}(n;\theta) = \sum_{j=o}^{r} \Psi_j(\theta)e(n-j), v_{2r}(n;\theta) = \sum_{j=o}^{r} \Phi_j(\theta)x(n-j),$$

$$V_{1r}(n) = (\partial/\partial\theta')v_{1r}(n;\theta_o), \ V_{2r}(n) = (\partial/\partial\theta')v_{2r}(n;\theta_o).$$

For this decomposition of $V(n)$ S5.(ii)(a) is satisfied.

$$E(\|\frac{\partial v(n;\theta_o)}{\partial \theta_i}\|^2) = \operatorname{tr} \left\{ \sum_{j,k=o}^{\infty} \frac{\partial \Phi_j(\theta_o)}{\partial \theta_i} x(n-j)x(n-k)' \frac{\partial \Phi_k(\theta_o)'}{\partial \theta_i} + \right.$$

$$\left. + \sum_{j=o}^{\infty} \frac{\partial \Psi_j(\theta_o)}{\partial \theta_i} \Omega_o \frac{\partial \Psi_j(\theta_o)'}{\partial \theta_i} \right\} .$$

Therefore $\lim_N N^{-1} \Sigma_n E(\|(\partial/\partial\theta_i)v(n;\theta_o)\|^2)$ equals the right side of (7.40) and so is bounded. Hence S5.(i) is satisfied and it is evident that S5.(ii)(b) is also. S5.(ii)(c) and (d) may be shown similarly using a proof analogous to that given in (a). □

We have now shown that assumptions A1.-A6. imply that conditions S1.-S6. hold. It follows from Chapter 3 that Theorems 2 and 3 also hold.

Proof of Theorem 4 Only a brief discussion is given as the proof of the theorem follows readily from the above analysis. From (1.10) and assumption A1.'

$$y(n) = -\sum_{k=o}^{n-1} H_k\left\{\sum_{j=n-k}^{r}{}' B_j(\theta_o)y(n-j-k)\right\}+$$

$$+\sum_{k=o}^{n-1} H_k\left\{\sum_{j=o}^{\infty} C_j(\theta_o)x(n-j-k) + \sum_{j=o}^{\infty} A_j(\theta_o)e(n-j-k)\right\},$$

and we can rewrite this as

$$y(n) = \sum_{k=o}^{n-1} D_k\, x(n-k) + \sum_{k=o}^{n-1} K_k\, e(n-k) + q(n)$$

where $q(n) = -\sum_{k=n-r}^{n-1} H_k\left\{\sum_{j=n-k}^{r} B_j(\theta_o)y(n-j-k)\right\} +$

$$+\sum_{\ell=n}^{\infty}\left\{\sum_{k=o,k\leq n-1}^{\ell} H_k C_{\ell-k}(\theta_o)\right\}x(n-\ell)+\sum_{\ell=n}^{n+t}\left\{\sum_{k=o,k\leq n-1}^{\ell} H_k A_{\ell-k}(\theta_o)\right\}e(n-\ell)$$

Put $w(n) = y(n) - q(n)$. It is not difficult to show that $\lim_n q(n) = 0$ a.s., $\lim_n n^{-1}\|w(-n)\|^2 = 0$ a.s. and

$$N^{-1}\sum_{n=1}^{N} w(n)w(n+j)' \quad \text{and} \quad N^{-1}\sum_{n=1}^{N} w(n)x(n+j)'$$

converge a.s. to the right sides of (7.16) and (7.17), respectively.

We can rewrite (3.25) as

$$v(n;\theta) = \sum_{j=o}^{n-1} G_j(\theta)w(n-k) - \sum_{j=o}^{n-1} F_k(\theta)x(n-k)+p(n;\theta)$$

and it is not hard to show that

$$\overline{\lim}_n \sup_{\theta\in\Theta}\|p(n;\theta)\| = 0 \text{ a.s.}, \quad \overline{\lim}_n \sup_{\theta\in U_{\alpha_1}}\left\|\frac{\partial p(n;\theta)}{\partial\theta_i}\right\| = 0 \text{ a.s.},$$

$$\overline{\lim}_n \sup_{\theta\epsilon U_{\alpha_1}} \left\| \frac{\partial^2 p(n;\theta)}{\partial\theta_i \partial\theta_j} \right\| = 0 \text{ a.s.}, \quad (i,j = 1,\ldots,m).$$

The rest of the proof now proceeds in the same way as, but is simpler than, the proof given for Theorems 1-3. □

We will deduce Theorem 5 from Theorems 1-3. To do this we will need the results given below. Alternatively, we could prove the strong consistency result in Theorem 5 directly by using the techniques of Chapters 7 and 8. See also Remark 5.2(iv).

Suppose that $\{a(n),\ b(n);\ n = 0,\pm 1,\ldots\}$ are two sequences of vector random variables, the element ξ belongs to the set Ξ and $\{P_j(\xi),\ Q_j(\xi);\ j \geq 0\}$ are two sequences of matrix functions of ξ.

We put

$$p(n;\xi) = \sum_{j=o}^{\infty} P_j(\xi)a(n-j), q(n;\xi) = \sum_{j=o}^{\infty} Q_j(\xi)b(n-j),$$

$$P(z;\xi) = \sum_{j=o}^{\infty} P_j(\xi)z^j,\ Q(z;\xi) = \sum_{j=o}^{\infty} Q_j(\xi)z^j;$$

$w_p(\lambda;\xi) = N^{-\frac{1}{2}} \sum_{n=1}^{N} p(n;\xi)e^{in\lambda}$, and define $w_a(\lambda)$, $w_b(\lambda)$ and $w_q(\lambda;\xi)$ similarly to $w_p(\lambda;\xi)$, but with respect to the $a(n)$, $b(n)$ and $q(n;\xi)$ terms, respectively.

Put
$$\nu_N(\lambda;\xi) = N^{\frac{1}{2}}\{w_p(\lambda;\xi) - P(\lambda;\xi)w_a(\lambda)\}$$
$$\zeta_N(\lambda;\xi) = N^{\frac{1}{2}}\{w_q(\lambda;\xi) - Q(\lambda;\xi)w_b(\lambda)\}.$$

For convenience, we will often omit to indicate dependence on N and ξ below.

Lemma 8 $\nu(\lambda) = \sum_{j=o}^{\infty} P_j\, e^{ij\lambda} R_{j,N}(\lambda)$, where

$$R_{j,N}(\lambda) = \sum_{m=-j+1}^{o} a(m)e^{im\lambda} - \sum_{m=N-j+1}^{N} a(m)e^{im\lambda}, \quad 0 \leq j \leq N$$

$$= \sum_{m=-j+1}^{N-j} a(m)e^{im\lambda} - \sum_{m=1}^{N} a(m)e^{im\lambda}, \quad j > N$$

A similar decomposition can be obtained for $\zeta(\lambda)$.

Proof

$$\sum_{n=1}^{n} p(n)e^{in\lambda} = \sum_{j=o}^{\infty} P_j e^{ij\lambda} S_{j,N}(\lambda), \quad \text{where}$$

$S_{j,N}(\lambda) = \sum_{n=1}^{N} a(n-j)e^{i(n-j)\lambda}$. Put $u_n = \sum_{k=-\infty}^{n} a(k)e^{ik\lambda}$.

Then $S_{j,N}(\lambda) = u_{N-j} - u_{-j}$. i.e.

$$S_{j,N}(\lambda) = u_N - u_o + u_o - u_{-j} - (u_N - u_{N-j}), \quad 0 \leq j \leq N$$

$$= u_N - u_o + (u_{N-j} - u_{-j}) - (u_N - u_o), \quad j > N.$$

The required result follows. □

Below, we will denote $\sup_{\xi\in\Xi}$ by $\sup_\xi$.

Lemma 9 If: (i) $\sup_\xi \|P_j(\xi)\| \leq m_j$ and $\sup_\xi \|Q_j(\xi)\| \leq m_j$, $\forall\, j \geq 0$, where $\{m_j,\ j \geq 0\}$ is a sequence of non-negative numbers satisfying $\Sigma_j m_j < \infty$,

(ii) a.s.

$$\lim_N N^{-1} \sum_{n=1}^{N} a(n)a(n)' \quad \text{and} \quad \lim_N N^{-1} \sum_{n=1}^{N} b(n)b(n)'$$

exist and are finite,

(iii) $\lim_n n^{-\frac{1}{2}}\|a(-n)\| = 0$ a.s., $\lim_n n^{-\frac{1}{2}}\|b(-n)\| = 0$ a.s.,

then $\lim\sup_N \sup_\xi N^{-2} \sum_{t=1}^{N} \|\nu(\lambda_t)\|^2 = 0$ a.s., $\lambda_t = 2\pi t/N$.

(iv) If, in addition, $\exists$ a finite positive constant c s.t. $E(\|a(n)\|^2) \leq c$ and $E(\|b(n)\|^2) \leq c$ for $n = 0, \pm 1, \ldots,$ then $\forall\ \varepsilon > 0\ \exists\, K = K(\varepsilon)$ s.t.

$$P(\sup_\xi N^{-1} \sum_{t=1}^{N} \|\nu(\lambda_t)\|^2 > K) < \varepsilon.$$

Similar results hold for $\zeta(\lambda)$.

Proof $N^{-1} \sum_{t=1}^{N} \nu(\lambda_t)\nu(\lambda_t)^* = N^{-1} \sum_t \Big\{ \sum_{j,k=o}^{N} T_{j,k,N}(t) +$

$$+ \sum_{j=o}^{N} \sum_{k=N+1}^{\infty} T_{j,k,N}(t) + \sum_{j=N+1}^{\infty} \sum_{k=o}^{N} T_{j,k,N}(t) +$$

$$+ \sum_{j=N+1}^{\infty} \sum_{k=N+1}^{\infty} T_{j,k,N}(t) \Big\} =$$

$$= I_N + II_N + III_N + IV_N \text{ say,}$$

where $T_{j,k,N}(t) = P_j R_{j,N}(\lambda_t) R_{k,N}(\lambda_t)^* P_k' \, e^{i(j-k)\lambda_t}$ with $R_{j,N}(\lambda)$ defined in Lemma 8 above. Putting

$$S_{j,k,m,n}(t) = P_j \; a(m)a(n)'P_k' \exp\{i(j-k+m-n)\lambda_t\}$$

we can write I_N as $I_N = I_N^{(1)} - I_N^{(2)} - I_N^{(3)} + I_N^{(4)}$ where

$$I_N^{(1)} = N^{-1} \sum_t \sum_{j,k=o}^{N} \sum_{m=-j+1}^{o} \sum_{n=-k+1}^{o} S_{j,k,m,n}(t) =$$

$$= \sum_{j,k=o}^{N} \sum_{f=-\infty}^{\infty} {\textstyle\ddagger_m} P_j \; a(m)a(j-k+m-fN)'P_k'$$

and $\ddagger_m$ is a sum over $-j+1 \le m \le 0$, $-k+1 \le j-k+m-fN \le 0$.
It follows that

$$\|I_N^{(1)}\| \le \sum_{j,k=o}^{N} \sum_{m=-j+1}^{\min(o,k-j)} \|P_j\|\|P_k\|\|a(m)\|\|a(j-k+m\|$$

$$\le c \sum_{j,k=o}^{N} j^{\frac12}k^{\frac12}m_j m_k \{(j+1)(k+1)\}^{-\frac12} \sum_{m=-j+1}^{\min(o,k-j)} \{\|a(m)\|^2 +$$

$$+ \|a(j-k+m)\|^2\}$$

Below we assume that conditions (i) - (iv) of the lemma hold unless stated otherwise.

Using Chebyshev's inequality, it follows from the above that $\forall\ \varepsilon > 0\ \ \exists K_1 = K_1(\varepsilon)$ s.t.

$$P(\sup_\xi \|I_N^{(1)}\| > K) < \varepsilon \quad \text{for} \quad K \geq K_1 \tag{7.45}$$

Furthermore, if only conditions (i) - (iii) hold, then it is easily shown that

$$\lim\sup_N \sup_\xi N^{-1}\|I_N^{(1)}\| = 0 \quad \text{a.s.} \tag{7.46}$$

$$I_N^{(2)} = N^{-1} \sum_t \sum_{j,k=o}^{N} \sum_{m=-j+1}^{o} \sum_{n=N-k+1}^{N} S_{j,k,m,n}(t) =$$

$$= \sum_{j,k=o}^{N} \sum_{m=-j+1}^{\min(o,k-j)} P_j\ a(m)a(j-k+m+N)'P_k'$$

$$I_N^{(3)} = N^{-1} \sum_t \sum_{j,k=o}^{N} \sum_{m=N-j+1}^{N} \sum_{n=-k+1}^{o} S_{j,k,m,n}(t)$$

$$I_N^{(4)} = N^{-1} \sum_t \sum_{j,k=o}^{N} \sum_{m=N-j+1}^{N} \sum_{n=N-k+1}^{N} S_{j,k,m,n}(t) =$$

$$= \sum_{j,k=o}^{N} P_j \left\{ \sum_{m=N-j+1}^{\min(N,N-j+k)} a(m)a(j-k+m)' \right\} P_k'$$

and it is not hard to show that results analogous to (7.45) and (7.46) hold for $I_N^{(2)},\ldots,I_N^{(4)}$ and hence for I_N.

We can write II_N as $II_N = II_N^{(1)} - II_N^{(2)} - II_N^{(3)} + II_N^{(4)}$, where

$$II_N^{(1)} = N^{-1} \sum_t \sum_{j=o}^{N} \sum_{k=N+1}^{\infty} \sum_{m=-j+1}^{o} \sum_{n=-k+1}^{N-k} S_{j,k,m,n}(t) =$$

$$= \sum_{j=o}^{N} \sum_{k=N+1}^{\infty} \sum_{n=-k+1}^{j-k} P_j\ a(n+k-j)a(n)'P_k'$$

$$II_N^{(2)} = N^{-1} \sum_t \sum_{j=o}^{N} \sum_{k=N+1}^{\infty} \sum_{m=-j+1}^{o} \sum_{n=1}^{N} S_{j,k,m,n}(t).$$

Let $k = sN + k'$ ($s = 1,2,\ldots;\ k' = 1,\ldots,N$). Hence

$$II_N^{(2)} = \sum_{j=o}^{N} \sum_{s=1}^{\infty} \sum_{k'=1}^{N} P_j \left\{ \sum_{n=1}^{j-k'} a(n+k'-j)a(n)' + \right.$$

$$\left. + \sum_{n=N-k'+1}^{N+j-k'} a(n-N+k'-j)a(n)' \right\} P_{sN+k'}'.$$

$$II_N^{(3)} = N^{-1}\sum_t \sum_{j=o}^{N} \sum_{k=N+1}^{\infty} \sum_{m=N-j+1}^{N} \sum_{n=-k+1}^{N-k} S_{j,k,m,n}(t) =$$

$$= \sum_{j=o}^{N} \sum_{k=N+1}^{\infty} P_j \left\{ \sum_{n=-k+1}^{j-k} a(n+N+k-j)a(n)' \right\} P_k'$$

$$II_N^{(4)} = N^{-1} \sum_t \sum_{j=o}^{N} \sum_{k=N+1}^{\infty} \sum_{m=N-j+1}^{N} \sum_{n=1}^{N} S_{j,k,m,n}(t).$$

Let $k = sN + k'$ ($s = 1,2,\ldots$; $k' = 1,\ldots,N$). Hence

$$II_N^{(4)} = \sum_{j=o}^{N} \sum_{s=1}^{\infty} \sum_{k'=1}^{N} P_j \left\{ \sum_{n=1}^{j-k'} a(n+k'-j+N)a(n)' + \right.$$

$$\left. + \sum_{n=N-k'+1}^{\min(N,N+j-k')} a(n+k'-j)a(n)' \right\} P'_{sN+k'}$$

and we will have results analogous to (7.45) and (7.46) holding for each of $II_N^{(1)},\ldots,II_N^{(4)}$ and hence for II_N. By the symmetry between j and k, similar results will also hold for III_N.

We can write IV_N as $IV_N = IV_N^{(1)} - IV_N^{(2)} - IV_N^{(3)} + IV_N^{(4)}$, where

$$IV_N^{(1)} = N^{-1} \sum_t \sum_{j,k=N+1}^{\infty} \sum_{m=-j+1}^{N-j} \sum_{n=-k+1}^{N-k} S_{j,k,m,n}(t) =$$

$$= \sum_{j,k=N+1}^{\infty} \sum_{m=-j+1}^{N-j} P_j' \; a(m)a(j-k+m)'P_k'$$

$$IV_N^{(2)} = N^{-1} \sum_t \sum_{j,k=N+1}^{\infty} \sum_{m=-j+1}^{N-j} \sum_{n=1}^{N} S_{j,k,m,n}(t).$$

Let $k = sN + k'$ ($s \geq -1$; $k' = 1,\ldots,N$). Hence

$$IV_N^{(2)} = \sum_{j=N+1}^{\infty} \sum_{s=1}^{\infty} \sum_{k'=1}^{N} P_j \left\{ \sum_{m=-j+1+k'}^{N-j} a(m)a(j-k'+m)' + \right.$$

$$\left. + \sum_{m=-j+1}^{-j+k'} a(m)a(j-k'+m+N)' \right\} P'_{sN+k'}.$$

$$IV_N^{(3)} = N^{-1} \sum_t \sum_{j,k=N+1}^{\infty} \sum_{m=1}^{N} \sum_{n=-k+1}^{N-k} S_{j,k,m,n}(t)$$

$$IV_N^{(4)} = N^{-1} \sum_t \sum_{j,k=N+1}^{\infty} \sum_{m=1}^{N} \sum_{n=1}^{N} S_{j,k,m,n}(t).$$

Let $j = rN+j'$, $k = sN+k'$ $(r,s \geq 1;\ j',k' = 1,\ldots,N)$. Thus

$$IV_N^{(4)} = \sum_{r,s=1}^{\infty} \sum_{j',k'=1}^{N} P_{rN+j'} \left\{ \sum_{m=\max(1,1+k'-j')}^{\min(N,N+k'-j')} a(m)a(j'-k'+m)' \right.$$

$$+ \sum_{m=N+1+k'-j'}^{N} a(m)a(j'-k'+m-N)' +$$

$$\left. + \sum_{m=1}^{k'-j'} a(m)a(j'-k'+m+N)' \right\} P_{sN+k'}$$

and we shall have results analogous to (7.45) and (7.46) holding for each of $IV_N^{(1)} - IV_N^{(4)}$ and hence for IV_N. This completes the proof of the lemma. □

Theorem 8 Put $\lambda_t = 2\pi t/N$ and

$$T_N(\xi) = \sum_{t=1}^{N} w_p(\lambda_t;\xi)w_q(\lambda_t;\xi)^* - \sum_{t=1}^{N} P(\lambda_t;\xi)w_a(\lambda_t)w_b(\lambda_t)^*Q(\lambda_t;\xi)^*$$

(i) If conditions (i) - (iii) of Lemma 9 hold, then

$$\lim_N \sup_\xi N^{-1} \| T_N(\xi) \| = 0 \text{ a.s.}$$

(ii) If, in addition, condition (iv) of Lemma 9 holds, then

$$\sup_\xi N^{-\frac{1}{2}} \| T_N(\xi) \| = o_p(1).$$

Proof of Theorem 8

$$T_N(\xi) = N^{-\frac{1}{2}} \sum_t \nu(\lambda_t)w_b(\lambda_t)^*Q(\lambda_t)^* +$$

$$+ N^{-\frac{1}{2}} \sum_t P(\lambda_t)w_a(\lambda_t)\zeta(\lambda_t)^* + N^{-1} \sum_t \nu(\lambda_t)\zeta(\lambda_t)^* \quad (7.50)$$

and it is obvious from Lemma 9 that (i) holds. Let $Q^{(M)}(\lambda;\xi)$ be the Cesaro sum to M terms of the Fourier series for $Q(\lambda;\xi)$. Then (Zygmund [1959, Theorem 3.4, p.89]), $\forall\ \varepsilon < 0$ there exists an $M_1 = M_1(\varepsilon)$ such that

$$\Lambda_M = \sup_{\xi\in\Xi}\ \sup_{\lambda\in[-\pi,\pi]} \|Q^{(M)}(\lambda;\xi)-Q(\lambda;\xi)\| < \varepsilon \quad \text{for} \quad M \geq M_1 \tag{7.51}$$

Let $S_{N,M} = \sup_\xi \|N^{-1} \sum_t \nu(\lambda_t)w_b(\lambda_t)*\{Q(\lambda_t)-Q^{(M)}(\lambda_t)\}\|$. Then

$$S_{N,M} \leq c\ \Lambda_M \sup_\xi N^{-1} \sum_t \|\nu(\lambda_t)\|^2 + c\ \Lambda_M N^{-1} \sum_{n=1}^{N} \|b(n)\|^2$$

so that

$$P(S_{N,M} > 2\varepsilon) \leq P(\sup_\xi N^{-1} \sum_t \|\nu(\lambda_t)\|^2 > \frac{\varepsilon}{c\Lambda_M}) + c\Lambda_M/\varepsilon$$

and by (7.51) and Lemma 9 the right side of this last inequality can be made arbitrarily small by increasing M. Hence, to show that $\forall\ \varepsilon > 0$

$$\lim\ \sup_N P(\sup_\xi \|N^{-1} \sum_t \nu(\lambda_t)w_b(\lambda_t)*Q(\lambda_t)*\| > \varepsilon) = 0 \tag{7.52}$$

it is sufficient to show that $\forall\ \varepsilon > 0$

$$\lim\ \sup_N P(\sup_\xi \|N^{-1} \sum_t \nu(\lambda_t)w_b(\lambda_t)*e^{iu\lambda_t}\| > \varepsilon) = 0 \tag{7.53}$$

for all finite integers u. To be specific we take $u \geq 0$, where u is fixed and $N >> u$. The analysis for $u \leq 0$ is very similar and is omitted.

$$N^{-1} \sum_t \nu(\lambda_t)w_b(\lambda_t)*e^{iu\lambda_t} = N^{-3/2} \sum_t \sum_{n=1}^{N} \left\{ \sum_{j=o}^{N} T_{j,n}(t) + \right.$$
$$\left. + \sum_{j=N+1}^{\infty} T_{j,n}(t) \right\} = I_N + II_N$$

with $T_{j,n}(t) = P_j R_{j,N}(\lambda_t) b(n)' \exp\{i(j+u-n)\lambda_t\}$

$$U_{j,m,n}(t) = P_j\ a(m)b(n)'\exp\{i(j+u+m-n)\lambda_t\}.$$

We can write I_N as $I_N^{(1)} - I_N^{(2)}$, where

$$I_N^{(1)} = N^{-3/2} \sum_t \sum_{n=1}^{N} \sum_{j=o}^{N} \sum_{m=-j+1}^{o} U_{j,m,n}(t)$$

$$= N^{-\frac{1}{2}} \sum_{j=o}^{N} P_j \left\{ \sum_{m=1-j}^{\min(o,N-j-u)} a(m)b(j+m+u)' + \sum_{m=N+1-j-u}^{o} a(m)b(j+m+u-N)' \right\}$$

and it follows that

$$\lim \sup_N P(\sup_\xi \| I_N^{(1)} \| > \varepsilon) = 0 \tag{7.54}$$

$$I_N^{(2)} = N^{-3/2} \sum_t \sum_{n=1}^{N} \sum_{j=o}^{N} \sum_{m=N-j+1}^{N} U_{j,m,n}(t) =$$

$$= N^{-\frac{1}{2}} \sum_{j=o}^{N} P_j \left\{ \sum_{m=N-j+1}^{\min(N,2N-j-u)} a(m)b(j+m+u-N)' + \right.$$

$$\left. + \sum_{m=2N-j-u+1}^{N} a(m)b(j+m+u-2N)' \right\}$$

and a result of the type (7.54) will hold for $I_N^{(2)}$ and hence for I_N. We can write II_N as $II_N^{(1)} - II_N^{(2)}$ with

$$II_N^{(1)} = N^{-3/2} \sum_t \sum_{n=1}^{N} \sum_{j=N+1}^{\infty} \sum_{m=-j+1}^{N-j} U_{j,m,n}(t) =$$

$$= N^{\frac{1}{2}} \sum_{j=N+1}^{\infty} P_j \left\{ \sum_{m=1-j}^{N-j-u} a(m)b(j-m+u)' + \sum_{m=N-j-u+1}^{N-j} a(m)b(j+m+u-N)' \right\}$$

$$II_N^{(2)} = N^{-3/2} \sum_t \sum_{n=1}^{N} \sum_{j=N+1}^{\infty} \sum_{m=1}^{N} U_{j,m,n}(t).$$

Let $j = rN+j'$ $(r \geq 1;\ j = 1,\dots,N)$. Hence

$$II_N^{(2)} = N^{-\frac{1}{2}} \sum_{r=1}^{\infty} \sum_{j'=1}^{N} P_{rN+j'} \left\{ \sum_{n=1}^{N-j-u} a(m)b(j'+u+m)' + \right.$$

$$+ \sum_{m=\max(1,N-u-j'+1)}^{\min(N,2N-u-j')} a(m)b(j'+m+u-N)' +$$

$$\left. + \sum_{m=2N-j'-u+1}^{N} a(m)b(j'+m+u-2N)' \right\}.$$

Hence a relation of the type (7.54) will hold for $II_N^{(1)}$ and $II_N^{(2)}$, and thus also for II_N. It follows that (7.53), and hence (7.52), hold. A result similar to (7.52) may be obtained for the second term on the right in (7.50) and, by Lemma 9, also for the third term on the right. This completes the proof of Theorem 9. □

Corollary 1 Let $\nu(\lambda)$ and $w_b(\lambda)$ be defined as above. If $R(\lambda)$ is a continuous periodic matrix function of λ of period 2π, and the following matrix-vector product is compatible, then

$$\sup_{\xi\in\Xi} \left\| N^{-1} \sum_{t=1}^{N} R(\lambda_t)\mathrm{vec}\{\nu(\lambda_t)w_b(\lambda_t)^*\}\right\| = o_p(1), \quad \lambda_t = 2\pi t/N.$$

Proof This follows directly from the proof of Theorem 8. □

Proof of Theorem 5 By using Theorem 8 we can deduce Theorem 5 from Theorems 1-3. □

Proof of Theorem 6 It can be readily verified that the assumptions of Theorem 6 imply that assumptions A1., A2.(i) or (i'), A3., A4., A5. and A6.(i) hold. Because assumptions A2.(ii) and A6.(ii) are only used to show that condition S6.(ii) holds, the proof of the theorem will be complete if we can show that assumptions B1. - B3. imply that S6.(ii) holds. We now do so.

For a fixed θ, suppose that $h(\theta) = 0$ and $\Omega(\theta)=\Omega(\theta_o)$. From the proof of Lemma 4 we will have that $\Psi(z;\theta) = I_{(p)}$ $\forall z$ and

$$\int_{-\pi}^{\pi} \Phi(\lambda;\theta)dF_x(\lambda)\Phi(\lambda;\theta)^* = 0 \qquad (7.60)$$

But $\Psi(z;\theta) = I_{(p)}$ means that $B(z;\theta) = B(z;\theta_o)$ and hence $\Phi(z;\theta) = C(z;\theta_o) - C(z;\theta)$. Therefore $\Phi(z;\theta)$ is a polynomial of degree s at most. It now follows from (7.60) and assumption B1. that $\Phi(z;\theta) = 0$ $\forall z$, i.e. $C(z;\theta) = C(z;\theta_o)$.

Thus, we have shown that if $\Omega(\theta) = \Omega(\theta_o)$, then $[B(z;\theta), C(z;\theta)] = [B(z;\theta_o), C(z;\theta_o)]$. If this last relation holds, then

$$[\underline{B}(z;\theta), \underline{C}(z;\theta)] = U(z;\theta)[\underline{B}(z;\theta_o), \underline{C}(z;\theta_o)]$$

where $U(z;\theta)$ is a rational function of z. By Lemma 2.3 and B2.(iii) $U(z;\theta)$ must be a polynomial matrix in z. It now follows from B2.(iv) and B3. that $U(z;\theta) = I_{(p)}$, i.e. $\theta = \theta_o$. □

Proof of Theorem 7 We omit the proof as it is similar to, but easier than, the proof of Theorem 6. □

COMMENT

Usually $v(n;\theta)$ as defined in (3.10) involves the infinite past (an exception to this is when assumption A1' of §3 holds) so that $\ell_N(\theta)$, given θ, is not computable. Using the lemma below, it is straightforward to show, however, that Theorems 1 to 3 still hold if $v(n;\theta)$ of (3.10) is replaced by

$$\sum_{j=o}^{n-1} G_j(\theta)y(n-j) - \sum_{j=o}^{n-1} F_j(\theta)x(n-j).$$

Lemma Suppose that: (a) Ξ is a subset of Euclidean space with typical element ξ. (b) $\{a(n), n = 0, \pm 1,\dots\}$ is a possibly stochastic vector sequence satisfying

$$\lim_N N^{-1} \| a(-N) \|^2 = 0 \quad \text{a.s.}$$

and

$$\lim_N N^{-1} \sum_{n=1}^{N} a(n)a(n+k)' = \Gamma(k) \quad \text{a.s.}, \ k = 0, \pm 1,\dots$$

where $\Gamma(k)$ is a finite nonstochastic matrix $\forall$ k.

(c) $\{P_j(\xi),\ j \geq 0\}$ is a matrix sequence with the P_j matrix functions of ξ such that

$$\sup_{\xi \in \Xi} \|P_j(\xi)\| \leq m_j \quad \text{for} \quad j \geq 0,$$

where $\{m_j,\ j \geq 0\}$ is a sequence of scalars satisfying $\Sigma_{j=o}^{\infty} jm_j < \infty$.

Let

$$p(n;\xi) = \sum_{j=o}^{\infty} P_j(\xi)a(n-j), \quad q(n;\xi) = \sum_{j=o}^{n-1} P_j(\xi)a(n-j).$$

Then, uniformly in ξ, $\xi_+ \in \Xi$, we have:

$$\text{(i)} \quad N^{-1}\sum_{n=1}^{N} p(n;\xi)p(n+\ell;\xi_+)' - N^{-1}\sum_{n=1}^{N} q(n;\xi)q(n+\ell;\xi_+)' \xrightarrow{a.s.} 0 \quad \text{as} \quad N \to \infty.$$

If, furthermore,

$$\sup_{\xi \in \Xi} \sum_{n=o}^{\infty} \left\{ \sum_{k=o}^{\infty} k^{\frac{1}{2}}\|P_{k+n}(\xi)\| \right\}^2 < \infty ,$$

then

$$\text{(ii)} \quad \lim_N N^{-\frac{1}{2}} \sum_{n=1}^{N} q(n;\xi)\{p(n;\xi_+) - q(n;\xi_+)\}' \overset{a.s.}{=} 0$$

Proof (i) This follows immediately from the proof of Lemma 1(c).

$$\text{(ii)} \quad N^{-\frac{1}{2}}\sum_{n=1}^{N} q(n;\xi)\{p(n;\xi_+) - q(n;\xi_+)\}' =$$

$$= N^{-\frac{1}{2}}\sum_{n=1}^{N} q(n;\xi)\left\{\sum_{j=o}^{\infty} P_{j+n}(\xi_+)a(-j)\right\}' +$$

$$+ N^{-\frac{1}{2}}\sum_{n=M+1}^{\infty} q(n;\xi)\left\{\sum_{j=o}^{\infty} P_{j+n}(\xi_+)a(-j)\right\}'.$$

For M a fixed positive integer, the first term on the right obviously tends to zero. We look at the second term.

$$\| N^{-\frac{1}{2}} \sum_{n=M+1}^{N} q(n;\xi) \left\{ \sum_{j=o}^{\infty} P_{j+n}(\xi_+) a(-j) \right\} \|^2 \leq$$

$$\leq \left\{ N^{-1} \sum_{n=M+1}^{N} \| q(n;\xi) \|^2 \right\} \times \sum_{n=M+1}^{N} \left\{ \sum_{j=o}^{\infty} j^{\frac{1}{2}} \| P_{j+n}(\xi_+) \| \times \right.$$

$$\left. \times\ j^{-\frac{1}{2}} \| a(-j) \| \right\}$$

$$\xrightarrow{a.s.} 0 \quad \text{as} \quad M,N \to \infty$$

by assumption. Thus we obtain (ii). □

CHAPTER 5

DISCUSSION OF THE GAUSS-NEWTON ALGORITHM

1. INTRODUCTION

This chapter has two aims. The first is to briefly and informally discuss the properties of the G.N. algorithm from a computational, rather than a statistical, point of view. The second aim is to relate the G.N. algorithm to some of the estimators proposed in the literature.

2. DERIVATION OF THE GAUSS-NEWTON ITERATES

Suppose that we minimize the criterion function

$$\ell(\theta) = (2N)^{-1} \sum_{n=1}^{N} \mathrm{tr}\{\Phi(n)v(n;\theta)v(n;\theta)'\}$$

over some permissible parameter space, Θ say, to obtain an estimate of an unknown parameter value θ_o. The $v(n;\theta)$, $1 \leq n \leq N$, are twice differentiable $p\times 1$ vector functions of θ with continuous second derivatives; the $\Phi(n)$, $1 \leq n \leq N$, are $p\times p$ symmetric p.d. matrices; θ is an $m\times 1$ vector and we assume that θ_o is an interior point of Θ.

To obtain the minimizing value of θ, $\tilde{\theta}$ say, we would normally try to solve the first order equations

$$\partial\ell(\theta)/\partial\theta = 0 \tag{2.11}$$

Put $k(\theta) = \partial\ell(\theta)/\partial\theta$. If $k(\theta)$ is highly nonlinear in θ, then a closed form solution of (2.11) cannot usually be obtained; instead, we often use some iterative procedure to obtain $\tilde{\theta}$. A common iterative algorithm to determine $\tilde{\theta}$ is the Newton-Raphson algorithm (e.g. Kowalik and Osborne [1968, p.65]) which is described by

$$\theta^{(j+1)}-\theta^{(j)} = -[\partial^2\ell(\theta)/\partial\theta\partial\theta']^{-1}k(\theta)\Big|_{\theta=\theta^{(j)}},\ j \geq 0 \qquad (2.12)$$

with $\theta^{(0)}$ initializing the iteration. Let θ_i, $1 \leq i \leq m$, be the i^{th} element of θ. Then

$$\partial^2\ell(\theta)/\partial\theta_i\partial\theta_k = N^{-1}\sum_{n=1}^{N} tr[\Phi(n)\{\partial v(n;\theta)/\partial\theta_i\}\{\partial v(n;\theta)'/\partial\theta_k\}]+$$

$$+ N^{-1}\sum_{n=1}^{N} tr[\Phi(n)v(n;\theta)\{\partial^2 v(n;\theta)'/\partial\theta_i\partial\theta_k\}] \qquad (2.13)$$

If we modify (2.12) by omitting the second summation on the right in (2.13), then we obtain the G.N. iteration

$$\theta^{(j+1)} - \theta^{(j)} = -\Delta(\theta)^{-1}k(\theta)\Big|_{\theta=\theta^{(j)}},\ j \geq 0 \qquad (2.14)$$

where $\Delta(\theta) = N^{-1}\Sigma_n V(n;\theta)'\Phi(n)V(n;\theta)$ and $V(n;\theta)=\partial v(n;\theta)/\partial\theta'$.

An instructive alternative way of obtaining (2.14) is the following. Given $\theta^{(0)}$ to $\theta^{(j)}$ we obtain $\theta^{(j+1)}$ by minimizing

$$(2N)^{-1}\sum_{n=1}^{N} tr\{\Phi(n)v(n;\theta,\theta^{(j)})v(n;\theta,\theta^{(j)})'\} \qquad (2.15)$$

where $v(n;\theta,\theta^{(j)})$ is the linearization of $v(n;\theta)$ about $\theta^{(j)}$; i.e. $v(n;\theta,\theta^{(j)}) = v(n;\theta^{(j)}) + V(n;\theta^{(j)})(\theta-\theta^{(j)})$. It is easy to check that the minimization of (2.15) yields (2.14). This second approach shows that in obtaining the G.N. iterates (2.14) we perform a sequence of linear regressions, regressing $v(n;\theta^{(j)})$ on $\{-V(n;\theta^{(j)})\}$, with weighting matrix $\Phi(n)$, to obtain the corrections $(\theta^{(j+1)} - \theta^{(j)})$. This last observation means that if $\Phi(n)$ is constant for $n=1$ to N, as is usually the case, then we can use ordinary linear regression programs to update the iterates once we have written subroutines to compute $v(n;\theta)$ and $V(n;\theta)$.

We now discuss some useful simple extensions of the basic G.N. algorithm (2.14).

1. We can permit the $\Phi(n)$ to be functions of θ. We would then use $\Phi(n;\theta^{(j)})$ in (2.14).

2. Suppose we want to minimize $\ell(\theta)$ subject to the continuously differentiable constraints $h(\theta) = 0$. Constrained minimization of $\ell(\theta)$ gives the first order conditions

$$\partial\ell(\theta)/\partial\theta + H'(\theta)\lambda = 0,\ h(\theta) = 0 \tag{2.16}$$

where $H(\theta) = \partial h(\theta)/\partial\theta'$ and λ is a vector of Lagrange multipliers. Hence a simple constrained version of (2.14) is

$$\begin{bmatrix} \theta^{(j+1)} \\ \cdots\cdots \\ \lambda^{(j+1)} \end{bmatrix} = \begin{bmatrix} \theta^{(j)} \\ \cdots\cdots \\ 0 \end{bmatrix} - \begin{bmatrix} \Delta(\theta) & \vdots & H'(\theta) \\ \cdots\cdots & \vdots & \cdots \\ H(\theta) & \vdots & 0 \end{bmatrix}^{-1} \begin{bmatrix} k(\theta) \\ \cdots \\ h(\theta) \end{bmatrix}_{\Big|\theta=\theta^{(j)}}, \quad j \geq 0 \tag{2.17}$$

We have already met the above algorithm in 3.§3.

3. The criterion function (2.10) and the algorithms (2.14) and (2.17) are mainly suitable for time domain estimation. We can, however, obtain direct analogues to deal with frequency domain estimation. See for example 4.§4 and 7.§6.

3. FURTHER PROPERTIES AND MODIFICATIONS

Δ is obviously p.s.d. and hence it is p.d. if it is non-singular. We can take advantage of this property in an unconstrained application of the G.N. algorithm, i.e. (2.14), by using the efficient Cholesky decomposition to invert Δ. See Wilkinson [1971, p.2].

We now assume that if there are any constraints on θ, then they are linear, with H of full row rank. We call $\{\theta^{(j)} - \theta^{(j-1)}\}$ the j^{th} step of the G.N. iteration (2.17).

For convenience we let $\bar{\theta} = \theta^{(j-1)}$, $\hat{\theta} = \theta^{(j)}$ and put $\theta_t = \bar{\theta} + t(\hat{\theta} - \bar{\theta})$, $t > 0$. Given that $\bar{\theta}$ and $\hat{\theta}$ satisfy the linear restrictions, we will say that the direction of the j^{th} step of the iteration is downhill if for t sufficiently small and positive $\ell(\hat{\theta}_t) \leq \ell(\bar{\theta})$, with equality if and only if $\hat{\theta}$ satisfies the first order conditions (2.16). The following lemma is a generalization of (I) in Kowalik and Osborne [1968, p.72].

Lemma 1 Suppose that the restrictions $h(\theta) = 0$ are linear in θ and the matrix

$$\begin{pmatrix} \Delta(\theta) & \vdots & H' \\ \cdots & & \cdots \\ H & \vdots & 0 \end{pmatrix}$$

is invertible. If $h(\theta^{(0)}) = 0$, then $h(\theta^{(j)}) = 0$ for $j \geq 1$ and (2.17) has its j^{th} step downhill for $j \geq 1$; furthermore, if $\theta^{(j)}$ satisfies (2.16), then $\theta^{(j+1)} = \theta^{(j)}$.

Proof It is obvious from (2.17) that if $h(\theta)$ is linear in θ and $h(\theta^{(0)}) = 0$, then $h(\theta^{(j)}) = 0$ for $j \geq 1$. Put $E = I_{(m)} - H'(HH')^{-1}H$. Because rank $(E\Delta E)$ = rank (E) by Theorem 2.13(iii), it follows from the proof of Lemma 3.5 that

$$\hat{\theta} - \bar{\theta} = -(E\Delta(\theta)E)^{+}Ek(\theta)\big|_{\theta=\bar{\theta}} \tag{3.10}$$

so that

$$\begin{aligned} \ell(\theta_t) - \ell(\bar{\theta}) &= (\theta_t - \bar{\theta})'k(\bar{\theta}) + O(t^2) \\ &= -t\{Ek(\bar{\theta})\}'\{E\Delta(\bar{\theta})E\}^{+}\{Ek(\bar{\theta})\} + O(t^2). \end{aligned}$$

Hence $\ell(\theta_t) < \ell(\bar{\theta})$ for t positive and sufficiently small unless $Ek(\bar{\theta}) = 0$. $Ek(\bar{\theta}) = 0$ means that there exists a λ such that $k(\bar{\theta}) + H'\lambda = 0$ so that $\bar{\theta}$ satisfies (2.16) because $h(\bar{\theta}) = 0$. Furthermore if $Ek(\bar{\theta}) = 0$, then it follows from (3.10) that $\hat{\theta} = \bar{\theta}$. □

If the restrictions are linear in θ and the iterations (2.14) or (2.17) converge very slowly or not at all, then the result of the above lemma leads naturally to the modified Gauss-Newton algorithm described below. For convenience we discuss the unconstrained case. Following Kowalik and Osborne [1968, p.66] we compute the modified G.N. iterates as follows. For $i \geq 1$:

(i) compute $\delta^{(i)} = -\Delta(\theta^{(i)})^{-1}k(\theta^{(i)})$;

(ii) define $\nu^{(i)} = \delta^{(i)}/\|\delta^{(i)}\|$;

(iii) let ρ^* minimize $\ell(\theta^{(i)} + \rho\nu^{(i)})$;

(iv) let $\theta^{(i+1)} = \theta^{(i)} + \rho^*\nu^{(i)}$.

This modification, which retains the G.N. direction but modifies the step length, usually minimizes total computation time and ensures convergence. Its success is mainly due to the fact that the G.N. direction is downhill so that each step of the modified iteration decreases $\ell(\theta)$. For a further discussion on the choice of steplength in the G.N. algorithm see Bard [1974, pp.110-116], and Kowalik and Osborne [1968, p.71].

If the above modification of the G.N. algorithm is used, then

$$T_j = [\ell(\theta^{(j-1)})-\ell(\theta^{(j)})]/\{1+\ell(\theta^{(j-1)})\}$$

could be used as a termination criterion. If $T_j \leq$ some small prescribed positive value on two or more occasions, then the iteration is terminated. See also Box, Davies and Swann [1969, p.8] and Bard [1974, p.114].

A different modification of the G.N. algorithm, which is also designed to increase computational efficiency and enhance convergence, is due to Marquardt, Levenberg and Morrison (see Bard [1974, p.94]) and is described by

$$\theta^{(j+1)} - \theta^{(j)} = -\{\Delta(\theta)+\rho_j I_{(m)}\}^{-1} k(\theta)\Big|_{\theta=\theta^{(j)}}, \quad j \geq 0$$

with $\rho_j > 0$. See also Kowalik and Osborne [1968, p.74].

4. RELATING THE G.N. ESTIMATOR TO ESTIMATORS PROPOSED IN THE LITERATURE

In this section and the following two we relate the G.N. estimator to some of the estimators proposed in the literature. Before doing so however we obtain some elementary but useful perturbation results.

Lemma 2 Suppose that A, $\hat{A}$, B, $\hat{B}$, C, $\hat{C}$, D and $\hat{D}$ are matrices of dimension $p\times p$, $p\times p$, $p\times q$, $p\times q$, $p\times r$, $p\times r$, $p\times p$ and $p\times p$ respectively, $\hat{A}$, $\hat{B}$, $\hat{C}$ and $\hat{D}$ may be regarded as perturbed values of A, B, C and D. If A, $\hat{A}$, D and $\hat{D}$ are nonsingular, then omitting terms of second order or higher in $(A-\hat{A})$ etc. we have

(i) $\quad A^{-1}B=\hat{A}^{-1}\hat{B}+\hat{A}^{-1}(B-\hat{B})-\hat{A}^{-1}(A-\hat{A})\hat{A}^{-1}\hat{B}.$

(ii) $\quad AB=\hat{A}\hat{B}+\hat{A}(B-\hat{B})+(A-\hat{A})\hat{B}=\hat{A}B+(A-\hat{A})\hat{B}.$

(iii) $\quad A^{-1}BC=\hat{A}^{-1}\hat{B}\hat{C}+\hat{A}^{-1}\hat{B}(C-\hat{C})+\hat{A}^{-1}(B-\hat{B})\hat{C}-\hat{A}^{-1}(A-\hat{A})\hat{A}^{-1}\hat{B}\hat{C}$

(iv) $\quad A^{-1}D^{-1}B=\hat{A}^{-1}\hat{D}^{-1}\hat{B}+\hat{A}^{-1}\hat{D}^{-1}(B-\hat{B})-\hat{A}^{-1}\hat{D}^{-1}(D-\hat{D})\hat{D}^{-1}\hat{B}-$
$$-\hat{A}^{-1}(A-\hat{A})\hat{A}^{-1}\hat{D}^{-1}\hat{B}.$$

The proof is straightforward and is omitted. □

The first model we look at is the simultaneous equations model with AR residuals

$$\sum_{j=o}^{r} B_j y(n-j) = \sum_{j=o}^{s} C_j x(n-j)+u(n), \quad \sum_{j=o}^{t} A_j u(n-j)=e(n); \; A_o = I_{(p)}. \tag{4.10}$$

We have already discussed this model in 4.§5. Let

$$B(z) = \sum_{j=o}^{r} B_j z^j, \; C(z) = \sum_{j=o}^{s} C_j z^j \quad \text{and} \quad A(z) = \sum_{j=o}^{t} A_j z^j$$

and put $\theta = \text{vec}(B_o,\ldots,B_r,C_o,\ldots,C_s,A_o,\ldots,A_t)$ and $v(n;\theta) = B_o^{-1}A(L)\{B(L)y(n) - C(L)x(n)\}$, with L the lag operator.

It is clear from Chapter 4 why $v(n;\theta)$ has been defined in this way. We will assume below that the G.N. iterates are constructed as described in Chapter 4.

Using Lemma 2(iii), the linearization of $v(n;\theta)$ about $\hat{\theta}$ is given by

$$v(n;\theta,\hat{\theta}) = p(n) + \hat{B}_o^{-1}\hat{A}(L)[\{B(L)-\hat{B}(L)\}y(n)-\{C(L)-\hat{C}(L)\}x(n)] + $$
$$+ \hat{B}_o^{-1}\{A(L)-\hat{A}(L)\}q(n) - \hat{B}_o^{-1}(B_o-\hat{B}_o)p(n) \tag{4.11}$$

where $p(n) = v(n;\hat{\theta}), q(n) = \hat{B}(L)y(n)-\hat{C}(L)x(n)$; we write $A(L)$ evaluated at $\hat{\theta}$ as $\hat{A}(L)$, with $\hat{B}(L)$ and $\hat{C}(L)$ defined similarly.

If (4.10) is a reduced form system, i.e. $B_o \equiv I_{(p)}$, then we redefine θ by dropping B_o. (4.11) reduces to

$$v(n;\theta,\hat{\theta}) = p(n) + \hat{A}(L)[\{B(L)-\hat{B}(L)\}y(n) - \{C(L)-\hat{C}(L)\}x(n)] + $$
$$+ \{A(L) - \hat{A}(L)\}q(n) \tag{4.12}$$

We now note the following.

(i) We can rewrite (4.12) as

$$v(n;\theta,\hat{\theta}) = \hat{A}(L)\{B(L)y(n) - C(L)x(n)\} + \{A(L)-\hat{A}(L)\}q(n)$$

and it is clear from this last equation that Hatanaka's [1974] two-step estimator is just the first iterate of a G.N. iteration scheme. This has already been observed by Hendry [1976, p.83] and Byron and Pagan [1975] but the derivations in the references just cited seem to be more cumbersome than our derivation.

(ii) Put $\hat{y}(n) = y(n) - p(n)$. We can rewrite (4.11) as

$$v(n;\theta,\hat{\theta}) = \hat{B}_o^{-1}[\{A(L)-\hat{A}(L)\}q(n) + \hat{A}(L)q(n)-\hat{A}(L)\{C(L)-\hat{C}(L)\}x(n)+$$
$$+\{\hat{A}(L) - I_{(p)}\}\{B(L)-\hat{B}(L)\}y(n) + \{B(L)-B_o-(\hat{B}(L)-\hat{B}_o)\}y(n) +$$
$$+ (B_o-\hat{B}_o)\hat{y}(n)] \tag{4.13}$$

Identifying $\hat{y}(n)$, $q(n)$ respectively with $\hat{\tilde{Y}}^{(2)}$ and $\tilde{u}$ of Hatanaka [1976, pp.194-196] it is not difficult to see from (4.13) that Hatanaka's [1976, p.196] estimator (c) is exactly the first iterate of the G.N. algorithm.

(iii) Hatanaka's estimators (a) and (b) are also closely related to the G.N. algorithm and it is easy to prove that they are asymptotically efficient once we have proved that the G.N. estimator is asymptotically efficient. We first discuss the estimator (b) (Hatanaka [1976, p.195]), using Hatanaka's notation for convenience.

It is not difficult to verify that

$$y - (\tilde{R}'\otimes I_T)\mathcal{L}y = \tilde{\Psi}\tilde{\delta} + \tilde{u} - (\tilde{R}'\otimes I_T)\mathcal{L}\tilde{u}. \tag{4.14}$$

Therefore, we can rewrite the expression for the estimator (b) as

$$(\hat{\theta}^{(2)} - \tilde{\theta}) = \Delta(\hat{\Psi}^{(2)}, \tilde{\Psi})^{-1}k(\hat{\Psi}^{(2)}) \tag{4.15}$$

where $\Delta(\Psi,\Phi) = [(\Psi, I_m\otimes\mathcal{L}\tilde{U})'(\tilde{\ddagger}\otimes I_T)^{-1}(\Phi,I_m\otimes\mathcal{L}\tilde{U})]$,

$k(\Psi) = (\Psi, I_m\otimes \mathcal{L}\tilde{U})'(\tilde{\ddagger}\otimes I_T)^{-1}\{\tilde{u} - (\tilde{R}'\otimes I_T) \mathcal{L}\tilde{u}\}$,

$\tilde{\theta} = \mathrm{vec}(\tilde{\delta},\tilde{R})$, $\hat{\theta}^{(2)} = \mathrm{vec}(\hat{\delta}^{(2)},\hat{R}^{(2)})$.

Because the expression for the G.N. algorithm is (Hatanaka, p.196)

$$(\hat{\theta}^{(3)} - \tilde{\theta}) = \Delta(\hat{\Psi}^{(2)}, \hat{\Psi}^{(2)})^{-1}k(\hat{\Psi}^{(2)})$$

it follows that $\hat{\theta}^{(2)} - \hat{\theta}^{(3)} = o_p(N^{-\frac{1}{2}})$. Hence Hatanaka's estimator (b) is asymptotically efficient.

Because $\Delta(\hat{\Psi}^{(2)},\tilde{\Psi})$ is not p.d.,it seems to me that (4.15) is computationally inferior to the G.N. algorithm for the following reasons:(1) We now require a general matrix inversion routine to invert $\Delta(\hat{\Psi}^{(2)}, \tilde{\Psi})$; (2) the direction of a step of (4.15) is no longer downhill.

We also note that if we iterate (4.15) till convergence, then we will obtain a solution of the first order equations (2.11).

(iv) Next we look at Hatanaka's estimator (a) (p.194). We again use Hatanaka's notation. It is not difficult to check that

$$\Delta(\hat{\Psi}^{(1)},\tilde{\Psi}) = \Delta(\hat{\Psi}^{(1)},\hat{\Psi}^{(1)}).$$

Using (4.14) we can now rewrite Hatanaka's estimator (a) as

$$\hat{\theta}^{(1)} - \tilde{\theta} = \Delta(\hat{\Psi}^{(1)},\hat{\Psi}^{(1)})^{-1}k(\hat{\Psi}^{(1)})$$

and again we see the close relationship between Hatanaka's estimator (a) and the G.N. estimator.

We note that the estimator proposed in Dhrymes and Taylor [1976, eqn(17), p.372] is exactly Hatanaka's estimator (a).

5. CONTINUATION - THE ARMA MODEL

The second model we consider is the ARMA model. It will be convenient to consider the scalar case and to depart, for this section only, from our usual notation. We consider

$$\sum_{j=o}^{q} \beta_j x(n-j) = \sum_{j=o}^{p} \alpha_j e(n-j) \qquad (n=1,\ldots,N) \qquad (5.10)$$

where $\alpha_o = \beta_o = 1$. The observed scalar sequence $\{x(n), n \geq 1\}$ forms a stationary process, while the $e(n)$ are independent and identically distributed with mean zero and variance σ^2. We put

$$a(z) = \sum_{j=o}^{p} \alpha_j z^j, \quad b(z) = \sum_{j=o}^{q} \beta_j z^j$$

$\alpha' = (\alpha_1,\ldots,\alpha_p)$, $\beta' = (\beta_1,\ldots,\beta_q)$, $\theta' = (\alpha',\beta')$, $\chi_\alpha'(\lambda) =$
$= (e^{i\lambda},\ldots,e^{ip\lambda})$, $\chi_\beta'(\lambda) = (e^{i\lambda},\ldots,e^{iq\lambda})$

$$w(\lambda) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} x(n)e^{in\lambda}, \quad I(\lambda) = |w(\lambda)|^2$$

We assume that $a(z)$ and $b(z)$ have all their zeros outside the unit circle and denote $a(e^{i\lambda})$ and $b(e^{i\lambda})$ by $a(\lambda)$ and $b(\lambda)$ respectively. Let

$$\ell(\theta) = (2N)^{-1} \sum_{n=1}^{N} |v(n;\theta)|^2$$

where $v(n;\theta) = a^{-1}(\lambda_n)b(\lambda_n)w(\lambda_n)$, with $\lambda_n = 2\pi n/N$ $(n=1,\ldots,N)$. For a motivation of the criterion function $\ell(\theta)$ see 4.§4.

We will now show that: (i) the estimator proposed by Akaike [1973] for the model (5.10) is exactly the first step of a G.N. iteration applied to $\ell(\theta)$; (ii) in general, the estimator proposed by Hannan [1969b] for (5.10) is not the same as that of Akaike but differs from it by a term of order $o_p(N^{-\frac{1}{2}})$. For the pure moving average case, however, Hannan's estimator will be the same as Akaike's if the initial estimate of α is obtained by factorizing the estimated spectrum of the $x(n)$ process.

Proof An application of the G.N. algorithm to $\ell(\theta)$ yields as a first step,

$$\tilde{\theta} - \hat{\theta} = -\left[\left\{N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial \theta} \frac{\partial \bar{v}(n;\theta)}{\partial \theta'}\right\}^{-1} \frac{\partial \ell(\theta)}{\partial \theta}\right]_{\theta=\hat{\theta}} =$$

$$= -\begin{pmatrix} U & . & -W' \\ & \cdots & \\ -W & . & V \end{pmatrix}^{-1} \times \begin{pmatrix} -c - U\alpha \\ \cdots\cdot \\ d + V\beta \end{pmatrix}_{\theta=\hat{\theta}} \tag{5.11}$$

where

$$U(\theta) = N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial \alpha} \frac{\partial \bar{v}(n;\theta)}{\partial \alpha'} = N^{-1} \sum_{n=1}^{N} \chi_\alpha \chi_\alpha^* b\bar{b}(a\bar{a})^{-2} I \tag{5.12}$$

$$W'(\theta) = -N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial \alpha} \frac{\partial \bar{v}(n;\theta)}{\partial \beta'} = N^{-1} \sum_{n=1}^{N} \chi_\alpha \chi_\beta^* b\bar{a}(a\bar{a})^{-2} I \tag{5.13}$$

$$V(\theta) = N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\theta)}{\partial \beta} \frac{\partial \bar{v}(n;\theta)}{\partial \beta'} = N^{-1} \sum_{n=1}^{N} \chi_\beta \chi_\beta^* (a\bar{a})^{-1} I \tag{5.14}$$

$$c(\theta) = N^{-1} \sum_{n=1}^{N} \chi_\alpha b\bar{b}(a\bar{a})^{-2} I, \quad d(\theta) = N^{-1} \sum_{n=1}^{N} \chi_\beta (a\bar{a})^{-1} I \tag{5.15}$$

with $\hat{\theta}$ an initial consistent estimate of θ_o, the true value of θ. In (5.12) to (5.15) χ_α stands for $\chi_\alpha(\lambda_n)$ with a similar explanation for χ_β, a, b and I. Comparing (5.11) to (5.15) with (3.9) to (3.13) and (4.1) of Akaike [1973] it is easily seen that Akaike's estimator corresponds exactly to the estimator just obtained. Hence (i) is proved.

Let $\beta^{(i)} = -\{V^{-1}d\}_{\theta=\hat{\theta}}$, $\alpha^{(i)}-\hat{\alpha}=[U^{-1}\{W'(\beta^{(i)}-\hat{\beta})+c+U\hat{\alpha}]_{\theta=\hat{\theta}}$,

$P = (U-W'V^{-1}W)^{-1}$ and $R = V^{-1}WP$. It follows that

$$\tilde{\alpha}-\hat{\alpha}=\{PU(\alpha^{(i)}-\hat{\alpha})\}_{\theta=\hat{\theta}},\ \tilde{\beta}-\hat{\beta}=\{RU(\alpha^{(i)}-\hat{\alpha})+\beta^{(i)}-\hat{\beta}\}_{\theta=\hat{\theta}} \tag{5.16}$$

To prove (ii) we restrict outselves to the case where Hannan's estimator is obtained by factorizing the spectrum $\hat{f}_y(\lambda)$ (Hannan [1969b, p.585]) to yield an initial estimate of α, $\hat{\alpha}$ say. The case where $\hat{f}_y(\lambda)$ is not factored follows similarly. The derivation of Hannan's estimator may be described as follows. Starting with initial consistent estimates $\hat{\alpha}$ and $\hat{\beta}$ satisfying

$$\hat{\theta} - \theta = O_p(N^{-\frac{1}{2}}),\ \text{let}\ \beta_H^{(i)} = -V^{-1}(\hat{\theta})d(\hat{\theta})\ \text{and}$$

$$\alpha_H^{(i)} = -U^{-1}(\hat{\alpha},\beta_H^{(i)})c(\hat{\alpha},\beta_H^{(i)}).$$

It follows from above that $\beta_H^{(i)} = \beta^{(i)}$ and

$$\alpha_H^{(i)}-\hat{\alpha} = \left(U^{-1}\frac{\partial \ell}{\partial \alpha}\right)_{\theta=\hat{\alpha},\beta^{(i)}} = \left[U^{-1}\left\{\frac{\partial \ell}{\partial \alpha} + \frac{\partial^2 \ell}{\partial\alpha\partial\beta'}(\beta^{(i)}-\hat{\beta})\right\}\right]_{\theta=\hat{\theta}} + O_p(N^{-1})$$

$$= [U^{-1}\{-c-U\hat{\alpha}-W'(\beta^{(i)}-\hat{\beta})\}]_{\theta=\hat{\theta}} + o_p(N^{-\frac{1}{2}})$$

because

$$\partial^2\ell(\hat{\theta})/\partial\alpha\partial\beta' = -W'(\hat{\theta}) + o_p(1) \tag{5.17}$$

It follows that $\alpha_H^{(i)} - \hat{\alpha} = -(\alpha^{(i)}-\hat{\alpha}) + o_p(N^{-\frac{1}{2}})$. Next we put

$$\alpha_H = \hat{\alpha} + P(\hat{\theta})U(\hat{\theta})(\hat{\alpha}-\alpha_H^{(i)}),\ \beta_H = -V^{-1}(\alpha_H)d(\alpha_H).$$

By (5.16), $\alpha_H = \tilde{\alpha} + o_p(N^{-\frac{1}{2}})$,

$$\beta_H - \beta^{(i)} = -\{V^{-1}\ \frac{\partial \ell}{\partial \beta}\}_{\theta=\alpha_H,\beta^{(i)}}$$

$$= -V^{-1}(\hat{\alpha})\{\frac{\partial \ell}{\partial \beta} + \frac{\partial^2 \ell}{\partial \beta \partial \alpha^\tau}(\alpha_H - \hat{\alpha})\}_{\theta=\hat{\alpha},\beta^{(i)}} + 0_p(N^{-1})$$

$$= V^{-1}(\hat{\alpha})W(\hat{\theta})(\alpha_H - \hat{\alpha}) + o_p(N^{-\frac{1}{2}}),$$

because of (5.17) and the fact that $\partial \ell(\hat{\alpha}, \beta^{(i)})/\partial \beta = 0$. Hence by (5.16)

$$\beta_H - \beta^{(i)} = V^{-1}(\hat{\alpha})W(\hat{\theta})(\tilde{\alpha}-\hat{\alpha}) + o_p(N^{-\frac{1}{2}}) = R(\hat{\theta})U(\hat{\theta})(\alpha^{(i)}-\hat{\alpha}) +$$

$+ o_p(N^{-\frac{1}{2}})$ and it follows that $\beta_H = \tilde{\beta} + o_p(N^{-\frac{1}{2}})$, again by (5.16). The proof of the first part of (ii) is completed by identifying $\hat{\alpha}$, $\hat{\beta}$, $\alpha_H^{(i)}$, $\beta_H^{(i)}$, α_H and β_H with $\hat{\alpha}$, $\hat{\beta}$, $\hat{\alpha}^{(2)}$, $\hat{\beta}^{(1)}$, $\tilde{\alpha}^{(1)}$ and $\tilde{\beta}^{(1)}$ respectively of Hannan [1969b, p.585]. For the pure moving average case, i.e. $b(z) \equiv 1$, $\tilde{\alpha} = 2\hat{\alpha} + U^{-1}(\hat{\alpha})c(\hat{\alpha})$, since $\tilde{\alpha} = \alpha^{(i)}$ by (5.16). The proof of (ii) is completed by identifying $\hat{\alpha}$, $-U^{-1}(\hat{\alpha})c(\hat{\alpha})$ and $\tilde{\alpha}$ with $\hat{\alpha}$, $\hat{\alpha}^{(1)}$ and $\tilde{\alpha}$ respectively of Hannan [1969b, p.582].

Remark 5.1 (i) Because the Akaike and Hannan estimators are not exactly the same it should be verified that $\tilde{\delta}$ is a real vector. To prove that it is we first note that if $z = e^{i2\pi/N}$, then $z^N = 1$ and $(z^m)^{N-n}$ is the complex conjugate of $(z^m)^n$ for m and n integers. Hence $(z^m)^{N-n}+(z^m)^n$ is real for $1 \leq n \leq [\frac{1}{2}N]$ and it is now easy to verify that the right sides of (5.12) to (5.15) are real. The required result follows.

(ii) It is clear that the matrix inverted in (5.11) is p.s.d., a result obtained more indirectly by Akaike [1973, p.261]. This remark extends immediately to the vector case dealt with by Akaike.

(iii) We now amplify on a comment made at the bottom of p.262 of Akaike [1973]. Using the notation of Hannan [1969b, eqn (2) and p.591] the Normal Equations for the vector moving average case become

$$N^{-1} \sum_{\ell=o}^{p} \sum_{n=1}^{N} \exp\{i(\ell-k)\lambda_n\} g^{-1*} G^{-1}(\alpha) g^{-1} A(\ell) g^{-1} I g^{-1*} = 0 \qquad k=1,\ldots,p \qquad (5.18)$$

where $G(\alpha) = 2\pi N^{-1} \Sigma_n g^{-1} I g^{-1*}$ and $\alpha = \text{vec}\{A(1),\ldots,A(p)\}$. Assuming again that the spectrum is factorized to yield $\hat{A}(j)$ $(j=1,\ldots,p)$, the linearized version of (5.18) becomes

$$N^{-1} \sum_{\ell=o}^{p} \sum_{n=1}^{N} \exp\{i(\ell-k)\lambda_n\} \hat{f}^{-1} \hat{A}^{(1)}(\ell) \check{g}^{-1} I \check{g}^{-1*} = 0 \qquad (5.19)$$

where

$$\check{g}(\lambda_j) = \sum_{k=o}^{p} \hat{A}(k) \exp\{ik\lambda_j\}.$$

Hence (5.19) would replace Hannan's eqn (12) if the treatment there was a direct generalization of the scalar case. A similar remark extends to Hannan's eqn (13). Therefore, unlike the scalar case, it is harder to compare Hannan's estimator for the vector moving average to the G.N. estimator.

6. CONTINUATION - THE ARMAX MODEL

The final model we look at is the ARMAX model

$$\sum_{j=o}^{r} B_j y(n-j) = \sum_{j=o}^{s} C_j x(n-j) + \sum_{j=o}^{t} A_j e(n-j), \quad A_o = I_{(p)} \qquad (6.10)$$

We have already discussed this model in 4.§5. Let $B(z)$, $C(z)$, $A(z)$ and θ be defined as in §4 and put

$v(n;\theta) = B_o^{-1}A(L)^{-1}\{B(L)y(n)-C(L)x(n)\}$. It is clear from Chapter 4 why $v(n;\theta)$ has been defined in this way. Below we assume that the G.N. iterates are constructed as described in Chapter 4.

Using Lemma 2(iv), the linearization of $v(n;\theta)$ about $\hat{\theta}$ is given by

$$v(n;\theta,\hat{\theta}) = p(n)+\hat{B}_o^{-1}\hat{A}(L)^{-1}[\{B(L)-\hat{B}(L)\}y(n)-\{C(L)-\hat{C}(L)\}x(n)- -\{A(L)-\hat{A}(L)\}\hat{B}_o p(n)] - \hat{B}_o^{-1}(B_o-\hat{B}_o)p(n) \qquad (6.11)$$

where $p(n) = v(n;\hat{\theta})$; we write $A(L)$ evaluated at $\hat{\theta}$ as $\hat{A}(L)$, with $\hat{B}(L)$ and $\hat{C}(L)$ defined similarly.

If (6.10) is a reduced form system, then we redefine θ by dropping B_o. (6.11) reduces to

$$v(n;\theta,\hat{\theta}) = p(n)+\hat{A}(L)^{-1}[\{B(L)-\hat{B}(L)\}y(n)-\{C(L)-\hat{C}(L)\}x(n)- -\{A(L)-\hat{A}(L)\}p(n)] \qquad (6.12)$$

We note the following

(i) Comparing (6.11) and (6.12) with (3.50) and (2.25) respectively in Phillips [1966] we see that the estimators derived by Phillips are each precisely the first step of a G.N. scheme for the simultaneous ARMAX model and the reduced form ARMAX models respectively. Although this fact is well known for the reduced form case, it may not be (as well?) known for the simultaneous equations case.

(ii) Hannan and Nicholls [1972] and Nicholls [1976] extend to the reduced form ARMAX case Hannan's [1969b] ARMA estimator, while Nicholls [1977] extends Akaike's [1973] ARMA estimator, also to the reduced form ARMAX case. Put $\alpha = \text{vec}(A_1,\ldots,A_t)$, $\gamma = \text{vec}(C_o,\ldots,C_t)$ and

$\Omega(\theta) = N^{-1}\Sigma_n v(n;\theta)v(n;\theta)'$.

Despite an assertion to the contrary in Nicholls [1977] we can easily deduce from §5 that the estimator given in Nicholls [1977] is just the first step of a G.N. scheme where, in addition,

$$N^{-1} \sum_{n=1}^{N} \frac{\partial v(n;\hat{\theta})'}{\partial \alpha} \Omega(\hat{\theta})^{-1} \frac{\partial v(n;\hat{\theta})}{\partial \gamma'}$$

is replaced by its asymptotic limit which is zero. (if $\hat{\theta} - \theta_o = o_p(1)$). Therefore the Nicholls [1977] estimator corresponds to the scoring method. We see from §5 that it is not identical to the Nicholls [1976] estimator (as claimed in Nicholls [1977]) but differs from it by a term of order $o_p(N^{-\frac{1}{2}})$.

It seems to me that the scoring estimator will be computationally inferior to the G.N. estimator for the reasons given in (iii) of §4.

CHAPTER 6

RECURSIVE ESTIMATION AND THE GAUSS-NEWTON ALGORITHM

1. INTRODUCTION

To introduce the idea of recursive estimation we first look at the multivariate linear regression model

$$y(n) = V(n)\theta^{(0)} + e(n) \qquad n = 1,2,... \tag{1.10}$$

where $V(n)$ is a $p\times q$ nonstochastic observed matrix, $\theta^{(0)}$ a $q\times 1$ vector of parameters, $y(n)$ a $p\times 1$ vector of observations and $\{e(n), n \geq 1\}$ a sequence of independent and identically distributed random variables having zero mean and a nonsingular covariance matrix $\Omega^{(0)}$. If θ_N is the generalized least squares estimator of $\theta^{(0)}$ based on the first N observations, then

$$\theta_N = P_N\{\sum_{n=1}^{N} V(n)'\Omega^{(0)-1}y(n)\} \tag{1.11}$$

where

$$P_N = \{\Sigma_n V(n)'\Omega^{(0)-1}V(n)\}^{-1}.$$

It can be readily verified that (see Appendix 1)

$$\theta_N = \theta_{N-1}+P_N V(N)'\Omega^{(0)-1}\{y(N)-V(N)\theta_{N-1}\} \tag{1.12}$$

where P_N can be updated by

$$P_N = P_{N-1} - P_{N-1}V(N)'[V(N)P_{N-1}V(N)'+\Omega^{(0)}]^{-1}V(N)P_{N-1} \tag{1.13}$$

In accordance with current usage, we call (1.12) the recursive version of the least squares estimator (1.11). Let $\hat{e}(n) = y(n) - V(n)\theta_{n-1}$; the $\hat{e}(n)$ are called the recursive residuals and have nice properties (Brown, Durbin and Evans [1975, p.152]) if the $e(n)$ are Normally distributed and the $V(n)$ are nonstochastic. If $\Omega^{(0)}$ is unknown, then it may be estimated by

$$\Omega_N = N^{-1} \sum_{n=1}^{N} \hat{e}(n)\hat{e}(n)'$$

and this estimate substituted for $\Omega^{(0)}$ in (1.12) and (1.13). To avoid inverting Ω_N for each N we can update Ω_N^{-1} by using

$$\Omega_N^{-1} = N\{(N-1)^{-1}\Omega_{N-1}^{-1} - \tau_N^{-1}(N-1)^{-2}\Omega_{N-1}^{-1}\hat{e}(n)\hat{e}(n)'\Omega_{N-1}^{-1}\} \quad (1.14)$$

where

$$\tau_N = 1 + (N-1)\hat{e}(N)'\Omega_{N-1}^{-1}\hat{e}(N).$$

The recursive estimation technique may be described as follows: we obtain the sequence of estimates $\{\theta_n; M < n \leq N\}$ by using the updating formulae (1.12) to (1.14) with M some prescribed non-negative integer and with θ_M, P_M and Ω_M prescribed prior to estimation in order to initialize the recursion. Depending on the choice of θ_M, P_M and Ω_M, the sequence of recursive estimators $\{\theta_n, n > M\}$ may or may not coincide with the sequence of non-recursive generalized least squares estimators (1.11). For the single equation version of (1.10), the recursions (1.12) and (1.13) were first given by Plackett [1950].

We will now discuss briefly recursive instrumental variable estimation for the model (1.10). Let $\{Z(n), n \geq 1\}$ be a sequence of $q \times p$ matrices of instrumental variables. Then

$$\theta_N^{(IV)} = Q_N \sum_{n=1}^{N} Z(n)y(n) \quad (1.20)$$

is an instrumental variable estimator of $\theta^{(0)}$ based on N observations, where $Q_N = \{\Sigma Z(n)V(n)\}^{-1}$. A recursive version of

(1.20) is given by

$$\theta_n = \theta_{n-1} + Q_n Z(n)\{y(n) - V(n)\theta_{n-1}\} \quad (1.21)$$

$$Q_n = Q_{n-1} - Q_{n-1} Z(n)\{V(n) Q_{n-1} Z(n) + I_{(p)}\}^{-1} V(n) Q_{n-1} \quad (1.22)$$

In this chapter we show how to obtain recursive estimators for multivariate time series models by using the G.N. algorithm. As will be explained below, the non-recursive versions of our class of recursive estimators have desirable statistical and computational properties. We will not give any consistency or convergence in distribution results for the recursive estimators as, given present techniques, they seem very hard to get.

A heuristic discussion of almost sure convergence for recursive estimators is given in Söderström, Ljung and Gustavsson [1974] (see also references therein), while Hannan [1976c] and Hannan and Tanaka [1976] give rigorous proofs of a.s. convergence for some simple time series models.

We will also relate our class of recursive estimators to some of the estimators appearing in the literature.

We conclude this section by explaining why it seems likely that recursive estimation will play an important role in the empirical analysis of time series models:

(a) To examine the stability of parameters and variances in the model by looking at the sequence of recursive parameter estimates and functions of the recursive residuals. Various test criteria have already been suggested in Brown, Durbin and Evans [1975] for the single equation version of (1.10).

(b) To estimate on-line (particularly in engineering) where it would be impractical to re-estimate the whole model every time a new observation came in. Furthermore we may continue estimating till a desired degree of accuracy is achieved.

(c) Tracking parameters in real time (see Gertler and Bányász [1974, fig. 2].

(d) Closely related to (b) is the updating of estimates in off-line (ordinary) estimation when it may be too costly to re-estimate the whole model every time a new observation came in; e.g. monthly observations in a large econometric model.

2. RECURSIVE ESTIMATION FOR NONLINEAR TIME SERIES MODELS

Suppose we have the model

$$v(n;\theta^{(0)}) = e(n) \qquad n = 1,2.... \tag{2.10}$$

subject to the assumptions of 3.§2. For convenience we denote the true values of θ and Ω by $\theta^{(0)}$ and $\Omega^{(0)}$ respectively. This is a departure from the notation of previous chapters.

In Chapters 3 to 5 we obtained estimates of $\theta^{(0)}$ and $\Omega^{(0)}$ by doing the following: (i) we linearized about θ_o, an initial estimate (guess) of $\theta^{(0)}$ to obtain

$$v(n;\theta) = v(n;\theta_o) + V(n;\theta_o)(\theta-\theta_o)$$

where $V(n;\theta) = \partial v(n;\theta)/\partial\theta'$; (ii) we then performed a generalized least squares regression based on N observations, using the estimated covariance matrix $\Omega_N(\theta_o)$, to give the first step of the G.N. algorithm

$$\hat{\theta}-\theta_o = -P_N(\theta_o) \sum_{n=1}^{N} V(n;\theta_o)'\Omega_N^{-1}(\theta_o)v(n;\theta_o) \tag{2.20}$$

where

$$P_N(\theta) = \{\Sigma V(n;\theta)'\Omega_N^{-1}(\theta)V(n;\theta)\}$$

By analogy with §1, a recursive version of (2.20) would be

$$\theta_n = \theta_{n-1} - P_n V(n;\theta_{n-1})'\Omega_n^{-1} v(n;\theta_{n-1}) \tag{2.21}$$

where P_n and Ω_n^{-1} are updated by

$$P_n = P_{n-1} - P_{n-1}V(n;\theta_{n-1})'\{V(n;\theta_{n-1})P_{n-1}V(n;\theta_{n-1})' + \\ + \Omega_n\}^{-1}V(n;\theta_{n-1})P_{n-1} \tag{2.22}$$

$$\Omega_n^{-1} = n\{(n-1)^{-1}\Omega_{n-1}^{-1} - (n-1)^{-2}\tau_n^{-1}\Omega_{n-1}^{-1}v(n;\theta_{n-1})v(n;\theta_{n-1})'\Omega_{n-1}^{-1}\} \tag{2.23}$$

with

$$\tau_n = 1 + (n-1)^{-1}v(n;\theta_{n-1})'\Omega_{n-1}^{-1}v(n;\theta_{n-1}).$$

Remark 2.1 We showed in 3.§3, in particular Theorem 3.4 and Corollary 3.1, and in 5.§2 and 5.§3 that the non-recursive versions of our class of recursive estimators had desirable statistical and computational properties.

(ii) Closely related to our sequence of recursive estimators (2.21) is the class of recursive estimators based on the Newton-Raphson algorithm which was suggested by Gertler and Bányász [1974]. It seems that our estimators are simpler to compute, especially in the multivariate case.

3. RECURSION IN MULTIVARIATE LINEAR TIME SERIES MODELS

For expository purposes we give a detailed derivation of the recursive estimator for the ARMAX model (3.10) (see below) and for the vector linear simultaneous equations ARMAX model (3.30). Before doing so however it is useful to have the result of the following lemma.

Lemma 1 Suppose

$$G(z) = \sum_{j=d}^{d'} G_j z^j$$

is a $b\times c$ polynomial matrix and r is a c dimensional vector. Let $\chi(z)' = (z^d,\dots,z^{d'})$ and $\gamma = \text{vec}(G_d,\dots,G_{d'})$. Then $G(z)r = (\chi(z)'\otimes r'\otimes I_{(b)})\gamma$.

The proof of the above lemma is straightforward and hence omitted.

We first look at the ARMAX model

$$y(n) + \sum_{j=1}^{r} B_j(\theta^{(0)})y(n-j) = \sum_{j=o}^{s} C_j(\theta^{(0)})x(n-j)+e(n) + \\ + \sum_{j=1}^{t} A_j(\theta^{(0)})e(n-j) \tag{3.10}$$

subject to the assumptions of 4.§2. Let

$$A(z;\theta)=I_{(p)}+ \sum_{j=1}^{t} A_j(\theta)z^j,\ B(z;\theta)=I_{(p)}+ \sum_{j=1}^{r} B_j(\theta)z^j,$$

$$C(z;\theta) = \sum_{j=o}^{s} C_j(\theta)z^j$$

$$\chi_\beta(z)' = (z,\dots,z^r),\ \chi_\gamma(z)' = (1,\dots,z^s),\ \chi_\alpha(z)' = (z,\dots,z^t),$$

$$\alpha(\theta) = \text{vec}\{A_1(\theta),\dots,A_t(\theta)\}$$

$$\beta(\theta) = \text{vec}\{B_1(\theta),\dots,B_r(\theta)\},\gamma(\theta) = \text{vec}\{C_o(\theta),\dots,C_s(\theta)\},$$

$$\underline{a}(\theta) = \partial\alpha(\theta)/\partial\theta',$$

$$\underline{b}(\theta) = \partial\beta(\theta)/\partial\theta',\ \underline{c}(\theta) = \partial\gamma(\theta)/\partial\theta'.$$

Putting $v(n;\theta) = A(L;\theta)^{-1}\{B(L;\theta)y(n)-C(L;\theta)x(n)\}$, where L is the lag operator, we can re-express (3.10) in the form (2.10). By Lemma 5.2(i)

$$v(n;\theta) = v(n;\tilde\theta)+A(L;\tilde\theta)^{-1}[\{B(L;\theta)-B(L;\tilde\theta)\}y(n)-\{C(L;\theta)- \\ -C(L;\tilde\theta)\}x(n)-\{A(L;\theta)-A(L;\tilde\theta)\}v(n;\tilde\theta)]+O(\|\theta-\tilde\theta\|^2)$$

and it now follows from Lemma 1 and the last equation that

$$V(n;\theta) = \frac{\partial v(n;\theta)}{\partial\theta'} = A(L;\theta)^{-1}[\{\chi_\beta(L)'\otimes y(n)'\otimes I_{(p)}\}\underline{b}(\theta) -$$

$$-\{\chi_\gamma(L)'\otimes x(n)'\otimes I_{(p)}\}\underline{c}(\theta)-\{\chi_\alpha(L)'\otimes v(n;\theta)'\otimes I_{(p)}\}\underline{a}(\theta)].$$

We obtain $v(n;\theta_{n-1})$ and $V(n;\theta_{n-1})$ recursively by using the relations

$$v(n;\theta_{n-1})=-\sum_{j=1}^{t} A_j(\theta_{n-1})v(n-j;\theta_{n-j-1})+B(L;\theta_{n-1})y(n)-C(L;\theta_{n-1})x(n) \tag{3.20}$$

$$\begin{aligned}V(n;\theta_{n-1})= &-\sum_{j=1}^{t} A_j(\theta_{n-1})V(n-j;\theta_{n-j-1})+\\ &+\{\chi_\beta(L)'\otimes y(n)'\otimes I_{(p)}\}\underline{b}(\theta_{n-1}) -\\ &-\{\chi_\gamma(L)'\otimes x(n)'\otimes I_{(p)}\}\underline{c}(\theta_{n-1})-\{\chi_\alpha(L)'\otimes v(n;\theta_{n-1})'\otimes I_{(p)}\}\underline{a}(\theta_{n-1})\end{aligned} \tag{3.21}$$

with the j^{th} block column of $\{\chi_\alpha(L)'\otimes v(n;\theta_{n-1})'\otimes I_{(p)}\}$ given by $\{v(n-j);\theta_{n-j-1})'\otimes I_{(p)}\}(j = 1,\dots,t)$. A recursive estimator of $\theta^{(0)}$ can now be constructed as in (2.21) - (2.23).

We now discuss recursive estimation for the linear simultaneous equations model

$$\sum_{j=o}^{r} \underline{B}_j y(n-j) = \sum_{j=o}^{s} \underline{C}_j x(n-j)+\underline{e}(n) + \sum_{j=1}^{t} \underline{A}_j \underline{e}(n-j) \tag{3.30}$$

where (3.30) is subject to the assumptions made for Model 2 of 4.§5. For (3.30) to be identified it is necessary (but not sufficient) to impose some constraints on the $\underline{B}_j$, $\underline{C}_j$ and $\underline{A}_j$ matrices. A common set of constraints is the following:

(a) The diagonal elements of $\underline{B}_o$ are unity - a normalization constraint.

(b) Certain elements of the matrices $\underline{B}_j$ $(j = 0,\dots,r)$ and $\underline{C}_j$ $(j = 0,\dots,s)$ are prescribed as zero.

Below we assume that the indices i,j and k have the following ranges

$$1 \le i \le p,\ 0 \le j \le r,\ 0 \le k \le s,\ 1 \le \ell \le t.$$

Let $b_j(i)'$ be the i^{th} row of $\underline{B}_j$, $\beta_j(i)'$ the row vector containing the unconstrained elements of $b_j(i)'$ in their natural order and define $c_k(i)'$ and $\gamma_k(i)'$ similarly but with respect to the $\underline{C}_k$ matrices; let

$$\beta_j = \text{vec}\{\beta_j(1),\ldots,\beta_j(p)\},\ \gamma_k = \text{vec}\{\gamma_k(1),\ldots,\gamma_k(p)\},$$

$\beta = \text{vec}\{\beta_o,\ldots,\beta_r\}$, $\gamma = \text{vec}\{\gamma_o,\ldots,\gamma_s\}$, $\alpha = \text{vec}(\underline{A}_1,\ldots,\underline{A}_t)$ and $\theta = \text{vec}(\beta,\gamma,\alpha)$.

We define $Y_j(n)$, $X_k(n)$ and $Q(n;\theta)$ by

$\underline{B}_j y(n) = Y_j(n)\beta_j$, $j > 0$; $(\underline{B}_o - I_{(p)})y(n) = Y_o(n)\beta_o$, $\underline{C}_k x(n) = X_k(n)\gamma_k$,

$(\underline{B}_o - I_{(p)})v(n;\theta) = Q(n;\theta)\beta_o$ and let $q(n;\theta) = \underline{B}_o v(n;\theta)$,

$X(n) = \{X_o(n),\ldots,X_s(n-s)\}$, $Y(n) = \{Y_o(n),\ldots,Y_r(n-r)\}$,

$S(n;\theta) = Y(n) - \{A(L)Q(n;\theta),0\}$, $B_j(\theta) = \underline{B}_o^{-1}\underline{B}_j$, $C_k(\theta) = \underline{B}_o^{-1}\underline{C}_k$,

$A_\ell(\theta) = \underline{B}_o^{-1}A_\ell\underline{B}_o$, $e(n) = \underline{B}_o^{-1}\underline{e}(n)$ and $\Omega^{(0)} = \underline{B}_o^{-1}\Omega^{(0)}\underline{B}_o^{-1'}$;

Putting $\underline{B}(z) = \sum_{j=o}^{r} \underline{B}_j z^j$, $\underline{C}(z) = \sum_{j=o}^{s} \underline{C}_j z^j$ and $\underline{A}(z) = I_{(p)} + \sum_{j=1}^{t} \underline{A}_j z^j$

we can rewrite (3.30) in the standard form (3.10) so that $v(n;\theta)$ becomes

$$v(n;\theta) = \underline{B}_o^{-1}\underline{A}(L)^{-1}\{\underline{B}(L)y(n) - \underline{C}(L)x(n)\} \tag{3.35}$$

It follows from (5.6.11) and Lemma 1 that

$$V(n;\theta) = \underline{B}_o^{-1}\underline{A}(L)^{-1}[S(n;\theta),-X(n),-\{\chi_\alpha(L)'\otimes q(n;\theta)'\otimes I_{(p)}\}]$$

$S(n;\theta_{n-1})$, $v(n;\theta_{n-1})$ and $V(n;\theta_{n-1})$ are now obtained recursively by

$$S(n;\theta_{n-1}) = Y(n) - \{Q(n;\theta_{n-1}) + \sum_{j=1}^{t}\underline{A}_j Q(n-j;\theta_{n-j-1}),0\} \tag{3.40}$$

$$v(n;\theta_{n-1}) = \underline{B}_o^{-1}\{\underline{B}(L)y(n) - \underline{C}(L)x(n)\} - \sum_{j=1}^{t} \underline{A}_j \underline{B}_o v(n-j;\theta_{n-j-1})\} \quad (3.41)$$

$$V(n;\theta_{n-1}) = \underline{B}_o^{-1}[\{S(n;\theta_{n-1}), -X(n), -(\chi_\alpha(L)'\otimes q(n;\theta_{n-1})'\otimes I_{(p)})\} - \\ - \sum_{j=1}^{t} \underline{A}_j \underline{B}_o V(n-j;\theta_{n-j-1})] \quad (3.42)$$

with $\underline{B}_j$, $\underline{C}_k$ and $\underline{A}_\ell$ evaluated at θ_{n-1} in (3.40) to (3.42). In (3.42) the ℓth block column of $\{\chi_\alpha(L)'\otimes q(n;\theta_{n-1})'\otimes I_{(p)}\}$ is evaluated as $\{q(n-\ell;\theta_{n-\ell-1})'\otimes I_{(p)}\}$. Recursive estimators for (3.30) may now be constructed as in (2.21) to (2.23).

Remark 3.1 (i) Because of the linear structure of the models (3.10) and (3.30) we were able to obtain $v(n;\theta_{n-1})$ and $V(n;\theta_{n-1})$ using (3.20), (3.21), (3.41) and (3.42) without doing excessive calculation. An alternative to obtaining $v(n;\theta_{n-1})$ by (3.20) would be to use

$$v(n;\theta_{n-1}) = -\sum_{j=1}^{t} A_j(\theta_{n-1})v(n-j;\theta_{n-1}) + B(L;\theta_{n-1})y(n) - \\ - C(L;\theta_{n-1})x(n)$$

but this involves much more calculation as we would have to recompute $v(j;\theta_{n-1})$ $(j = 1,\ldots,n-1)$ for each n.

(ii) As in (1.31) and (1.32) we may obtain recursive instrumental variable estimators for the models (3.10) and (3.30). Instruments that work well in non-recursive estimation should be suitable candidates for recursive estimation.

(iii) With $y(n)$, $x(n)$ and $\underline{e}(n)$ scalars and $\underline{B}_o$ equal to unity, Führt [1973] seems to have been the first to derive the algorithm just obtained for the model (3.30) and for this case the recursive algorithm is sometimes known in the literature as RML2 (see Söderström, Ljung and Gustavsson [1974, p.16]. The direct link of the algorithm with the Gauss-Newton linearization does not seem to have been noted by Führt.

(iv) We can similarly obtain recursive estimators for Model 1 of 4.§5. With $y(n)$, $x(n)$ and $\underline{e}(n)$ scalars and $\underline{B}_o$ equal to unity Hannan and Tanaka [1976] derived the recursive G.N. algorithm for this model. Again the direct link with the G.N. algorithm does not seem to have been noted.

4. SOME FINAL COMMENTS

(i) It is clear from (2.21) that if P_n tends to zero as n increases, which is normally the case, new observations will affect the sequence of recursive estimators $\{\theta_n, n \geq 1\}$ less and less. A common practice (e.g. Söderström, Ljung and Gustavsson [1974, p.22]) to make the recursive algorithm more responsive to new data is to compute P_n by (2.22) with P_{n-1} replaced by $\lambda(n)^{-1}P_{n-1}$ on the right in (2.22) where the sequence $\{\lambda(n), n \geq 1\}$ satisfies the relation $\lambda(n+1) = \lambda_o\lambda(n) + (1-\lambda_o)$ with λ_o chosen close to 1, e.g. 0.99.

(ii) When P_n becomes small we may approximate

$$\{V(n;\theta_{n-1})P_{n-1}V(n;\theta_{n-1})' + \Omega_n\}^{-1}$$

in (2.22) by $\Omega_n^{-1} - \Omega_n^{-1}V(n;\theta_{n-1})P_{n-1}V(n;\theta_{n-1})'\Omega_n^{-1}$ thus eliminating the need to do any matrix inversions when updating the recursions.

(iii) To initialize the recursions (2.21)-(2.23) we need to specify θ_o, P_o and Ω_o apriori. The choice of these parameters should, if possible, be based on prior knowledge. Bearing in mind the remarks in (i) above, strong uncertainty about θ_o ought to be reflected in a choice of a large P_o so that (2.21) is responsive to new data.

(iv) We now briefly recapitulate the reasons for recommending the class of recursive estimators given in the thesis. As mentioned in Remark 2.1(i) the non-recursive analogues of these estimators have desirable statistical and computational properties. Theoretical results (Hannan [1976c] and Hannan and Tanaka [1976] for very simple models indicate that our class of recursive estimators is strongly consistent and simulation results (Hannan and Tanaka [1976] and Söderström *et.al* [1974]; see also the references in the last cited paper), again for very simple models, indicate that the Gauss-Newton recursive estimates are well behaved.

(v) It seems likely that even asymptotic results on how the recursive estimators and functions of the recursive residuals are distributed will be very difficult to obtain. None have been obtained to date. Nevertheless I feel that a judicious use of analogy between the model discussed in §1, where some yardsticks have been given by Brown, Durbin and Evans [1975], and more complicated time series models, together with simulation results, ought to give us some yardsticks in the time series case as well.

(vi) For an interesting discussion of recursive estimators see Young [1976].

APPENDIX 1

We now check that (1.12) and (1.13) hold. For convenience we omit the zero superscript in $\Omega^{(0)}$.

$$[P_{N-1}^{-1} + V(N)'\Omega^{-1}V(N)][P_{N-1}-P_{N-1}V(N)'\{V(N)P_{N-1}V(N)' + \Omega\}^{-1}\times$$
$$\times V(N)P_{N-1}]$$

$$= I_{(q)}-V(N)'[I_{(p)}+\Omega^{-1}V(N)P_{N-1}V(N)' -$$

$$-\Omega^{-1}\{\Omega+V(N)P_{N-1}V(N)'\}]\{V(N)P_{N-1}V(N)'+\Omega\}^{-1}V(N)P_{N-1}$$

$$= I_{(q)}.$$

Let $b_n = \sum_{j=1}^{n} V(j)'\Omega^{-1}y(j)$. Then

$$\theta_N = [P_{N-1}-P_{N-1}V(N)'\{V(N)P_{N-1}V(N)'+\Omega\}^{-1}V(N)P_{N-1}]\{b_{N-1}+$$
$$+ V(N)'\Omega^{-1}y(N)\}$$

$$= \theta_{N-1}+P_{N-1}V(N)'\{V(N)P_{N-1}V(N)'+\Omega\}^{-1}[\{V(N)P_{N-1}V(N)'+\Omega\}\Omega^{-1}y(N)-$$

$$-V(N)\theta_{N-1}-V(N)P_{N-1}V(N)'\Omega^{-1}y(N)]$$

$$= \theta_{N-1}+P_{N-1}V(N)'\{V(N)P_{N-1}V(N)'+\Omega\}^{-1}\{y(N)-V(N)\theta_{N-1}\}$$

$$= \theta_{N-1}+P_NV(N)'\Omega^{-1}\{y(N)-V(N)\theta_{N-1}\}$$

because

$$P_NV(N)'\Omega^{-1} = P_{N-1}V(N)'\{V(N)P_{N-1}V(N)'+\Omega\}^{-1}.$$

Hence (1.12) holds.

CHAPTER 7

ESTIMATION IN REGRESSION MODELS WITH STATIONARY RESIDUALS

1. INTRODUCTION

In this chapter and the next we consider the estimation of the unknown parameter vector θ_o appearing in the multivariate regression model

$$y(n) = \sum_{j=o}^{\infty} D_j(\theta_o)x(n-j) + v(n), \qquad n \geq 1, \tag{1.10}$$

with $v(n)$ a stationary p dimensional residual. We observe the $y(n)$ and $x(n)$, but <u>not</u> the $v(n)$. Unlike Chapters 3 and 4 all estimation takes place in the frequency domain rather than in the time domain. Given $\{y(n), x(n);\ 1 \leq n \leq N\}$ let $w_x(\lambda)$ and $w_y(\lambda)$ be the FFT's (see 1.§2) of the $\{x(n)\}$and $\{y(n)\}$ respectively. Let

$$D(z;\theta) = \sum_{j=o}^{\infty} D_j(\theta)z^j, \quad D(\lambda;\theta) = D(e^{i\lambda};\theta),$$

$$V(\lambda;\theta) = w_y(\lambda) - D(\lambda;\theta)w_x(\lambda)$$

and put $\lambda_t = 2\pi t/N$ for $-\tfrac{1}{2}N < t \leq [\tfrac{1}{2}N]$.

Suppose that: (i)[1] $\mathcal{B}$ is a subset of $(-\pi,\pi)$ and is composed of a finite number of disjoint open intervals that are symmetrical about $\lambda = 0$, i.e. if $\lambda \in \mathcal{B}$, then $-\lambda \in \mathcal{B}$; (ii) $\Phi(\lambda)$ is a Hermitian p.s.d. matrix function of λ, $\lambda \in [-\pi,\pi]$, with $\Phi(-\lambda) = \Phi(\lambda)'$. An estimate of θ_o is obtained by minimizing the criterion function

---

[1] Such a set $\mathcal{B}$ is also considered by Hannan and Robinson [1973].

$$L_N(\theta;\Phi) = (2N)^{-1} \operatorname{tr} \sum_{B} \Phi(\lambda_t) V(\lambda_t;\theta) V(\lambda_t;\theta)^* \tag{1.15}$$

over all θ lying in a compact space and satisfying restrictions of the form $h(\theta) = 0$. Here, and in the sequel, $\sum_{B}$ is a sum over all $\lambda_t \in B$. Subject to the assumptions stated below we shall show that $\hat{\theta}_N$, that value of θ minimizing (1.15), is strongly consistent and asymptotically normal. It is also efficient if $\Phi = f_v^{-1}$ and $B = (-\pi,\pi)$, where f_v is the spectral density matrix of v(n). By efficient we mean that $\hat{\theta}_N$ will have the same <u>asymptotic</u> distribution as the maximum likelihood estimator of θ_o when the v(n) are Gaussian and f_v is known. We remind the reader that <u>efficiency</u> in this chapter and the next means <u>asymptotic</u> <u>efficiency</u>.

Because it will usually be difficult to obtain the closed form solution of the Normal Equations we again give an iterative G.N. algorithm with iterates that converge to $\hat{\theta}_N$. As in previous chapters, the iterates themselves will be consistent and have the same asymptotic distribution as $\hat{\theta}_N$.

In the next chapter we show that if $\Phi(\lambda)$ is unknown but a consistent estimate of it is available, then we can substitute it for $\Phi(\lambda)$ in (1.10) and the results of the present chapter will still hold. We will also apply the theory to the linear regression model

$$\sum_{j=o}^{r} B_j\, y(n-j) = \sum_{j=o}^{s} C_j\, x(n-j) + u(n), \tag{1.20}$$

where $\{u(n)\}$ is a stationary residual sequence. We have already seen in Chapters 4 and 5 that when $B_o \neq I_{(p)}$ and u(n) has a rational spectrum, i.e. u(n) is generated by an ARMA model, (1.20) is frequently used in Econometric work. For a further discussion of this model see 8.§2.

Our model (1.10) is a particular case of a more general continuous time model considered in Robinson [1972b,1975]. Using a criterion function of the form (1.15) Robinson obtains consistency and asymptotic normality results for $\hat{\theta}_N$ similar to ours. Because our model is more specific than Robinson's we have been able to weaken his conditions in some respects; in other respects, however, his results are more general than ours. A fuller discussion of the difference between Robinson's results and ours is given in the text below as they arise. To help compare Robinson's work with ours his assumptions are listed in Appendix 2 of this chapter.

2. MOTIVATION AND CHAPTER STRUCTURE

We now motivate the use of the criterion function (1.15) and give reasons for estimating in the frequency domain. Suppose that the $v(n)$ are p dimensional, stationary, Gaussian, random variables and $f_v(\lambda)$ is nonsingular for $\lambda \in [-\pi,\pi]$. Let $v_N' = \{v(1)',\ldots,v(N)'\}$ and $\Gamma_N = E(v_N v_N')$. Then, omitting additive constants, $-{}^1/N \times \log$ - likelihood of v_N is

$$(2N)^{-1} \log \det \Gamma_N + (2N)^{-1} v_N' \Gamma_N^{-1} v_N \qquad (2.10).$$

Let S_N be the $Np \times Np$ block matrix having $(t,n)^{th}$ block $N^{-\frac{1}{2}}.\exp(in\lambda_t).I_{(p)}$ and w_N the block vector with entries $(2\pi)^{\frac{1}{2}} w_v(\lambda_t)$ $(1 \le n \le N,\ -\frac{1}{2}N < t \le [\frac{1}{2}N])$. It is easy to check that $S_N^* S_N = I_{(Np)}$ and we know that (Hannan[1970,p.378]) $S_N \Gamma_N S_N^* \approx \Lambda_N + o(1)$; where Λ_N is a block diagonal matrix with entries $2\pi \times f_v(\lambda_t)$ on the diagonal. Since $S_N v_N = w_N$, it follows that for N large, a good approximation to (2.10) is

$$(2N)^{-1} \log \det \Lambda_N + (2N)^{-1} w_N^* \Lambda_N^{-1} w_N =$$

$$=^2 (2N)^{-1} \sum_t \log \det f_v(\lambda_t) + (2N)^{-1} \mathrm{tr} \sum_t f_v^{-1}(\lambda_t) w_v(\lambda_t) w_v(\lambda_t)^*;$$

here, and in the sequel, $\sum_t$ means summation over $-\frac{1}{2}N < t \leq [\frac{1}{2}N]$. Hence a good approximation to $-{}^1/N \times \log$ - likelihood[2] of vec$\{y(1),\ldots,y(N)\}$ is

$$(2N)^{-1} \sum_t \log \det 2\pi f_v(\lambda_t) + (2N)^{-1} \mathrm{tr} \sum_t f_v^{-1}(\lambda_t) V(\lambda_t;\theta) V(\lambda_t;\theta)^* \quad (2.12)$$

where we have used the approximation

$$(2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} \{y(n) - \sum_{j=o}^{\infty} D_j(\theta) x(n-j)\} e^{in\lambda_t} \approx V(\lambda_t;\theta).$$

Replacing f_v^{-1} by Φ in the second term of (2.12) and restricting the summation over the λ_t to the set $\mathcal{B}$ described in §1, we obtain the criterion function $L_N(\theta;\Phi)$. Hence our estimation procedure consists of minimizing a "generalized exponent" of a Gaussian likelihood.

The above heuristic derivation of (1.15) depended on Gaussianity and the approximation of (2.10) by (2.12). Our results, however, do not depend on Gaussianity nor on justifying rigorously the approximation of (2.10) by (2.12). Purely as a matter of interest, however, we show in Appendix 1 that the difference between (2.10) and (2.12) tends to zero as N increases.

We now briefly discuss why replacing f_v^{-1} by Φ and allowing $\mathcal{B}$ to be a strict subset of $[-\pi,\pi]$ permits a greater flexibility in the estimation of θ_o. It is true, however, that the choice of $\Phi = f_v^{-1}$ and $\mathcal{B} = [-\pi,\pi]$, if permissible, is optimal asymptotically in a minimum variance sense. This is shown in §5.

[2] Omitting additive constants.

Dealing with Φ first, it may be that f_v is singular, or nearly so, for some values of λ, or that it is unknown and difficult to estimate. To avoid numerical instability in the estimates, associated with a nearly singular f_v, or the problems associated with estimating f_v, we may prefer to choose, if possible, a $\Phi \neq f_v^{-1}$ that avoids such difficulties and results in only a small loss in efficiency. For a further discussion of this point see 8.§6.

When working in the frequency domain it is very easy to filter off certain frequencies or bands of frequencies by simply choosing $\mathcal{B} \neq (-\pi,\pi)$. Reasons for doing such filtering may be:

(i) the model (1.10) is not valid over the full frequency range $(-\pi,\pi)$;

(ii) mean correction and trend and seasonal adjustment are easily done by dropping off appropriate frequencies;

(iii) the signal to noise ratio may be very low for some frequency bands. The estimation procedure may then be simplified, without too much loss in efficiency, by dropping these frequency bands. Knowledge of the signal to noise characteristics may come from an understanding of the recording device, previous estimation experience with like data or preliminarly examination of the current data.

We will now give reasons for working in the frequency domain rather than, or in complement with, the time domain.

(i) For the filtering reasons advanced when discussing $\mathcal{B}$.

(ii) If $D(z;\theta)$ is not a rational function of z, or if $v(n)$ cannot be represented adequately by an ARMA process, then it seems to be more natural to estimate in the frequency domain than in the time domain.

$$\text{e.g. if} \quad y(n) = \sum_{j=o}^{\infty} \frac{\theta^j}{j!} x(n-j) + v(n)$$

then $D(z;\theta) = e^{\theta z}$.

(iii) Even if we could estimate in the time domain, the frequency domain estimates may still be quite useful because:

(a) of the reasons advanced in (i) above

(b) f_v may not be known (this is usually the case). The estimates of θ_o which are obtained by using a nonparametric estimate of f_v can be used as benchmark against which to compare estimates obtained by fitting an ARMA model to the residuals.

(c) The fitting of an AR or ARMA model to the residuals may be extremely tedious, especially in the vector case, if the lag lengths are large, and we may prefer to estimate f_v non-parametrically.

(d) We may want to obtain separate estimates of θ_o for each of several frequency bands and then see whether the estimates stay stable over these bands. Such an exercise would be quite useful in model building.

We should mention, however, that for small samples, time domain estimators, if applicable, may have smaller mean square errors than their frequency domain equivalents.

Finally, we give a brief outline of the remaining sections of this chapter. §3 deals with the Strong Law of Large Numbers, §4 with the Central Limit Theorem, §5 considers the choice of an optimal Φ, while §6 deals with the G.N. iteration. All proofs are placed in §7. There are two appendices attached to this chapter. The first gives a more rigorous derivation of (2.12) from (2.10), while the second gives the model and conditions of Robinson [1972b,1975].

3. STRONG CONSISTENCY

For the model (1.10) we assume that:

A1. $(v(n), n = 0, \pm 1, \ldots)$ is a p dimensional, regular, stationary sequence having zero mean and finite second moments. For any integer j, $E\{v(n)v(n+j)'\} = \int_{-\pi}^{\pi} e^{ij\lambda} f_v(\lambda)d\lambda$

with $f_v(\lambda)$ continuous on $[-\pi,\pi]$. Furthermore, the best linear predictor of $v(n)$ is the best predictor.[3]

A2.(i) $\sum_{j=o}^{\infty} j\|D_j(\theta_o)\| < \infty$.

(ii) $\{x(n), n = 0,\pm 1,\ldots\}$ is a q dimensional sequence, possibly stochastic, of real exogenous variables such that for any integer $j \geq 0$

$$\lim_N N^{-1} \sum_{n=1}^{N-j} x(n)x(n+j)' \overset{a.s.}{=} \Gamma_x(j)$$

with $\Gamma_x(j)$ a finite constant matrix. We also assume $\lim_N N^{-\frac{1}{2}}\|x(-N)\| \overset{a.s.}{=} 0$. If the $\{x(n)\}$ sequence is stochastic we will also assume that it is independent of the $\{v(n)\}$ sequence.

(iii) $\forall\, j, \Gamma_x(j) = \int_{-\pi}^{\pi} e^{ij\lambda} f_x(\lambda)d\lambda$, with $f_x(\lambda)$ continuous on $[-\pi,\pi]$ and nonsingular on an open subset of $\mathcal{B}$, $\mathcal{B}_1$ say, having Lebesgue measure greater than zero.

With $D(z;\theta)$ defined as in §1 we will assume that it can be written in the form $D(z;\theta) = B(z;\theta)^{-1}C(z;\theta)$ where the assumptions placed on $B(z;\theta)$ and $C(z;\theta)$ are given below.

[3] See 1.§2.

A3.(i) Θ, the permissible parameter space, is a bounded subset of m dimensional Euclidean space. θ_o, the true parameter point, belongs to Θ.

Let $\bar{\Theta}$ be the closure of Θ.

(ii) For each $\theta \in \bar{\Theta}$, $B(z;\theta)$ is an analytic function of z for $|z| \leq R (R > 1)$ and $C(z;\theta)$ is a regular function of z for $|z| < 1$. In addition $B(z;\theta)$ and $C(z;\theta)$ are continuous in $(z,\theta) \in (z;|z| \leq 1) \times \bar{\Theta}$.

(iii) $\forall \theta \in \bar{\Theta}$, det $B(0;\theta) \neq 0$; $\forall \theta \in \Theta$, $|z| \leq 1$ det $B(z;\theta) \neq 0$.

A4.(i) For $\theta \in \bar{\Theta}$, $h(\theta)$ is an f dimensional $(f < m)$ continuous vector function of θ. $h(\theta_o) = 0$.

Put $\bar{\Theta}_h = \bar{\Theta} \cap (\theta; h(\theta) = 0)$.

(ii) If $\theta \in \bar{\Theta}_h$ and $D(\lambda;\theta) = D(\lambda;\theta_o)$ for λ lying in some open subset of $[-\pi,\pi]$, then $\theta = \theta_o$. □

For convenience, we will often omit to indicate dependence on Φ. We also adopt the convention that if the θ argument is omitted from an expression, then it is assumed to be θ_o. e.g., we will write $L_N(\theta;\Phi)$ as $L_N(\theta)$ and $D_j(\theta_o)$ as D_j.

Theorem 1 Given assumptions A1.-A4., suppose that:

(i) For $\lambda \in [-\pi,\pi]$, $\Phi(\lambda)$ is a $p \times p$ Hermitian p.s.d., continuous, matrix function of λ. It is also nonsingular on $\mathcal{B}_1$ (see assumption A2.(iii)).

(ii) a.s. and $\forall$ $N \geq 1$, $\exists$ $\hat{\theta}_N \in \bar{\Theta}_h$ such that

$$L_N(\hat{\theta}_N) = \inf_{\theta \in \bar{\Theta}_h} L_N(\theta)$$

If for each N we choose one such $\hat{\theta}_N$, then we shall have that $\hat{\theta}_N \to \theta_o$ a.s. □

Remark 3.1 (i) Because Θ is bounded and $h(\cdot)$ is continuous, both $\bar{\Theta}$ and $\bar{\Theta}_h$ are compact sets. Requiring Θ to be bounded is not a severe restriction in practice because we can make the bound arbitrarily large.

(ii) It was necessary to impose condition (ii) of Theorem 1 because $L_N(\theta)$ may be discontinuous on $\bar{\Theta} - \Theta$. We note, however, that, except in very pathological situations, a minimizing $\hat{\theta}_N$ is likely to exist because $L_N(\theta) \geq 0$ $\forall\ \theta \in \bar{\Theta}$. This conjecture is supported by the results of Dunsmuir and Hannan [1976] for their model. Moreover, Wald's [1949] classic proof of the strong consistency of the maximum likelihood estimator is based on the assumption that the maximum likelihood estimator exists.

(iii) The main way in which our assumptions are weaker than Robinson's [1972b] is that we do not require $D(e^{i\lambda};\theta)$ to be continuous on $\bar{\Theta}$. e.g. if $D(e^{i\lambda};\theta) = (1+\theta e^{i\lambda})^{-1}$, then Robinson would require $|\theta| \leq 1-\delta$ for some $\delta > 0$ whereas we can have $\Theta = \{\theta: |\theta| < 1\}$. See also 8.§2. On the other hand, Robinson [1972b,1975] does not require the factorization $D(z;\theta) = B(z;\theta)^{-1}C(z;\theta)$. If $L_N(\theta)$ was continuous on $\bar{\Theta}$, then the existence of a minimizing $\hat{\theta}_N$ follows as in Theorem 3.1 and the strong consistency result is given by Robinson [1972b, Theorem 2.3; 1975, Theorem 1]. (see Appendix 2 for Robinson's assumptions).

4. THE CLT

To prove a CLT we need the following additional assumptions given below.

$\forall\ \alpha \gtrsim 0$ put $U_\alpha = (\theta: \|\theta - \theta_o\| < \alpha)$ and let $\bar{U}_\alpha$ be the closure of U_α.

A5.(i) θ_o is an interior point of Θ.

(ii) $\exists\ \alpha 1 > 0$ s.t. $U_{\alpha 1} \subseteq \Theta$ and $D(\lambda;\theta)$ has elements that are twice differentiable with respect to θ for $\theta \in U_{\alpha 1}$, these second derivatives being continuous functions of $(\lambda,\theta) \in [-\pi,\pi]\times U_{\alpha 1}$.

(iii) $h(\theta)$ has continuous first and second partial derivatives for $\theta \in U_{\alpha 1}$.

Let $\alpha 2 = \frac{1}{2}\alpha 1$. For $\theta \in \bar{U}_{\alpha 2}$ put

$$H(\theta) = (\partial/\partial\theta')h(\theta);\ A(\lambda;\theta)=(\partial/\partial\theta')\{\text{vec } D(\lambda;\theta)\}, \lambda \in[-\pi,\pi];$$

$$L(\theta;\Phi) = \lim_N L_N(\theta;\Phi).$$

By Lemma 1 of §7 the last limit exists a.s., and uniformly in $\theta \in \bar{U}_{\alpha 2}$. Put $\Delta(\theta;\Phi)=(\partial^2/\partial\theta\partial\theta')L(\theta;\Phi)$. It readily follows that

$$\Delta(\Phi) = (2\pi)^{-1}\int_B A^*(\lambda)\{f_x'(\lambda)\otimes\Phi(\lambda)\}A(\lambda)d\lambda.$$

A5.(iv) H is of rank f.

(v) θ_o is a regular point of the matrix $[\Delta(\theta;\Phi)',H(\theta)']$; i.e. the matrix has constant rank in some neighbourhood of θ_o.

A6. There exists a positive constant c s.t. $\|x(n)\|^2 \le c$, $(n = 0,\pm 1,\ldots)$, if the sequence $\{x(n)\}$ is nonstochastic, and $E(\|x(n)\|^2) \le c$, $(n = 0,\pm 1,\ldots)$, if it is stochastic. □

By Theorem 1 and assumption A5., $\hat{\theta}_N \in \bar{U}_{\alpha 2}$ for N sufficiently large (a.s.), and hence there exists a sequence of Lagrange multiplier vectors, $\{\hat{\lambda}_N\}$, such that

$$k_N(\hat{\theta}_N) + H'(\hat{\theta}_N)\hat{\lambda}_N = 0$$
$$h(\hat{\theta}_N) = 0 \qquad (4.10)$$

where $k_N(\theta) = \frac{\partial}{\partial\theta} L_N(\theta)$. We will call (4.10) the <u>Normal Equations</u>.

Let $\hat{\tau}_N' = (\hat{\theta}_N', \hat{\lambda}_N')$, $\tau_o' = (\theta_o', 0')$,

$$J(\Phi) = \begin{pmatrix} \Delta(\Phi) & \vdots & H' \\ \cdots & \vdots & \cdots \\ H & \vdots & 0 \end{pmatrix} \text{ and } \ddagger(\Phi) = J(\Phi)^{-1}\begin{pmatrix} \Delta(\Phi f_v \Phi) & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{pmatrix} J(\Phi)^{-1}$$

Because θ_o is an isolated constrained local minimum of $L(\theta;\Phi)$, it follows from Lemma 9 of 3.§4 that $J(\Phi)$ is nonsingular.

<u>*Theorem 2*</u> Subject to assumptions A1. - A6.

$$\lim_N N^{\frac{1}{2}}(\hat{\tau}_N - \tau_o) \overset{\mathcal{D}}{=} \mathcal{N}[0, \ddagger(\Phi)].\square$$

<u>*Remark 4.1*</u> (i) The main difference between Theorem 2 and Robinson [1972b, Theorem 2.6; 1975, Theorem 2] is that Robinson requires the exogenous variables to be stationary and ergodic with finite fourth moments whereas we do not. (see Appendix 2 conditions R5. and R6.). On the other hand, Robinson places weaker conditions on $D(z;\theta)$ than we do.

(ii) Let $J_N(\theta)$ and $\Delta_N(\theta;\Phi)$ be defined as in §6, and let

$$\ddagger_N(\theta) = J_N(\theta)^{-1}\begin{pmatrix} \Delta_N(\theta;\Phi f_v \Phi) & \cdot & 0 \\ \cdots & & \cdots \\ 0 & \cdot & 0 \end{pmatrix}^{-1} J_N(\theta)^{-1}.$$

Given Φ and f_V, it follows from the proof of Theorem 2 that $\ddagger_N(\hat{\theta}_N)$ is a strongly consistent estimate of $\ddagger(\Phi)$. If either Φ or f_V is unknown we replace the unknown function(s) by a suitably constructed estimate; the modified $\ddagger_N$ will again be strongly consistent. Further discussion of conditions under which the theory continues to hold when either Φ or f_V is replaced by its estimate is given in 8.§6.

5. EFFICIENT ESTIMATION

The asymptotic efficiency of $\hat{\theta}_N$ obviously depends on the choice of Φ. Put

$$J(\Phi)^{-1} = \begin{bmatrix} J^{11}(\Phi) & . & J^{12}(\Phi) \\ . & . & . \\ J^{21}(\Phi) & . & J^{22}(\Phi) \end{bmatrix} .$$

It follows from Theorem 2 that $N^{\frac{1}{2}}(\hat{\theta}_N - \theta_o) \xrightarrow{\mathcal{D}} N[0, \ddagger_\theta(\Phi)]$, where $\ddagger_\theta(\Phi) = J^{11}(\Phi)\Delta(\Phi f_v \Phi)J^{11}(\Phi)$. Subject to assumption A7., given below, Robinson [1972b] shows that f_v^{-1} is an optimal choice of Φ in the sense that

$$\ddagger_\theta(\Phi) \geq \ddagger_\theta(f_v^{-1}) \tag{5.10}$$

The following theorem, using a proof which is statistical in nature and different from Robinson's, gives necessary and sufficient conditions for equality to hold in (5.10). (5.10) is obtained during the proof.

A7. $f_v(\lambda)$ is nonsingular for $\lambda \in \bar{\mathcal{B}}$.

Theorem 3 Subject to assumptions A1.-A7., (5.10) holds. Equality is attained if and only if

$$LA^*(\lambda)\{f_x(\lambda)^{\frac{1}{2}'} \otimes \Phi f_v(\lambda)\} = TLA^*(\lambda)\{f_x^{\frac{1}{2}'}(\lambda) \otimes I_{(p)}\} \; \forall \lambda \in \mathcal{B},$$

where T is some real matrix independent of λ and $L = I_{(m)} - H'(HH')^{-1}H$. $\square$

In the proof of Theorem 3 we show that $\ddagger_\theta(f_v^{-1}) = [L\Delta(f_v^{-1})L]^+$. It follows from this that the optimal choice of $\mathcal{B}$, in a minimum variance sense, is $\mathcal{B} = (-\pi, \pi)$.

6. GAUSS-NEWTON ITERATION

A natural way of obtaining $\hat{\theta}_N$ would be to solve the Normal Equations (4.10). Unfortunately they are likely to be highly non-linear in general. We therefore again resort to G.N. iteration. As expected, the results are similar to those of Chapter 3. Let $I_x(\lambda) = w_x(\lambda)w_x(\lambda)^*$,

$$\Delta_N(\theta;\Phi) = N^{-1}\sum_{B} A^*(\lambda_t;\theta)\{I_x'(\lambda_t)\otimes\Phi(\lambda_t)\}A(\lambda_t;\theta), \qquad (6.05)$$

$$J_N(\theta) = \begin{bmatrix} \Delta_N(\theta;\Phi) & \vdots & H'(\theta) \\ \cdots & \cdots & \cdots \\ H(\theta) & \vdots & 0 \end{bmatrix}, \quad K_N(\theta) = \begin{bmatrix} k_N(\theta) \\ \cdots \\ h(\theta) \end{bmatrix}$$

and define the iteration scheme by

$$\tau_N^{(j+1)} = \begin{bmatrix} \theta_N^{(j)} \\ \cdots \\ 0 \end{bmatrix} - J_N(\theta_N^{(j)})^{-1} K_N(\theta_N^{(j)}) \quad , j \geq 0, \quad (6.10)$$

where $\tau_N^{(j)'} = [\theta_N^{(j)'}, \lambda_N^{(j)'}]$, with $\tau_N^{(o)}$ a given initial estimate.

Theorem 4 In (6.10) put $\tau_N^{(o)'} = (\theta_N^{(o)'}, 0')$. Then:

(i) $\exists\, \alpha 3, 0 < \alpha 3 \leq \alpha 1$, such that in the set $U_{\alpha 3} \times (\lambda: \|\lambda\| \leq \alpha 3)$ there exists, for N sufficiently large and a.s., a unique solution to the Normal Equations, $\hat{\tau}_N$ say.

(ii) If $\lim_N \theta_N^{(o)} \overset{p}{=} \theta_o$, then, with probability tending to 1 as $N \to \infty$, $\lim_j \tau_N^{(j)} = \hat{\tau}_N$.

(iii) If $\lim_N \theta_N^{(o)} \overset{a.s.}{=} \theta_o$, then, for N sufficiently large (a.s.), $\lim_j \tau_N^{(j)} = \hat{\tau}_N$.

(iv) If, in addition, in both (ii) and (iii), $\theta_N^{(o)} = \theta_o + O_p(N^{-\frac{1}{2}})$, then $\lim_N N^{\frac{1}{2}}(\tau_N^{(j)} - \tau_o) \overset{\mathcal{D}}{=} N[0, \ddagger(\Phi)]$ for $j \geq 1$. □

Thus, we again have that for $j \geq 1$ the $\theta_N^{(j)}$ converge to, and have the same asymptotic distribution as, $\hat{\theta}_N$. With $\hat{\ddagger}_N(\theta)$ defined in §4, $\hat{\ddagger}_N(\theta_N^{(j)})$, $j \geq 1$, will be a strongly consistent estimate of $\ddagger(\Phi)$.

If $B(z;\theta)$, $C(z;\theta)$ and $h(\theta)$ are linear in θ, then the requirement that $\theta_N^{(o)}$ be consistent and $\sqrt{N}$-consistent in Theorem 4 can be relaxed. By generalizing a result due to Hannan and Terrell [1973] we show that (iii) and (iv) of Theorem 4 still hold if we just have, respectively, uniformly in $\lambda \in \mathcal{B}$, consistent and $\sqrt{N}$-consistent estimates of $D(e^{i\lambda})$. Because it may be simpler to obtain an estimate of D than to obtain an estimate of θ_o, this may lessen the computational burden of the algorithm. An application of the next theorem is given in 8.§2 and 8.§3.

We now suppose that $\text{vec } B(z;\theta) = F(z)\theta$, $\text{vec } C(z;\theta) = G(z)\theta$ with $F(z)$ and $G(z)$ independent of θ. Let $d(z;\theta) = -\{D(z;\theta)' \otimes I_{(p)}\}F(z) + G(z)$.

It follows easily that $A(z;\theta) = \{I_{(q)} \otimes B(z;\theta)^{-1}\}d(z;\theta)$. Let $\theta_N^{(o)}$ be independent of N (we write it henceforth as $\theta^{(o)}$), lie in Θ, and satisfy $h(\theta^{(o)}) = 0$. Given that $h(\theta)$ is linear in θ, we will have that

$$h(\theta) = H(\theta - \theta^{(o)})$$

with H a constant matrix independent of θ. Let $\hat{B}(\lambda) = B(e^{i\lambda};\ \theta^{(o)})$, $\hat{C}(\lambda) = C(e^{i\lambda};\theta^{(o)})$. The first step of the G.N. iteration (6.10) is modified by putting

$$\hat{\Delta}_N = N^{-1} \sum_{\mathcal{B}} \hat{d}_N^*[I_x' \otimes \hat{B}^{-1*}\Phi\ \hat{B}^{-1}]\hat{d}_N$$

$$\hat{k}_N = -N^{-1} \sum_{\mathcal{B}} \hat{d}_N^*[\bar{w}_x \otimes \hat{B}^{-1*}\Phi](w_y - \hat{B}^{-1}\hat{C}\ w_x)$$

where

$$\hat{d}_N = -[\hat{D}_N' \otimes I_{(p)}]F + G;$$

for convenience the λ argument is omitted in the above summations. This will often be done in the sequel. We note that $\hat{D}_N \neq \hat{B}^{-1}\hat{C}$.

Theorem 5 Suppose that the first step of the G.N. iteration is modified as described above and that $B(z;\theta)$, $C(z;\theta)$ and $h(\theta)$ are linear in $\theta; \theta^{(o)}$ is independent of N, belongs to Θ, and satisfies $h(\theta^{(o)}) = 0$. Then:

(i) If $\limsup_N \sup_{\lambda \in \mathcal{B}} \|\hat{D}_N(\lambda) - D(\lambda)\| = 0$ a.s. (p), then $\lim_N \theta_N^{(1)} = \theta_o$ a.s. (p).

(ii) If, in addition, $\sup_{\lambda \in \mathcal{B}} N^{\frac{1}{2}} \|\hat{D}_N(\lambda) - D(\lambda)\| = O_p(1)$, then $N^{\frac{1}{2}}(\theta_N^{(1)} - \theta_o) = O_p(1)$. □

Once $\theta_N^{(1)}$ is obtained we proceed with the G.N. iteration as usual. From Theorem 4 we know that the $\theta_N^{(j)}$ $(j \geq 2)$ will have the properties described in Theorem 4. Because $\theta^{(o)}$ can be any $\theta \in \Theta$ that satisfies the linear restrictions, this gives us great flexibility in the choice of $\theta^{(o)}$ to start the iteration off.

7. PROOFS

Throughout this section c is a universal constant, not always the same. Unless stated otherwise it should be assumed that all convergence results below mean a.s. convergence and the mode of convergence will not be explicitly stated; for nonstochastic functions we will just have convergence in the ordinary sense.

In the proof of Theorem 1 we will need the result of the following lemma due to Hannan and Robinson [1973, pp. 256-257].

Lemma 1 Let $\{u_j(n);\ n \geq 1;\ j = 1,2\}$ be two real sequences and define

$$c_{jk}(n) = N^{-1} \sum_{m=1}^{N-n} u_j(m) u_k(m+n) = c_{kj}(-n),\quad 0 \leq n < N;$$

$j,k = 1,2.$

$w_k(s) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} u_k(n) e^{in\lambda_s}$, $I_{jk}(s) = w_j(s)\overline{w_k(s)}$, $j,k = 1,2,$

$\lambda_s = 2\pi s N^{-1}$ and $-\frac{1}{2}N < s \leq [\frac{1}{2}N]$.

Put $g_{jk}(N) = N^{-1} \sum_C I_{jk}(s)\ \psi\ (\lambda_s)$; $C \subseteq [-\pi,\pi]$

and is a union of a finite number of disjoint open intervals.

If for each n, $\lim_N c_{jk}(n) \overset{a.s.}{=} \gamma_{jk}(n)$, $j,k = 1,2$ and $\psi(\lambda)$ is a complex valued continuous function of λ for $\lambda \in C$, then

$$\lim_N g_{jk}(N) = (2\pi)^{-1} \int_C \psi(\lambda) dF_{jk}(\lambda) \tag{7.10}$$

where $F_{jk}(\lambda)$ is defined by $\gamma_{jk}(n) = \int_{-\pi}^{\pi} e^{in\lambda}\, dF_{jk}(\lambda)$.

If, in addition, ψ is also a function of a vector parameter τ, lying in a compact set, T say, of some finite dimensional Euclidean space, so that $\psi(\lambda;\tau)$ is continuous in $(\lambda,\tau) \in \overline{C} \times T$, then the limit in (7.10) is uniform in $\tau \in T$ ($\overline{C}$ is the closure of C). □

Proof of Theorem 1 The proof utilizes techniques developed in Dunsmuir and Hannan [1976]. Let $z(n) = \Sigma_j D_j\ x(n-j)$. We know from Hannan [1973a, Theorem 1] that $\forall j \geq 0$

$$\lim_N N^{-1} \sum_{m=1}^{N-j} x(n)v(n+j) = 0.$$

It now follows from Lemma 1 of 4.§7 that

$$\lim_N N^{-1} \sum_{n=1}^{N-j} z(n)z(n+j)' = \int_{-\pi}^{\pi} e^{ij\lambda} D(e^{i\lambda}) f_x(\lambda) D(e^{i\lambda})^* d\lambda,$$

$$\lim_N N^{-1} \sum_{n=1}^{N-j} y(n)x(n+j)' = \int_{-\pi}^{\pi} e^{ij\lambda} D(e^{i\lambda}) f_x(\lambda) d\lambda$$

and

$$\lim_N N^{-1} \sum_{n=1}^{N-j} y(n)y(n+j)' = \int_{-\pi}^{\pi} e^{ij\lambda} D(e^{i\lambda}) f_x(\lambda) D(e^{i\lambda})^* d\lambda + \\ + \int_{-\pi}^{\pi} f_v(\lambda) d\lambda.$$

Applying Lemma 1 we have

$$\limsup_N L_N(\hat{\theta}_N) \leq \lim_N L_N(\theta_o) = (2\pi)^{-1} \int_B \operatorname{tr}\{\Phi(\lambda) f_v(\lambda)\} d\lambda \tag{7.20}$$

Let $\eta > 0$ and define $b(z;\theta) = \det B(z;\theta)$,
$F(z;\theta) = [\operatorname{adj} B(z;\theta)]C(z;\theta)$, $\Phi_\eta(\lambda;\theta) = \Phi(\lambda)\{|b(e^{i\lambda};\theta)|^2 + \eta\}^{-1}$,
$G(\lambda;\theta) = b(e^{i\lambda};\theta)D(e^{i\lambda}) - F(e^{i\lambda};\theta)$,
$g_N(\lambda;\theta) = b(e^{i\lambda};\theta)w_y(\lambda) - F(e^{i\lambda};\theta)w_x(\lambda)$ and

$$L_{N;\eta}(\theta) = N^{-1} \operatorname{tr} \sum_B \Phi_\eta(\lambda_t;\theta) g_N(\lambda_t;\theta) g_N(\lambda_t;\theta)^*.$$

If $\hat{\theta}_N \to \theta_o$, then the proof is complete. If it does not, then there exists a $\dot{\theta} \in \bar{\Theta}_h - \theta_o$ and a subsequence $\{\hat{\theta}_{m(n)}, n \geq 1\}$ of $\{\hat{\theta}_n\}$ such that $\hat{\theta}_{m(n)} \to \dot{\theta}$. Using Lemma 1 we will have that

$$\liminf_n L_{m(n)}(\hat{\theta}_{m(n)}) \geq \liminf_n L_{m(n);\eta}(\hat{\theta}_{m(n)}) = \\ = (2\pi)^{-1} \operatorname{tr} \int_B \Phi_\eta(\lambda;\dot{\theta}) G(\lambda;\dot{\theta}) f_x(\lambda) G(\lambda;\dot{\theta})^* d\lambda + \\ + (2\pi)^{-1} \operatorname{tr} \int_B \Phi_\eta(\lambda;\dot{\theta}) |b(e^{i\lambda};\dot{\theta})|^2 f_v(\lambda) d\lambda . \tag{7.21}$$

Because $b(0;\dot{\theta}) \neq 0$ and $b(z;\dot{\theta})$ is regular in $|z|<R$, $R>1$,

$b(z;\dot\theta)$ can have at most a finite number of zeros on $|z| = 1$.

Hence,

$$\sup_{\eta>0} (2\pi)^{-1}\mathrm{tr}\int_{\mathcal{B}} \Phi_\eta(\lambda;\dot\theta)|b(e^{i\lambda};\dot\theta)|^2 f_v(\lambda)d\lambda = (2\pi)^{-1}\mathrm{tr}\int_{\mathcal{B}} \Phi(\lambda)f_v(\lambda)d\lambda \tag{7.22}$$

It now follows from (7.20) - (7.22) that

$$(2\pi)^{-1}\mathrm{tr}\int_{\mathcal{B}} \Phi_\eta(\lambda;\dot\theta)G(\lambda;\dot\theta)f_x(\lambda)G(\lambda;\dot\theta)^*d\lambda = 0$$

$\Rightarrow \Phi(\lambda)\{b(e^{i\lambda};\dot\theta)D(e^{i\lambda})-F(e^{i\lambda};\dot\theta)\}f_x(\lambda) = 0 \;\forall\; \lambda \in \mathcal{B}$, and hence that

$$b(e^{i\lambda};\dot\theta)D(e^{i\lambda})-F(e^{i\lambda};\dot\theta) = 0 \quad , \forall\; \lambda \in \mathcal{B}_1,$$

by assumption A2.(iii). Therefore $\dot\theta = \theta_o$ by assumption A4.(ii), giving us a contradiction. Hence $\hat\theta_N \to \theta_o$. □

To prove Theorem 2 we need the results of the following lemmas. The next lemma is due to Hannan [1973b, Theorem 1 and pp.168-169] when $\mathcal{B} = [-\pi,\pi]$. The modification for $\mathcal{B}$ a strict subset of $[-\pi,\pi]$ is due to Robinson [1972b, Theorem 2.4].

Lemma 2 Let $T(\lambda)$ be a pq columned matrix that is continuous in λ over $\mathcal{B}$ and such that $T(-\lambda) = \overline{T}(\lambda)$. If assumptions A1. and A2. hold, then

$$\lim_N N^{-\frac{1}{2}}\sum_{\mathcal{B}} T(\lambda_t)\{\overline{w_x}(\lambda_t \otimes w_v(\lambda_t)\} \overset{\mathcal{D}}{=} N(0,\ddagger),$$

with $\ddagger = (2\pi)^{-1} \int_{\mathcal{B}} T(\lambda)\{f_x(\lambda)' \otimes f_v(\lambda)\}T(\lambda)^*d\lambda.$ □

Let $N^{-\frac{1}{2}}\zeta_N(\lambda) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} \{ \sum_{j=o}^{\infty} D_j x(n-j)\}e^{in\lambda} - D(e^{i\lambda})w_x(\lambda)$

Lemma 3 Subject to the assumptions of Theorem 2 ,

$$N^{-1}\sum_{\mathcal{B}} A(\lambda_t)^*[I_{(q)} \otimes \Phi(\lambda_t)]\mathrm{vec}\{\zeta_N(\lambda_t)w_x(\lambda_t)^*\} = o_p(1)$$

Proof Let $\phi(\lambda) = A(\lambda)^*[I_{(q)} \otimes \Phi(\lambda)]$ and let

$$\begin{aligned}\tilde\phi(\lambda) &= \phi(\lambda), \quad \lambda \in \mathcal{B}\\ &= 0 \quad , \lambda \in c\,\mathcal{B} \cap [-\pi,\pi].\end{aligned}$$

Then $\tilde{\phi}$ is discontinuous at a finite number of points at most, namely the end points of the intervals making up $\mathcal{B}$. We now employ an argument similar to Hannan and Robinson [1973, p.257]. For each positive integer m let C_m be a finite union of open intervals of total length m^{-1} such that C_m covers all points of discontinuity of $\tilde{\phi}$; let ψ_m be a continuous periodic function, of period 2π, which is equal to $\tilde{\phi}$ on $[-\pi,\pi] \cap c\ C_m$ and is bounded, uniformly in m, on $[-\pi,\pi]$.

$$N^{-1} \sum_{\mathcal{B}} \phi \text{ vec } \zeta_N w_x^* = N^{-1}\sum_t \psi_m \text{ vec } \zeta_N w_x^* + N^{-1}\sum_t (\tilde{\phi}-\psi_m)\text{vec } \zeta_N w_x^* \quad (7.30).$$

The first term on the right is $o_p(1)$ by Corollary 1 of 4.§7 We now look at the second term. Let $\rho_m = \tilde{\phi} - \psi_m$.

$$\|N^{-1} \sum_t \rho_m \text{ vec } \zeta_N w_x^*\|^2 \leq c\{N^{-1}\sum_t \|\rho_m\|^2 \|w_x\|^2\}\{N^{-1}\sum_t \|\zeta_N\|^2\}.$$

$$N^{-1} \sum_t \|\zeta_N\|^2 = O_p(1) \quad \text{by Lemma 9 of 4.§7}.$$

It follows that we can choose the sequence of functions $\{\psi_m\}$ in such a way that $\forall\ \varepsilon > 0\ \exists\ M = M(\varepsilon)$ s.t.

$$\lim\sup_N P(\|N^{-1} \sum_t \rho_m \text{ vec } \zeta_N w_x^*\| > \varepsilon) < \varepsilon \quad \text{for } m \geq M.$$

The result of the lemma now follows. □

Lemma 4 Subject to the conditions of Theorem 2

$$\lim_N N^{\frac{1}{2}} k_N(\theta_o) \overset{\mathcal{D}}{=} \mathcal{N}[0,\Delta(\Phi f_v \Phi)]$$

Proof $-N^{\frac{1}{2}}k_N(\theta_o) = N^{-\frac{1}{2}}\sum_{\mathcal{B}} A(\lambda_t)^*(\bar{w}_x \otimes \Phi)\{w_y - D(e^{i\lambda_t})w_x\} =$

$$= N^{-\frac{1}{2}}\sum_{\mathcal{B}} A(\lambda_t)^*(I_{(q)} \otimes \Phi)\text{vec}\{w_v w_x^*\} + N^{-1}\sum_{\mathcal{B}} A(\lambda_t)^*(I_{(q)} \otimes \Phi)\text{vec}\{\zeta_N w_x^*\}$$

By Lemma 2 the first term tends to $\mathcal{N}[0,\Delta(\Phi f_v \Phi)]$; the second term is $o_p(1)$ by Lemma 3. The result now follows. □

Let $A_j(\lambda;\theta)$ be the j^{th} column of $A(\lambda;\theta)$, $(1 \le j \le m)$.

Lemma 5 Suppose assumptions A1.-A6. hold. If $\theta_N^\dagger \to \theta_o$, then

$$\lim_N \frac{\partial k_N(\theta_N^\dagger)}{\partial \theta_j} = (2\pi)^{-1} \int_B A^*(\lambda)[f_x'(\lambda) \otimes \Phi(\lambda)]A_j(\lambda)d\lambda$$

Proof $k_N(\theta) = -N^{-1}\sum_B A^*(\lambda_t;\theta)\{I_{(q)} \otimes \Phi(\lambda_t)\}(\bar{w}_x(\lambda_t) \otimes I_{(p)})$
$\times \{w_y - D(e^{i\lambda_t};\theta)w_x\}$

$$\frac{\partial k_N(\theta)}{\partial \theta_j} = -N^{-1}\sum_B \frac{\partial A^*}{\partial \theta_j}(\lambda_t;\theta)\{I_{(q)}\otimes\Phi(\lambda_t)\}\mathrm{vec}[\{w_y - D(e^{i\lambda_t};\theta)w_x\}w_x^*] +$$

$$+ N^{-1}\sum_B A^*(\lambda_t;\theta)\{I_{(q)}\otimes\Phi(\lambda_t)\}\{I_x'(\lambda_t)\otimes I_{(p)}\}A_j(\lambda_t;\theta)$$

$$= II_N(\theta) + III_N(\theta), \text{ say.}$$

By Lemma 1,

$$II_N(\theta) \to (2\pi)^{-1} \int_B \frac{\partial A^*}{\partial \theta_j}(\lambda;\theta)\{I_{(q)}\otimes\Phi(\lambda)\}\mathrm{vec}[\{D(e^{i\lambda}) - D(e^{i\lambda};\theta)\} \times$$
$$\times f_x(\lambda)]d\lambda,$$

$$III_N(\theta) \to (2\pi)^{-1}\int_B A^*(\lambda;\theta)\{I_{(q)}\otimes\Phi(\lambda)\}A_j(\lambda;\theta)d\lambda,$$

uniformly in $\theta \in U_{\alpha 2}$. The required result follows. □

Proof of Theorem 2 The proof is virtually the same as that of Theorem 3.2. As in the proof of Theorem 3.2 we will have that

$$\begin{bmatrix} \dfrac{\partial k_N^\dagger(\theta_N^\dagger)}{\partial \theta'} & \vdots & H'(\hat{\theta}_N) \\ \cdots\cdots & \vdots & \cdots\cdots \\ H^{\dagger\dagger}(\theta_N^{\dagger\dagger}) & \vdots & 0 \end{bmatrix} \times N^{\frac{1}{2}}(\hat{\tau}_N - \tau_o) = - \begin{bmatrix} N^{\frac{1}{2}}k_N(\theta_o) \\ \cdots\cdots \\ 0 \end{bmatrix}$$

$\dfrac{\partial k_N^\dagger(\theta_N^\dagger)}{\partial \theta'}$ is a symbolic derivative and indicates that each row

$\dfrac{\partial k_N^\dagger(\theta_N^\dagger)}{\partial \theta'}$ is to be evaluated at a possibly different value $\theta_N^\dagger$

such that $\|\theta_N^\dagger - \theta_o\| \le \|\hat{\theta}_N - \theta_o\|$. A similar explanation holds for $H^{\dagger\dagger}(\theta_N^{\dagger\dagger})$. The matrix above on the left tends to $J(\Phi)$ by

Lemma 5, while

$$N^{\frac{1}{2}}k_N(\theta_o) \xrightarrow{\mathcal{D}} N[0,\Delta(\Phi f_v\Phi)]$$

by Lemma 4. The result of the theorem follows. □

Proof of Theorem 3 Let L be defined as in §5. It follows from Lemma 3.5 that

$$J^{11}(\Phi) = [L\Delta(\Phi)L]^+$$

so that

$$\ddagger_\theta(\Phi) = [L\Delta(\Phi)L]^+\Delta(\Phi f_v\Phi)[L\Delta(\Phi)L]^+$$

and

$$\ddagger_\theta(f_v^{-1}) = [L\Delta(f_v^{-1})L]^+ \tag{7.40}$$

It is now not difficult to show that

$$\ddagger_\theta(\Phi) \geq \ddagger_\theta(f_v^{-1}) \Longleftrightarrow L\Delta(\Phi f_v\Phi)L \geq L\Delta(\Phi)L(L\Delta(f_v^{-1})L)^+L\Delta(\Phi)L$$

and

$$\ddagger_\theta(\Phi) = \ddagger_\theta(f_v^{-1}) \Longleftrightarrow L\Delta(\Phi f_v\Phi)L = L\Delta(\Phi)L(L\Delta(f_v^{-1})L)^+L\Delta(\Phi)L$$

To complete the proof we apply Lemma 9.3. Let $\chi(\lambda)$, $\lambda \in [\pi,\pi]$, be a p dimensional, zero mean, process with orthogonal increments[4] s.t.

$$E\{\chi(d\lambda)\chi(d\mu)^*\} = I_{(p)}d\lambda, \quad \lambda = \mu$$
$$= 0 \qquad \lambda \neq \mu; \ [\lambda,\mu]\in[-\pi,\pi]$$

and define

$$Z_1 = (2\pi)^{-\frac{1}{2}}L\int_B A^*(\lambda)(f_x'^{\frac{1}{2}}\otimes \Phi f_v^{\frac{1}{2}})\chi(d\lambda),$$

$$Z_2 = (2\pi)^{-\frac{1}{2}}L\int_B A^*(\lambda)(f_x'^{\frac{1}{2}}\otimes f_v^{-\frac{1}{2}})\chi(d\lambda)$$

It follows that

$$E(Z_1Z_1^*) = L\Delta(\Phi f_v\Phi)L, \quad E(Z_1Z_2^*) = L\Delta(\Phi)L,$$
$$E(Z_2Z_2^*) = L\Delta(f_v^{-1})L.$$

By applying Lemma 3 of 9.§5 we obtain the result of the lemma. □

Proof of Theorem 4 The proof is very similar to the proof of Theorem 3.4. It is clearly sufficient to prove the result for j = 1, the proof for j > 1 following similarly. Let $\tau' = (\theta',\lambda')$ and let

$$F_N(\tau) = \begin{pmatrix}\theta\\ \cdots\\ 0\end{pmatrix} - J_N(\theta)^{-1}\begin{pmatrix}k_N(\theta)\\ \cdots\\ h(\theta)\end{pmatrix}$$

(a) We know that

$$k_N(\theta) \to (2\pi)^{-1}\int_B A^*(\lambda;\theta)(f_x'\otimes\Phi)\mathrm{vec}\{D(e^{i\lambda};\theta)-D(e^{i\lambda})\}d\lambda$$

uniformly in $\theta \in \overline{U}_{\alpha 2}$ by Lemma 1.

Hence $k_N(\theta_o) \to 0$ and for $\alpha(>0)$ sufficiently small $\limsup_N \sup_{\theta\in U_\alpha} \|k_N(\theta)\| < \frac{1}{4}$.

(b) It follows from the proof of Lemma 5 that $\forall\ \varepsilon > 0,\ \exists\ \alpha,\ 0 < \alpha \leq \alpha 1$, s.t.

$$\limsup_N \sup_{\theta\in U_\alpha} \|\Delta_N(\theta) - \Delta(\theta)\| < \varepsilon\ ;$$

[4] See 1.§2.

$$\lim\sup_N \sup_{\theta\in U_\alpha} \left\| \frac{\partial k_N(\theta)}{\partial\theta'} - \Delta_N(\theta) \right\| < \varepsilon$$

and

$$\lim\inf_N \inf_{\theta\in U_\alpha} \lambda_{min}\{J_N(\theta)'J_N(\theta)\} \geq c > 0.$$

From (a) and (b) it can be readily shown that $\exists\, \alpha$, $0 < \alpha \leq \alpha 1$, s.t. $\lim\sup_N \sup_{\theta\in U_\alpha} \left\| \frac{\partial F_N(\tau)}{\partial\theta'} \right\| < \frac{1}{2}$. Since $\frac{\partial F_N(\tau)}{\partial\lambda'} = 0$ and $F_N(\tau_o) \to \tau_o$, because $k_N(\theta_o) \to 0$ by the above and $h(\theta_o) = 0$, (i), (ii) and (iii) of Theorem 4 follow from Corollary 3.2. To prove (iv) we note that

$$N^{\frac{1}{2}} \begin{pmatrix} \theta_N^{(1)} - \theta_o \\ \cdots\cdots \\ \lambda_N^{(1)} \end{pmatrix} = \left[\begin{pmatrix} I_{(m)} \\ \cdots\cdots \\ 0 \end{pmatrix} - J_N(\theta_N^{(o)})^{-1} \begin{pmatrix} \frac{\partial k_N^{\dagger}(\theta_N^{\dagger})}{\partial\theta'} \\ \cdots\cdots \\ H^{\dagger\dagger}(\theta_N^{\dagger\dagger}) \end{pmatrix} \right] \cdot N^{\frac{1}{2}}(\theta_N^{(o)} - \theta_o) +$$

$$+ J_N(\theta_N^{(o)})^{-1} \begin{pmatrix} N^{\frac{1}{2}} k_N(\theta_o) \\ \cdots\cdots \\ 0 \end{pmatrix},$$

where $\frac{\partial k_N^{\dagger}(\theta_N^{\dagger})}{\partial\theta'}$, $H^{\dagger\dagger}(\theta_N^{\dagger\dagger})$ are symbolic derivatives having the same meaning as in the proof of Theorem 2. The result (iv) of the theorem now follows because the matrix in $[\cdot]$ above is $o_p(1)$, $N^{\frac{1}{2}}(\theta_N^{(o)} - \theta_o) = O_p(1)$, and $N^{\frac{1}{2}} k_N(\theta_o) \xrightarrow{\mathcal{D}} N[0, \Delta(\Phi f_v \Phi)]$ by Lemma 4. □

In the proof of Theorem 5 we need the following lemma whose proof is an adaptation of Hannan [1971b, Lemma p.776].

Lemma 6 Let $T(\lambda)$ be a pq columned matrix function of λ. Subject to assumptions A1., A2. and A6.

$$N^{-\frac{1}{2}}\sum_{B} T(\lambda_t)\text{vec } w_v(\lambda_t)w_x(\lambda_t)^* = \sup_{\lambda\in B} \|T(\lambda)\|^2 . o_p(1)$$

Proof Let $Q(\lambda_t) = T(\lambda_t)\{\overline{w_x}(\lambda_t)\otimes I_{(p)}\}$ and let $E_x(.)$ mean expectation conditional on the $\{x(n)\}$. Without loss of generality we can take $B = [-\pi,\pi]$; if it is not, then we can always redefine T so that $\sum_B T \text{ vec } w_v w_x^* = \sum_t T \text{ vec } w_v w_x^*$.

$$\|N^{-\frac{1}{2}}\sum_t Q(\lambda_t)w_v(\lambda_t)\|^2 = cN^{-2}\text{tr}\sum_s\sum_t\sum_{m=1}^{N}\sum_{n=1}^{N} Q(\lambda_s)v(m)v(n)'Q(\lambda_t)^* . \exp\{im\lambda_s - in\lambda_t\}$$

Hence

$$E_x\{\|N^{-\frac{1}{2}}\sum_t Q(\lambda_t)w_v(\lambda_t)\|^2\} = cN^{-2}\text{tr}\sum_s\sum_t\sum_m\sum_n Q(\lambda_s)\{\int_{-\pi}^{\pi} e^{i(n-m)\lambda}f_v(\lambda)d\lambda\}\times$$

$$\times Q(\lambda_t)^*.\exp\{im\lambda_s - in\lambda_t\} \leq$$

$$\leq cN^{-2}\int_{-\pi}^{\pi}\|\sum_s\sum_m Q(\lambda_s)e^{im(\lambda_s-\lambda)}\|^2\|f_v(\lambda)\|d\lambda \leq$$

$$\leq cN^{-2}\int_{-\pi}^{\pi}\|\sum_s\sum_m Q(\lambda_s)e^{im(\lambda_s-\lambda)}\|^2 d\lambda$$

$$= cN^{-2}\text{ tr }\sum_{s,t}\sum_m Q(\lambda_s)Q(\lambda_t)^*.\exp\{im\lambda_s - im\lambda_t\}$$

$$= cN^{-1}\text{ tr }\sum_s Q(\lambda_s)Q(\lambda_s)^*$$

$$\leq c\sup_\lambda \|T(\lambda)\|^2 . N^{-1}\sum_{n=1}^{N}\|x(n)\|^2$$

Hence

$$E\{\|N^{-\frac{1}{2}}\sum_B T(\lambda_t)\text{ vec } w_v(\lambda_t)w_x(\lambda_t)^*\|^2\} \leq c\sup_\lambda\|T(\lambda)\|^2$$

The proof of the lemma can now be completed by an application of Chebyshev's inequality. □

Proof of Theorem 5 Let $\hat{B}$, $\hat{C}$, $\hat{\Delta}_N$, $\hat{d}_N$, $\hat{D}_N$ and $\hat{k}_N$ be defined as in §6.

$$w_y - \hat{B}^{-1}\hat{C}w_x = -\hat{B}^{-1}(B-\hat{B})\hat{D}_N w_x + \hat{B}^{-1}(C-\hat{C})w_x + \hat{B}^{-1}(\hat{B}-B)(w_y - \hat{D}_N w_x) +$$

$$+ \hat{B}^{-1}B(w_y - Dw_x) =$$

$$= (w_x' \otimes \hat{B}^{-1})\hat{d}_N\{\theta_o - \theta^{(o)}\} + [\hat{B}^{-1}(\hat{B}-B)\{(D-\hat{D}_N)w_x + N^{-\frac{1}{2}}\zeta_N + w_v\} +$$

$$+ \hat{B}^{-1}B\{N^{-\frac{1}{2}}\zeta_N + w_v\}],$$

where ζ_N is defined above Lemma 3. Call the expression in $[\cdot]$ above $g_N(\lambda)$. Then $\hat{k}_N$ is given by

$$\hat{k}_N = -\hat{\Delta}_N(\theta_o - \theta^{(o)}) - N^{-1}\sum_{\mathcal{B}} \hat{d}_N^*(\overline{w_x} \otimes \hat{B}^{-1*}\Phi)g_N$$

$$= -\hat{\Delta}_N(\theta_o - \theta^{(o)}) + q_N, \text{ say}.$$

Therefore $\tau_N^{(1)} = \begin{pmatrix} \theta^{(o)} \\ \cdots\cdots \\ 0 \end{pmatrix} - \begin{pmatrix} \hat{\Delta}_N & H' \\ \cdots & \cdots \\ H & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\hat{\Delta}_N(\theta_o - \theta^{(o)}) + q_N \\ \cdots\cdots \\ -H(\theta_o - \theta^{(o)}) \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} \theta_N^{(1)} - \theta_o \\ \cdots\cdots \\ \lambda_N^{(1)} \end{pmatrix} = -\begin{pmatrix} \hat{\Delta}_N & \vdots & H' \\ \cdots & \vdots & \cdots \\ H & \vdots & 0 \end{pmatrix}^{-1} \begin{pmatrix} q_N \\ \cdots \\ 0 \end{pmatrix} \qquad (7.60)$$

We now show that Theorem 5(i) holds. It is not hard to see that

$$\hat{\Delta}_N \to (2\pi)^{-1}\int_{\mathcal{B}} d(e^{i\lambda})^*(f_x' \otimes \hat{B}^{-1*}\Phi\hat{B}^{-1})d(e^{i\lambda})d\lambda = \hat{\Delta}, \text{ say}.$$

As in Theorem 2, it follows that

$$\begin{pmatrix} \hat{\Delta}_N & \vdots & H' \\ \cdots & \vdots & \cdots \\ H & \vdots & 0 \end{pmatrix}$$

is nonsingular for N sufficiently large.

Next we show that $q_N \to 0$.

Let

$$T_{1N} = N^{-1}\Sigma\hat{d}_N^*(\overline{w_x} \otimes \hat{B}^{-1*}\Phi)\hat{B}^{-1}(\hat{B}-B)(D-\hat{D}_N)w_x$$

$$\|T_{1N}\| \leq c \sup_{\lambda\in\mathcal{B}} \|D - \hat{D}_N\| . \sup_{\lambda\in\mathcal{B}} \|\hat{d}_N\| . \sup_{\lambda\in\mathcal{B}} \|\hat{B} - B\| S_N \qquad (7.70)$$

where $\quad S_N = N^{-1}\sum_{B} \|(I_x' \otimes \hat{B}^{-1*}\Phi\hat{B}^{-1})\| = O(1).$

Since

$$\sup_{\lambda\in B} \|\hat{d}_N\| = O(1) \quad \text{and} \quad \sup_{\lambda\in B} \|\hat{B}-B\| = O(1)$$

it follows that $T_{1N} \to 0$.

Let

$$T_{2N} = N^{-\frac{3}{2}} \sum_{B} \hat{d}_N^*(\overline{w_x} \otimes \hat{B}^{-1*}\Phi)\hat{B}^{-1}(\hat{B}-B)\zeta_N$$

$$\|T_{2N}\|^2 \leq c.\sup_{B}\|\hat{d}_N(\lambda)\|^2.N^{-1}\sum_t\|I_x\|.\ N^{-2}\sum_t\|\zeta_N\|^2 = o(1)$$

by Lemma 9 of 4.§7. It follows from Lemma 1 that

$$N^{-1}\sum_{B} \hat{d}_N^*(\overline{w_x} \otimes \hat{B}^{-1*}\Phi)\hat{B}^{-1}(\hat{B}-B)w_v \to 0$$

because $f_{vx} = 0$. The rest of the terms in q_N may similarly be shown to tend to zero. From (7.60) we have that $\theta_N^{(1)} \to \theta_o$, $\lambda_N^{(1)} \to 0$. This completes the proof of (i) when

$$\sup_{\lambda\in B} \|\hat{D}_N - D\| \xrightarrow{a.s.} 0.$$

The proof of (i) when

$$\sup_{\lambda\in B} \|\hat{D}_N - D\| \xrightarrow{p} 0$$

is very similar and is omitted.

We now show that Theorem 5(ii) holds. Let T_{1N} and T_{2N} be defined as above. It follows from (7.70) that $N^{\frac{1}{2}}T_{1N} = O_p(1)$. We will also have that

$$\|N^{\frac{1}{2}}T_{2N}\|^2 \leq c\cdot\sup_{B}\|\hat{d}_N\|^2.N^{-1}\sum_t\|I_x\|^2.N^{-1}\sum_t\|\zeta_N\|^2 = O_p(1) \text{ by}$$

Lemma 9 of 4.§7. Since

$$N^{-\frac{1}{2}}\sum_{B}\{d-\hat{d}_N\}^*\ (\overline{w_x} \otimes \hat{B}^{-1*}\Phi)\hat{B}^{-1}(\hat{B}-B)w_v = O_p(1) \quad \text{and}$$

$$N^{-\frac{1}{2}}\sum_{B} d^*(\overline{w_x} \otimes \hat{B}^{-1*}\Phi)\hat{B}^{-1}(\hat{B}-B)w_v = O_p(1)$$

by Lemma 6,it follows that

$$N^{-\frac{1}{2}}\sum_{B} \hat{d}_N^*(\overline{w}_x \otimes \hat{B}^{-1*}\Phi)\hat{B}^{-1}(\hat{B}-B)w_v = O_p(1).$$

The rest of the terms in $N^{\frac{1}{2}}q_N$ may similarly be shown to be $O_p(1)$. It follows from (7.60) that $N^{\frac{1}{2}}(\theta_N^{(1)}-\theta_o) = O_p(1)$. □

APPENDIX 1

We now provide a more rigorous justification for the approximation of the log-likelihood by (2.12). Although this more rigorous approach is not needed in obtaining our asymptotic results it seems both interesting and instructive to follow it through, as it captures the essence of many of the asymptotic proofs encountered when working with stationary variables. As we will see, the required proof is obtained by approximating the inverse of the spectral density matrix by a trigonometric polynomial or, viewed in another way, by approximating the stationary process $\{v(n)\}$ by a finite length autoregression. The derivation we give below follows, to a large extent, Dunsmuir and Hannan [1976, pp.350-352].

For a $p\times p$ matrix function $U(\lambda)$, let $\Gamma(n,U) = \int_{-\pi}^{\pi} e^{in\lambda}U(\lambda)d\lambda$, and let $\Gamma_N(U)$ be the $Np\times Np$ block matrix with $\Gamma(n-m,U)$ in the $(m,n)^{th}$ block. If $U_1 \geq U_2$ then $\Gamma_N(U_1) \geq \Gamma_N(U_2)$ in the usual partial ordering of Hermitian matrices.

If $\phi(\lambda)$ is a $p\times p$ matrix that is continuous in $\lambda \in[-\pi,\pi]$ with $\phi(-\pi) = \phi(\pi)$, then it may be approximated arbitrarily well, uniformly in λ, by a matrix of trigonometric polynomials; to do this it suffices to take the Cesaro sum of the Fourier series for $\phi(\lambda)$.

Let $v(n;\theta)=y(n)-\sum_{j=o}^{\infty} D_j(\theta)x(n-j)$, $(1\leq n\leq N)$, $v_m(\theta)' = \{v(1;\theta)',\ldots,v(m;\theta)'\}$, $(1 \leq m \leq N)$. $-2N^{-1} \times$ log-likelihood of $\{y(j),\ 1 \leq j \leq N\}$ based on Gaussian assumptions is

$$\ell_N(\theta) = N^{-1} \log \det \Gamma_N(f) + N^{-1}v_N(\theta)'\Gamma_N^{-1}(f)v_N(\theta) \quad (A1.10)$$

with $f(\lambda)$ standing for $f_v(\lambda)$; we will assume that $f(\lambda)$ is p.d. for $\lambda\in[-\pi,\pi]$ and $f(-\pi) = f(\pi)$. We show below that for fixed $\theta \in \Theta$, Θ defined by assumption A3.,

$$\ell_N(\theta)-N^{-1}\sum_t \log\det 2\pi f(\lambda_t)-N^{-1}\sum_t \mathrm{tr}\, f^{-1}(\lambda_t)V(\lambda_t;\theta)V(\lambda_t;\theta)^* \to 0, \tag{A1.12}$$

as $N \to \infty$. With a little more care, and under slightly more restrictive conditions we could show that the convergence in (A1.12) is uniform in θ. We will not do this, however.

To give the derivation we need the results of the following lemmas.

Lemma A1.1 If A and B are p.d. matrices of the same dimension, $A \leq B$ and $\lambda_{min}(A)$, $\lambda_{min}(B) \geq \delta > 0$, then

$$\lambda_{max}(A^{-1}-B^{-1}) \leq \lambda_{max}(B-A).\delta^{-2}.$$

Proof $x'(A^{-1}-B^{-1})x=(B^{-1}x)'(B-A)(A^{-1}x) \leq$

$$\leq \|B-A\|_S \|B^{-1}x\| \|A^{-1}x\| \leq \|B-A\|_S.\delta^{-2}.\|x\|^2.$$

$$\Rightarrow \|A^{-1}-B^{-1}\|_S = \sup_{\|x\|=1} x'(A^{-1}-B^{-1})x \leq \|B-A\|_S.\delta^{-2}. \quad \square$$

Lemma A1.2 Suppose $P(\lambda)$ is a $p\times p$ Hermitian p.s.d. matrix. Then

$$\lambda_{min}[\Gamma_N(P)] \geq 2\pi \inf_{\lambda\in[-\pi,\pi]} \lambda_{min}[P(\lambda)] \tag{A1.20}$$

$$\|\Gamma_N(P)\|_S \leq 2\pi \sup_{\lambda\in[-\pi,\pi]} \|P(\lambda)\|_S \tag{A1.21}$$

Proof Let $\{s(j),\ 1 \leq j \leq N\}$ be p dimensional vectors and let $s = \mathrm{vec}\{s(1),\dots,s(N)\}$.

$$\lambda_{min}[\Gamma_N(P)] = \min_{\|s\|=1} \int_{-\pi}^{\pi} (\Sigma\, s(j)e^{ij\lambda})^*P(\lambda)(\Sigma s(j)e^{ij\lambda})d\lambda \geq$$

$$\geq 2\pi.\inf_{\lambda\in[-\pi,\pi]} \lambda_{min}[P(\lambda)].$$

We can similarly show that (A1.21) holds. □

We now show (A1.12) holds. We know that $\forall\ \varepsilon > 0$ there exists a matrix of trigonometric polynomials $P(\lambda)$ such that

$$f^{-1} \leq P \leq f^{-1} + \varepsilon I_{(p)} \tag{A1.25}$$,

for all $\lambda \in [\pi,\pi]$. Then P^{-1} may be taken as the spectrum of an autoregressive process

$$\sum_{j=o}^{M} B_j u(n-j) = \xi(n),\ E\{\xi(m)\xi(n)'\} = \delta_{mn} K_P;\ B_o \equiv I_p$$

i.e. $$P^{-1}(\lambda) = (2\pi)^{-1} B(\lambda)^{-1} K_p B(\lambda)^{-1*},\ B(\lambda) = \sum_{j=o}^{M} B_j e^{ij\lambda}.$$

Let A_N be the $Np\times Np$ matrix

$$A_N = \begin{bmatrix} I_{(Mp)} & & & & & O \\ B_M & \cdots & B_o & & & \\ & B_M & \cdots & B_o & & \\ & & \ddots & & \ddots & \\ O & & & B_M & \cdots & B_o \end{bmatrix}$$

Then

$$A_N \Gamma_N(P^{-1}) A_N' = \begin{bmatrix} \Gamma_M(P^{-1}) & & & O \\ & K_P & & \\ & & \ddots & \\ O & & & K_P \end{bmatrix} \tag{A1.30}$$

We first deal with the $\log\det \Gamma_N(f)$ term of (A1.10).

From (A1.25) and (A1.30)

$$N^{-1} \log\det \Gamma_N(f) \geq N^{-1} \log\det \Gamma_N(P^{-1}) = (N-M)N^{-1} \log\det K_P + \frac{1}{N} \log\det \Gamma_M(P^{-1}) =$$

$$= \log\det K_P + o(1)$$

$$\log\det K_P = (2\pi)^{-1}\int_{-\pi}^{\pi}\log\det[2\pi P^{-1}(\lambda)]d\lambda =$$

$$= N^{-1}\sum_t \log\det[2\pi P^{-1}(\lambda_t)] + o(1)$$

$$= N^{-1}\sum_t \log\det\{2\pi f(\lambda_t)\} + O(\varepsilon) + o(1),$$

because $\log\det f \geq \log\det P^{-1} \geq \log\det f - \log\det[I_{(p)}+\varepsilon f]$.

It now follows that

$$N^{-1}\log\det\Gamma_N(f) \geq N^{-1}\sum_t \log\det 2\pi f(\lambda_t) + O(\varepsilon) + o(1).$$

We can similarly show that

$$N^{-1}\log\det\Gamma_N([f^{-1}+\varepsilon I_{(p)}]^{-1}) \leq N^{-1}\sum_t \log\det 2\pi f(\lambda_t) + o(1).$$

Hence

$$N^{-1}[\log\det\Gamma_N(f) - \sum_t \log\det 2\pi f(\lambda_t)] \to 0 \text{ as } N \to \infty. \tag{A1.35}$$

Next we deal with the exponent term of (A1.10). By Lemma A1.2 $\exists\, \delta > 0$ s.t.

$$\lambda_{min}[\Gamma_N(f)] \geq \delta,\ \lambda_{min}[\Gamma_N(P^{-1})] \geq \delta,\ \|\Gamma_N(f)-\Gamma_N(P^{-1})\|_S \leq c\varepsilon = O(\varepsilon),$$

where c is independent of N and ε. We now have by Lemma A1.1 that

$$\|\Gamma_N(P^{-1})^{-1} - \Gamma_N(f)^{-1}\|_S = O(\varepsilon)$$

Hence

$$0 \leq N^{-1}v_N(\theta)'\Gamma_N(P^{-1})^{-1}v_N(\theta) - N^{-1}v_N(\theta)'\Gamma_N(f)^{-1}v_N(\theta) \leq$$

$$\leq c\,\varepsilon\, N^{-1}\sum_{n=1}^{N} v(n;\theta)'v(n;\theta) = O(\varepsilon).$$

Let $$a(n;\theta) = \sum_{j=o}^{M} B_j\, v(n-j;\theta) \quad , n > M,$$

$$= 0 \quad , 1 \leq n \leq M.$$

For convenience we will not indicate dependence on θ.

$$N^{-1}v_N'\Gamma_N(P^{-1})^{-1}v_N = N^{-1}v_M'\Gamma_M(P^{-1})^{-1}v_M + N^{-1}\sum_{n=M+1}^{N} a(n)'K_P^{-1}a(n) =$$

$$= 2\pi N^{-1}\sum_t \operatorname{tr} K_P^{-1} w_a(\lambda_t)w_a(\lambda_t)^* + o(1)$$

$$= 2\pi N^{-1}\sum_t \operatorname{tr} K_P^{-1} B(\lambda_t)w_v(\lambda_t)w_v(\lambda_t)^*B(\lambda_t)^* + o(1)$$

$$= N^{-1}\sum_t \operatorname{tr} P(\lambda_t)V(\lambda_t)V(\lambda_t)^* + o(1)$$

by Theorem 4.8(i),

$$= N^{-1}\sum_t \operatorname{tr} f(\lambda_t)^{-1}V(\lambda_t)V(\lambda_t)^* + O(\varepsilon) + o(1).$$

It now follows that

$$N^{-1}v_N(\theta)'\Gamma_N(f)^{-1}v_N(\theta) - N^{-1}\sum_t \operatorname{tr} f(\lambda_t)^{-1}V(\lambda_t;\theta)V(\lambda_t;\theta)^* \to 0 \qquad \text{(A1.40)}$$

Putting together (A1.35) and (A1.40) we obtain (A1.12). □

APPENDIX 2

Conditions imposed by Robinson [1972b,1975]

Robinson considers the continuous time model

$$y(t)=B_o \int_{-\infty}^{\infty} D(\tau;\theta_o)x(t-\tau)d\tau+v(t),\ -\infty<t<\infty, \tag{A2.10}$$

Here $v(t)$ is a $p\times 1$ residual vector; $y(t)$ is a $p\times 1$ vector of endogenous variables; $x(t)$ is a $q\times 1$ vector of exogenous variables; B_o is a $p\times s$ matrix of real parameters; θ_o is an $m\times 1$ vector of real parameters; $\Gamma(t;\theta)$ is a $s\times q$ matrix of possibly non-linear functions or generalized functions of t and θ. Let $\delta(t)$ be the Dirac delta function defined by

$$\delta(t) = 0,\ t \neq 0;\ \int_{-\infty}^{\infty} \delta(t)dt = 1.$$

If $B_o D(\tau;\theta_o) = \sum_{j=o}^{\infty} D_j(\theta_o)\delta(\tau-j)$, then (A2.10) becomes, for $t = n$, the model we consider in this chapter, i.e.

$$y(n) = \sum_{j=o}^{\infty} D_j(\theta_o)x(n-j) + v(n).$$

Let

$$\tilde{\Gamma}(\lambda;\theta) = \int_{-\infty}^{\infty} e^{i\lambda t}\ \Gamma(t;\theta)dt,\ -\infty < \lambda < \infty, \tag{A2.15}$$

and take $\hat{\theta}_N$, $\hat{B}_N$ to be the values, out of the set of all admissible θ, B, that absolutely minimize

$$N^{-1}\sum_{\mathcal{B}} \|\Phi(\lambda_s)\ \{w_y(\lambda_s)-B\tilde{\Gamma}(\lambda_s;\theta)w_x(\lambda_s)\}\|^2,$$

with $\Phi(\lambda)$, $\mathcal{B}$ defined as in §1.

To obtain strong consistency Robinson imposes the following conditions:

R1. Θ, the set of all admissible θ, is the compact closure of a $(m-f)$ dimensional C^k manifold, $m > f \geq 0$, $k \geq 2$.

R2. The processes $x(t)$, $v(t)$ have null mean vectors, are mutually incoherent, are stationary and ergodic and have absolutely continuous spectra and continuous spectral density matrices $f_x(\lambda)$, $f_v(\lambda)$ respectively, $|\lambda| \leq \pi$.

R3. Uniformly in $-\infty < \lambda < \infty$, $\theta \in \Theta$, $\tilde{\Gamma}(\lambda;\theta)$ exists and is given by (A2.15). Uniformly in $\theta \in \Theta$, $\tilde{\Gamma}(\lambda;\theta)$ is continuous in λ, satisfying a Lipschitz condition of order greater than one half. (see Zygmund [1959, p.43]).

R4. Uniformly in admissible $(\theta,B) \neq (\theta_o,B_o)$

$$\mathrm{tr}\{(2\pi)^{-1}\int_B [B_o\tilde{\Gamma}(\lambda;\theta_o)-B\tilde{\Gamma}(\lambda;\theta)]f_x(\lambda)[B_o\tilde{\Gamma}(\lambda;\theta_o)-B\tilde{\Gamma}(\lambda;\theta)]^*\Phi(\lambda)d\lambda\}>0$$

B may also be subject to some linear restrictions but is otherwise unconstrained. We have omitted an aliasing condition imposed by Robinson as it is of no interest to us; its purpose is to enable inferences to be made about a continuous time process from discrete data.

To obtain a central limit theorem result Robinson imposes the following additional conditions:

R5. $v(t)$ satisfies the strong mixing condition and, uniformly in $k_j \in \{k: k = 1,\ldots,p\}$, $j = 0,1,2,3$,

$$\sum_{n_1=o}^{\infty}\sum_{n_2=o}^{\infty}\sum_{n_3=o}^{\infty} |\kappa_{k_o k_1 k_2 k_3}(0,n_1,n_2,n_3)| < \infty$$

the summand being the absolute value of the fourth cumulant of $v_{k_o}(n)$, $v_{k_1}(n+n_1)$, $v_{k_2}(n+n_2)$ and $v_{k_3}(n+n_3)$.

R6. Uniformly in $-\infty < t_j < \infty$ and $k_j \in \{k:k=1,\ldots,q\}$, $j = 1,\ldots,4$, the cumulant function of $x_{k_1}(t_1)$, $x_{k_2}(t_2)$, $x_{k_3}(t_3)$ and $x_{k_4}(t_4)$ is finite and given by

$$\underset{\Sigma\lambda_j=0}{\iiiint} f_{k_1k_2k_3k_4}(\lambda_1,\lambda_2,\lambda_3,\lambda_4)\exp\{i\ \Sigma\ t_j\lambda_j\}\Pi d\lambda_j$$

where $f_{k_1k_2k_3k_4}$ is continuous over $\Sigma\lambda_j = 0$.

θ is constrained to satisfy the f, possibly non-linear, restrictions $h(\theta)= 0$, $(0 \leq f < m)$, while B satisfies the linear restrictions $L[\text{vec } B] = \text{vec } B$, with L a symmetric idempotent matrix.

R7. Uniformly in $\theta \in \Theta$ the derivatives

$$[\frac{\partial}{\partial\theta'}, \frac{\partial}{\partial(\text{vec } B')}]h(\theta) = H(\theta)$$

exist and are continuous in θ; θ_o is a regular point of H and $\text{rank}\{H(\theta_o)\} = f$.

R8. Uniformly in $-\infty < \lambda < \infty$, $\theta \in \Theta$, the derivatives $(\partial/\partial\theta_k)\tilde{\Gamma}(\lambda;\theta)$, $(\partial^2/\partial\theta_k\partial\theta_\ell)\tilde{\Gamma}(\lambda;\theta)$, $k, \ell = 1,\ldots,m$, exist, being continuous in λ, uniformly in $\theta \in \Theta$, and continuous in θ, uniformly in $\lambda \in \mathcal{B}$; (θ_o, B_o) is a regular point of $[\Psi(\Phi), H']$, where,

$$\Psi(\Phi)=(2\pi)^{-1}\int_{\mathcal{B}} T(\lambda;\theta,B)^*[f_x(\lambda)'\otimes\Phi(\lambda)]T(\lambda;\theta,B)d\lambda,$$

with

$$T(\lambda;\theta,B) = [(\partial/\partial\theta')\text{vec}\{B\tilde{\Gamma}(\lambda;\theta)\},\{\tilde{\Gamma}(\lambda;\theta)'\otimes I_{(p)}\}L].$$

CHAPTER 8

ESTIMATION IN REGRESSION MODELS WITH STATIONARY RESIDUALS: APPLICATIONS AND FURTHER EXTENSIONS

1. INTRODUCTION

This chapter is an extension of the last. A large part of it is devoted to a discussion of the linear model

$$\sum_{j=o}^{r} B_j y(n-j) = \sum_{j=o}^{s} C_j x(n-j) + u(n) \qquad (1.10)$$

where $u(n)$ is a stationary p dimensional residual. We know from previous chapters that if $u(n)$ can be modelled as an autoregressive, MA or ARMA process,then numerous algorithms have been suggested to estimate the unknown system and noise parameters. When $\{u(n)\}$ is a general stationary process whose spectrum is not necessarily rational,Hannan and Terrell [1973] give estimators that are strongly consistent and asymptotically efficient (in the sense of 7.§5) when $r = s = 0$ and the parameters are subject to linear restrictions only; they also allow identities to be present. For a different, but closely related, model Robinson [1976b] obtains consistent and asymptotically efficient instrumental variable estimators; the analogous estimators for the model (1.10) are discussed by Engle [1976] and Espasa and Sargan [1975]. Espasa and Sargan [1975] also suggest a ML type estimator based on the full log-likelihood (7.2.12) rather than on the exponent only; this is discussed further in §9.

We now give a brief outline of the contents of this chapter. §2 applies the results of Chapter 7 to the model (1.10); §3 allows identities to be present. In §4 we discuss the instrumental variable estimator mentioned above and compare it to the G.N. estimator, while §5 looks at the estimation of a subset of the equations in (1.10). §6 shows that the results of Chapter 7 will still hold if Φ is unknown, but an estimate of it, either parametric or non-parametric, is available; an explanation of what these terms mean is given in §6. §7 discusses in detail the construction of the G.N. iterates for the model (1.10) when only the standard exclusion and normalization restrictions are present. Because the restrictions are imposed explicitly rather than through Lagrange multipliers, a considerable saving in computer storage can be made. §8 significantly relaxes the boundedness assumptions previously imposed on the parameters of (1.10) in order that strong consistency hold. Finally, in §9 we consider estimation based on the full log-likelihood (7.2.12) rather than on the exponent only. All proofs are given in §10.

<u>It is important that the reader take note of the following three conventions adopted in this chapter</u>.

1. z will be used both as a lag operator, i.e. $zy(n) = y(n-1)$, and as a complex variable. Its usage will be clear from the context.

2. We will indicate dependence on θ, even for the linear model (1.10), explicitly; if the θ argument is omitted from an expression, then it is assumed to be θ_o, the true parameter point; e.g. we will write $B_j(\theta)$, $B(z;\theta)$ ordinarily but write $B_j(\theta_o)$ and $B(z;\theta_o)$ as B_j and $B(z)$, respectively.

3. For a function of the complex variable z, $g(z)$ say, we will write $g(e^{i\lambda})$ as $g(\lambda)$.

2. THE LINEAR MODEL

In this section we apply the results of Chapter 7 to the model (1.10). We make the following assumptions:

B1. $\{u(n), n = 0,\pm 1,\ldots\}$ is a p dimensional, regular, stationary sequence having zero mean and finite second moments. For any integer j we assume that

$E\{u(n)u(n+j)'\} = \int_{-\pi}^{\pi} e^{ij\lambda} f_u(\lambda)\,d\lambda$, with $f_u(\lambda)$ continuous

for $\lambda \in [-\pi,\pi]$. Furthermore, the best linear predictor of $\{u(n)\}$ is the best predictor.

B2. This consists of assumptions A2.(i) and (ii) of 7.§3.

B3.(i) r and s are given non-negative integers.

Let $\theta = \mathrm{vec}(B_o,\ldots,B_r,C_o,\ldots,C_s)$. The dimension, m, of θ is $(r+1)p^2 + (s+1)pq$. R_1 and ε_1 are two positive numbers chosen before the start of the estimation; R_1 can be taken very large and ε_1 will be chosen close to zero. We define the permissible parameter space Θ as the set of all θ such that:

B3.(ii) $\|\theta\| < R_1$

(iii) $\det\left(\sum_{j=o}^{r} B_j z^j\right) \neq 0$ for $|z| \leq 1$

(iv) $|\det(B_o)| > \varepsilon_1$

(v) $\sum_{j=o}^{r} B_j z^j$ and $\sum_{j=o}^{s} C_j z^j$ are left coprime

(vi) $[B_r, C_s]$ is of rank p.

The true parameter point, θ_o, belongs to Θ. $\overline{\Theta}$ is the closure of Θ.

B4.(i) For $\theta \in \overline{\Theta}$, $h(\theta)$ is an f dimensional, $(f < m)$, continuous vector function of θ. In addition $h(\theta_o) = 0$.

(ii) If $\theta \in \overline{\Theta}$, $\theta = [I_{(t)} \otimes V]\theta_o$ and $h(\theta) = 0$, then $\theta = \theta_o$; V is a $p \times p$ matrix and $t = p(r+1) + q(s+1)$.

B5.(i) $\forall\ \theta \in \overline{\Theta}$, $h(\theta)$ has continuous first and second partial derivatives. Put $H(\theta) = (\partial/\partial\theta')h(\theta)$.

Let the matrix $\Delta(\theta;\Phi)$ be defined as in 7.§4.

(ii) θ_o is a regular point of the matrix $[\Delta(\theta;\Phi)', H(\theta)']$.

B6. This is the same as assumption A6. of 7.§4. □

Put $B(z;\theta) = \sum_{j=o}^{r} B_j(\theta)z^j$; $C(z;\theta) = \sum_{j=o}^{s} C_j(\theta)z^j$; $D(z;\theta) =$

$$= B(z;\theta)^{-1}C(z;\theta).$$

We define the stationary sequence $\{v(n)\}$ constructively by

$$\sum_{j=o}^{r} B_j\, v(n-j) = u(n) \qquad (n = 0,\pm1,\ldots).$$

It follows readily that $\{v(n)\}$ satisfies assumption A1. of 7.§3 and that $f_v(\lambda) = B(\lambda)^{-1}f_u(\lambda)B(\lambda)^{-1*}$. We can rewrite (1.10) as

$$y(n) = D(z)x(n) + v(n)$$

possibly omitting negligible transient terms that decay exponentially to zero. This suggests that we can apply Theorems 7.1 - 7.5 to (1.10). With $D(z;\theta)$ as defined above we minimize (7.1.15) to obtain an estimate of θ_o. We now have the following result:

Theorem 1 Given conditions B1. - B4., Theorem 7.1 holds for the model (1.10). If B5. and B6. are also given, then Theorems 7.2 - 7.5 will also hold. □

Remark 2.1 (i) For practical application condition B2.(ii) will not be a serious restriction because we can make R_1 as large as we like. Due to the limitations of computer storage, any minimization algorithm must, of necessity, also place an upper bound on $\|\theta\|$.

(ii) We would like to replace B3.(iv) by

B3.(iv)' $|\det B_o| > 0$.

Unfortunately our method of proof requires B3.(iv) in order that $|\det B_o(\theta)| > 0 \quad \forall\ \theta \in \overline{\Theta}$. If (1.10) is a reduced form system, i.e. $B_o(\theta) \equiv I_{(p)}$, then B3.(iv) is automatically satisfied if we take $\varepsilon_1 < 1$. We also note that for such a system, with θ redefined as $\text{vec}\{B_1,\ldots,B_r, C_o,\ldots,C_s\}$, B4.(ii) is unnecessary.

(iii) From Chapter 2 we recognize B3.(v) and (vi) as being time series identification conditions, while B4.(ii) is a classical identification requirement. We know from 2.§7 that we could substitute other, essentially equivalent, conditions for B3.(vi). For a further discussion of time series identification conditions for the model (1.10) we refer the reader to Chapter 2, in particular 2.§7.

(iv) We now show how Theorem 7.5 can be applied to the estimation of (1.10). First we note that vec B(z) and vec C(z) are linear in θ. Put $\beta = \text{vec}\{B_o,\ldots,B_r\}$; $\gamma = \text{vec}\{C_o,\ldots,C_s\}$; $\chi_\beta(z)' = (1,z,\ldots,z^r)$, $\chi_\gamma(z)'=(1,z,\ldots,z^s)$. Next we let $\underline{B}_j = B_o^{-1}B_j$, $\underline{C}_k = B_o^{-1}C_k (0 \le j \le r;\ 0 \le k \le s)$. Because

$$y(n) = -\sum_{j=1}^{r} \underline{B}_j\, y(n-j) + \sum_{j=o}^{s} \underline{C}_j\, x(n-j) + B_o^{-1}u(n)$$

we can obtain instrumental variable estimates, $\hat{\underline{B}}_j$ and $\hat{\underline{C}}_k$, of the $\underline{B}_j$ and $\underline{C}_k$ such that $\hat{\underline{B}}_j \to \underline{B}_j$ a.s., $\hat{\underline{C}}_k \to \underline{C}_k$ a.s.,

$$N^{1/2}(\hat{\underline{B}}_j-\underline{B}_j) = O_p(1),\ N^{1/2}(\hat{\underline{C}}_k-\underline{C}_k) = O_p(1).$$

Let $\hat{D}_N(\lambda) = (\Sigma_j \underline{\hat{B}}_j e^{ij\lambda})^{-1}(\Sigma_j \underline{\hat{C}}_j e^{ij\lambda})$, $\lambda \in [-\pi,\pi]$. Because $D(z) = B(z)^{-1}C(z) = (\Sigma_j \underline{B}_j z^j)^{-1}(\Sigma_j \underline{C}_j z^j)$, it follows that

$$\sup_{\lambda\in[-\pi,\pi]} \|\hat{D}_N(\lambda) - D(\lambda)\| \to 0 \text{ a.s.}$$

and

$$\sup_{\lambda\in[-\pi,\pi]} N^{\frac{1}{2}}\|\hat{D}_N(\lambda) - D(\lambda)\| = O_p(1).$$

Thus conditions (i) and (ii) of Theorem 7.5 are satisfied.

(v) Let $\hat{B}_j$ and $\hat{C}_k$ be perturbed values of B_j and C_k and put $\hat{B}(z) = \Sigma_j \hat{B}_j z^j$; $\hat{C}(z) = \Sigma_j \hat{C}_j z^j$. We know from Lemma 5.2 that

$$B(z)^{-1}C(z) = \hat{B}^{-1}(z)\hat{C}(z) - \hat{B}^{-1}(z)\{B(z)-\hat{B}(z)\}\hat{B}^{-1}(z)\hat{C}(z) + $$
$+ \hat{B}(z)^{-1}\{C(z) - \hat{C}(z)\}+$ second order terms in $(B_j-\hat{B}_j)$ and $(C_k-\hat{C}_k)$. Hence,

$$y(n) - B^{-1}(z)C(z)x(n) = y(n)-\hat{B}^{-1}(z)\hat{C}(z)x(n) +$$
$$+ \hat{B}^{-1}(z)\{B(z)-\hat{B}(z)\}\hat{B}^{-1}(z)\hat{C}(z)x(n)-\hat{B}(z)^{-1}\{C(z)-\hat{C}(z)\}x(n)$$
$$+ \mu(n) = v(n) \qquad (2.10)$$

with $\mu(n)$ consisting of second order terms in $(B_j-\hat{B}_j)$ and $(C_k-\hat{C}_k)$.

It is now apparent from (2.10), 5.§2 , and equations (12) and (13) in Hannan and Terrell [1973], that the Hannan and Terrell algorithm (pp.303-309 in Hannan and Terrell [1973]) is just the G.N. algorithm with the first step modified as described in Theorem 7.5 and with $\Phi = \hat{f}_v^{-1}$ (see §6) and $\mathcal{B} = (-\pi,\pi)$. The construction of the Hannan - Terrell estimator differs slightly from the G.N. estimator as given in 7.§6 because Hannan and Terrell average periodograms over bands (see §7 for further details).

(vi) It follows as in Deistler, Dunsmuir and Hannan [1977] that Θ is an open subset of m dimensional Euclidean space.

Hence θ_o is an interior point of Θ and $\bar{\Theta} - \Theta$ is not empty.

For $\theta \in \bar{\Theta} - \Theta$, det $B(z;\theta)$ may be zero on $|z| = 1$, and $B(z;\theta)$ and $C(z;\theta)$ are not necessarily left prime. Hence $L_N(\theta;\Phi)$ may not be continuous on $\bar{\Theta}$. (See Remarks 3.1(i) and (ii) of Chapter 7).

3. EXTENSION TO IDENTITIES

We now consider the linear model (1.10) when identities are present. To clearly indicate that identities are present we rewrite (1.10) as

$$\sum_{j=o}^{r} \begin{pmatrix} B_{11,j} & \cdot & B_{12,j} \\ \cdot & \cdot & \cdot \\ B_{21,j} & \cdot & B_{22,j} \end{pmatrix} \begin{pmatrix} y(n-j) \\ \cdots \\ \tilde{y}(n-j) \end{pmatrix} = \sum_{j=o}^{s} \begin{pmatrix} C_{1,j} \\ \cdots \\ C_{2,j} \end{pmatrix} x(n-j) + \begin{pmatrix} u(n) \\ \cdots \\ 0 \end{pmatrix} \quad (3.10)$$

$B_{11,j}$ is $p\times p$; $B_{12,j}$ is $p\times\tilde{p}$; $B_{21,j}$ is $\tilde{p}\times p$; $B_{22,j}$ is $\tilde{p}\times\tilde{p}$; $C_{1,k}$ is $p\times q$; $C_{2,k}$ is $\tilde{p}\times q$; $B_{21,j}$, $B_{22,j}$, $C_{2,k}$ are <u>known</u> matrices; $0 \leq j \leq r$ and $0 \leq k \leq s$. $y(n)$ is $p\times 1$; $\tilde{y}(n)$ is $\tilde{p}\times 1$; $x(n)$ is $q\times 1$; $u(n)$ is $p\times 1$. Put

$$B_I(z;\theta) = \sum_{j=o}^{r} \begin{pmatrix} B_{11,j}(\theta) & \cdot & B_{12,j}(\theta) \\ \cdot & \cdots & \cdot \\ B_{21,j}(\theta) & \cdot & B_{22,j}(\theta) \end{pmatrix} z^j; C_I(z;\theta) = \sum_{j=o}^{s} \begin{pmatrix} C_{1,j}(\theta) \\ \cdots \\ C_{2,j}(\theta) \end{pmatrix} z^j;$$

$\Gamma_1(z) = \{\Sigma_j B_{22,j} z^j\}^{-1}\{\Sigma_j B_{21,j} z^j\}$, $\Gamma_2(z) = \{\Sigma_j B_{22,j} z^j\}^{-1}\{\Sigma_j C_{2,j} z^j\}$

$B(z;\theta) = \Sigma_j B_{11,j}(\theta) z^j - \{\Sigma_j B_{12,j}(\theta) z^j\}\Gamma_1(z)$;

$C(z;\theta) = \Sigma_j C_{1,j}(\theta) z^j - \{\Sigma_j B_{12,j}(\theta) z^j\}\Gamma_2(z)$;

$D_I(z;\theta) = B_I(z;\theta)^{-1} C_I(z;\theta)$.

It is not hard to show that (3.10) can be rewritten as

$$B(z)y(n) = C(z)x(n) + u(n) \quad (3.15)$$

We make the following assumptions:

C1. and C2. are the same as B1. and B2.

Let $\theta = \mathrm{vec}\{B_{11,o}, B_{12,o}, \ldots, B_{11,r}, B_{12,r}, C_{1,o}, \ldots, C_{1,s}\}$. R_1 and ε_1 are given positive numbers having the same meaning as in §2. We take the permissible parameter space Θ to be the set of all θ such that:

C3.(i) $r,s \geq 0$ are given. (ii) $\|\theta\| < R_1$.

(iii) $\det B_I(z;\theta) \neq 0$ for $|z| \leq 1$; $\det(\Sigma_j B_{22,j} z^j) \neq 0$ for $|z| \leq 1$. (iv) $|\det B_I(0;\theta)| \geq \varepsilon_1$.

(v) $B_I(z;\theta)$ and $C_I(z;\theta)$ are left coprime.

(vi) The following matrix is of rank $(p+\tilde{p})$.

$$\left\| \begin{Bmatrix} B_{11,r}(\theta) & \cdot & B_{12,r}(\theta) \\ \cdot\ \cdot & \cdot & \cdot\ \cdot \\ B_{21,r}(\theta) & \cdot & B_{22,r}(\theta) \end{Bmatrix} , \begin{Bmatrix} C_{1,s}(\theta) \\ \cdot\ \cdot\ \cdot \\ C_{2,s}(\theta) \end{Bmatrix} \right\|$$

Let $\rho = \mathrm{vec}\{B_{21,o}, B_{22,o}, \ldots, B_{21,r}, B_{22,r}, C_{2,o}, \ldots, C_{2,s}\}$.

C4.(i), C5.(i) and (ii) and C6. are the same as B4.(i), B5.(i) and (ii) and B6., respectively.

C4.(ii) If $\theta \in \overline{\Theta}$ and $\theta = [I_{(t)} \otimes V_1]\theta_o + [I_{(t)} \otimes V_2]\rho$ and $h(\theta) = 0$, then $\theta = \theta_o$; $t = (r+1)(p+\tilde{p}) + q(s+1)$; V_1 is a $p \times p$ matrix and V_2 is a $p \times \tilde{p}$ matrix. □

Defining the stationary sequence $\{v(n)\}$ by

$$B(z)v(n) = u(n) \quad (n = 0, \pm 1, \ldots)$$

we will again have that $\{v(n)\}$ satisfies assumption A1. of 7.§3 and that $f_v(\lambda) = \tilde{B}(\lambda)^{-1} f_u(\lambda) B(\lambda)^{-1*}$. Putting $D(z;\theta) = B(z;\theta)^{-1} C(z;\theta)$ we can rewrite (3.15) as

$$y(n) = D(z)x(n) + v(n).$$

This suggests that we can again apply Theorems 7.1 - 7.5 to (3.15). With $D(z;\theta)$ as defined above we will again minimize a criterion function of the form (7.1.15) to obtain estimates of θ_o. Formally, we have the following result:

Theorem 2 If conditions C1. - C4. are given, then the result of Theorem 7.1 holds for the model (3.10). If conditions C5. and C6. are also given, then the results of Theorems 7.2 - 7.5 also hold for (3.10). □

Remark 3.1 (i) Remarks 2.1(i) - (vi) apply equally well here.

(ii) We motivate condition C4.(ii) by showing that it is a classical identification condition. If θ and θ_o are observationally equivalent, then, because θ_o satisfies C3.(v) and (vi), we know from 2.§9 that there exist matrices V_1 and V_2, having dimensions $p\times p$ and $p\times\tilde{p}$, respectively, such that

$$\{B_{11,o}(\theta),B_{12,o}(\theta),\ldots,C_{1,o}(\theta),\ldots,C_{1,s}(\theta)\}=$$

$$=\{V_1,V_2\}\times\left[\begin{Bmatrix} B_{11,o} & \vdots & B_{12,o} \\ \cdot & \cdot & \cdot \\ B_{21,o} & \vdots & B_{22,o}\end{Bmatrix},\ldots,\begin{Bmatrix} C_{1,s} \\ \\ C_{2,s}\end{Bmatrix}\right]$$

Vectorizing both sides yields $\theta=[I_{(t)}\otimes V_1]\theta_o+[I_{(t)}\otimes V_2]\rho$.

(iii) Because $\Sigma_j B_{22,j}z^j$ is known we can check the condition $\det(\Sigma_j B_{22,j}z^j)\neq 0$ for $|z|\leq 1$ prior to estimation.

(iv) Let $\chi_\beta(z)'=(1,z,\ldots,z^r)$, $\chi_\gamma(z)'=(1,z,\ldots,z^s)$. We now show that
$A(z;\theta)=\frac{\partial}{\partial\theta'}\{\mathrm{vec}\,D(z;\theta)\}=\{I_{(q)}\otimes B^{-1}(z;\theta)\}d(z;\theta)$ with

$$d(z;\theta)=[\chi_\beta(z)'\otimes\{-D_I(z;\theta)\}'\otimes I_{(p)},\chi_\gamma(z)'\otimes I_{(q)}] \qquad (3.20)$$

This gives us a representation of $A(z;\theta)$ in terms of $D_I(z;\theta)$ rather than $D(z;\theta)$. The usual representation in terms of $D(z;\theta)$ is given in 7.§6 (see also (3.30) below). Because

$A(z;\theta)$ is needed in the construction of the G.N. algorithm, the representation (3.20) will be of some importance if $D_I(z;\theta)$ is computationally easier to obtain than $D(z;\theta)$. To obtain (3.20) we first note that

$$D_I(z;\theta)' = \{D(z;\theta)', \; - D(z;\theta)'\Gamma_1(z)'+\Gamma_2(z)'\} \tag{3.25}$$

(3.25) can be verified by direct substitution.

Because

$$B(z;\theta) = \{B_{11,o},B_{12,o},\ldots,B_{11,r},B_{12,r}\}\left\{\chi_\beta(z)\otimes \begin{bmatrix} I_{(p)} \\ \cdots \\ -\Gamma_1(z) \end{bmatrix}\right\}$$

and

$$C(z;\theta) = \{B_{11,o},B_{12,o},\ldots,B_{11,r},B_{12,r}\}\left\{\chi_\beta(z)\otimes \begin{bmatrix} 0_{p\times q} \\ \cdots \\ -\Gamma_2(z) \end{bmatrix}\right\} +$$

$$+ \{C_{1,o},\ldots,C_{1,s}\}\{\chi_\gamma(z)\otimes I_{(q)}\}$$

we can write $\operatorname{vec} B(z;\theta) = F(z)\theta$, $\operatorname{vec} C(z;\theta) = G(z)\theta$ with

$$F(z) = [\chi_\beta(z)'\otimes\{I_{(p)},-\Gamma_1(z)'\}\otimes I_{(p)}, \; 0_{p^2\times pq(s+1)}] \tag{3.26}$$

and

$$G(z) = [\chi_\beta(z)'\otimes\{0'_{p\times q}, \; - \Gamma_2(z)'\}\otimes I_{(p)},\chi_\gamma(z)'\otimes I_{(pq)}] \tag{3.27}$$

From 7.§6

$$d(z;\theta)' = -F(z)'\{D(z;\theta)\otimes I_{(p)}\} + G(z)' \tag{3.30}$$

We can now obtain (3.20) from (3.25) - (3.27) by some elementary manipulation.

(v) We showed above that $\operatorname{vec} B(z;\theta)$ and $\operatorname{vec} C(z;\theta)$ are linear in θ, even though $B(z;\theta)$ and $C(z;\theta)$ are not necessarily polynomials in z now. Thus Theorem 7.5 is relevant to the model (3.10).

We can obtain a consistent (a.s.) and $\sqrt{N}$ - consistent estimator of $D_I(z)$ by using instrumental variables as outlined in Remark 2.1(iv). Using (3.20) we can now obtain a consistent and $\sqrt{N}$ - consistent estimator of $d(z)$ which is required when applying Theorem 7.5.

4. COMPARING THE G.N. ESTIMATOR WITH AN INSTRUMENTAL VARIABLE ESTIMATOR

We now go back to model (1.10), i.e. the no identities case, and look at an estimator which is closely related to the G.N. estimator. For convenience of exposition we assume that $B_o \equiv I_{(p)}$ and that there are no constraints on θ. We redefine θ and $\chi_\beta(z)$, respectively, as $\theta = \mathrm{vec}\{B_1,\ldots,B_r\}$ and $\chi_\beta(z)' = (z,z^2,\ldots,z^r)$. It is not hard to show that

$$A(z;\theta) = \frac{\partial}{\partial\theta'}\{\mathrm{vec}\, D(z;\theta)\} = [-\{\chi_\beta(z)'\otimes D(z;\theta)'\otimes B^{-1}(z;\theta)\},\ \{\chi_\gamma(z)'\otimes I_{(q)}\otimes B^{-1}(z;\theta)\}] \qquad (4.10).$$

It follows that $\Delta_N(\theta;\Phi)$, as given in (7.6.05), can be written as

$$\Delta_N(\theta)=N^{-1}\sum_{B}\begin{bmatrix}\{\chi_\beta\chi_\beta^*\otimes DI_x D^*\otimes\phi\} & -\{\chi_\beta\chi_\gamma^*\otimes DI_x\otimes\phi\}\\ -\{\chi_\gamma\chi_\beta^*\otimes I_x D^*\otimes\phi\} & \{\chi_\gamma\chi_\gamma^*\otimes I_x\otimes\phi\}\end{bmatrix} \qquad (4.11)$$

where $\phi(\lambda;\theta) = \{B(\lambda;\theta)^{-1*}\Phi(\lambda)B(\lambda;\theta)^{-1}\}'$

Put

$$\Psi_N(\theta)=N^{-1}\sum_{B}\begin{bmatrix}\{\chi_\beta\chi_\beta\otimes DI_{xy}\otimes\phi\} & -\{\chi_\beta\chi_\gamma^*\otimes DI_x\otimes\phi\}\\ -\{\chi_\gamma\chi_\beta^*\otimes I_{xy}\otimes\phi\} & \{\chi_\gamma\chi_\gamma^*\otimes I_x\otimes\phi\}\end{bmatrix} \qquad (4.12)$$

For convenience the arguments λ and θ are omitted from the summations in (4.11) and (4.12); dependence on Φ is also not shown. With $k_N(\theta)$ as defined in 7.§4, consider the iteration scheme

$$\theta_N^{(j+1)} = \theta_N^{(j)} - \Psi_N(\theta_N^{(j)})^{-1} k_N(\theta_N^{(j)}),\ j \geq 0 \qquad (4.20)$$

If Ψ_N is replaced by Δ_N in (4.20) we obtain the G.N. iteration. Because $\lim_N \Psi_N(\theta_N^{(j)}) = \lim_N \Delta_N(\theta_N^{(j)})$ if $\theta_N^{(j)} \to \theta_o$, it follows that the iterates in (4.20) have the same asymptotic properties as the G.N. iterates. Following Robinson [1976b](see below) we call the sequence of iterates (4.20) iterated instrumental variable estimators (IIVE).

We note that

$$B(\lambda;\theta)w_y - C(\lambda;\theta)w_x = w_y(\lambda) + [\{\chi_\beta'\otimes w_y'\otimes I_{(p)}\}, -\{\chi_\gamma'\otimes w_x'\otimes I_{(p)}\}]\theta$$

so that

$$k_N(\theta) = -N^{-1}\sum_{\mathcal{B}} A^*\{\bar{w}_x\otimes\Phi\}\{w_y - Dw_x\}$$

$$= \Psi_N(\theta)\theta - \rho_N(\theta) \qquad (4.25)$$

using (4.10), where

$$\rho_N(\theta) = N^{-1}\sum_{\mathcal{B}} \begin{pmatrix} -\{\chi_\beta\otimes Dw_x \otimes \phi'\} \\ \cdots\cdots \\ \{\chi_\gamma\otimes w_x \otimes \phi'\} \end{pmatrix} \bar{w}_y$$

Substituting (4.25) into (4.20) we obtain

$$\theta_N^{(j+1)} = \Psi_N(\theta_N^{(j)})^{-1} \rho_N(\theta_N^{(j)}) \qquad j \geq 0 \qquad (4.30)$$

As mentioned in the introduction, Robinson [1976b, p.771] proposes an estimator of the form (4.30) for a linear differential equations model and suggests iterating. For a derivation of (4.30) using instrumental variables see Robinson [1976b]. It is clear from (4.25) and (4.30) that an alternative way of obtaining (4.30) is to linearize $k_N(\theta)$. Thus, equating $k_N(\theta)$ to zero and solving

$$\Psi_N(\theta_N^{(j)})\theta_N^{(j+1)} - \rho_N(\theta_N^{(j)}) = 0$$

for $\theta_N^{(j+1)}$ we obtain (4.30).

Because $\Psi_N(\theta)$ is unlikely to be p.d., whereas $\Delta_N(\theta)$ almost certainly will be, the IIVE seems to be computationally inferior to the G.N. estimator for two reasons.

1. As explained in 5.§3, it will be easier to invert Δ_N than Ψ_N because the former is p.d.

2. In 5.§3 we showed that the G.N. direction is downhill. This means that the G.N. iteration can be modified so that each step of the iteration is acceptable; by acceptable we mean that $L_N(\theta_N^{(j+1)}) < L_N(\theta_N^{(j)})$ except at the minimum. The IIVE direction is <u>not</u> downhill because Ψ_N is not p.d.; we cannot, therefore, modify (4.20) as in 5.§3 to ensure the acceptability of the modified iterates. Because acceptability aids convergence in non-linear minimization we prefer the modified G.N. iteration to IIVE.

5. ESTIMATING A SUBSET OF EQUATIONS

In this section we show how to obtain relatively efficient estimates of the parameters in a subset of the equations of (1.10) without having to estimate all the structural parameters in the system; we will require, however, estimates of the <u>reduced form</u> parameters for the whole system. Without loss of generality we can consider this subset to consist of the first p_1 equations ($p_1 < p$) in (1.10). We write these as

$$\sum_{j=o}^{r} B_{1,j}\, y(n-j) = \sum_{j=o}^{s} C_{1,j}\, x(n-j) + u_1(n)$$

<u>For this section only</u> we put $\sigma = \mathrm{vec}\{B_{1,o}, \ldots, B_{1,r}, C_{1,o}, \ldots, C_{1,s}\}$. This is done to differentiate σ from the total parameter vector θ dealt with in §2. We will also assume that the restrictions on σ are independent of the other parameters in (1.10) and that they are linear in σ; we write them as $h_S(\sigma) = H_S(\sigma-\sigma_o) = 0$, H_S being a constant matrix of full row rank and σ_o is the true value of σ. Put $B_S(z;\sigma) = \Sigma_j B_{1,j}(\sigma)z^j$; $C_S(z;\sigma) = \Sigma_j C_{1,j}(\sigma)z^j$; the subscript S stands for subset.

The estimator to be given shortly is motivated by the linearization (2.10). Premultiplying both sides of (2.10) by $\hat{B}(z)$ we obtain

$$\hat{B}(z)y(n) - \hat{C}(z)x(n)+\{B(z)-\hat{B}(z)\}\hat{B}(z)^{-1}\hat{C}(z)x(n)-$$
$$-\hat{B}(z)^{-1}\{C(z)-\hat{C}(z)\}x(n)+\hat{B}(z)\mu(n) = \hat{B}(z)v(n). \tag{5.10}$$

$\mu(n)$ consists of second order terms in $B_j - \hat{B}_j$, $C_k - \hat{C}_k$. The first p_1 rows of (5.10) are

$$\hat{B}_S(z)y(n) - \hat{C}_S(z)x(n)+\{B_S(z)-\hat{B}_S(z)\}\hat{B}^{-1}(z)\hat{C}(z)x(n) -$$
$$- \{C_S(z) - \hat{C}_S(z)\}x(n) + \hat{B}_S(z)\mu(n) = \hat{B}_S(z)v(n).$$

The last equation suggests the following iteration scheme:

(i) First obtain $\hat{D}_N(\lambda)$ as outlined in Remark 2.1(iv) of §2. Put $w_S(\lambda) = \hat{D}_N(\lambda)w_x(\lambda)$.

(ii) For $j \geq 0$, given that $\sigma_N^{(j)}$ is the j^{th} iterate, the $(j+1)^{st}$ iterate is obtained by minimizing

$$L_{S;N}(\sigma;\Phi)=(2N)^{-1}\mathrm{tr} \sum_{\mathcal{B}}\Phi(\lambda_t)q(\lambda_t;\sigma,\sigma_N^{(j)})q(\lambda_t;\sigma,\sigma_N^{(j)})^* \tag{5.20}$$

for σ, subject to the linear restrictions $h_S(\sigma) = 0$;

$$q(\sigma;\sigma_N^{(j)}) = B_S(\sigma_N^{(j)})w_y-C_S(\sigma_N^{(j)})w_x+\{B_S(\sigma)-B_S(\sigma_N^{(j)})\}w_S -$$
$$- \{C_S(\sigma) - C_S(\sigma_N^{(j)})\}w_x.$$

For convenience we have omitted to indicate dependence on λ in the last equation. Φ is a $p_1 \times p_1$ matrix function of λ satisfying similar conditions to the Φ matrix introduced in Chapter 7.

Let

$$I_S(\lambda) = w_S(\lambda)w_S(\lambda)^*, \quad I_{Sx}(\lambda) = w_S(\lambda)w_x(\lambda)^* = I_{xS}(\lambda)^*,$$

$$\Delta_{S;N} = N^{-1} \sum_{\mathcal{B}} \begin{pmatrix} \{\chi_\beta \chi_\beta^* \otimes I_S \otimes \Phi'\} & \vdots & -\{\chi_\beta \chi_\gamma^* \otimes I_{Sx} \otimes \Phi'\} \\ \cdots & \vdots & \cdots \\ -\{\chi_\gamma \chi_\beta^* \otimes I_{xS} \otimes \Phi'\} & \vdots & \{\chi_\gamma \chi_\gamma^* \otimes I_x \otimes \Phi'\} \end{pmatrix}$$

and

$$k_{S;N}(\sigma) = N^{-1} \sum_{\mathcal{B}} \begin{pmatrix} \{\chi_\beta \otimes w_S \otimes \Phi'\} \\ \cdots \\ -\{\chi_\gamma \otimes w_x \otimes \Phi'\} \end{pmatrix} \{\overline{B}_S(\lambda_t;\sigma)\overline{w}_y - \overline{C}_S(\lambda_t;\sigma)\overline{w}_x\}$$

It is not difficult to verify now that the constrained minimization of (5.20) gives

$$\begin{pmatrix} \sigma_N^{(j+1)} \\ \cdots \\ \lambda_N^{(j+1)} \end{pmatrix} = \begin{pmatrix} \sigma_N^{(j)} \\ \cdots \\ 0 \end{pmatrix} - \begin{pmatrix} \Delta_{S;N} & \vdots & H_S' \\ \cdots & \vdots & \cdots \\ H_S & \vdots & 0 \end{pmatrix}^{-1} \begin{pmatrix} k_{S;N}(\sigma_N^{(j)}) \\ \cdots \\ h_S(\sigma_N^{(j)}) \end{pmatrix} \tag{5.30}$$

where $(\lambda_N^{(j)}, j \geq 1)$ is a sequence of Lagrange multiplier vectors.

Put

$$\Delta_S(\Psi) = (2\pi)^{-1} \int_{\mathcal{B}} \begin{bmatrix} \{\chi_\beta \chi_\beta^* \otimes D f_x D^* \otimes \Psi'\} & \vdots & -\{\chi_\beta \chi_\gamma^* \otimes D f_x \otimes \Psi'\} \\ \cdots & \vdots & \cdots \\ -\{\chi_\gamma \chi_\beta^* \otimes f_x D^* \otimes \Psi'\} & \vdots & \{\chi_\gamma \chi_\gamma^* \otimes f_x \otimes \Psi'\} \end{bmatrix} d\lambda;$$

$$J_S(\Phi) = \begin{pmatrix} \Delta_S(\Phi) & \vdots & H_S' \\ \cdots & \vdots & \cdots \\ H_S & \vdots & 0 \end{pmatrix}; \quad \ddagger_S(\Phi) = J_S(\Phi)^{-1} \begin{pmatrix} \Delta_S(\Phi B_S f_v B_S^* \Phi) & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{pmatrix} J_S(\Phi)^{-1};$$

$$J_S(\Phi)^{-1} = \begin{bmatrix} J_S^{11}(\Phi) & J_S^{12}(\Phi) \\ \cdots & \cdots \\ J_S^{21}(\Phi) & J_S^{22}(\Phi) \end{bmatrix} \; ; \; \ddagger_{S;\sigma}(\Phi) = J_S^{11}(\Phi)\Delta_S(\Phi B_S f_V B_S^* \Phi) J_S^{11}(\Phi).$$

In order to obtain results similar to Theorem 1 we will assume that:

D1. Conditions B1. and B2. of §2 hold.

D2.(i) r and s are given non-negative integers.

(ii) The true parameter point, θ_o, satisfies B3.(iii) - (vi).

D3.(i) $h_S(\sigma_o) = 0$

(ii) If $\sigma = [I_{(t)} \otimes V]\theta_o$ and $h_S(\sigma) = 0$, then $\sigma = \sigma_o$; V is a $p \times p_1$ matrix and $t = (r+1)p + (s+1)q$.

(iii) σ_o is a regular point of the matrix $\Delta_S[\Phi B_S(\sigma) f_V B_S(\sigma)^* \Phi]$.

D4. This is the same as condition A6. of 7.§4. □

Subject to the above assumptions the following result holds.

Theorem 3 For the iteration scheme (5.30) above put

$$\tau_N^{(j)\prime} = (\sigma_N^{(j)\prime}, \lambda_N^{(j)\prime}), \; j \geq 0, \text{ with } \tau_o' = (\theta_o', 0').$$

Then,

(i) if $\sigma_N^{(o)} = O(1)$ a.s. $(O_p(1))$ and $h_S(\sigma_N^{(o)}) = 0$, then $\sigma_N^{(j)} \to \sigma_o$ a.s. (p) $\forall\, j \geq 1$ and $h_S(\sigma_N^{(j)}) = 0 \quad \forall\, j \geq 1$.

(ii) If $\sigma_N^{(o)} - \sigma_o = o_p(1)$, then

$$\lim_N N^{\frac{1}{2}}(\tau_N^{(j)} - \tau_o) \overset{\mathcal{D}}{=} N[0, \ddagger_S(\Phi)] \; \forall \; j \geq 1.$$

(iii) Alternatively, suppose that $\sigma_N^{(o)}$ is nonstochastic and $h_S(\sigma_N^{(o)}) = 0$. Then we shall have that $\sigma_N^{(j)} \to \sigma_o$ a.s. and $h_S(\sigma_N^{(j)}) = 0 \;\forall\; j \geq 1$. In addition,

$$\lim_N N^{\frac{1}{2}}(\tau_N^{(j)} - \tau_o) \overset{\mathcal{D}}{=} N[0, \ddagger_S(\Phi)], \quad j \geq 2.$$

Let $f_{u_1}(\lambda)$ be the spectral density matrix of $\{u_1(n)\}$. In (ii) and (iii) we will have that

$$\lim_N N^{\frac{1}{2}}(\sigma_N^{(j)} - \sigma_o) \overset{\mathcal{D}}{=} N[0, \ddagger_{S;\sigma}(\Phi)].$$

The following result is analogous to Theorem 7.3.

(iv) If f_{u_1} is nonsingular on $\overline{\mathcal{B}}$, then $\ddagger_{S;\sigma}(f_{u_1}^{-1}) \leq \ddagger_{S;\sigma}(\Phi)$ i.e. $f_{u_1}^{-1}$ is the optimal choice of Φ. □

Remark 5.1 For N large the iteration (5.30) will be well defined because the matrix being inverted will be non-singular. This follows from Lemma 2 of §10. □

6. USING AN ESTIMATED Φ

We now return to the general model considered in Chapter 7. There are two ways in which our treatment so far is deficient as far as practical estimation is concerned. First f_v is unlikely to be known; in order to construct the estimate $\hat{\ddagger}_N(\hat{\theta}_N)$ of $\ddagger(\Phi)$ (see Remark 4.1 of 7.§4) in terms of known quantities the replacement of f_v by an estimate of it must be justified.

The second way in which it is important to extend the analysis of Chapter 7 is to consider the situation where Φ is unknown but an estimate of it is available. The following theorems provide the basis for such extensions of previous results. Their application to some important examples is given later in the section.

For each positive integer M we define the 2M sets

$S(j)$[1] $(-M < j \leq M)$ as

$$S(j)=\{\lambda:\omega_j-\pi/2M < \lambda \leq \omega_j+\pi/2M;\ \omega_j = \pi j/M;\ \lambda\in\mathcal{B}\},\ |j|\leq M-1$$

$$S(M)=\{\lambda:-\pi < \lambda \leq -\pi(1-\frac{1}{2M}),\ \pi(1-\frac{1}{2M})< \lambda \leq \pi;\lambda \in \mathcal{B}\} \qquad (6.05)$$

We have omitted to indicate the dependence of the $S(j)$ on M; this will be clear from the usage. We first deal with the case where Φ is estimated non-parametrically, i.e. Φ is not assumed to be a function of an auxiliary set of parameters.

We now suppose that for each pair of positive integers M and N, $M << N$, $\hat{\Phi}_{M,N}(\lambda)$ is a $p\times p$ matrix function of $\lambda \in \overline{\mathcal{B}}$ and has the following properties:

(i) $\hat{\Phi}_{M,N}(\lambda)$ is a Hermitian p.s.d. matrix with $\hat{\Phi}_{M,N}(-\lambda) = \hat{\Phi}_{M,N}(\lambda)'$.

(ii) $\hat{\Phi}_{M,N}(\lambda)$ is a step-function on $\mathcal{B}$, being constant on each of the sets $S(j)$ defined above.

(iii) There exists an Hermitian, nonstochastic, p.s.d. matrix function $\Phi_M(\lambda)$, defined on $\lambda \in \mathcal{B}$, such that

$$\lim_N \sup_{\lambda\in\mathcal{B}} \|\hat{\Phi}_{M,N}(\lambda)- \Phi_M(\lambda)\|= 0 \text{ a.s.} \qquad (6.10)$$

and

$$\lim_M \sup_{\lambda\in\mathcal{B}} \|\Phi_M(\lambda) - \Phi(\lambda)\| = 0 \text{ a.s.} \qquad (6.11)$$

We note that Φ_M must be a step-function, constant on the sets $S(j)$.

The following theorem generalizes, in an obvious way, Hannan and Robinson [1973, pp.260-262].

---

[1] Such a decomposition of $\mathcal{B}$ follows Hannan and Robinson [1973].

Theorem 4 Suppose that assumptions A1.-A6. of Chapter 7 hold and $\hat{\Phi}_{M,N}$ is defined as above. Then, there exists a sequence of positive integers, $\{M(N), N \geq 1\}$, such that $M(N) \uparrow \infty$ as $N \to \infty$ and the results of Theorems 7.1 and 7.2 hold when $\hat{\Phi}_{M(N),N}$ is substituted for Φ. □

It is clear that Theorems 7.4, 7.5 and Theorems 1-3 of this chapter can be similarly extended. Subject to the assumptions made in the respective theorems we have:

Corollary 1 There exists a sequence of positive integers, $\{M(N), N \geq 1\}$, such that $M(N) \uparrow \infty$ as $N \to \infty$ and the results of Theorems 7.1, 7.2, 7.4, 7.5 hold when $\hat{\Phi}_{M(N),N}$ is substituted for Φ. □

Remark 6.1(i) As pointed out by Hannan [1973b,pp.169-170] there are no results available, at present, to guide us in choosing the appropriate rate of increase of M with N. □

The next theorem deals with the case where $\Phi(\lambda)$ can be parameterized by a fixed and finite number of parameters so that $\Phi(\lambda) = \Phi(\lambda;\psi_o)$, with ψ_o an $\tilde{m}$ dimensional vector. We assume that:

A8. There exists a compact neighbourhood of ψ_o, Ψ say, in which $\Phi(\lambda;\psi)$ is differentiable w.r.t. ψ. $\Phi(\lambda;\psi)$ and $\Gamma(\lambda;\psi) = \frac{\partial}{\partial\psi'}\{\text{vec } \Phi(\lambda;\psi)\}$ are continuous in $(\lambda,\psi)\in\bar{B}\times\Psi$. □

Suppose that $\hat{\psi}_N \to \psi_o$ a.s. and $N^{\frac{1}{2}}(\hat{\psi}_N-\psi_o) = 0_p(1)$. We now have the following result:

Theorem 5 Given conditions A1.-A6. of Chapter 7 and A8., the results of Theorems 7.1 and 7.2 continue to hold when $\Phi(\lambda)$ is replaced by $\Phi(\lambda;\hat{\psi}_N)$. □

Theorems 7.4, 7.5 and Theorems 1-3 of this chapter can be similarly extended. Subject to the assumptions made in the respective theorems we have:

Corollary 2 The results of Theorems 7.4, 7.5 continue to hold when $\Phi(\lambda)$ is replaced by $\Phi(\lambda, \hat{\mu}_N)$. □

Remark 6.2(i) The main difference between Theorems 4 and 5 is that when Φ is a parametric function it is no longer necessary to have $\hat{\Phi}$ constant on the sets $S(j)$ for the CLT result to remain valid when Φ is replaced by $\hat{\Phi}$. In addition, given the sample size N, we eliminate the need to choose an appropriate M. □

In proving Theorem 4 we need the result of the following lemma. It is placed here, rather than in §10, because the result is of independent interest, being a generalization of a lemma due to Bernstein (e.g. Hannan [1970], p.242).

Lemma 1 Suppose $\{Z_{j,k}, Y_{j,k}, X_{j,k};\ j,k \geq 1\}$ are sequences of vector random variables and

$$Z_{j,k} = Y_{j,k} + X_{j,k} \qquad j,k \geq 1.$$

If (i) $\lim_k X_{j,k} \overset{\mathcal{D}}{=} N(0,\sigma_j)$

(ii) $\lim_j \sigma_j = \sigma,\ \sigma > 0$

(iii) $\forall\ \varepsilon > 0,\ \delta > 0\ \exists K = K(\delta,\varepsilon)$ s.t. $\lim_k P(\|Y_{j,k}\| > \delta) < \varepsilon$ for $k \geq K$,

then, for any sequence of positive integers $\{j_1(k), k \geq 1\}$ s.t. $j_1(k) \uparrow \infty$ as $k \to \infty$, there exists a sequence of positive integers $\{j(k),\ k \geq 1\}$ s.t. $j(k) \leq j_1(k)$, $k \geq 1$, $j(k) \uparrow \infty$ as $k \to \infty$ and

$$\lim_k Z_{j(k),k} \overset{\mathcal{D}}{=} N(0,\sigma). \ \square$$

Applications For each positive integer M let the sets $S(j)$ be defined as in (6.05) above and suppose $\bar{\theta}_n \to \theta_o$ a.s..

For $M << N$ and $\lambda \in S(j)$ put (following Hannan and Robinson [1973])

$$\hat{f}_{M,N}(\lambda) = 2MN^{-1} \sum_{s(j)} \{w_y - D(\bar{\theta}_N)w_x\}\{w_y - D(\bar{\theta}_N)w_x\}^*,$$

omitting the argument λ for convenience; $s(j)$ means summation over all $\lambda_t \in S(j)$. One way of obtaining a strongly consistent estimate of θ_o would be to minimize $L_N(\theta;\Phi)$, with $\Phi \equiv I_{(p)}$.

(a) It follows from Lemmas 4 and 10 that there exists a sequence of positive integers $M(N) \uparrow \infty$ as $N \to \infty$ such that

$$\lim_N \sup_{\lambda \in \mathcal{B}} \| \hat{f}_{M(N),N}(\lambda) - f_v(\lambda) \| = 0 \text{ a.s..}$$

Let $\ddagger(\Phi)$ be defined as in 7.§4 and $\Delta_N(\theta;\Phi)$, $J_N(\theta)$ as in 7.§6. If Φ is known but f_v is not, then a strongly consistent estimate of $\ddagger(\Phi)$ is

$$J_N(\hat{\theta}_N)^{-1} \begin{bmatrix} \Delta_N(\hat{\theta}_N; \Phi\hat{f}_{M(N),N}\Phi) & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} J_N(\hat{\theta}_N)^{-1},$$

where $\hat{\theta}_N$ is the estimate obtained by minimizing $L_N(\theta;\Phi)$.

(b) If f_v satisfies condition A7. of 7.§5, then we can put $\hat{\Phi}_{M,N} = \hat{f}_{M,N}$ and apply Theorem 4 and Corollary 1. In practice N is fixed and we would decrease M till $\hat{f}_{M,N}$ became non-singular over each band.

(c) Irrespective of whether or not f_v satisfies A7. we can choose $\rho(\lambda) > 0$ $(\lambda \in \bar{\mathcal{B}})$ and put

$$\hat{\Phi}_{M,N}(\lambda) = \{\hat{f}_{M,N}(\lambda) + \rho(\lambda) I_{(p)}\}^{-1} \qquad (6.20)$$

If $\rho(\lambda)$ is chosen uniformly small then the estimates obtained would be highly efficient asymptotically. In practice N is fixed and f_v is unknown. If we suspect that $f_v(\lambda)$ is singular, or close to singularity, over some frequency bands, then we can use (6.20) to eliminate the numerical instability inherent in inverting nearly singular matrices; this procedure is an alternative to decreasing M as suggested in (b).

(d) We now give an application of Theorem 5 and Corollary 2. Suppose $v(n)$ is generated by an ARMA process, i.e.

$$v(n) + \sum_{j=1}^{n_1} F_j \, v(n-j) = e(n) + \sum_{j=1}^{n_2} G_j \, e(n-j)$$

with the $\{e(n)\}$ i.i.d. having $E(e) = 0$ and $E(ee') = \Omega_e$;

$\det(I_{(p)} + \Sigma_j F_j z^j) \neq 0$ for $|z| \leq 1$;

$\det(I_{(p)} + \Sigma_j G_j z^j) \neq 0$ for $|z| \leq 1$.

Put $\psi_o = \mathrm{vec}(F_1,\ldots,F_{n_1}, G_1,\ldots,G_{n_2})$. If $N^{\frac{1}{2}}(\overline{\theta}_N - \theta_o) = 0_p(1)$ and $\overline{\theta}_N \to \theta_o$ a.s., then we can use the residuals $\hat{v}(n) = y(n) - D(z;\overline{\theta}_N)x(n)$ to obtain an estimate $\hat{\psi}_N$ (of ψ_o) such that $\hat{\psi}_N \to \psi_o$ a.s. and $N^{\frac{1}{2}}(\hat{\psi}_N - \psi_o) = 0_p(1)$. We then put

$$\Phi(\lambda;\hat{\psi}_N) = f_v(\lambda;\hat{\psi}_N)^{-1}$$

and apply Theorem 5. $\hat{\psi}_N$ may be obtained by using either the time domain or frequency domain estimators described in Chapter 5.

7. COMPUTATIONAL ASPECTS OF THE G.N. ALGORITHM

In this section we look at the computational side of the G.N. algorithm for the model (1.10). Rather than discuss estimation when we have general nonlinear constraints, we will confine ourselves to the following two types of restrictions.

(a) The diagonal elements of B_o are unity - a normalization restriction.

(b) Certain elements of the matrices B_j $(0 \leq j \leq r)$ and C_j $(0 \leq j \leq s)$ are set to zero prior to estimation.

Although (a) and (b) are very specialized they occur frequently in Econometric work. Rather than use Lagrange multipliers we will impose the constraints explicitly. Doing this enables a significant reduction in the dimension of the parameter vector, thus greatly reducing the computer storage needed for the iterative algorithm.

More generally, we also note, and we show this below, that when the constraints are linear in θ we can reduce the estimation to the case just discussed by redefining the parameters using a <u>known</u> orthogonal transformation.

We first show how the constrained G.N. iterates (7.6.10) can be reduced to the projection - operator form used in Hannan and Terrell [1972, 1973] if the restrictions are linear in θ, i.e. H is a constant matrix. From (7.6.10)

$$\begin{bmatrix} \theta_N^{(j+1)} - \theta_N^{(j)} \\ \cdots\cdots \\ \lambda_N^{(j+1)} \end{bmatrix} = - \begin{bmatrix} \Delta_N(\theta) & \vdots & H' \\ \cdots & \vdots & \cdots \\ H & \vdots & 0 \end{bmatrix}^{-1} \begin{bmatrix} k_N(\theta) \\ \cdots\cdots \\ h(\theta) \end{bmatrix}_{\theta_N^{(j)}} \qquad (7.10)$$

It is easy to verify that $h(\theta_N^{(j)}) = 0$ for $j \geq 1$ if $h(\theta_N^{(o)}) = 0$. We henceforth assume that this is so. Let $L = I_{(m)} - H'(HH')^{-1}H$. Then, by Lemma 3.5

$$\begin{pmatrix} \Delta_N & \vdots & H' \\ \cdots & \cdot & \cdots \\ H & \vdots & 0 \end{pmatrix}^{-1} = \begin{pmatrix} (L\Delta_N L)^+ & \vdots & x \\ \cdots & \cdot & \cdots \\ x & \vdots & x \end{pmatrix} \tag{7.11}$$

where for the remainder of this section we adopt the convention that submatrices denoted by x's are of no concern to us and therefore will not be given explicitly. In addition, since N can be regarded as fixed in this section, we will often omit to indicate dependence on N.

It follows from (7.10) and (7.11) that

$$\theta^{(j+1)} - \theta^{(j)} = -(L\Delta L)^+ \, k_{|\theta^{(j)}} \quad j \geq 0; \tag{7.15}$$

the linear restrictions can be rewritten as

$$L(\theta - \theta_o) = \theta - \theta_o \tag{7.16}$$

Because L is a symmetric, idempotent matrix we call (7.15) the projection - operator form of the G.N. algorithm. There exists an orthogonal matrix $\underline{O}$ such that

$$\underline{O}\, L\, \underline{O}' = \begin{pmatrix} I_{(m-f)} & \vdots & 0 \\ \cdots & \cdot & \cdots \\ 0 & \vdots & 0 \end{pmatrix}$$

where f is the rank of H. Let

$$\underline{O}\Delta\underline{O}' = \begin{pmatrix} \tilde{\Delta} & \vdots & x \\ \cdots & \cdot & \cdots \\ x & \vdots & x \end{pmatrix}, \quad \underline{O}k = \begin{pmatrix} \tilde{k} \\ \cdots \\ x \end{pmatrix}, \quad \begin{pmatrix} \theta^u \\ \cdots \\ \theta^c \end{pmatrix} = \underline{O}\theta, \quad \begin{pmatrix} \theta_o^u \\ \cdots \\ \theta^c \end{pmatrix} = \underline{O}\,\theta_o$$

where $\tilde{\Delta}$ is $(m-f)\times(m-f)$, $\tilde{k}$ is $(m-f)\times 1$ and θ^u is $(m-f)\times 1$.

It follows from (7.16) that θ^u is unrestricted while θ^c is just a known constant. Thus we need only update θ^u when iterating and (7.15) becomes

$$(\theta^u)^{(j+1)} - (\theta^u)^{(j)} = -\tilde{\Delta}(\theta^u)^{-1}\tilde{k}(\theta^u)\Big|_{(\theta^u)^{(j)}} \tag{7.20}$$

By Theorem 2.13(iii) $\tilde{\Delta}(\theta^u)$ will be nonsingular for N large and θ^u close to θ_o^u.

If only the constraints (a) and (b) are present, then L *will already be diagonal with just zeros and ones on the diagonal and we only need to permute the elements of* θ *to obtain the representation (7.20).* We now take a closer look at the construction of $\tilde{\Delta}$ and $\tilde{k}$ for this particular case.

Below we assume that the indices i,j and k have the following ranges

$$1 \leq i \leq p,\ 0 \leq j \leq r,\ 0 \leq k \leq s.$$

Let $b_j(i)'$ be the i^{th} row of B_j, $\beta_j(i)'$ the row vector containing the unconstrained elements of $b_j(i)'$ in their natural order and define $c_k(i)'$ and $\gamma_k(i)'$ similarly but with respect to the C_k matrices; let

$$\beta_j = \mathrm{vec}\{\beta_j(1), \ldots, \beta_j(p)\},\ \gamma_k = \mathrm{vec}\{\gamma_k(1), \ldots, \gamma_k(p)\},$$

$\beta = \mathrm{vec}\{\beta_o, \ldots, \beta_r\}$, $\gamma = \mathrm{vec}\{\gamma_o, \ldots, \gamma_s\}$ and $\theta = \mathrm{vec}\{\beta, \gamma\}$.

We note that θ, as just defined, is unconstrained and will therefore differ from that used in §§1 and 2. Following Dhyrmes [1973] we define the selector matrices S_{ij}, T_{ik} by

$$S_{ij} b_j(i) = \beta_j(i),\ T_{ik} c_k(i) = \gamma_k(i).$$

Put $z(\lambda) = D(\lambda;\theta) w_x(\lambda)$; $Z_j(i;\lambda) = \{S_{ij} z(\lambda)\}'$; $X_k(i;\lambda) = \{T_{ik} w_x(\lambda)\}'$;

$$Z_j(\lambda) = \begin{bmatrix} Z_j(1;\lambda) & & 0 \\ & \ddots & \\ 0 & & Z_j(p;\lambda) \end{bmatrix} e^{ij\lambda},\quad X_k(\lambda) = \begin{bmatrix} X_k(1;\lambda) & & 0 \\ & \ddots & \\ 0 & & X_k(p;\lambda) \end{bmatrix} e^{ik\lambda};$$

$Z(\lambda) = \{Z_o(\lambda),\ldots,Z_r(\lambda)\}$; $X(\lambda) = \{X_o(\lambda),\ldots,X_s(\lambda)\}$.

It is easy to verify that

$$\{\sum_{j=o}^{r} B_j e^{ij\lambda}\}z(\lambda) = z(\lambda)+Z(\lambda)\beta,\{\sum_{j=o}^{s} C_j e^{ij\lambda}\}w_x(\lambda)=X(\lambda)\gamma$$

and it therefore follows that

$$\frac{\partial}{\partial\theta'}\{B^{-1}(\lambda;\theta)C(\lambda;\theta)w_x(\lambda)\}= B^{-1}(\lambda;\theta)[-Z(\lambda),X(\lambda)].$$

Choosing a positive integer M, $M << N$, let the sets $S(j)$ and the frequencies ω_j, $-M+1 \le j \le M$, be defined as in (6.05); take $\hat{\Phi}_M$ to be an estimate of Φ, with $\hat{\Phi}_M$ taking the constant value $\hat{\Phi}_M(\omega_j)$ on the set $S(j)$. Some suggestions on the choice of Φ are given in the last section. Because M will remain fixed until indicated otherwise we will omit to indicate dependence on M below.

Put

$$\Psi(\lambda_t) = B^{-1}(\omega_j;\theta)*\Phi(\omega_j)B^{-1}(\omega_j;\theta) \quad \text{for} \quad \lambda_t \in S(j);$$

$$w_u(\lambda_t;\theta) = B(\lambda_t;\theta)w_y(\lambda_t) - C(\lambda_t;\theta)w_x(\lambda_t);$$

$$\hat{f}_x(\omega_j) = 2MN^{-1}\sum_{s(j)} I_x(\lambda_t);\ \hat{f}_{xu}(\omega_j) = 2MN^{-1}\sum_{s(j)} w_x(\lambda_t)w_u(\lambda_t)*;$$

where $\sum_{s(j)}$ is a sum over $\lambda_t \in S(j)$, $-M+1 \le j \le M$. We note that $\Psi(\lambda_t)$ is constant over each of the sets $S(j)$.

Put $\Delta(\theta) = \sum_{B}\{-Z(\lambda_t),X(\lambda_t)\}*\Psi(\lambda_t)\{-Z(\lambda_t),X(\lambda_t)\}$;

$k(\theta) = \sum_{B}\{-Z(\lambda_t),X(\lambda_t)\}*\Psi(\lambda_t)w_u(\lambda_t)$.

The G.N. iteration (7.20) becomes

$$\theta^{(j+1)} - \theta^{(j)} = -\Delta(\theta)^{-1}k(\theta)\Big|_{\theta^{(j)}} \qquad j \ge 0.$$

We note at this point that the quantities $z(\lambda)$, $Z_j(i;\lambda)$, $Z_j(\lambda)$, $Z(\lambda)$, $\Psi(\lambda)$ *and* $w_u(\lambda;\theta)$ *all depend on the unknown parameter* θ *and so will have to be updated after each iteration.*

We now give a detailed description on how to construct $\Delta(\theta)$ and $k(\theta)$. For notational convenience we omit to indicate dependence on θ.

<u>*Constructing* Δ</u> It is sufficient to consider the construction of

$$\Delta^Z = \sum_{\mathcal{B}} Z(\lambda_t)^*\Psi(\lambda_t)Z(\lambda_t).$$

as the other submatrices of Δ are obtained similarly. Δ^Z is a block matrix having $(j,k)^{th}$ submatrix

$\Delta^Z_{jk} = \Sigma_{\mathcal{B}} Z_j(\lambda_t)^*\Psi(\lambda_t)Z_k(\lambda_t)$. Δ^Z_{jk} is also a block matrix with $(\ell,n)^{th}$ submatrix

$$\begin{aligned}\Delta^Z_{jk,\ell n} &= \Sigma_{\mathcal{B}} Z_j(\ell;\lambda_t)^*\Psi_{\ell n}(\lambda_t)Z_k(n;\lambda_t).\exp\{i(k-j)\lambda_t\} \\ &= \sum_{m=-M+1}^{M} \Psi_{\ell n}(\omega_m) \sum_{s(m)} Z_j(\ell;\lambda_t)^*Z_k(n;\lambda_t).\exp\{i(k-j)\lambda_t\} \end{aligned} \tag{7.30}$$

with $\Psi_{\ell n}$ the $(\ell,n)^{th}$ element of Ψ.

$$\begin{aligned}&\sum_{s(m)} Z_j(\ell;\lambda_t)^*Z_k(n;\lambda_t).\exp\{i(k-j)\lambda_t\} = \\ &= S_{\ell j}[\sum_{s(m)} \overline{D}(\lambda_t)I_x(\lambda_t)'D(\lambda_t)'.\exp\{i(k-j)\lambda_t\}]S'_{nk} \end{aligned} \tag{7.31}$$

We can approximate this last expression in $[\cdot]$ by

$$(2M)^{-1}N\ \overline{D}(\omega_m)\hat{f}'_x(\omega_m)D(\omega_m)'\exp\{i(k-j)\omega_m\}$$

for easier computation. It is easy to check that the effect of the matrices $S_{\ell j}$, S_{nk} on the matrix in $[\cdot]$ in (7.31) is to delete certain of its rows and columns. Working backwards we first construct the matrices $D(\omega_j)f_x(\omega_j)D(\omega_j)^*, -M < j \le M$; the matrices $\Delta^Z_{jk,\ell n}$ are then obtained by appropriate row and column deletions. Next we obtain the Δ^Z_{jk} matrices and finally Δ^Z.

Constructing $k(\theta)$ It is again sufficient to discuss the construction of $k^Z = \Sigma_B Z(\lambda_t)^*\Psi(\lambda_t)w_u(\lambda_t)$.
k^Z has j^{th} block row submatrix $k^Z_j = \Sigma_B Z_j(\lambda_t)^*\Psi(\lambda_t)w_u(\lambda_t)$.
k^z_j has an ℓ^{th} block row submatrix

$$k^z_{j\ell} = \sum_B Z_j(\ell;\lambda_t)^*\Psi_{\ell 1}(\lambda_t)w_1(\lambda_t)e^{-ij\lambda_t} + \ldots + \sum_B Z_j(\ell;\lambda_t)^*\Psi_{\ell p}(\lambda_t)w_p(\lambda_t)e^{-ij\lambda_t}$$

with w_α the α^{th} element of w_u, $1 \le \alpha \le p$. A typical sum in the above equation is

$$\sum_{m=-M+1}^{M} \Psi_{\ell\alpha}(\omega_m) \sum_{s(m)} Z_j(\ell;\lambda_t)^*w_\alpha(\lambda_t)e^{-ij\lambda_t} \qquad (7.40)$$

with

$$\sum_{s(m)} Z_j(\ell;\lambda_t)^*w_\alpha(\lambda_t)e^{-ij\lambda_t} = S_{\ell j}\Big[\sum_{s(m)} \bar{D}(\lambda_t)\bar{w}_x(\lambda_t)w_\alpha(\lambda_t)e^{-ij\lambda_t}\Big] \qquad (7.41)$$

We can approximate the matrix in $[\cdot]$ by

$$(2M)^{-1}N\, \bar{D}(\omega_m)[\bar{\hat{f}}_{xu}(\omega_m)]_\alpha \cdot \exp(-ij\omega_m)$$

where $[\hat{f}_{xu}]_\alpha$ denotes the α^{th} column of $\hat{f}_{xu}$. Premultiplication by $S_{\ell j}$ has the effect of deleting certain rows of the matrix in $[\cdot]$ in (7.41). As in the construction of Δ^Z we can now work backwards to obtain k^Z.

Remark 7.1 Many of the following comments will apply equally well to the general model (7.1.10).

(i) In constructing $\Delta(\theta)$ and $k(\theta)$ we have dealt with terms of the form (see (7.30) and (7.40)):

$\sum_{j=-M+1}^{M} g(\omega_j)$ with $g(-\omega_j) = \bar{g}(\omega_j)$ and $g(0)$ and $g(\pi)$ real.

Hence, $\sum_{j=-M+1}^{M} g(\omega_j) = 2R\ell\{\sum_{j=1}^{M-1} g(\omega_j)\} + g(0) + g(\pi)$.

It follows that we need only use real summands in the computation. For further remarks on this aspect of the computation see Hannan and Terrell [1973, pp.316-317].

(ii) Suppose $\bar{\theta}_N$ is a strongly consistent estimate of θ_o. If $\hat{\Phi} = \hat{f}_v^{-1}$, then $\Psi = \hat{f}_u^{-1}$, and $\hat{f}_u$ can be obtained directly by putting

$$\hat{f}_u(\lambda_t) = 2MN^{-1} \sum_{s(j)} w_u(\lambda_s;\bar{\theta}_N)w_u(\lambda_s;\bar{\theta}_N)^*, \lambda_t \in S(j).$$

Ψ will certainly be Hermitian p.d. if $\hat{f}_u$ is nonsingular. Let m_j be the number of $\lambda_t \in S(j)$, $0 \le j \le M$. A necessary condition for $\hat{f}_u$ to be nonsingular is for the m_j to satisfy the following inequalities:

$$2m_j \ge p \quad j = 1,\ldots,M-1; \; 2m_o - 1 - \delta \ge p; \quad 2m_M - 1 \ge p$$

where $\delta = 1$ if mean correction has been made and $\delta = 0$ if it has not. We note that mean correcting the data is equivalent to dropping the zero frequency in $\Sigma_{s(o)}$.

In practice $\bar{\theta}_N$ could be chosen as an inconsistent estimator, e.g. the OLS estimator if we are dealing with the linear system (1.10). $\hat{f}_u$ would then be updated for the first few iterations by replacing $\bar{\theta}_N$ by $\theta_N^{(j)}$.

(iii) We have already noted in Remark 6.1(i) that there are, at present, no useful results on the rate at which M should increase with N. It is therefore advisable to try out several different values of M and pick that value of M at which the parameter estimates stabilize.

(iv) If the convergence of the G.N. iteration is slow, or the iteration fails to converge, it is advisable to modify the algorithm as suggested in 5.§3. Alternatively we may prefer to use the Marquardt type iteration described in 5.§3. For some remarks on termination criteria for the iteration see 5.§3.

8. A FURTHER RESULT ON STRONG CONSISTENCY

For the linear model (1.10) the requirement that Θ is a bounded set is now relaxed; this will require the strengthening of some of the other conditions stated in §2. The following assumptions are made:

E0. $\mathcal{B} = [-\pi,\pi]$.

E1.(i) This is B1. (of §2).

(ii) $\det f_u(\lambda) > 0 \quad \forall\ \lambda \in [-\pi,\pi]$, except possibly at a finite number of points.

E2.(i) This is B2.

(ii) $\det f_x(\lambda) > 0 \quad \forall\ \lambda \in [-\pi,\pi]$.

Let $\theta = \text{vec}(B_o,\ldots,B_r, C_o,\ldots,C_s)$. The dimension, m, of θ, is $(r+1)p^2 + (s+1)pq$. R_1 and ε_1 are two positive numbers chosen before the start of the estimation; R_1 can be taken very large and ε_1 can be chosen close to zero.

The permissible parameter space is defined by the following conditions:

E3.(i) $\varepsilon_1 < |\det B_o| < R_1$

(ii) $\det [B(z)] \neq 0 \quad$ for $|z| \leq 1$.

(iii) $B(z)$ and $C(z)$ are left coprime.

(iv) $[B_r, C_s]$ is of rank p.

E4.(i) The constraints $h(\theta) = 0$ are linear in θ and $H = (\partial/\partial\theta')h$ has full row rank f.

(ii) If $\|\theta\| < \infty$, $h(\theta) = 0$ and $\theta = \{I_{(t)} \otimes V\}\theta_o$, then $\theta = \theta_o$; V is a $p \times p$ constant matrix and $t = p(r+1)+q(s+1)$.□

Put $\overline{\Theta}_h = \overline{\Theta} \cap (\theta\colon h(\theta) = 0)$.

Remark 8.1 (i) Our assumptions do not exclude $u(n)$ being generated by an ARMA process with the moving average polynomial having roots on the unit circle.

(ii) There are no boundedness assumptions placed on the B_j, C_k $(1 \leq j \leq r,\ 0 \leq k \leq s)$. Unfortunately, we still need to impose E3.(i). In the reduced form case, i.e. $B_o \equiv I_{(p)}$, E3.(i) is automatically satisfied. □

The following theorem strengthens the strong consistency result given in Theorem 1.

Theorem 6 We take conditions E0.-E4. as given. Suppose that:

(i) $\forall\ \lambda \in [-\pi,\pi]$, $\Phi(\lambda)$ is a Hermitian, p.d. and continuous matrix function of λ, with $\Phi(-\lambda) = \Phi(\lambda)'$.

(ii) a.s. and $\forall\ N \geq 1$ $\exists\ \hat{\theta}_N \in \overline{\Theta}_h$ with $\|\hat{\theta}_N\| < \infty$ s.t.

$$L_N(\hat{\theta}_N) = \inf_{\theta \in \overline{\Theta}_h} L_N(\theta).$$

If for each N we choose one such $\hat{\theta}_N$, then $\hat{\theta}_N \to \theta_o$ a.s. □

Remark 8.2 (i) Clearly the rest of Theorem 1 goes through if we assume, in addition, that condition B6. holds.

9. SOME RESULTS ON GAUSSIAN ESTIMATION

We now go back to the general non-linear model (7.1.10). The notation used in this section will be that of Chapter 7. In 7.§2 and Appendix 1 of Chapter 7 we showed that if the $y(n)$ are generated by (7.1.10) and the $v(n)$ are stationary and Gaussian, then, neglecting additive constants,

$$(2N)^{-1} \sum_t \log f_v(\lambda_t) + (2N)^{-1} \operatorname{tr} \sum_t f_v^{-1} V(\lambda_t;\theta)V(\lambda_t;\theta)^* \quad (9.10)$$

is a good approximation to $-N^{-1} \times$log-likelihood of $\{y(1),\dots,y(N)\}$; $V(\lambda_t;\theta) = w_y(\lambda_t) - D(\lambda_t;\theta)w_x(\lambda_t)$ and $\sum_t$ indicates summation over all λ_t for $-\frac{1}{2}N < t \le [\frac{1}{2}N]$. Let M be a positive integer, $M << N$, and define the sets $S(j)$ and the frequencies ω_j $(-M < j \le M)$ as in §6. Put

$$\hat{f}_{M,N}(\lambda;\theta) = 2MN^{-1} \sum_{s(j)} V(\lambda_t;\theta)V(\lambda_t;\theta)^* \text{ for } \lambda \in S(j),$$

$-M < j \le M$, and suppose, for the moment, that $f_v(\lambda)$ is a step-function, constant on each of the sets $S(j)$. We follow Sargan and Espasa [1975] and rewrite (9.10) as

$$L_{M,N}(\theta;f_v) = (4M)^{-1} \sum_{j=-M+1}^{M} \log f_v(\omega_j) + (4M)^{-1} \sum_{j=-M+1}^{M} \operatorname{tr} f_v^{-1}(\omega_j) \times \hat{f}_{M,N}(\omega_j;\theta) \quad (9.15)$$

where $f_v(\omega_j)$ is the common value of all the $f_v(\lambda_t)$ for $\lambda_t \in S(j)$. We note that in going from (9.10) to (9.15) we have replaced the summation Σ_t by $\Sigma_{\mathcal{B}}$ in order to give extra flexibility to our estimation. Estimates of θ and the $f_v(\omega_j)$ can be obtained by minimizing $L_{M,N}(\theta;f_v)$ with respect to these terms. Concentrating the $f_v(\omega_j)$ out of (9.15) we obtain, omitting additive constants,

$$\ell_{M,N}(\theta) = (4M)^{-1} \sum_{j=-M+1}^{M} \log \det \hat{f}_{M,N}(\omega_j;\theta) \qquad (9.20)$$

Let $\hat{\theta}_{M,N}$ give the constrained minimum of (9.20) for fixed M and N. We will show that for fixed M, $\hat{\theta}_{M,N} \to \theta_o$ a.s. and that $\hat{\theta}_{M,N}$ is asymptotically normal. Because f_v will rarely be a step-function it is desirable to be able to increase M with N, i.e. make M a monotonically increasing function of N, and still retain the consistency and asymptotic normality properties of our estimator. This is done in the following theorems.

The first theorem gives some strong consistency results.

Theorem 7 We take assumptions A1.-A4. and A7. of Chapter 7 as given. Suppose that, given $M, \exists N_1 = N_1(M)$ s.t. for each $N \geq N_1 \ \exists \hat{\theta}_{M,N} \in \overline{\Theta}_h$ satisfying

$$\ell_{M,N}(\hat{\theta}_{M,N}) = \inf_{\theta \in \overline{\Theta}_h} \ell_{M,N}(\theta).$$

Then,

(i) $\lim_N \hat{\theta}_{M,N} = \theta_o$ a.s.

(ii) $\exists$ a sequence of positive integers $\{M(N), N \geq 1\}$ s.t. $M(N) \uparrow \infty$ as $N \to \infty$ and $\hat{\theta}_{M(N),N} \to \theta_o$ a.s.. If $\{M_1(N), N \geq 1\}$ is another sequence of positive integers s.t. $M_1(N) \uparrow \infty$ as $N \to \infty$ and $M_1(N) \leq M(N)\ \forall N$, then $\hat{\theta}_{M_1(N),N} \to \theta_o$ a.s. □

Next, we give a CLT result. Let the neighbourhoods $(U_\alpha, \alpha > 0)$ be defined as in 7.§4. Put $H(\theta) = (\partial/\partial\theta')h(\theta)$; $A(\lambda;\theta) = (\partial/\partial\theta')\mathrm{vec}\, D(\lambda;\theta)$; $k_{M,N}(\theta) = (\partial/\partial\theta')\ell_{M,N}(\theta)$. For the $\hat{\theta}_{M,N}$ defined in Theorem 7, we will have, analogously to 7.§4, that for N sufficiently large

$$k_{M,N}(\hat{\theta}_{M,N}) + H'(\hat{\theta}_{M,N})\hat{\lambda}_{M,N} = 0$$

$$h(\hat{\theta}_{M,N}) = 0 \tag{9.30}$$

(9.30) are the Normal Equations with the $\hat{\lambda}_{M,N}$ Lagrange multipliers. Put $\hat{\tau}'_{M,N} = (\hat{\theta}'_{M,N}, \hat{\lambda}'_{M,N})$; $\tau'_o = (\theta'_o, 0')$;

$$\hat{f}_M(\lambda) = M/\pi \int_{S(j)} f_v(\lambda)d\lambda \quad \text{for} \quad \lambda \in S(j),\ -M < j \leq M.$$

$J(\Phi)$ and $\ddagger(\Phi)$ are defined as in 7.§4.

Theorem 8 We take conditions A1.-A7. of Chapter 7 as given. Then:

(i) For each M, $\lim_N N^{\frac{1}{2}}(\hat{\tau}_{M,N} - \tau_o) \overset{\mathcal{D}}{=} N[0, \ddagger(\hat{f}_M^{-1})]$

(ii) $\exists$ a sequence of positive integers $\{M(N), N \geq 1\}$ s.t. $M(N) \uparrow \infty$ as $N \to \infty$ and

$$\lim_N N^{\frac{1}{2}}(\hat{\tau}_{M(N),N} - \tau_o) \overset{\mathcal{D}}{=} N[0, \ddagger(f_v^{-1})]$$

If $\{M_1(N), N \geq 1\}$ is another sequence of positive integers s.t. $M_1(N) \uparrow \infty$ as $N \to \infty$ and $M_1(N) \leq M(N)\ \forall\ N$, then

$$\lim_N N^{\frac{1}{2}}(\hat{\tau}_{M_1(N),N} - \tau_o) \overset{\mathcal{D}}{=} N[0, \ddagger(f_v^{-1})]. \ \square$$

Because the Normal Equations (9.30) are likely to be highly nonlinear, we will again set up a G.N. iteration.

Put

$$\Delta_{M,N}(\theta) = N^{-1} \sum_{j=-M+1}^{M} \sum_{s(j)} A^*(\lambda_t;\theta)\{I'_x(\lambda_t) \otimes \hat{f}_{M,N}(\omega_j;\theta)^{-1}\}A(\lambda_t;\theta);$$

$$J_{M,N}(\theta) = \begin{pmatrix} \Delta_{M,N}(\theta) & \vdots & H'(\theta) \\ \cdots & \cdots & \cdots \\ H(\theta) & \vdots & \vec{0} \end{pmatrix};$$

The next theorem gives an analogous result to Theorem 7.4.

Theorem 9 We take conditions A1.-A7. of Chapter 7 as given. There exists a sequence of positive integers, $\{M(N),\ N \geq 1\}$, s.t. $M(N) \uparrow \infty$ as $N \to \infty$ and the following results hold:

(i) $\hat{\theta}_{M(N),N} \to \theta_o$ a.s. and $N^{\frac{1}{2}}(\hat{\tau}_{M(N),N} - \tau_o) \xrightarrow{\mathcal{D}} N[0, \ddagger(f_v^{-1})]$.

Let $U_{\alpha 1}$ be defined as in 7.§4.

(ii) $\exists\ \alpha$, $0 < \alpha \leq \alpha 1$, s.t. in the set $U_\alpha \times (\lambda : \|\lambda\| \leq \alpha)$ there is (a.s.), for N sufficiently large, a unique solution to the Normal equations (9.30) (with $M=M(N)$), $\hat{\tau}_N$ say.

Put $\tau_N^{(o)'} = (\theta_N^{(o)'}, 0'),\ \tau_N^{(j)'} = (\theta_N^{(j)'}, \lambda_N^{(j)'})$ $j \geq 1$ and define the G.N. iteration by

$$\tau_N^{(j+1)} = \begin{pmatrix} \theta_N^{(j)} \\ \cdots \\ 0 \end{pmatrix} - J_{M(N),N}(\theta)^{-1} \begin{pmatrix} k_{M(N),N}(\theta) \\ \cdots\cdots \\ h(\theta) \end{pmatrix}\Bigg|_{\theta_N^{(j)}} \qquad j \geq 0$$

(iii) If $\theta_N^{(o)} \xrightarrow{P} \theta_o$, then, with probability tending to 1 as $N \to \infty$, $\lim_j \tau_N^{(j)} = \hat{\tau}_N$.

(iv) If $\theta_N^{(o)} \to \theta_o$ a.s., then (a.s.) for N sufficiently large, $\lim_j \tau_N^{(j)} = \hat{\tau}_N$.

(v) If, in addition, in either (iii) or (iv), $N^{\frac{1}{2}}(\theta_N^{(o)} - \theta_o) = O_p(1)$, then $\lim_N N^{\frac{1}{2}}(\tau_N^{(j)} - \tau_o) \overset{\mathcal{D}}{=} N[0, \ddagger(f_v^{-1})]$ for $j \geq 1$. □

We now comment briefly on the computation of the G.N. iterates.

Remark 9.1 (i) The G.N. iterates can be constructed as outlined in §7; now, however, $\hat{f}_u$ will be updated at the end of each iteration rather than for the first few iterations only.

(see Remark 7.1(ii) in §7).

(ii) Let $\ell^j = \ell(\theta^{(j)})$, omitting the M and N subscripts for convenience. If the G.N. iteration is modified as in 5.§3, then we could use

$$T_j = (\ell^{j-1} - \ell^j)/(1+|\ell^{j-1}|)$$

as a termination criterion (see 5.§3). If, however, some of the $\hat{f}(\omega_j;\theta)$ are nearly singular, then it may be more satisfactory to use

$$T'_j = \{\exp(\ell^{j-1}) - \exp(\ell^j)\}/\{1+ \exp(\ell^{j-1})\}.$$

This is because small changes in x result in large changes in $\log x$ for $0 < x < 1$. Using T_j rather than T'_j may lead to unnecessary additional iterations.

A ridge regression analogue

If in minimizing $\ell_{M,N}(\theta)$ one (or more) of the matrices $\hat{f}_{M,N}(\omega_j;\hat{\theta}_{M,N})$ is nearly singular, possibly due to the near singularity of $f_v(\omega_j)$, the minimization process may have the following undesirable features:

(i) the estimation may become numerically unstable in the sense that small perturbations in the data produce marked changes in the parameter estimates

(ii) the suggested G.N. iteration may take extremely long to converge.

(iii) the effect (on the parameter estimates) of the bands where $\hat{f}_{M,N}$ is close to singularity may be overwhelming.

To avoid such unpleasantness we could alternatively minimize

$$\ell_{M,N;\eta}(\theta) = (4M)^{-1} \sum_{j=-M+1}^{M} \log\det[\hat{f}_{M,N}(\omega_j;\theta)+\eta I_{(p)}]$$

where η is some small positive number. Put $\hat{f}_{M,\eta}(\lambda) = \hat{f}_M(\lambda) + \eta I_{(p)}$; $f_{v;\eta} = f_v + \eta I_{(p)}$. If, in the above, we replace $\hat{f}_M$ and f_v, respectively, by $\hat{f}_{M;\eta}$ and $f_{v;\eta}$, then all the results of this section will still hold when working with $\ell_{M,N;\eta}$ rather than $\ell_{M,N}$. A procedure for choosing an optimal η, based on a suggestion by Hoerl and Kennard [1970a, b] is to plot the parameter estimates for several different values of η and choose η in the region where the parameter estimates stabilize. The idea is to choose an η such that small changes in the data produce only small changes in the parameter estimates. Hopefully, by choosing an η such that small changes in η produce only small changes in the parameter estimates the above requirement is fulfilled.

10. PROOFS

Throughout this section c is a universal constant, not always the same. Unless stated otherwise, it should be assumed that all convergence results below mean a.s. convergence; the mode of convergence will often not be indicated.

Proof of Theorem 1 To show that Theorems 7.1-7.5 can be applied we need to check that conditions B1.-B4. imply conditions A1.-A4. of Chapter 7 and conditions B1.-B6. imply conditions A1.-A6. With the exception of A4.(ii) it is easy to check that the B. conditions imply the A. conditions. We now show that B3.(v), (vi) and B4.(ii) imply A4.(ii). If $D(\lambda;\theta) = D(\lambda)$ and B3.(v), (vi) hold, then $\{B(\lambda;\theta), C(\lambda;\theta)\} =$ $= V\{B(\lambda), C(\lambda)\}$ by Lemma 3 of 2.§4, where V is a constant matrix. By B4.(ii), $\theta = \theta_o$. □

Proof of Theorem 2 To apply Theorems 7.1-7.5 we need to check that conditions C1.-C4. imply conditions A1.-A4. and B1.-B6. imply A1.-A6. With the exception of conditions A3.(iii) and A4.(ii) it is easy to show that the C. conditions imply the A. conditions.

Let $B(z;\theta)$, $C(z;\theta)$ and $D(z;\theta)$ be defined as in §3. We first show that C3.(iii), (iv) imply A3.(iii). Because

$$\det B_I(z;\theta) = \det B(z;\theta).\det(\Sigma_j B_{22,j} z^j)$$

it follows from C3.(iii) that $\forall\ \theta \in \Theta, \det B(z;\theta) \neq 0$ for $|z| \leq 1$. From C3.(iv)

$$|\det B(0;\theta)| = |\det B_I(0;\theta)/\det B_{22,o}| \geq \varepsilon_1/|\det B_{22,o}| > 0$$

$\forall\ \theta \in \overline{\Theta}$.

Hence A3.(iii) holds. Next we show that C3.(v), (vi) and C4.(ii) imply A4.(ii). If $D(\lambda;\theta) = D(\lambda)$, then $B_I(\lambda;\theta)^{-1}C_I(\lambda;\theta) = B_I(\lambda)^{-1}C_I(\lambda)$ from (3.25). By C3.(v), (vi) and Lemma 2.3.

$$\{B_I(\lambda;\theta),\ C_I(\lambda;\theta)\} = V\{B_I(\lambda), C_I(\lambda)\}$$

where V is a constant matrix. It now follows from C4.(ii) that $\theta = \theta_o$. □

Let σ, $\Delta_S(\Phi)$, $J_S(\Phi)$, $B_S(\sigma;\Phi)$ and $C_S(\sigma;\Phi)$ be defined as in §5. Before giving a proof of Theorem 3 we obtain several preliminary results. We first show that $J_S(\Phi)$ is nonsingular.

Lemma 2 If assumptions D1.-D4. of §5 hold, then $J_S(\Phi)$ is nonsingular.

Proof Put $Q(\lambda;\sigma) = B_S(\lambda;\sigma)D(\lambda)-C_S(\lambda;\sigma)$;

$$P(\lambda)=[\{\chi_\beta'(\lambda)\otimes D'(\lambda)\otimes I_{(p_1)}\},-\{\chi_\gamma'(\lambda)\otimes I_{(q)}\otimes I_{(p_1)}\}];$$

$$M(\sigma;\Phi) = (4\pi)^{-1}\ \mathrm{tr}\int_{\mathcal{B}} \Phi\ Q\ f_x Q^*\ d\lambda.$$

It is easily shown that $\mathrm{vec}\ Q(\lambda;\sigma) = P(\lambda)\sigma$ and

$$\mathrm{tr}\ \Phi Q f_x Q^* = \{\mathrm{vec}(\Phi Q f_x)\}'\mathrm{vec}\ \overline{Q} = \sigma' P'(f_x\otimes\Phi')\overline{P}\sigma\ .$$

Hence, $M(\sigma;\Phi) = \tfrac{1}{2}\sigma'\Delta_S(\Phi)\sigma$. If $M(\sigma;\Phi) = 0$, then $Q(\lambda;\sigma) \equiv 0$. From D2. and D3. we have that $\sigma = \sigma_o$ if $Q(\lambda;\sigma) \equiv 0$ and $h_S(\sigma) = 0$. But $M(\sigma_o;\Phi) = 0$. Hence σ_o must be an isolated constrained local minimum of $M(\sigma;\Phi)$. An application of Lemma 3.9 completes the proof. □

Let $\underline{B}_j$, $\underline{C}_k$ $(j = 0,\ldots,r;\ k = 0,\ldots,s)$ and $\hat{D}_N$ be defined as in Remark 2.1(iv) of §2. Put $\xi = \mathrm{vec}\{\underline{B}_1,\ldots,\underline{B}_r, \underline{C}_o,\ldots,\underline{C}_s\}$ and write ξ_o for the true ξ. Put

$$N^{-\frac{1}{2}}\zeta_N(\lambda) = (2\pi N)^{-\frac{1}{2}}\sum_{n=1}^{N}\{\sum_{j=o}^{\infty} D_j x(n-j)\}e^{in\lambda} - D(\lambda)w_x(\lambda).$$

We know that $\hat{D}_N(\lambda) = D(\lambda;\hat{\xi}_N)$ with

$$\hat{\xi}_N \to \xi_o \text{ a.s. and } N^{\frac{1}{2}}(\hat{\xi}_N - \xi_o) = 0_p(1) \qquad (10.10)$$

Lemma 3 If $\phi(\lambda)$ is a p^2 columned matrix function of λ, continuous in λ for $\lambda\in\mathcal{B}$, then the following results hold:

(i) $N^{-1}\sum_{\mathcal{B}}\phi\{\hat{D}_N w_x\otimes I_{(p)}\}\overline{w}_v \to 0$ a.s.

(ii) $N^{-3/2}\sum_{\mathcal{B}}\phi\{\hat{D}_N w_x\otimes I_{(p)}\}\overline{\zeta}_N \to 0$ a.s.

(iii) $N^{-\frac{1}{2}}\sum_{\mathcal{B}}\phi\{(\hat{D}_N-D)w_x\otimes I_{(p)}\}\overline{w}_v = o_p(1)$

(iv) $N^{-1}\sum_{\mathcal{B}}\phi\{(\hat{D}_N-D)w_x\otimes I_{(p)}\}\overline{\zeta}_N = o_p(1)$

Proof (i) From Lemma 5 (see below)

$$N^{-1}\sum_{\mathcal{B}} \phi\{(\hat{D}_N - D)w_x \otimes I_{(p)}\}\overline{w}_v \to 0.$$

Because $N^{-1}\sum_{\mathcal{B}} \phi\,\{Dw_x \otimes I_{(p)}\}\overline{w}_v \to 0$ by Lemma 7.1, (i) follows.

(ii) This follows from Lemma 4.9(iii).

Put $Q(\lambda;\xi) = (\partial/\partial\xi')\text{vec}\{D(\lambda;\xi)'\}$; $I_{xv} = w_x w_v^*$;

$I_{x\zeta} = w_x \zeta_N^*$.

(iii) $$N^{-\frac{1}{2}}\sum_{\mathcal{B}}\phi\{\hat{D}_N - D)w_x \otimes I_{(p)}\}\overline{w}_v = [N^{-1}\sum_{\mathcal{B}}\phi\{I_{(p)} \otimes I'_{xv}\}Q(\lambda_t;\hat{\xi}_N^{\dagger})] \times \times N^{\frac{1}{2}}(\hat{\xi}_N - \xi_o)$$
$$= o(1).O_p(1) = o_p(1)$$

by Lemma 7.1 and (10.10); $\|\hat{\xi}_N^{\dagger} - \xi_o\| \leq \|\hat{\xi}_N - \xi_o\|$.

(iv) $$N^{-1}\sum_{\mathcal{B}} \phi\,\{(\hat{D}_n - D)w_x \otimes I_{(p)}\}\overline{\zeta}_N =$$
$$= [N^{-3/2}\sum_{\mathcal{B}} \phi\{I_{(p)} \otimes I'_{x\zeta}\}Q(\lambda_t;\xi_N^{\dagger})].N^{\frac{1}{2}}(\hat{\xi}_N - \xi_o)$$
$$= o(1).O_p(1) = o_p(1)$$

by Lemma 4.9 and (10.10). □

Proof of Theorem 3 It is sufficient to prove Theorem 3 for $\tau_N^{(1)}$, since the proof for $\tau_N^{(j)}$ $(j > 1)$ follows similarly. For convenience we put $\tilde{\sigma} = \sigma_N^{(o)}$, $\tilde{B}_S(\lambda) = B_S(\lambda;\sigma_N^{(o)})$, $\tilde{C}_S(\lambda) = C_S(\lambda;\sigma_N^{(o)})$. Then,

$$\tilde{B}_S w_y - \tilde{C}_S w_x = \tilde{B}_S D w_x + N^{-\frac{1}{2}}\tilde{B}_S \zeta_N + \tilde{B}_S w_v - \tilde{C}_S w_x =$$
$$= -(B_S - \tilde{B}_S)Dw_x + (C_S - \tilde{C}_S)w_x + \tilde{B}_S w_v + N^{-\frac{1}{2}}\tilde{B}_S\zeta_N =$$
$$= -(B_S - \tilde{B}_S)w_S + (C_S - \tilde{C}_S)w_x + B_S w_v + g_N(\lambda;\tilde{\sigma})$$

where $w_S = \hat{D}_N w_x$ and

$$g_N(\lambda;\tilde{\sigma}) = N^{-\frac{1}{2}}\tilde{B}_S\zeta_N + (\tilde{B}_S - B_S)w_v + (\tilde{B}_S - B_S)(D - \hat{D}_N)w_x.$$

Put

$$\hat{\eta}_N(\lambda)' = [\{\chi_\beta \otimes w_S \otimes \Phi'\}', -\{\chi_\gamma \otimes w_x \otimes \Phi'\}'];$$

$$\eta_N(\lambda)' = [\{\chi_\beta \otimes Dw_x \otimes \Phi'\}', \; -\{\chi_\gamma \otimes w_x \otimes \Phi'\}'],$$

$$\mu_N = N^{-1} \sum_{\mathcal{B}} \hat{\eta}_N \overline{B}_S \overline{w}_v, \quad \nu_N = N^{-1} \sum_{\mathcal{B}} \hat{\eta}_N \overline{g}_N(\lambda_t;\tilde{\sigma}).$$

It follows that

$$k_{S;N}(\tilde{\sigma}) = -\Delta_{S;N}(\sigma_o - \tilde{\sigma}) + \mu_N + \nu_N$$

so that

$$\begin{pmatrix} \sigma_N^{(1)} - \sigma_o \\ \cdots\cdots \\ \lambda_N^{(1)} \end{pmatrix} = \begin{pmatrix} \Delta_{S;N} & \vdots & H_S' \\ \cdots & \vdots & \cdots \\ H_S & \vdots & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mu_N + \nu_N \\ \cdots\cdots \\ 0 \end{pmatrix} \tag{10.20}$$

because $H_S(\tilde{\sigma}-\sigma_o) = 0$ by assumption.

(i) If $\tilde{\sigma} = O(1)$ a.s., then it follows from Lemma 3(i), (ii) that $\mu_N \to 0$ and $\nu_N \to 0$. It now follows from (10.20) that $\tau_N^{(1)} - \tau_o \to 0$. If $\tilde{\sigma} = O_p(1)$, then we can show, similarly, that $\tau_N^{(1)} - \tau_o \xrightarrow{p} 0$. Finally, it is clear from (10.20) that $h_S(\sigma_N^{(1)}) = 0$.

$$\text{(ii)} \quad \lim_N N^{\frac{1}{2}} \mu_N \overset{\mathcal{D}}{=} N[0, \Delta_S(\Phi B_S f_v B_S^* \Phi)] \tag{10.25}$$

because $N^{-\frac{1}{2}} \Sigma_{\mathcal{B}} (\hat{\eta}_N - \eta_N) \overline{B}_S w_v = o_p(1)$ by Lemma 3(iii) and $N^{-\frac{1}{2}} \Sigma_{\mathcal{B}} \eta_N \overline{B}_S \overline{w}_v \xrightarrow{\mathcal{D}} N[0, \Delta_S(\Phi B_S f_v B_S^* \Phi)]$ by Lemma 7.2.

We now show that

$$N^{\frac{1}{2}} \nu_N = o_p(1) \tag{10.26}$$

if $\tilde{\sigma} - \sigma_o = o_p(1)$.

$N^{-1} \Sigma_{\mathcal{B}} \hat{\eta}_N \bar{\bar{B}}_S \overline{\zeta}_N = o_p(1)$ because $N^{-1} \Sigma_{\mathcal{B}} (\hat{\eta}_N - \eta_N) \bar{\bar{B}}_S \overline{\zeta}_N = o_p(1)$ by Lemma 3(iv) and $N^{-1} \Sigma_{\mathcal{B}} \eta_N \bar{\bar{B}}_S \overline{\zeta}_N = o_p(1)$ from the proof of Lemma 7.3. $N^{-\frac{1}{2}} \Sigma_{\mathcal{B}} \hat{\eta}_N (\bar{\bar{B}}_S - \bar{B}_S) \overline{w}_v = o_p(1)$ using Lemma 3(iii) and Lemma 7.2. (10.26) now follows easily. From (10.20), (10.25) and (10.26) we will have that

$$\lim_N N^{\frac{1}{2}}(\tau_N^{(1)}-\tau_o) \overset{\mathcal{D}}{=} N[0,\ddagger_S(\Phi)]$$

(iii). (iii) of Theorem 3 follows easily from Theorem 3(i), (ii).

(iv) We first note that $\Phi B_S f_v B_S^* \Phi = f_{u_1}$. The proof of (iv) is now very similar to the proof of Theorem 7.3 and will not be given. □

We now give a proof of Lemma 1. The statement of the lemma is given in §6.

Proof of Lemma 1 Put $\alpha(t) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{t} e^{-\frac{1}{2}s^2} ds$. It follows from Rao [1973, p.123(xi)] that it is sufficient to prove the result for $X_{j,k}$, $Y_{j,k}$ and $Z_{j,k}$ all scalars. Let $\{\varepsilon_j, j \geq 1\}$ be a non-negative sequence of numbers such that $\varepsilon_j \downarrow 0$ as $j \to \infty$. By Polya's Theorem (Rao [1973, p.120 (vi)]) and conditions (i) and (iii) of Lemma 1, there exists a positive function $g(t)$ s.t. $g(t) \uparrow \infty$ as $t \to \infty$ and for $k \geq g(j), P(|Y_{j,k}| > \varepsilon_j) < \varepsilon_j$,

$P(X_{j,k} \leq t + \varepsilon_j) \leq \alpha[(t+ \varepsilon_j)/\sigma_j] + \varepsilon_j$,

$P(X_{j,k} \leq t - \varepsilon_j) \geq \alpha[(t- \varepsilon_j)/\sigma_j] - \varepsilon_j$

for <u>all</u> real numbers t.

Hence,

$$\begin{aligned} P(Z_{j,k} \leq t) &= P(Z_{j,k} \leq t, X_{j,k} \leq t+\varepsilon_j) + \\ &\quad + P(Z_{j,k} \leq t, X_{j,k} > t + \varepsilon_j) \\ &\leq P(X_{j,k} \leq t + \varepsilon_j) + P(|Y_{j,k}| > \varepsilon_j) \\ &\leq \alpha[(t + \varepsilon_j)/\sigma_j] + 2\varepsilon_j \qquad k \geq g(j) \end{aligned}$$

uniformly in t.

Similarly

$$P(X_{j,k} \leq t - \varepsilon_j) \leq P(Z_{j,k} \leq t) + P(|Y_{j,k}| > \varepsilon_j)$$

$\Rightarrow P(Z_{j,k} \leq t) \geq \alpha[(t-\varepsilon_j)/\sigma_j] - 2\varepsilon_j \qquad k \geq g(j)$

uniformly in t.

Put $j(k) = \min\{[g^{-1}(k)], j_1(k)\}$, where $[x]$ is the integral part of x. It follows from above that

$$P(Z_{j(k),k} \leq t) \leq \alpha[(t + \varepsilon_{j(k)})/\sigma_{j(k)}] + 2\varepsilon_{j(k)}$$

$$P(Z_{j(k),k} \leq t) \geq \alpha[(t - \varepsilon_{j(k)})/\sigma_{j(k)}] - 2\varepsilon_{j(k)}.$$

The proof of the lemma is completed by noting that $j(k) \uparrow \infty$ as $k \to \infty$, so that $\varepsilon_{j(k)} \downarrow 0$, $\sigma_{j(k)} \to \sigma$ as $k \to \infty$. □

The next few lemmas are needed in the proofs of Theorems 4 and 5.

Lemma 4 Suppose that $X_{j,k}(\alpha)$, $X_j(\alpha)$, $(j,k \geq 1)$, and $X(\alpha)$ are matrix functions of the variable α, where α belongs to the set A. Let $\sup_\alpha$ denote the supremum over all $\alpha \in A$. If $\lim_k \sup_\alpha \|X_{j,k}(\alpha) - X_j(\alpha)\| = 0$ a.s. and $\lim_j \sup_\alpha \|X_j(\alpha) - X(\alpha)\| = 0$ a.s., then, for any sequence of positive integers $\{j_1(k), k \geq 1\}$ s.t. $j_1(k) \uparrow \infty$ as $k \to \infty$, there exists a sequence of positive integers $\{j(k), k \geq 1\}$ s.t. $j(k) \leq j_1(k)$, $k \geq 1$, $j(k) \uparrow \infty$ as $k \to \infty$ and

$$\lim_k \sup_\alpha \|X_{j(k),k}(\alpha) - X(\alpha)\| = 0 \quad \text{a.s.}$$

The proof is a generalization of Robinson [1972b, p. 51].

Proof We put $G_{j,k} = \sup_\alpha \|X_{j,k}(\alpha) - X_j(\alpha)\|$, $j,k \geq 1$, and choose a sequence $\{\varepsilon_j, j \geq 1\}$ of positive real numbers s.t. $\varepsilon_j \downarrow 0$ as $j \to \infty$ and $\Sigma_j \varepsilon_j < \infty$. From the assumptions of the lemma, there exists a positive function $g(t)$ s.t. $g(t) \uparrow \infty$ as $t \to \infty$ and

$$P(\sup_{k \geq K} G_{j,k} > \varepsilon_j) < \varepsilon_j \qquad \text{for } K \geq g(j).$$

Put

$$j(k) = \min\{j_1(k),[g^{-1}(k)]\}. \quad \text{It follows that}$$

$$\sum_{k=1}^{\infty} P(G_{j(k),k} > \varepsilon_{j(k)}) < \infty$$

so that $\forall\ \varepsilon > 0$

$$\sum_{k=1}^{\infty} P(G_{j(k),k} > \varepsilon) < \infty.$$

By the Borel-Cantelli lemma, $G_{j(k),k} \to 0$ a.s. as $k \to \infty$. The result of the lemma follows readily. □

Lemma 5 We take the assumptions and notation of Lemma 7.1 as given. Put

$$h_{jk}(N) = N^{-1} \sum_{C} I_{jk}(s)\ \psi(\lambda_s)\phi_N(\lambda_s).$$

If $\sup_{\lambda\in C} |\phi_N(\lambda)| \to 0$ a.s., then $\lim_N h_{jk}(N) = 0$ a.s.; if ψ is a function of τ, then the convergence is uniform in $\tau \in \Gamma$(see the statement of Lemma 7.1 for a definition of τ and Γ).

Proof $|h_{jk}(N)| \leq \sup_{\lambda\in C}|\phi_N(\lambda)|.N^{-1}\sum_{C}\{I_{jj}(s)+I_{kk}(s)\}|\psi(\lambda_s)|=o(1)$. $.O(1) = o(1)$ using Lemma 1 of 7.§7. □

Let the matrices $\hat{\Phi}_{M,N}$, $\Phi_M(M,N \geq 1)$ be as described in §6, and let $L_N(\theta;\Phi)$ be as defined in (7.1.15). Suppose that $\hat{\theta}_{M,N}$ minimizes $L_N(\theta;\hat{\Phi}_{M,N})$ over $\theta \in \overline{\Theta}_h$. Because $\Phi_M \to \Phi$, there exists an M_o such that Φ_M is nonsingular on an open subset of $\mathcal{B}_1$($\mathcal{B}_1$ is defined in assumption A2.(iii) of 7.§3).

Lemma 6 Suppose assumptions A1.-A6. of Chapter 7 hold. For $M \geq M_o$ we have that:

(i) $\hat{\theta}_{M,N} \to \theta_o$ a.s.

(ii) Put $\hat{\tau}'_{M,N} = (\hat{\theta}'_{M,N},\hat{\lambda}'_{M,N})$, where $\hat{\lambda}_{M,N}$ is a vector of

Lagrange multipliers, and let $\tau_o' = (\theta_o', 0')$. Then, $N^{\frac{1}{2}}(\hat{\tau}_{M,N} - \tau_o) \xrightarrow{\mathcal{D}} N[0, \ddagger(\Phi_M)]$.

Proof The proofs of (i) and (ii) are essentially the same as the proofs of Theorems 7.1 and 7.2 respectively. The method of proof consists in showing that we can replace $\hat{\Phi}_{M,N}$ by Φ_M. Once this is done (i) and (ii) will follow because Φ_M has, essentially, the same properties as Φ.

(i) Let $L_{N;\eta}(\theta;\Phi)$ be defined as in the proof of Theorem 7.1. From Lemma 5 we will have that

$$L_N[\theta_o;(\hat{\Phi}_{M,N}-\Phi_M)] \to 0, \ \sup_{\theta\in\overline{\Theta}} L_{N;\eta}[\theta;(\hat{\Phi}_{M,N} - \Phi_M)] \to 0 \ \forall\ \eta > 0.$$

The proof of (i) now follows from the proof of Theorem 7.1.

(ii) Put $A(\lambda;\theta) = (\partial/\partial\theta')\mathrm{vec}\ D(\lambda;\theta)$, $k_N(\theta;\Phi) = (\partial/\partial\theta)\ell_N(\theta;\Phi)$; $\phi(\lambda;\theta;\Phi) = A(\lambda;\theta)^*\{I_{(q)} \otimes \Phi(\lambda)\}$.

We first show that

$$\lim_N N^{-1}\sum_{\mathcal{B}}\phi(\lambda_t;\theta_o;\hat{\Phi}_{M,N})\mathrm{vec}\ \zeta_N w_x^* \overset{p}{=} 0 \qquad (10.40)$$

The above expression can be rewritten as

$$\sum_{j=-M+1}^{M} N^{-1} \sum_{s(j)} \phi(\lambda_t;\theta_o;\hat{\Phi}_{M,N})\mathrm{vec}\ \zeta_N w_x^* .$$

Noting that $\hat{\Phi}_{M,N}$ is constant on each of the sets $S(j)$, we can deduce from the proof of Lemma 7.3 that for each j

$$N^{-1} \sum_{s(j)} \phi(\lambda_t;\theta_o;\hat{\Phi}_{M,N})\mathrm{vec}\ \zeta_N w_x^* = o_p(1).$$

Since M is fixed, (10.40) follows.

For α small and positive

$$\lim_N(\partial/\partial\theta')k_N[\theta;(\hat{\Phi}_{M,N} - \Phi_M)] = 0,$$

uniformly in $\theta \in U_\alpha$, by Lemma 5. It follows from Lemma 7.5 that if $\theta_N^\dagger \to \theta_o$, then

$$\lim_N (\partial/\partial\theta_j) k_N(\theta_N^\dagger;\hat{\Phi}_{M,N}) = (2\pi)^{-1} \int_B A^*(\lambda)[f_x'(\lambda)\otimes\Phi_M(\lambda)]A_j(\lambda)d\lambda \tag{10.41}$$

where $A_j(\lambda)$ is the j^{th} column of $A(\lambda)$.

Finally, we show that

$$\lim_N N^{\frac{1}{2}} k_N(\theta_o;\hat{\Phi}_{M,N}) \overset{\mathcal{D}}{=} \mathcal{N}[0,\Delta(\Phi_M f_v \Phi_M)] \tag{10.42}$$

From the proof of Lemma 7.4 we know that

$$N^{\frac{1}{2}} k_N[\theta_o;(\hat{\Phi}_{M,N} - \Phi_M)] = \sum_{j=-M+1}^{M} N^{-\frac{1}{2}} \sum_{s(j)} \phi[\lambda_t;\theta_o;(\hat{\Phi}_{M,N}-\Phi_M)]\text{vec } w_v w_x^* +$$

$$+ N^{-1}\sum_B \phi[\lambda_t;\theta_o;(\hat{\Phi}_{M,N}-\Phi_M)]\text{vec } \zeta_N w_x^* = II_{M,N} + III_{M,N}, \text{ say.}$$

We can deduce from the proof of Lemma 7.6 that for each j

$$\lim_N N^{-\frac{1}{2}}\Sigma_{s(j)} \phi[\lambda_t;\theta_o;(\hat{\Phi}_{M,N}-\Phi_M)]\text{vec } w_v w_x^* \overset{p}{=} 0.$$

Hence, $\lim_N II_{M,N} \overset{p}{=} 0$. Furthermore, $\lim_N III_{M,N} \overset{p}{=} 0$ from (10.40). i.e., $\lim_N N^{\frac{1}{2}} k_N[\theta_o;(\hat{\Phi}_{M,N}-\Phi_M)] \overset{p}{=} 0$.

Because $\lim_N N^{\frac{1}{2}} k_N(\theta_o;\Phi_M) \overset{\mathcal{D}}{=} \mathcal{N}[0,\ddagger(\Phi_M f_v \Phi_M)]$ by Lemma 7.4, (10.42) follows. Given (10.41) and (10.42), the proof of (ii) is the same as that of Theorem 7.2. □

Proof of Theorem 4 Applying Lemma 4, there exists a sequence of positive integers $M_1(N) \uparrow \infty$ as $N \to \infty$ s.t. $\hat{\theta}_{M_1(N),N} \to \theta_o$. Next we note that $\lim_M \ddagger(\Phi_M) = \ddagger(\Phi)$. It follows from Lemma 1 that there exists a sequence of positive integers $M_2(N) \uparrow \infty$ as $N \to \infty$ s.t. $N^{\frac{1}{2}}(\theta_{M_2(N),N} - \theta_o) \overset{\mathcal{D}}{\to} \mathcal{N}[0,\ddagger(\Phi)]$.

Putting $M(N) = \min\{M_1(N), M_2(N)\}$, the result of the theorem follows. □

<u>*Proof of Corollary 1*</u> By Lemmas 1, 4 and 6, there exists a sequence of positive integers $M(N) \uparrow \infty$ as $N \to \infty$ s.t.

$$\sup_{\lambda\in\mathcal{B}}\|\hat{\Phi}_{M(N),N}(\lambda)-\Phi(\lambda)\| \to 0,\ \hat{\theta}_{M(N),N}\to\theta_o,\ N^{\frac{1}{2}}(\hat{\theta}_{M(N),N}-\theta_o)\xrightarrow{\mathcal{D}}N[0,\ddagger(\Phi)].$$

The proof of the analogues of Theorems 7.4 and 7.5 follow along the same lines as the respective proofs of these theorems. The method of proof consists in applying Lemma 5 to show that $\hat{\Phi}_{M(N),N}$ can be replaced by Φ. We omit the details. □

<u>*Proof of Theorem 5*</u> We note that

$$\lim_N \sup_{\lambda\in\mathcal{B}}\| \Phi(\lambda;\psi_o) - \Phi(\lambda;\hat{\psi}_N)\| = 0 \quad \text{a.s.} \tag{10.50}$$

Put $\hat{\Phi}_N(\lambda) = \Phi(\lambda;\ \hat{\psi}_N)$. We first discuss strong consistency. From Lemma 5 we will have that

$$L_N[\theta_o;(\hat{\Phi}_N-\Phi)] \to 0,\ \sup_{\theta\in\bar{\Theta}} L_{N;\eta}[\theta;(\hat{\Phi}_N-\Phi)] \to 0 \quad \forall\ \eta > 0.$$

The proof of strong consistency now follows from the proof of Theorem 7.1. We now discuss the proof of the CLT. Put $k_N(\theta;\Phi) = (\partial/\partial\theta)\ell_N(\theta;\Phi)$. From the proof of Lemma 7.4

$$-N^{\frac{1}{2}}k_N[\theta_o;(\hat{\Phi}_N-\Phi)] = N^{-\frac{1}{2}}\sum_{\mathcal{B}}A(\lambda_t)^*[I_{(q)}\otimes\{\hat{\Phi}_N-\Phi\}]\mathrm{vec}\{(w_y-Dw_x)w_x^*\}=$$

$$= [N^{-1}\sum_{\mathcal{B}}A(\lambda_t)^*[\{(w_y-Dw_x)w_x^*\}'\otimes I_{(p)}]\Gamma^\dagger(\lambda;\psi_N^\dagger)].N^{\frac{1}{2}}(\hat{\psi}_N-\psi_o)$$

$$= o(1).O_p(1) = o_p(1);$$

$\Gamma(\lambda;\psi)$ is defined in §6; $\Gamma^\dagger(\lambda;\psi_N^\dagger)$ is a symbolic derivative, in the sense of p. 84 , and $\|\psi_N^\dagger-\psi_o\| \le \|\hat{\psi}_N-\psi_o\|$. It follows from the above and Lemma 7.4 that

$$N^{\frac{1}{2}}k_N(\theta_o;\hat{\Phi}_N) \xrightarrow{\mathcal{D}} N[0,\Delta(\Phi f_v\Phi)] \tag{10.51}$$

For α small and positive, $\lim_N(\partial/\partial\theta')k_N[\theta;(\hat{\Phi}_N-\Phi)] = 0$

uniformly in $\theta \in U_\alpha$, by (10.50) and Lemma 5. It follows from Lemma 7.5 that if $\theta_N^\dagger \to \theta_o$, then

$$\lim_N (\partial/\partial\theta_j) k_N(\theta_N^\dagger;\hat{\Phi}_N) = (2\pi)^{-1}\int_B A^*(\lambda)[f_x'(\lambda)\otimes\Phi(\lambda)]A_j(\lambda)d\lambda \tag{10.52}$$

where $A_j(\lambda)$ is the j^{th} column of $A(\lambda)$.
Given (10.51) and (10.52), the proof of the CLT is the same as that of Theorem 7.2. □

Proof of Corollary 2 The proofs of the analogues of Theorems 7.4 and 7.5 follow along the same lines as the proofs of Theorems 7.4 and 7.5 respectively. The method of proof consists in applying Lemma 5 to show that $\hat{\Phi}_N$ can be replaced by Φ. We omit the details. □

To prove Theorem 6 we need the following results.

Lemma 7 (Deistler, Dunsmuir and Hannan [1977]). If $r(z)$ is a rational function of z which is analytic within a circle containing the unit circle, then

$$\int_{-\pi}^{\pi} r(e^{i\lambda})e^{-ij\lambda}d\lambda = 0, \; 0 \le j \le m$$

implies $r(z) \equiv 0$, where m is the degree of the numerator of $r(z)$.

Proof Let $r(z) = (\alpha_o+\alpha_1 z+\ldots+\alpha_m z^m)(\sum_{j=o}^{\infty} \beta_j z^j)$ for $|z| \le 1$ with $\beta_o = 1$. Suppose that $\alpha_j = 0$ $(0 \le j \le k-1)$ for $k < m$. Then,

$$(2\pi)^{-1}\int_{-\pi}^{\pi} r(e^{i\lambda})e^{-ik\lambda}d\lambda = \alpha_k\beta_o = 0 \Rightarrow \alpha_k = 0.$$

Since it is easy to verify that $\alpha_o = 0$, it follows that $\alpha_j = 0$ $(0 \le j \le m)$. □

<u>*Lemma 8*</u> (Dunsmuir [1976, p.90]). Let $f(\lambda)$ be a $p\times p$ Hermitian p.s.d. matrix function of λ, which is nonsingular for $\lambda \in [-\pi,\pi]$ except possibly at a finite number of points. For $V \geq 0$ let Γ_V be the block-matrix having $\int_{-\pi}^{\pi} e^{i(k-j)\lambda} f(\lambda)d\lambda$ as $(j,k)^{th}$ submatrix, $(j,k = 1,\ldots,V)$. Then, $\forall\ V \geq 1$, Γ_V is a p.d. matrix.

<u>*Proof*</u> Let $\{a(j),\ 1 \leq j \leq V\}$ be p dimensional vectors and put $a = \text{vec}\{a(1),\ldots,a(V)\}$

$$\lambda_{min}(\Gamma_V) = \inf_{\|a\|=1} a^*\Gamma_V a = \inf_{\|a\|=1} \int_{-\pi}^{\pi} \left(\sum_1^V a(j)e^{ij^\lambda}\right)^* f(\lambda) \left(\sum_1^V a(j)e^{ij^\lambda}\right) d\lambda \tag{10.60}$$

Now, $\forall\ \varepsilon > 0$, define $A_\varepsilon = [\lambda\in[-\pi,\pi],\ \lambda_{min}\{f(\lambda)\} \geq \varepsilon]$ and let ω_j be the j^{th} of the, R say, λ-values at which $\det f(\lambda) = 0$. Then ε may be chosen so that $[-\pi,\pi] - A_\varepsilon$ has arbitrarily small Lebesgue measure δ. This is possible because $\lambda_{min}\{f(\lambda)\}$ is a continuous function of λ. Thus the right hand side of (10.60) is not less than

$$\varepsilon \inf_{\|a\|=1} \int_{A_\varepsilon} \|\sum_1^V a(j)e^{ij^\lambda}\|^2 \geq$$

$$\geq \varepsilon \inf_{\|a\|=1} \{1 - \delta.R \sup_{1\leq j\leq R}\ \sup_{\lambda\in[\omega_j-\delta,\omega_j+\delta]} \|\sum_1^V a(j)e^{ij\lambda}\|^2\} \geq$$

$$\geq \varepsilon(1-2.V.\delta.R)$$

since $\|\sum_1^V a(j)e^{ij\lambda}\|^2 \leq (\sum_1^V \|a(j)\|)^2 \leq 2(\sum_1^V \|a(j)\|^2)$. By choosing δ s.t. $(1-2.\delta.R.V) > 0$, the required result follows. □

Let $\{P_n(z),\ n \geq 1\}$ be a sequence of matrix functions of z and $P(z)$ another matrix function of z, all matrices being of the same dimension.

Definition We shall say that the sequence $\{P_n, n \geq 1\}$ converges point wise to P if $\forall\ \rho,\ 0 < \rho < 1$,

$$\lim_n \sup_{|z| \leq \rho} \|P_n(z) - P(z)\| = 0$$

We indicate this mode of convergence by $P_n \overset{P_t}{\to} P$. The above definition follows Hannan and Dunsmuir [1976].

Suppose that $P_n(z)$ and $P(z)$ are analytic functions of z for $|z| < 1$. Then we can write

$$P_n(z) = \sum_{j=o}^{\infty} p_{j,n} z^j,\ P(z) = \sum_{j=o}^{\infty} p_j z^j \qquad |z| < 1.$$

The following result holds.

Lemma 9 $\forall\ j \geq 0$ $\lim_n \|p_{j,n} - p_j\| = 0$ if $P_n \overset{P_t}{\to} P$.

Proof Choose $0 < \rho < 1$. Then

$$p_{j,n} - p_j = (2\pi i)^{-1} \int_{|z|=\rho} \{P_n(z) - P(z)\} z^{-j-1} dz$$

$$\forall\ \varepsilon > 0,\ \exists\ N_1 = N_1(\varepsilon) \text{ s.t. } \sup_{|z| \leq \rho} \|P_n(z) - P(z)\| < \varepsilon \qquad n \geq N_1 .$$

It follows that $\|p_{j,n} - p_j\| \leq \varepsilon\, \rho^{-j}$ for $n \geq N_1$. The required result is now easily obtained. □

Proof of Theorem 6 The proof follows Dunsmuir and Hannan [1976, pp.354-355]. As in the proof of Theorem 1 we shall have that

$$\lim \sup_N L_N(\hat{\theta}_N) \leq (2\pi)^{-1} \int_{-\pi}^{\pi} \operatorname{tr}\{\Phi(\lambda) f_v(\lambda)\} d\lambda \qquad (10.70)$$

Put $b_n(z) = \det B(z;\hat{\theta}_n)$, $F_n(z) = \{\operatorname{adj} B(z;\hat{\theta}_n)\} C(z;\hat{\theta}_n)$, $P_n(\lambda) = \{b_n(\lambda) I_{(p)},\ - F_n(\lambda)\}$, $D_n(z) = D(z;\hat{\theta}_n)$ and $s(n)' = \{y(n)',\ x(n)'\}$. It follows from assumptions E3.(i),(ii) that

$$\sup_{\lambda \in [-\pi,\pi]} |b_n(\lambda)| \leq R_1 2^{pr}.$$

Hence

$$L_N(\hat{\theta}_N) = N^{-1}\sum_t \operatorname{tr}\{\Phi P_N w_s w_s^* P_N^* |b_N|^{-2}\} \geq cN^{-1}\sum_t \operatorname{tr}\{P_n w_s w_s^* P_N^*\},\ c > 0.$$

Since $P_N(\lambda)$ is a polynomial matrix in $e^{i\lambda}$ of degree $V=pr+s$ at most, i.e. $P_N(\lambda) = \sum_{o}^{V} p_N(j)e^{ij\lambda}$ say, we have that

$$L_N(\hat{\theta}_N) \geq cN^{-2}\operatorname{tr}\left[\sum_t \sum_{j,k=0}^{V} \sum_{m,n=1}^{N} p_N(j)s(m)s(n)'p_N(k)' \times \exp\{i(j-k+m-n)\lambda_t\}\right] \qquad (10.75)$$

$$N^{-2}\sum_t \sum_{j,k=o}^{V} \sum_{m,n=1}^{N} p_N(j)s(m)s(n)'p_N(k)'\exp\{i(j-k+m-n)\lambda_t\} =$$

$$= c\,N^{-1}\sum_{j,k=o}^{V} p_N(j)\left\{\sum_{m=\max(1,k-j+1)}^{\min(k-j+N,N)} s(m)s(m+j-k)' + \right.$$

$$\left. + \sum_{m=k-j+1+N}^{N} s(m)s(m+j-k-N)' + \sum_{m=1}^{k-j} s(m)s(m+j-k+N)'\right\}p_N(k)' =$$

$$= c\sum_{j,k=o}^{V} p_N(j)\{\Gamma(j-k)+G_N(j,k)\}p_N(k)',\ \text{say}$$

where $\Gamma(j) = \lim_N N^{-1}\sum_{n=1}^{N-j} s(n)s(n+j)'$. It follows from assumptions E1. - E4. that

$$\lim_N \|G_N(j,k)\| = 0 \qquad (j,k=0,\ldots,V)$$

Let Γ_V, $G_{V,N}$ be block-matrices having, respectively, $\Gamma(j-k)$ and $G_N(j,k)$ as $(j,k)^{th}$ submatrices $(j,k = 0,\ldots,V)$. Then,

$$\operatorname{tr}\left[\sum_{j,k=o}^{V} p_N(j)\{\Gamma(j-k)+G_N(j,k)\}p_N(k)'\right] \geq$$

$$\geq c\,\lambda_{\min}(\Gamma_V+G_{V,N}).\operatorname{tr}\left\{\sum_{j=o}^{V} p_N(j)p_N(j)'\right\},\ c > 0 \qquad (10.85)$$

It is not difficult to show that

$$\Gamma(j) = \int_{-\pi}^{\pi} e^{ij\lambda}f(\lambda)d\lambda$$

where

$$\dot{f}(\lambda) = \left(\begin{array}{c:c} D(\lambda)f_x(\lambda)D(\lambda)^* + f_v(\lambda) & D(\lambda)f_x(\lambda) \\ \hdashline f_x(\lambda)D(\lambda)^* & f_x(\lambda) \end{array} \right)$$

It follows that $\det \dot{f}(\lambda) = \det f_v(\lambda).\det f_x(\lambda)$. Assumptions E1.(ii) and E2.(ii) now imply that $\lambda_{min}\{\dot{f}(\lambda)\} > 0$ for $\lambda \in [-\pi,\pi]$, except, possibly, at a finite number of points. It follows from Lemma 8 that $\lambda_{min}(\Gamma_V) > 0$ and hence that

$$\lambda_{min}(\Gamma_V + G_{V,N}) \geq c > 0$$

for all N sufficiently large. By (10.75) and (10.85)

$$L_N(\hat{\theta}_N) \geq c.\mathrm{tr}\left\{ \sum_{j=o}^{V} p_N(j)p_N(j)' \right\}, \; c > 0,$$

for all N sufficiently large. Combining this with (10.70) we have that $\mathrm{tr}\left\{ \sum_{j=o}^{V} p_N(j)p_N(j)' \right\} \leq c < \infty$ for all N sufficiently large. Next we show that

$$\lim_N D_N \overset{p_t}{=} D(\equiv D(z;\theta_o)) \qquad (10.90).$$

The proof proceeds by contradiction. Suppose that (10.90) does not hold. Then, because $[\{p_N(0),\ldots,p_N(V)\},N]$ is a bounded sequence for all N sufficiently large, it has a convergent subsequence $[\{p_{m(N)}(0),\ldots,p_{m(N)}(V)\},N \geq 1]$ with $\lim_N p_{m(N)}(j) = \dot{p}(j)$, say, $0 \leq j \leq V$. Put $\dot{b}(z)=\lim_N b_{m(N)}(z)$, $\dot{F}(z)=\lim_N F_{m(N)}(z)$ and $\dot{D} = \dot{b}^{-1}\dot{F}$. It follows that $D_{m(N)} = b_{m(N)}^{-1}F_{m(N)} \overset{p_t}{\to} \dot{D}$. As in the proof of Theorem 7.1, we can now show that $\dot{b}D - \dot{F} \equiv 0$ and this gives us a contradiction. Therefore (10.90) holds. To complete the proof of the theorem consider, for given ϕ, the following set of equations for θ

$$\int_{-\pi}^{\pi} \{B(\lambda;\theta)D(\lambda;\phi) - C(\lambda;\theta)\}e^{-ij\lambda}d\lambda = 0, \quad 0 \leq j \leq pr+s, \tag{10.95}$$

Assuming that $D(z;\phi)$ does not have any poles in or on the unit circle, (10.95) becomes

$$\sum_{j=o}^{r} B_j(\theta)D_{k-j}(\phi) - C_k(\theta) = 0, \quad 0 \leq k \leq pr+s, \tag{10.96}$$

which we will write more compactly in the form $Q(\phi)\theta = 0$, say; we take $C_k(\theta) \equiv 0$ for $k > s$ and $D_k(\phi) \equiv 0$ for $k < 0$ in (10.96). We know from Lemma 7 that if θ is a solution of (10.96), then it is a solution of

$$B(\lambda;\theta)D(\lambda;\phi) - C(\lambda;\theta) = 0 \quad \forall \quad \lambda \in [-\pi,\pi].$$

It now follows from assumptions E3. and E4. that θ_o is the unique solution of $[Q(\theta_o)',H']'\theta = [0',(H\theta_o)']'$. Therefore $[Q(\theta_o)',H']'$ has full column rank. Because $\hat{\theta}_N$ is a solution of (10.96) for $\phi = \hat{\theta}_N$, it must be a solution of

$$[Q(\hat{\theta}_N)',H']'\theta = [0',(H\theta_o)']'. \tag{10.100}$$

By (10.90) and Lemma 9 $\|Q(\hat{\theta}_N) - Q(\theta_o)\| \to 0$ implying that $[Q(\hat{\theta}_N)',H']'$ has full column rank for N sufficiently large. Hence $\hat{\theta}_N$ is the unique solution of (10.100) for N large and

$$\hat{\theta}_N = [Q(\hat{\theta}_N)'Q(\hat{\theta}_N)+H'H]^{-1}H'H\theta_o \to \theta_o. \; \square$$

For the final few results we adopt the notation of §9.

Proof of Theorem 7 (i) From Lemma 7.1,

$$\lim_N \ell_{M,N}(\theta_o) = (4M)^{-1} \sum_{j=-M+1}^{M} \log \det \hat{f}_M(\omega_j)$$

Hence,

$$\lim\sup_N \ell_{M,N}(\hat\theta_{M,N}) \leq (4M)^{-1} \sum_{j=-M+1}^{M} \log\det \hat f_M(\omega_j) \quad (10.120)$$

Put

$b(z;\theta) = \det B(z;\theta), F(z;\theta)=\{\mathrm{adj}\ B(z;\theta)\}C(z;\theta)$,

$G(z;\theta) = b(z;\theta)D(z) - F(z;\theta)$, $g_N(\lambda;\theta) = b(\lambda;\theta)w_y(\lambda) -$

$- F(\lambda;\theta)w_x(\lambda)$,

$$\hat f_{M,N;\eta}(\omega_j;\theta)=2MN^{-1}\sum_{s(j)} g_N(\lambda_t;\theta)g_N(\lambda_t;\theta)^*\{|b(\lambda_t;\theta)|^2+\eta\}^{-1}\ \forall\eta>0,$$

$$\ell_{M,N;\eta}(\theta) = (4M)^{-1}\sum_{j=-M+1}^{M} \log\det \hat f_{M,N;\eta}(\omega_j;\theta).$$

If $\lim_N \hat\theta_{M,N} \to \theta_o$, then the proof of (i) is complete. If it does not, then there is a $\dot\theta \in \overline{\Theta}_h - \theta_o$ and a subsequence of $\{\hat\theta_{M,N}, N \geq 1\}$, $\{\hat\theta_{M,n(N)}, N \geq 1\}$ say, s.t. $\lim_N \hat\theta_{M,n(N)} = \dot\theta$. Put $\dot G(\lambda) = G(\lambda;\dot\theta)$, $\dot b(\lambda) = b(\lambda;\dot\theta)$. Then,

$$\lim\inf_N \ell_{M,n(N)}(\hat\theta_{M,n(N)}) \geq \lim_N \ell_{M,n(N);\eta}(\hat\theta_{M,n(N)}) =$$

$$= (4M)^{-1}\sum_{j=-M+1}^{M} \log\det\{{}^M/_\pi \int_{S(j)} \dot G f_x \dot G^*(|\dot b|^2+\eta)^{-1}d\lambda +$$

$$+ {}^M/_\pi \int_{S(j)} |\dot b|^2 f_v(|\dot b|^2 + \eta)^{-1}d\lambda\}$$

Using exactly the same reasoning as in the proof of Theorem 7.1 we have that

$$\sup_{\eta>0} {}^M/_\pi \int_{S(j)} |\dot b|^2 f_v(\lambda)(|\dot b|^2+\eta)^{-1}d\lambda = \hat f_M(\omega_j)$$

so that

$$\lim\inf_N \ell_{M,n(N)}(\hat\theta_{M,n(N)}) \geq (4M)^{-1}\sum_{j=-M+1}^{M} \log\det\{{}^M/_\pi \times$$

$$\times \int_{S(j)} \dot G f_x \dot G^*(|\dot b|^2+\eta)^{-1}d\lambda + \hat f_M(\omega_j)\}$$

It now follows from (10.120) that $\dot G f_x \dot G^* = 0$ for $\lambda \in \mathcal{B}$ and hence $\dot G(\lambda) = 0$ for $\lambda \in \mathcal{B}_1$. By assumption A4.(ii)

$\dot{\theta} = \theta_o$, giving us a contradiction.

Hence $\lim_N \hat{\theta}_{M,N} = \theta_o$.

(ii) The proof of (ii) follows from Lemma 4. □

Lemma 10 Put $\Psi(\lambda;\theta)=\{D(\lambda)-D(\lambda;\theta)\}f_x(\lambda)\{D(\lambda)-D(\lambda;\theta)\}^*+$
$+ f_v(\lambda), T(\lambda;\theta)_\ell = \{(\partial/\partial\theta_\ell)D(\lambda;\theta)\}f_x(\lambda)\{D(\lambda)-D(\lambda;\theta)\}^* +$
$+ \{D(\lambda)-D(\lambda;\theta)\}f_x(\lambda)\{(\partial/\partial\theta_\ell)D^*(\lambda;\theta)\}$

$\Psi_M(\lambda;\theta) = {}^M/_\pi \int_{S(j)} \Psi(\lambda;\theta)d\lambda$ for $\lambda \in S(j)$,

$T_M(\lambda;\theta)_\ell = {}^M/_\pi \int_{S(j)} T(\lambda;\theta)_\ell d\lambda$ for $\lambda \in S(j)$,

$\alpha 2 = \frac{1}{2}\alpha 1$.

Given conditions A1. - A7. of §7 we have that

(i) $\lim_N \hat{f}_{M,N}(\lambda;\theta) = \Psi_M(\lambda;\theta)$ a.s.

(ii) $\lim_M \Psi_M(\lambda;\theta) = \Psi(\lambda;\theta)$

(iii) $\lim_N (\partial/\partial\theta_\ell)\hat{f}_{M,N}(\lambda;\theta) = -T_M(\lambda;\theta)_\ell$ a.s. $1 \le \ell \le m$

(iv) $\lim_M T_M(\lambda;\theta)_\ell = T(\lambda;\theta)_\ell$ $1 \le \ell \le m$

the convergence in (i) - (iv) being uniform in $\theta \in \overline{U}_{\alpha 2}$ and $\lambda \in \mathcal{B}$.

Proof (i) and (iii) follow from Lemma 7.1 while (ii) and (iv) are immediate. □

Proof of Theorem 8 It follows from the proof of Theorem 7.2 that the result of Theorem 8(i) holds if we can show that

$$\lim_N N^{\frac{1}{2}}k_{M,N}(\theta_o) \overset{\mathcal{D}}{=} N[0,\Delta(\hat{f}_M^{-1}f_v\hat{f}_M^{-1})] \qquad (10.130)$$

and

$$\lim_N (\partial/\partial\theta')k_{M,N}(\theta_N^\dagger) \overset{P}{=} \Delta(\hat{f}_M^{-1}) \qquad (10.131)$$

if $\theta_N^\dagger \overset{P}{\to} \theta_o$.

From Lemma 10(i) we have that

$$\lim_N \hat{f}_{M,N}(\lambda) = \hat{f}_M(\lambda) \qquad (10.132)$$

uniformly in $\lambda \in \mathcal{B}$.

$$(\partial/\partial\theta_\alpha)\ell_{M,N}(\theta) = (4M)^{-1} \sum_j \mathrm{tr}\, \hat{f}^{-1}_{M,N}(\omega_j;\theta)\{(\partial/\partial\theta_\alpha)\hat{f}_{M,N}(\omega_j;\theta)\} =$$

$$= -N^{-1} \sum_j \sum_{s(j)} \{(\partial/\partial\theta_\alpha)\mathrm{vec}\, D(\lambda_t;\theta)\}^*\{\bar{w}_x \otimes \hat{f}^{-1}_{M,N}(\lambda_t;\theta)\}.$$

$$.\{w_y - D(\lambda_t;\theta)w_x\}$$

$$\Rightarrow k_{M,N}(\theta_o) = -N^{-1} \sum_j \sum_{s(j)} A^*(\bar{w}_x \otimes \hat{f}^{-1}_{M,N})(w_y - Dw_x) =$$

$$= -N^{-\frac{1}{2}}[N^{-\frac{1}{2}} \sum_j \sum_{s(j)} A^*(\bar{w}_x \otimes \hat{f}^{-1}_M)(w_y - Dw_x) -$$

$$- N^{-\frac{1}{2}} \sum_j \sum_{s(j)} A^*\{\bar{w}_x \otimes (\hat{f}^{-1}_{M,N} - \hat{f}^{-1}_M)\}(w_y - Dw_x)] =$$

$$= N^{-\frac{1}{2}}(II_N + III_N) \text{ say.}$$

It follows from Lemma 7.4 that $\lim_N II_N \overset{\mathcal{D}}{=} N[0, (\hat{f}^{-1}_M f_v \hat{f}^{-1}_M)]$

To show that $III_N \overset{p}{\to} 0$ it is sufficient to show that for each j, $-M < j \le M$,

$$N^{-\frac{1}{2}} \sum_{s(j)} A^*\{\bar{w}_x \otimes (\hat{f}^{-1}_{M,N} - \hat{f}^{-1}_M)\}(w_y - Dw_x) =$$

$$N^{-\frac{1}{2}} \sum_{s(j)} [\{(w_y - Dw_x)w_x^*\}' \otimes I_{(p)}]\mathrm{vec}(\hat{f}^{-1}_{M,N} - \hat{f}^{-1}_M) \overset{p}{\to} 0 \qquad (10.140)$$

(10.140) follows from Lemma 7.2, Lemma 7.6 and (10.132) by noting that $(\hat{f}^{-1}_{M,N} - \hat{f}^{-1}_M)$ is constant on $S(j)$. Thus (10.130) holds.

Next we show that (10.131) holds

$$-\frac{\partial k_{M,N}(\theta)}{\partial\theta_\ell} = N^{-1} \sum_j \sum_{s(j)} \frac{\partial A^*}{\partial\theta_\ell}(\bar{w}_x \otimes \hat{f}^{-1}_{M,N})(w_y - Dw_x) -$$

$$- N^{-1} \sum_j \sum_{s(j)} A^*\left\{\bar{w}_x \otimes \hat{f}^{-1}_{M,N} \frac{\partial \hat{f}_{M,N}}{\partial\theta_\ell} \hat{f}^{-1}_{M,N}\right\}(w_y - Dw_x) -$$

$$- N^{-1} \sum_{j} \sum_{s(j)} A^*(I_x' \otimes \hat{f}_{M,N}^{-1}) \frac{\partial \text{ vec } D}{\partial \theta_\ell} \qquad (10.141)$$

$$= II_N(\theta) - III_N(\theta) - IV_N(\theta), \text{ say.}$$

For convenience we have omitted to indicate dependence on λ and θ in (10.141).

Let $\Psi_M(\lambda;\theta)$ be defined as in Lemma 10.

$$II_N(\theta) \to \sum_j {}^{M}/_{\pi} \int_{S(j)} \frac{\partial A^*(\lambda;\theta)}{\partial \theta_\ell} \text{vec}[\Psi_M(\lambda;\theta)\{D(\lambda)-D(\lambda;\theta)\}f_x(\lambda)]d\lambda,$$

uniformly in $\theta \in \overline{U}_{\alpha 2}$, by Lemma 7.1 and Lemma 10(i). It follows that $II_N(\theta_N^{\dagger}) \to 0$. We can similarly show that $III_N(\theta_N^{\dagger}) \to 0$ (use Lemma 10(i), (iii)).

Finally,

$$IV_N(\theta) \to \sum_j {}^{M}/_{\pi} \int_{S(j)} A^*(\lambda;\theta)\{f_x'(\lambda) \otimes \Psi_M(\lambda;\theta)\}A_\ell(\lambda;\theta)d\lambda$$

uniformly in $\theta \in \overline{U}_{\alpha 2}$, where $A_\ell(\lambda;\theta)$ is the ℓ^{th} column of $A(\lambda;\theta)$. (10.131) follows.

(ii) This follows from (i) and Lemma 1. □

Proof of Theorem 9 Let $\Psi(\lambda;\theta)$, $T(\lambda;\theta)_\ell$ $(1 \leq \ell \leq m)$ be defined as in Lemma 10. From Lemma 4 and Lemma 10 we know that there exists a sequence of positive integers $M_1(N) \uparrow \infty$ as $N \to \infty$ s.t.

$$\lim_N \hat{f}_{M_1(N),N}(\lambda;\theta) = \Psi(\lambda;\theta) \qquad (10.150)$$

$$\lim_N \frac{\partial f_{M_1(N),N}(\lambda;\theta)}{\partial \theta_\ell} = - T(\lambda;\theta)_\ell \quad 1 \leq \ell \leq m \qquad (10.151)$$

uniformly in $\theta \in \overline{U}_{\alpha 2}$, $\lambda \in \mathcal{B}$.

It now follows from Lemma 4 and Theorems 7 and 8 that there exists a sequence of positive integers $M(N) \uparrow \infty$ as $N \to \infty$ s.t.

$\hat{\theta}_{M(N),N} \to \theta_o$, $N^{\frac{1}{2}}(\hat{\theta}_{M(N),N}-\theta_o) \xrightarrow{\mathcal{D}} N[0,\ddagger(f_v^{-1})]$ and (10.150), (10.151) hold with $M_1(N)$ replaced by $M(N)$. Given the sequence $M(N)$, the following results can be easily deduced from the proof of Theorem 8.

$$k_{M(N),N}(\theta_o) \to 0$$

For $\alpha(>0)$ sufficiently small, $\limsup_N \sup_{\theta\in U_\alpha} \|k_{M(N),N}(\theta)\| < \frac{1}{4}$.

$\forall\ \varepsilon > 0$, $\exists\ \alpha$, $0 < \alpha \leq \alpha 1$ s.t.

$$\limsup_N \sup_{\theta\in U_\alpha} \|\Delta_{M(N),N}(\theta)-\Delta(\theta)\| < \varepsilon$$

$$\limsup_N \sup_{\theta\in U_\alpha} \left\|\frac{\partial k_{M(N),N}(\theta)}{\partial\theta'} - \Delta_{M(N),N}(\theta)\right\| < \varepsilon$$

Also, $\exists\ \alpha > 0$ s.t.

$$\liminf_N \inf_{\theta\in U_\alpha} \lambda_{\min}\{J_{M(N),N}(\theta)'J_{M(N),N}(\theta)\} \geq c > 0.$$

Given the above results, the proof of Theorem 9 can be completed similarly to the proof of Theorem 7.4. We omit the details. □

CHAPTER 9

ON THE RELATIVE EFFICIENCY OF TWO METHODS OF ESTIMATING A DYNAMIC SIMULTANEOUS EQUATIONS MODEL

1. INTRODUCTION

In this chapter we deal with an identified, dynamic, simultaneous equations model

$$\sum_{j=o}^{r} B_j y(n-j) = \sum_{j=o}^{s} C_j x(n-j) + u(n) \qquad (1.10)$$

and we assume that the residuals are generated by either a stationary MA

$$u(n) = \sum_{j=o}^{t} A_j \, e(n-j) \qquad (1.11)$$

or a stationary AR

$$\sum_{j=o}^{t} A_j \, u(n-j) = e(n) \qquad (1.12)$$

Given the particular residual structure, either (1.11) or (1.12), Chapters 4 and 5 dealt with the efficient estimation of the coefficient matrices A_j, B_j and C_j. We will call such estimators *efficient parametric estimators*, where the word parametric is meant to convey that we take into account the particular parametric nature of the residuals when estimating. We remind the reader (see 1.§1) that by an efficient estimator we mean one that has the same asymptotic distribution as the maximum likelihood estimator if the $e(n)$ are really Gaussian.

An alternative way of estimating the coefficient matrices B_j and C_j, which does not rely on the parameterization of the residuals $u(n)$ but just assumes that they are stationary, is given in Chapters 7 and 8. The estimators of the B_j and C_j

matrices obtained using this method of estimation will be called *nonparametric estimators* for the obvious reason.

Given two estimators $\delta_N^{(1)}$ and $\delta_N^{(2)}$ of a vector δ_o, with $N^{\frac{1}{2}}(\delta_N^{(1)}-\delta_o) \xrightarrow{\mathcal{D}} N(0,\Sigma_1)$ and $N^{\frac{1}{2}}(\delta_N^{(2)}-\delta_o) \xrightarrow{\mathcal{D}} N(0,\Sigma_2)$, we will say that $\delta_N^{(1)}$ is asymptotically more efficient than $\delta_N^{(2)}$ if $(\Sigma_2-\Sigma_1)$ is p.s.d. and not zero.

In the presence of general nonlinear restrictions on the parameters we will show that the parametric estimators of the coefficient matrices B_j and C_k are at least as efficient asymptotically as, and generally more efficient than, the non-parametric ones, *when the parametric treatment is correct*.

The results verify that the estimator given by Hannan and Terrell [1973] will be as asymptotically efficient as the parametric estimators. See Remark 4.2(i) of §4. Some of our results have already been obtained, in a different way, by Espasa [1975] and Espasa and Sargan [1975]. This is discussed in more detail in Remark 4.2(ii) of §4.

We now give a brief outline of the contents of the chapter. §2 states the assumptions required for the results to hold; §3 derives the asymptotic covariance matrices; §4 states the results of the chapter, while §5 contains all the proofs.

Finally, we give some notation and conventions used in the chapter.

1. z will be used both as a lag operator, i.e. $zy(n) = y(n-1)$, and as a complex variable. Its usage will be clear from the context.

2. For a function of the complex variable z, g(z) say, we will write $g(e^{i\lambda})$ as $g(\lambda)$.

3. $r(A)$ will mean the rank of the matrix A. If A is a p.s.d. matrix we will write $A > 0$ if $A \neq 0$; if A and B are matrices such that $(A-B)$ is p.s.d., then we will write $A > B$ if $(A-B) > 0$.

2. ASSUMPTIONS

In (1.10) - (1.12) the B_j matrices are $p \times p$, $0 \leq j \leq r$; the C_j matrices are $p \times q$, $0 \leq j \leq s$; the A_j matrices are $p \times p$, $0 \leq j \leq t$; the $\{y(n)\}$ are the p dimensional endogenous variables; the $\{x(n)\}$ are the q dimensional exogenous variables.

Put $B(z) = \sum_{j=o}^{r} B_j z^j$, $C(z) = \sum_{j=o}^{s} C_j z^j$ and $A(z) = \sum_{j=o}^{t} A_j z^j$.

As long as the asymptotic results outlined in the next section hold, it is immaterial what the exact assumptions on the $\{e(n)\}$ and $\{x(n)\}$ are. To be specific, however, we assume that:

A1. $(e(n), G_n, n = 0,\pm 1,\ldots)$ is a sequence of p dimensional, stationary, martingale differences satisfying $E(e(n)e(n)'|G_{n-1}) = \Omega$, where Ω is a constant nonsingular matrix.

A2. The $\{x(n)\}$ satisfy assumption A2. of 4.§2 and, in addition, $E(\|x(n)\|^2) \leq c < \infty \;\forall\; n$.

A3.(i) The degrees r, s and t of the polynomials $B(z)$, $C(z)$ and $A(z)$ respectively are chosen prior to estimation. If r', s' and t' are the degrees of the respective polynomials at the true parameter point, then we assume that $r' \leq r$, $s' \leq s$ and $t' \leq t$.

(ii) $\det B(z) \neq 0$ for $|z| \leq 1$; $\det A(z) \neq 0$ for $|z| \leq 1$

and $A_o \equiv I_{(p)}$. For the simultaneous equations case, i.e. $B_o \not\equiv I_{(p)}$, put $\theta = \text{vec}(A_1,\ldots,A_t, B_o,\ldots,B_r,C_o,\ldots,C_s)$ and $m = p^2(t+r+1) + pq(s+1)$. In the reduced form case, i.e. $B_o \equiv I_{(p)}$, put $\theta = \text{vec}(A_1,\ldots,A_t, B_1,\ldots,B_r,C_o,\ldots,C_s)$ and $m = p^2(r+t) + pq(s+1)$. Let θ_o be the true value of θ. m is the dimension of θ.

A4. The constraints $h(\theta) = 0$ are imposed during estimation. $h(\theta)$ is an f $(0 \leq f < m)$ vector function of θ with continuous first and second derivatives. Put $H = (\partial/\partial\theta')h(\theta_o)$. We assume H has rank f.

A5. In addition we assume that if $B(z)^{-1}C(z)$ is known, then, given A3. and A4., additional identification restrictions are available to determine $B(z)$ and $C(z)$ uniquely. □

A5. is made vague on purpose, because there are available a number of essentially equivalent identification requirements (see Chapter 2) and we leave it to the reader to choose the set he prefers.

For both the AR and MA cases assumption A5. ensures that $A(z)$ and Ω are also identified. This follows from Ch. 2

Put $\delta = \text{vec}(B_o,\ldots,B_r, C_o,\ldots,C_s)$ in the simultaneous equations case, $\delta = \text{vec}(B_1,\ldots,B_r, C_o,\ldots,C_s)$ in the non-simultaneous one, and let δ_o be the true value of δ. Without loss of generality we can restrict our attention to those non-linear constraints that only involve δ, as only these can be applied in both parametric and nonparametric estimation. The other nonlinear constraints involving elements of the A_j matrices can only enhance the relative efficiency of the parametric estimators. We will, therefore, assume in the sequel that $h = h(\delta)$ and put $H_1 = (\partial/\partial\delta')h(\delta_o)$. Hence $H = (\partial/\partial\theta')h(\delta_o) = (0,H_1)$ with H_1 of full row rank.

3. DERIVING THE COVARIANCE MATRICES

Put $\beta = \text{vec}(B_o,\ldots,B_r)$, $\chi_\beta(z)' = (1,z,\ldots,z^r)$ for $B_o \not\equiv I_{(p)}$ and $\beta = \text{vec}(B_1,\ldots,B_r)$, $\chi_\beta(z)' = (z,\ldots,z^r)$ for $B_o \equiv I_{(p)}$. Put $\gamma = \text{vec}(C_o,\ldots,C_s)$, $\chi_\gamma(z)' = (1,z,\ldots,z^s)$, $\alpha = \text{vec}(A_1,\ldots,A_t)$, $\chi_\alpha(z)' = (z,\ldots,z^t)$, $L = I_{(m)} - H'(HH')^{-1}H'$, $L_1 = I_{(\tilde{m})} - H_1'(H_1H_1')^{-1}H_1$ where $\tilde{m}$ is the dimension of δ, $D(z) = B(z)^{-1}C(z)$.

Then, it is easy to show that

$$L = \begin{pmatrix} I_{(p^2t)} & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & L_1 \end{pmatrix} \tag{3.10}$$

We will now derive the asymptotic covariance matrices.

Nonparametric Estimation We define the stationary sequence $\{v(n)\}$ constructively by

$$\sum_{j=o}^{r} B_j\, v(n-j) = u(n) \quad , \quad n = 0,\pm1,\ldots$$

and let $f_v(\lambda)$ be its spectral density matrix. Let $w_y(\lambda)$ and $w_x(\lambda)$ be the Finite Fourier Transforms of $y(n)$ and $x(n)$ respectively (see 1.§2). If N is the sample size, then the nonparametric estimate of δ_o, $\hat{\delta}_N^{(NP)}$ say, is obtained by minimizing

$$\sum_{-\frac{1}{2}N<t\le[\frac{1}{2}N]} \text{tr}\ f_v^{-1}\{w_y-D(\lambda_t)w_x(\lambda_t)\}\{w_y(\lambda_t)-D(\lambda_t)w_x(\lambda_t)\}^*, \lambda_t=2\pi t/N,$$

for δ, subject to the constraints $h(\delta) = 0$. It follows from 7.§2 and (7.7.40) that

$$N^{\frac{1}{2}}(\hat{\delta}_N^{(NP)} - \delta_o) \xrightarrow{\mathcal{D}} N(0,\Phi_\delta^{(NP)}) \tag{3.15}$$

where

$$\ddagger_{\delta}^{(NP)} = (L_1 \Delta_{\delta\delta}^{(NP)} L_1)^{+}, \tag{3.16}$$

and

$$\Delta_{\delta\delta}^{(NP)} = (2\pi)^{-1} \int_{-\pi}^{\pi} F(\lambda)'\{f_x(\lambda) \otimes f_v^{-1'}(\lambda)\}\overline{F(\lambda)}d\lambda; \tag{3.17}$$

$$F(\lambda) = (\partial/\partial\delta')\text{vec } D(\lambda).$$

If f_v is unknown, then, as shown in 8.§6, it can be replaced by a suitable estimate and (3.15) will still hold. It is also shown in 8.§6 how to construct the estimate of f_v.

We now obtain an explicit expression for $\Delta_{\delta\delta}^{(NP)}$. It is easy to show that

$$\text{vec } B(z)=\{\chi_\beta(z)' \otimes I_{(p^2)}\}\beta, \; B_o \not\equiv I_{(p)}, \tag{3.20}$$

$$\text{vec } B(z)=\{\chi_\beta(z)' \otimes I_{(p^2)}\}\beta + \text{vec } I_{(p)}, B_o \equiv I_{(p)}, \tag{3.21}$$

$$\text{vec } C(z)=\{\chi_\gamma(z)' \otimes I_{(pq)}\}\gamma \tag{3.22}$$

Let δ_j be the j^{th} element of δ. Then,

$$\frac{\partial \text{ vec } D(\lambda)}{\partial \delta_j} = -\text{vec}\left\{B^{-1}(\lambda)\frac{\partial B(\lambda)}{\partial \delta_j} D(\lambda)\right\} + \text{vec}\left\{B^{-1}(\lambda) \frac{\partial C(\lambda)}{\partial \delta_j}\right\} =$$

$$= -\{D(\lambda)' \otimes B^{-1}(\lambda)\}\frac{\partial \text{ vec } B(\lambda)}{\partial \delta_j} + \{I_{(q)} \otimes B^{-1}(\lambda)\}\frac{\partial \text{ vec } C(\lambda)}{\partial \delta_j}$$

Using (3.20) - (3.22) we have

$$F(\lambda)=[-\{\chi_\beta(\lambda)' \otimes D(\lambda)' \otimes B^{-1}(\lambda)\}, \; \{\chi_\gamma(\lambda)' \otimes I_{(q)} \otimes B^{-1}(\lambda)\}].$$

It now follows from (3.17) that

$$\Delta_{\delta\delta}^{(NP)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \left(\begin{array}{c:c} \{\chi_\beta \chi_\beta^* \otimes D f_x D^* \otimes f_u^{-1'}\} & -\{\chi_\beta \chi_\gamma^* \otimes D f_x \otimes f_u^{-1'}\} \\ \cdots & \cdots\cdots \\ -\{\chi_\gamma \chi_\beta^* \otimes f_x D^* \otimes f_u^{-1'}\} & \{\chi_\gamma \chi_\gamma^* \otimes f_x \otimes f_u^{-1'}\} \end{array} \right) d\lambda$$

where f_u is the spectral density matrix of the $u(n)$ sequence.

Parametric Estimation - MA Residuals

Put $\xi(n;\theta) = B_o^{-1}A(z)^{-1}\{B(z)y(n)-C(z)x(n)\}$, $\xi(n) = \xi(n;\theta_o) = B_o^{-1}e(n)$, $\Omega_\xi = B_o^{-1}\Omega B_o^{-1'}$. Let $\hat{\theta}_N^{(MA)} = (\hat{\alpha}_N^{(MA)'}, \hat{\delta}_N^{(MA)'})'$ be the estimate of θ_o obtained by minimizing

$$\log \det\left\{N^{-1} \sum_{n=1}^{N} \xi(n;\theta)\xi(n;\theta)'\right\}$$

with respect to θ, subject to the constraints $h(\delta) = 0$.

By Theorem 4.2

$$N^{\frac{1}{2}}(\hat{\theta}_N^{(MA)} - \theta_o) \xrightarrow{\mathcal{D}} N(0,\ddagger_\theta^{(MA)}) \tag{3.30}$$

where

$$\ddagger_\theta^{(MA)} = (L\Delta_{\theta\theta}^{(MA)}L)^+ \tag{3.31}$$

and $\Delta_{\theta\theta}^{(MA)}$ is given by

$$\lim_N N^{-1} \sum_{n=1}^{N} \frac{\partial\xi(n;\theta_o)'}{\partial\theta} \Omega_\xi^{-1} \frac{\partial\xi(n;\theta_o)}{\partial\theta'} .$$

It follows from (3.30) that $N^{\frac{1}{2}}(\hat{\delta}_N^{(MA)} - \delta_o) \xrightarrow{\mathcal{D}} N(0,\ddagger_\delta^{(MA)})$ where an expression for $\ddagger_\delta^{(MA)}$ is given in (3.46).

We now obtain an explicit expression for $\Delta_{\theta\theta}^{(MA)}$ when $B_o \neq I_{(p)}$. The expression for the reduced form case is then deduced from it.

It is easy to see that

$$\text{vec } A(z) = \{\chi_\alpha(z)'\otimes I_{(p^2)}\}\alpha + \text{vec } I_{(p)} \tag{3.35}$$

Let α_j, β_j be the j^{th} elements of α and β respectively.

Then,

$$(\partial/\partial\alpha_j)\xi(n;\theta_o) = -B_o^{-1}A(z)^{-1}\frac{\partial A(z)}{\partial\alpha_j}A(z)^{-1}\{B(z)y(n)-C(z)x(n)\}$$

$$= -\{e(n)'\otimes B_o^{-1}A(z)^{-1}\}\frac{\partial \text{ vec } A(z)}{\partial\alpha_j}.$$

From (3.35)

$$(\partial/\partial\alpha')\xi(n;\theta_o) = -\{\chi_\alpha(z)'\otimes e(n)'\otimes B_o^{-1}A(z)^{-1}\} \tag{3.36}$$

$$\frac{\partial\xi(n;\theta_o)}{\partial\beta_j} = -B_o^{-1}\frac{\partial B_o}{\partial\beta_j}\xi(n) + B_o^{-1}A(z)^{-1}\frac{\partial B(z)}{\partial\beta_j}y(n)$$

$$= -\{\xi(n)'\otimes B_o^{-1}\}\frac{\partial \text{ vec } B_o}{\partial\beta_j} + \{y(n)'\otimes B_o^{-1}A(z)^{-1}\}\frac{\partial \text{ vec } B(z)}{\partial\beta_j}$$

From (3.20)

$$(\partial/\partial\beta')\xi(n;\theta_o) = -[\{\xi(n)'\otimes B_o^{-1}\},0]+\{\chi_\beta(z)'\otimes y(n)'\otimes B_o^{-1}A(z)^{-1}\} \tag{3.37}$$

We can similarly show that

$$(\partial/\partial\gamma')\xi(n;\theta_o) = \{\chi_\gamma(z)'\otimes x(n)'\otimes B_o^{-1}A(z)^{-1}\} \tag{3.38}$$

Denote

$$\lim_N N^{-1}\sum_{n=1}^{N}\frac{\partial\xi(n;\theta_o)'}{\partial\alpha}\Omega_\xi^{-1}\frac{\partial\xi(n;\theta_o)}{\partial\alpha'},$$

$$\lim_N N^{-1}\sum_{n=1}^{N}\frac{\partial\xi(n;\theta_o)'}{\partial\alpha}\Omega_\xi^{-1}\frac{\partial\xi(n;\theta_o)}{\partial\delta'} \quad \text{etc. by } \Delta_{\alpha\alpha}^{(MA)},$$

$\Delta_{\alpha\delta}^{(MA)}$ etc., respectively. Let $f_e(\lambda)$, $f_\xi(\lambda)$ be the spectral densities of the $e(n)$ and $\xi(n)$ respectively, and let $f_{e\xi}(\lambda)$, $f_{eu}(\lambda)$ be the cross-spectra between $e(n)$, $\xi(n)$ and $e(n)$, $u(n)$ respectively.

From (3.36) - (3.38) we have that

$$\Delta_{\alpha\alpha}^{(MA)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\alpha \chi_\alpha^* \otimes f_e \otimes f_u^{-1'}\} d\lambda$$

$$\Delta_{\beta\alpha}^{(MA)'} = \Delta_{\alpha\beta}^{(MA)} = -(2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\alpha \chi_\beta^* \otimes f_{eu} \otimes f_u^{-1'}\} d\lambda +$$

$$+ [(2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\alpha \otimes f_{e\xi} \otimes A^{-1'} (\Omega/2\pi)^{-1}\} d\lambda,\ 0] \qquad (3.40)$$

$$\Delta_{\gamma\alpha}^{(MA)'} = \Delta_{\alpha\gamma}^{(MA)} = 0,\ \Delta_{\delta\alpha}^{(MA)'} = \Delta_{\alpha\delta}^{(MA)} = (\Delta_{\alpha\beta}^{(MA)}, \Delta_{\alpha\gamma}^{(MA)})$$

$$\Delta_{\delta\delta}^{(MA)} = \Delta_{\delta\delta}^{(NP)} + \Psi_\delta^{(MA)} \qquad (3.41)$$

where $$\Psi_\delta^{(MA)} = \begin{bmatrix} \Psi_\beta^{(MA)} & \vdots & 0 \\ \cdots & \cdot & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \qquad (3.42)$$

with

$$\Psi_\beta^{(MA)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\beta \chi_\beta^* \otimes B^{-1} f_u B^{-1*} \otimes f_u^{-1'}\} d\lambda -$$

$$-(\Phi_1^{(MA)}, 0) - (\Phi_1^{(MA)}, 0')' + \begin{bmatrix} \Phi_2^{(MA)} & \vdots & 0 \\ \cdots & \cdot & \cdots \\ 0 & \vdots & 0 \end{bmatrix};$$

$$\Phi_1^{(MA)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\beta \otimes B^{-1} f_{u\xi} \otimes A^{-1'} (\Omega/2\pi)^{-1}\} d\lambda$$

$$\Phi_2^{(MA)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{f_\xi \otimes (\Omega/2\pi)^{-1}\} d\lambda$$

Putting together the above results we can obtain $\Delta_{\theta\theta}^{(MA)}$. It follows from (3.10) that

$$L\Delta_{\theta\theta}^{(MA)} L = \begin{bmatrix} \Delta_{\alpha\alpha}^{(MA)} & \vdots & \Delta_{\alpha\delta}^{(MA)} L_1 \\ \cdots & \cdot & \cdots \\ L_1 \Delta_{\delta\alpha}^{(MA)} & \vdots & L_1 \Delta_{\delta\delta}^{(MA)} L_1 \end{bmatrix} \qquad (3.45)$$

To obtain $\ddagger_\delta^{(MA)}$ we need the following lemma due to Rohde [1965]

Lemma 1 Let H be a square p.s.d. matrix. If

$$H = \begin{pmatrix} H_{11} & \vdots & H_{12} \\ \cdots & \vdots & \cdots \\ H_{21} & \vdots & H_{22} \end{pmatrix} \quad \text{and} \quad H^+ = \begin{pmatrix} H^{11} & \vdots & H^{12} \\ \cdots & \vdots & \cdots \\ H^{21} & \vdots & H^{22} \end{pmatrix},$$

where H_{22} and H^{22} are square matrices of the same dimension, then $H^{22} = (H_{22} - H_{21}H_{11}^{+}H_{12})^{+}$. □

It follows from (3.31), (3.45) and Lemma 1 that

$$\ddagger_{\delta}^{(MA)} = [L_1(\Delta_{\delta\delta}^{(MA)} - \Delta_{\delta\alpha}^{(MA)}\Delta_{\alpha\alpha}^{(MA)+}\Delta_{\alpha\delta}^{(MA)})L_1]^{+} \tag{3.46}$$

In the reduced form case (3.45) and (3.46) still hold, with the second term on the right in (3.40) and $\Psi_{\delta}^{(MA)}$ dropped.

Parametric Estimation - AR Residuals

Put $\xi(n;\theta) = B_o^{-1}A(z)\{B(z)y(n) - C(z)x(n)\}$, $\xi(n) = \xi(n;\theta_o) = B_o^{-1}e(n)$, $\Omega_\xi = B_o^{-1}\Omega B_o^{-1'}$. Let $\hat{\theta}_N^{(AR)} = (\hat{\alpha}_N^{(AR)'}, \hat{\delta}_N^{(AR)'})'$ be the estimate of θ_o obtained by minimizing

$$\log\det\{N^{-1}\sum_{n=1}^{N}\xi(n;\theta)\xi(n;\theta)'\}$$

with respect to θ, subject to the constraints $h(\delta) = 0$.

By Theorem 4.2

$$N^{\frac{1}{2}}(\hat{\theta}_N^{(AR)} - \theta_o) \xrightarrow{\mathcal{D}} N(0, \ddagger_{\theta}^{(AR)})$$

where $\ddagger_{\theta}^{(AR)} = (L\Delta_{\theta\theta}^{(AR)}L)^{+}$. Expressions for $\Delta_{\alpha\alpha}^{(AR)}$, $\Delta_{\alpha\delta}^{(AR)}$, $\Delta_{\delta\delta}^{(AR)}$ are given below and are obtained similarly to the analogous terms in the MA case. From these we can deduce $\Delta_{\theta\theta}^{(AR)}$.

Let $f_{\xi u}$ be the cross-spectrum between $\xi(n)$ and $u(n)$. f_u and f_ξ are defined as above.

$$\Delta_{\alpha\alpha}^{(AR)} = (2\pi)^{-1}\int_{-\pi}^{\pi}\{\chi_\alpha\chi_\alpha^{*}\otimes f_u\otimes(\Omega/2\pi)^{-1}\}d\lambda$$

$$\Delta_{\beta\alpha}^{(AR)'} = \Delta_{\alpha\beta}^{(AR)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\alpha \chi_\beta^* \otimes f_u B^{-1*} \otimes (\Omega/2\pi)^{-1} \overline{A(\lambda)}\} d\lambda -$$

$$-[(2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\alpha \otimes f_{u\xi} \otimes (\Omega/2\pi)^{-1}\} d\lambda,\ 0] \qquad (3.50)$$

$$\Delta_{\gamma\alpha}^{(AR)'} = \Delta_{\alpha\gamma}^{(AR)} = 0,\ \Delta_{\delta\alpha}^{(AR)'} = \Delta_{\alpha\delta}^{(AR)} = (\Delta_{\alpha\beta}^{(AR)}, \Delta_{\alpha\gamma}^{(AR)})$$

$$\Delta_{\delta\delta}^{(AR)} = \Delta_{\delta\delta}^{(NP)} + \Psi_\delta^{(AR)},\ \Psi_\delta^{(AR)} = \begin{pmatrix} \Psi_\beta^{(AR)} & \cdot & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdot & 0 \end{pmatrix}$$

$$\Psi_\beta^{(AR)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\beta \chi_\beta^* \otimes B^{-1} f_u B^{-1*} \otimes f_u^{-1'}\} d\lambda - (\Phi_1^{(AR)}, 0) -$$

$$- (\Phi_1^{(AR)}, 0')' + \begin{pmatrix} \Phi_2^{(AR)} & \cdot & 0 \\ \cdots & \vdots & \cdots \\ 0 & \cdot & 0 \end{pmatrix}$$

with

$$\Phi_1^{(AR)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{\chi_\beta \otimes B^{-1} f_{u\xi} \otimes A'(\Omega/2\pi)^{-1}\} d\lambda$$

$$\Phi_2^{(AR)} = (2\pi)^{-1} \int_{-\pi}^{\pi} \{f_\xi \otimes (\Omega/2\pi)^{-1}\} d\lambda.$$

It is easy to see that (3.45) and (3.46) also hold for the AR case if the (MA) superscripts are replaced by the (AR) superscripts. In the reduced form case the second term on the right in (3.50) and $\Psi_\delta^{(AR)}$ are dropped.

4. RESULTS

We will state and discuss the results in this section and give the proofs in §5.

In estimating δ parametrically or nonparametrically we have allowed for the possibility that the degree of one or more of the polynomials $A(z)$, $B(z)$ or $C(z)$ is overstated (see assumption A3.(i)). The degrees of $B(z)$, $C(z)$ and $A(z)$ at the true parameter point are r', s', t' respectively, whereas we estimate on the assumption that they are r, s and t, with $r' \le r$, $s' \le s$, $t' \le t$.

Put

$$L_1 = \begin{bmatrix} L_{11} & \vdots & L_{12} \\ \cdots & \vdots & \cdots \\ L_{21} & \vdots & L_{22} \end{bmatrix} \qquad (4.10)$$

where L_{11} is a square $m \times m$ matrix with m the dimension of β.

We first look at the case where $u(n)$ is a MA.

Theorem 1 Suppose that $u(n)$ is generated by (1.11). The first three parts of the theorem deal with the simultaneous equations case, i.e. $B_o \not\equiv I_{(p)}$, although it may equal it. Part (iv) discusses the reduced form case.

(i) Even if $r' < r$ or $t' < t$, or both, $\ddagger_\delta^{(MA)} \le \ddagger_\delta^{(NP)}$. In general $\ddagger_\delta^{(MA)} < \ddagger_\delta^{(NP)}$ if $r' > 0$.

Put $G(z) = B(z)^{-1}A(z)$,

$$T_1(\lambda) = \{\chi_\beta(\lambda) \otimes G(\lambda) \otimes I_{(p)}\} - [\{B_o^{-1} \otimes A'(\lambda)\}', 0']', \lambda \in [-\pi, \pi] \qquad (4.11)$$

$$T_2(\lambda) = \{\chi_\alpha(\lambda) \otimes I_{(p^2)}\} \qquad , \lambda \in [-\pi, \pi]$$

and let L_{11} be defined by (4.10).

(ii) A necessary and sufficient condition for $\ddagger_\delta^{(MA)}$ to equal $\ddagger_\delta^{(NP)}$ is that there exists a real matrix S, independent of λ, such that

$$L_{11}T_1(\lambda) = ST_2(\lambda) \quad \forall \; \lambda \in [-\pi, \pi] \qquad (4.15)$$

(iii) A sufficient condition for $\ddagger_\delta^{(MA)}$ to equal $\ddagger_\delta^{(NP)}$ is that there exists a constant real matrix S_1 such that

$$T_1(\lambda) = S_1T_2(\lambda) \;, \; \forall \; \lambda \in [-\pi, \pi] \qquad (4.16)$$

(iv) In the reduced form case, the second term on the right in (4.11) drops out. (i) above still holds and (ii) and (iii) apply if there are restrictions. In the absence of restrictions (4.16) becomes necessary as well as sufficient.□

Remark 4.1 (i) Theorem 1(i) implies that even if the lags on the endogenous variables and the moving average are over specified, the parametric estimator will still be at least as efficient as the nonparametric one. If there really are lagged endogenous variables, i.e. $r' > 0$, then the parametric estimator will in general be more efficient than the non-parametric one. Although (ii) may be difficult to work with, (iii) is easy to apply as can be seen from the proof of the next corollary. □

Corollary 1 Let $G(z)$ be defined as above. If $G(z)$ is a polynomial of degree g and $r+g \leq t$, then $\ddagger_\delta^{(MA)} = \ddagger_\delta^{(NP)}$.

This result covers:

1. The distributed lag case where $B(z) = A(z)$ and $B_o = I_{(p)}$. Here $G(z) = I_{(p)}$, $g = 0$ and $r = t$.

2. The simultaneous equations case when there are no lagged endogenous variables, i.e. $r = 0$. Here $G(z)=B_o^{-1}A(z)$ and $g = t'$.

The condition $r+g \leq t$ does not automatically hold when $G(z)$ is a polynomial even if $r=r'$ and $t=t'$. As an example consider

$$B(z) = I_{(2)} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} z,\ A(z) = I_{(2)},\ r'=r=1,\ t'=t=0.$$

Because $\det A(z) = \det B(z) = 1 \ \forall\ z$, $A(z)$ and $B(z)$ certainly satisfy condition A3. of §2. $G(z) = B(z)^{-1} = I_{(2)} + \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} z$. Hence $g = 1$ and $r+g = 2 > t = 0$. □

Next we will deal with the AR case.

Theorem 2 Suppose that $u(n)$ is generated by (1.12). The first three parts of the theorem deal with the simultaneous equations case and part (iv) looks at the reduced form case.

(i) With $\ddagger_\delta^{(MA)}$ replaced by $\ddagger_\delta^{(AR)}$ this is the same as Theorem 1(i).

Put

$$T_1(\lambda) = \{\chi_\beta(\lambda)\otimes B^{-1}(\lambda)\otimes A'(\lambda)\}-[\{B_o^{-1}A(\lambda)\otimes I_{(p)}\}',0']',\lambda\in[-\pi,\pi] \tag{4.20}$$

$$T_2(\lambda) = \{\chi_\alpha(\lambda)\otimes I_{(p^2)}\} \qquad \lambda \in[\pi,\pi]$$

(ii) The necessary and sufficient condition for $\ddagger_\delta^{(AR)}$ to equal $\ddagger_\delta^{(NP)}$ is the same as given in Theorem 1(ii).

(iii) and (iv) are the same as (iii) and (iv) of Theorem 1 with $T_1(\lambda)$ given by (4.20). In the reduced form case, the second term on the right in (4.20) drops out. □

Corollary 2 Put $G(z) = B(z)^{-1}$. If $G(z)$ is a polynomial of degree g and $r+g+t' \leq t$, then $\ddagger_\delta^{(NP)} = \ddagger_\delta^{(AR)}$.

The condition $r+g+t' \leq t$ does not automatically hold when $G(z)$ is a polynomial, even if $r=r'$ and $t=t'$. This can be seen from the example of Corollary 1 where $r=r'=g=1$, $t=t'=0$. □

Remark 4.2 (i) Hannan and Terrell [1973, Theorem 1] obtain estimates of δ_o for $r=r'=s=s'=0$ which have the same asymptotic distribution as the nonparametric estimates discussed above. We know from Corollary 1 that if $u(n)$ is a MA, then under the Hannan - Terrell conditions $\ddagger_\delta^{(MA)} = \ddagger_\delta^{(NP)}$. By Corollary 2 a similar result also holds for the AR case.

(ii) Espasa [1975] demonstrates Theorem 2(i) in the simple scalar case with $u(n)$ following a first order AR by using the calculus of residues. Espasa and Sargan [1975] show that as the lag on the assumed AR tends to infinity, the efficiency of the parametric estimator tends to that of the nonparametric one. They also state, but do not prove Theorem 1(i).

(iii) It is clear that $\hat{\theta}_N^{(MA)}$ and $\hat{\theta}_N^{(AR)}$, which are defined in §3, have the same asymptotic distribution as the estimates obtained by maximizing the exact Gaussian likelihood rather than an approximation to it as we have done. Therefore Theorem 1 and 2 and Corollaries 1 and 2 will also apply to these quasi-maximum-likelihood estimates.

(iv) An alternative to estimating δ_o nonparametrically is to obtain quasi-ML estimates of it from the model

$$y(n) = B(z)^{-1}C(z)x(n) + F(z)^{-1}G(z)e(n) \qquad (4.25)$$

where the $e(n)$ have the properties described in assumption A1., $F(z) = \sum_{j=o}^{f} F_j z^j$, $G(z) = \sum_{j=o}^{g} G_j z^j$, and we do not assume that the coefficient matrices F_j and G_j are related to δ. Let $\ddagger_\delta$ be the asymptotic covariance matrix of the obtained estimate of δ_o. If f and g are taken sufficiently large to ensure that all estimates are consistent, then $\ddagger_\delta = \ddagger_\delta^{(NP)}$. It follows that the results of this chapter will also apply to the parametric estimation of (4.25) as described above.

5. PROOFS

To prove the results of §4 we need the following two lemmas.

<u>*Lemma 2*</u> Let A and B be two $n \times n$ p.s.d. matrices.

(i) If A and B are p.d., then $A \geq B \Leftrightarrow B^{-1} \geq A^{-1}$.

(ii) If $r(A) = r(B)$, then $A \geq B \Leftrightarrow B^{+} \geq A^{+}$.

(iii) If E is a symmetric matrix such that $r(E)=r(EBE)$, then $EAE \geq EBE \Rightarrow (EBE)^+ \geq (EAE)^+$.

Proof (i) By Rao [1973, p.41,(ii)], $\exists$ a nonsingular matrix R s.t. $R'AR = I_{(n)}$, $R'BR = \Lambda$, where Λ is a diagonal matrix. Let λ_j, $1 \leq j \leq n$, be the j^{th} diagonal element of Λ. Then $A \geq B \Rightarrow I_{(n)} \geq \Lambda \Rightarrow 1 \leq \lambda_j$, $1 \leq j \leq n$. It follows that $\Lambda^{-1} \geq I_{(n)} \Rightarrow B^{-1} \geq A^{-1}$. The reverse inequality is now obvious.

(ii) By Rao [1973, p.39(i)] $\exists$ an orthogonal matrix M having columns m_i, $1 \leq i \leq n$, such that $Am_i = 0$ for $i > r(A)$. If $A \geq B$, then $Bm_i = 0$ for $i > r(A)$. Thus

$$M'AM = \begin{pmatrix} A_1 & \cdot & 0 \\ \cdots & & \cdots \\ 0 & \cdot & 0 \end{pmatrix}, \quad M'BM = \begin{pmatrix} B_1 & \cdot & 0 \\ \cdots & & \cdots \\ 0 & \cdot & 0 \end{pmatrix}$$

say, where A_1 and B_1 are p.d. matrices of dimension $r(A)$. If $A \geq B$, then $A_1 \geq B_1$ so that $B_1^{-1} \geq A_1^{-1}$ by (i). Hence $B^+ \geq A^+$ because

$$M'B^+M = (M'BM)^+ = \begin{pmatrix} B_1^{-1} & \cdot & 0 \\ \cdots & & \cdots \\ 0 & \cdot & 0 \end{pmatrix} \geq \begin{pmatrix} A_1^{-1} & \cdot & 0 \\ \cdots & & \cdots \\ 0 & \cdot & 0 \end{pmatrix} = (M'AM)^+ = M'A^+M.$$

The reverse inequality is proved similarly.

(iii) $EAE \geq EBE \Rightarrow r(E)=r(EBE) \leq r(EAE) \leq r(E)$. Hence $r(EBE) = r(EAE)$. Applying (ii) we obtain $(EBE)^+ \geq (EAE)^+$. □

Lemma 3 Let Z_1, Z_2 be two complex random variables such that $E(Z_i) = 0$, $E(Z_i Z_j^*) = Q_{ij}$, $i,j = 1,2$. Put $S = Q_{12}Q_{22}^+$ and $Q = Q_{11} - Q_{12}Q_{22}^+Q_{21}$. Then,

(i) $0 \leq Q = E\{(Z_1-SZ_2)(Z_1-SZ_2)^*\}$

(ii) $Q \leq E\{(Z_1-\tilde{S}Z_2)(Z_1-\tilde{S}Z_2)^*\}$ for any arbitrary complex matrix $\tilde{S}$ having the same dimensions as S.

(iii) $Q = 0$ iff $\exists$ a complex matrix $\tilde{S}$ such that

$Z_1 = \tilde{S}Z_2$ a.s. iff $Z_1 = SZ_2$ a.s.

Proof (i) $E\{(Z_1-SZ_2)Z_2^*\} = 0$ because $Q_{12} = Q_{12}Q_{22}^+Q_{22}$.

Hence

$$0 \leq E\{(Z_1-SZ_2)(Z_1-SZ_2)^*\} = E\{(Z_1-SZ_2)Z_1^*\} + 0 = Q$$

and (i) follows

(ii) This follows from

$E\{(Z_1 - \tilde{S}Z_2)(Z_1-\tilde{S}Z_2)^*\} = Q + (S-\tilde{S})Q_{22}(S-\tilde{S})'$.

(iii) If there is an $\tilde{S}$ such that $Z_1 = \tilde{S}Z_2$ a.s., then $Q = 0$ from (i) and (ii) and hence $Z_1 = SZ_2$ a.s.. If $Q = 0$, then $Z_1 = SZ_2$ a.s. □

Proof of Theorem 1 From (3.16) and (3.46)

$$\ddagger_\delta^{(NP)} = (L_1\Delta_{\delta\delta}^{(NP)}L_1)^+, \ \ddagger_\delta^{(MA)} = [L_1(\Delta_{\delta\delta}^{(MA)} - \Phi_{\alpha,\delta})L_1]^+$$

where $\Phi_{\alpha,\delta} = \Delta_{\delta\alpha}^{(MA)}\ \Delta_{\alpha\alpha}^{(MA)+}\ \Delta_{\alpha\delta}^{(MA)}$. By Theorem 2.13(iii)

$$r(L_1\Delta_{\delta\delta}^{(NP)}L_1) = r(L_1) \tag{5.10}$$

because δ is identified.

From Lemma 2(iii) and (5.10) we obtain that

$$\ddagger_\delta^{(NP)} \geq \ddagger_\delta^{(MA)} \quad \text{if} \quad L_1(\Delta_{\delta\delta}^{(MA)} - \Phi_{\alpha,\delta})L_1 \geq L_1\Delta_{\delta\delta}^{(NP)}L_1.$$

In addition

$$\ddagger_\delta^{(NP)} = \ddagger_\delta^{(MA)} \text{ iff } L_1(\Delta_{\delta\delta}^{(MA)} - \Phi_{\alpha,\delta})L_1 = L_1\Delta_{\delta\delta}^{(NP)}L_1$$

by the uniqueness of the Moore-Penrose generalized inverse.

From (3.41) and (3.42)

$$L_1(\Delta_{\delta\delta}^{(MA)} - \Phi_{\alpha,\delta})L_1 \geq L_1\Delta_{\delta\delta}^{(NP)}L_1 \iff L_{11}(\Psi_\beta^{(MA)} - \Phi_{\alpha,\beta})L_{11} \geq 0 \tag{5.15}$$

and

$$L_1(\Delta_{\delta\delta}^{(MA)} - \Phi_{\alpha,\delta})L_1 = L_1\Delta_{\delta\delta}^{(NP)}L_1 \iff L_{11}(\Psi_\beta^{(MA)} - \Phi_{\alpha,\beta})L_{11} = 0 \tag{5.16}$$

where L_{11} is defined in (4.10) and

$$\Phi_{\alpha,\beta} = \Delta_{\beta\alpha}^{(MA)}\Delta_{\alpha\alpha}^{(MA)^+}\Delta_{\alpha\beta}^{(MA)}.$$

Let $\phi_e(\lambda)$, $\lambda \in[-\pi,\pi]$, be a p dimensional, zero mean, stochastic function of λ having orthogonal increments (see 1.§2) such that

$$\begin{aligned} E\{\phi_e(d\lambda)\phi_e(d\mu)^*\} &= 0 \quad , \quad \lambda \neq \mu, \\ &= f_e(\lambda)d\lambda, \ \lambda = \mu \quad , \ (\lambda,\mu)\in[-\pi,\pi]. \end{aligned}$$

Let τ be a p dimensional vector random variable which is independent of $\phi_e(\lambda)$ and has zero mean and covariance $(\Omega/2\pi)^{-1}$.

We now define the complex vector random variables

$$W_1 = (2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \{\chi_\beta(\lambda)\otimes B^{-1}(\lambda)A(\lambda)\phi_e(d\lambda)\otimes A^{-1}(\lambda)'\tau\} -$$

$$- \begin{bmatrix} (2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \{B_o^{-1}\phi_e(d\lambda)\otimes\tau\} \\ \cdots\cdots\cdots\cdots\cdots \\ 0 \end{bmatrix} \tag{5.17}$$

$$W_2 = -(2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \{\chi_\alpha(\lambda)\otimes\phi_e(d\lambda)\otimes A^{-1'}(\lambda)\tau\}.$$

Putting $Z_1 = L_{11}W_1$, $Z_2 = W_2$ we can readily show that $E(Z_1Z_1^*) = L_{11}\Psi_\beta^{(MA)}L_{11}$, $E(Z_1Z_2^*)=L_{11}\Delta_{\beta\alpha}^{(MA)}$, $E(Z_2Z_2^*) = \Delta_{\alpha\alpha}^{(MA)}$. Identifying Z_1 and Z_2 with the Z_1 and Z_2 of Lemma 3, (i) of that lemma implies that the inequality on the right side of (5.15) holds. Hence $\Phi_\delta^{(NP)} \geq \Phi_\delta^{(MA)}$.

Put $S = L_{11}\Delta_{\beta\alpha}^{(MA)}\Delta_{\alpha\alpha}^{(MA)^+}$. S is real because $\Delta_{\beta\alpha}^{(MA)}$ and $\Delta_{\alpha\alpha}^{(MA)}$ are real. By Lemma 3(iii) the equality on the right side of (5.16) holds iff

$$Z_1 = SZ_2 \text{ a.s.} \tag{5.20}$$

Hence

$$\ddagger_\delta^{(MA)} = \ddagger_\delta^{(NP)} \quad \text{iff} \quad Z_1 = SZ_2 \text{ a.s.} \tag{5.21}$$

By postmultiplying both sides of (5.20) by $\{\chi_e(d\lambda)^* \otimes \tau'\}$ and taking expectations we have that $Z_1 = SZ_2$ a.s. implies that

$$L_{11}\Big[\{\chi_\beta(\lambda)\otimes B^{-1}(\lambda)A(\lambda)f_e(\lambda)\otimes A^{-1'}(\lambda)(\Omega/2\pi)^{-1}\}-$$

$$-\{(B_o^{-1}f_e(\lambda)\otimes(\Omega/2\pi)^{-1})',0'\}'\Big]$$

$$= S\{\chi_\alpha(\lambda)\otimes f_e(\lambda)\otimes A^{-1'}(\lambda)(\Omega/2\pi)^{-1}\} \qquad \forall\ \lambda\in[-\pi,\pi]$$

$$\Rightarrow L_{11}T_1(\lambda) = ST_2(\lambda) \qquad \forall\ \lambda\in[-\pi,\pi]$$

with $T_1(\lambda)$ and $T_2(\lambda)$ defined in the statement of Theorem 1.

Conversely, if $L_{11}T_1(\lambda) = \tilde{S}T_2(\lambda)$ $\forall\ \lambda\in[-\pi,\pi]$, where $\tilde{S}$ is a constant complex matrix, then $Z_1 = \tilde{S}Z_2$ and it follows from Lemma 3 that $Z_1 = SZ_2$ a.s. From (5.21)

$$\ddagger_\delta^{(MA)} = \ddagger_\delta^{(NP)} \quad \text{iff} \quad L_{11}T_1(\lambda) = ST_2(\lambda)\ \forall\ \lambda\in[-\pi,\pi].$$

Taking $r > r'$ or $t > t'$ merely increases the dimension of the χ_α or χ_β vectors; we will still have $\ddagger_\delta^{(MA)} \leq \ddagger_\delta^{(NP)}$. If $r' > 0$, then, in general, $G(z)=B(z)^{-1}A(z)$ is not a polynomial but has the infinite series expansion $G(z) = \sum_{j=o}^{\infty} G_j z^j$ for $|z| \leq 1$ by condition A3.

The submatrices of $T_1(\lambda)$ are

$$\sum_{j=o}^{\infty}(G_j\otimes I_{(p)})z^j - \sum_{j=o}^{t'}(B_o^{-1}\otimes A_j')z^j \quad , \quad z = e^{i\lambda} \tag{5.30}$$

and

$$\sum_{j=o}^{\infty}(G_j\otimes I_{(p)})z^{j+\ell},\quad 1\leq \ell\leq r,\ z = e^{i\lambda} \tag{5.31}$$

The submatrices of $T_2(\lambda)$ are

$$I_{(p^2)} z^j, \; 1 \leq j \leq t, \; z = e^{i\lambda} \tag{5.32}$$

If $r' > 0$, then it is clear from (5.30) - (5.32) that there is usually no constant matrix $\tilde{S}$ s.t. $L_{11}T_1(\lambda) = \tilde{S}T_2(\lambda)$. Therefore we will usually have $\ddagger_\delta^{(NP)} > \ddagger_\delta^{(MA)}$ if $r' > 0$. This completes the proof of (i) and (ii) of Theorem 1.

(iii) If there is a constant matrix S_1 such that $T_1(\lambda) = S_1T_2(\lambda) \; \forall \lambda \in [-\pi,\pi]$, then $L_{11}T_1(\lambda) = (L_{11}S_1)T_2(\lambda)$, and putting $\tilde{S} = L_{11}S_1$, (iii) follows from (ii).

(iv) The reduced form case follows in exactly the same way as the simultaneous equations one, except that the second term on the right in (5.17) drops out. □

Proof of Corollary 1 If $G_j = 0$ for $j > g$, then (5.30) can be rewritten as

$$(G_o \otimes I_{(p)}) + \sum_{j=1}^{g} (G_j \otimes I_{(p)}) z^j - (B_o^{-1} \otimes I_{(p)}) - \sum_{j=1}^{t'} (B_o^{-1} \otimes A_j') z^j$$

$$= \sum_{j=1}^{g} (G_j \otimes I_{(p)}) - \sum_{j=1}^{t'} (B_o^{-1} \otimes A_j') z^j, \; z = e^{i\lambda} \tag{5.35}$$

where we have used $G_o = B_o^{-1}$ to cancel the first and third terms on the left side of (5.35). (5.31) becomes

$$\sum_{j=o}^{g} (G_j \otimes I_{(p)}) z^{j+\ell}, \; 1 \leq \ell \leq r, \; z = e^{i\lambda} \tag{5.36}$$

If $(r+g) \leq t$, then it is clear from (5.32), (5.35) and (5.36) that there is a constant matrix S such that $T_1(\lambda) = ST_2(\lambda) \;\; \forall \lambda \in [-\pi,\pi]$. An application of Theorem 1 (iii) completes the proof. □

Because the proofs for the AR case are very similar to the MA case we will not give them in detail.

Proof of Theorem 2 Let the random variables $\phi_e(\lambda)$

and τ be defined as in the proof of Theorem 1.

Put

$$W_1 = (2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \chi_\beta(\lambda) \otimes B^{-1}(\lambda) A^{-1}(\lambda) \phi_e(d\lambda) \otimes A'(\lambda)\tau -$$

$$- \begin{bmatrix} \{(2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} B_o^{-1} \phi_e(d\lambda) \otimes \tau\} \\ \cdots\cdots\cdots \\ 0 \end{bmatrix}, \qquad (5.40)$$

$$W_2 = (2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \{\chi_\alpha(d\lambda) \otimes A^{-1}(\lambda) \phi_e(d\lambda) \otimes \tau\}, \qquad (5.41)$$

$Z_1 = L_{11} W_1$, $Z_2 = W_2$. As in the proof of Theorem 1 it can be shown that $\ddagger_\delta^{(AR)} \leq \ddagger_\delta^{(NP)}$, and that $\ddagger_\delta^{(AR)} = \ddagger_\delta^{(NP)}$ if and only if there is a real, constant, matrix S such that $Z_1 = SZ_2$ a.s.

By postmultiplying (5.40) and (5.41) by $(\phi_e(d\lambda)^* \otimes \tau')$ and taking expectations we have that $Z_1 = SZ_2$ a.s. implies that

$$L_{11}[\{\chi_\beta(\lambda) \otimes B^{-1}(\lambda) A^{-1}(\lambda) f_e(\lambda) \otimes A'(\lambda)(\Omega/2\pi)^{-1}\} -$$

$$- \{(B_o^{-1} f_e(\lambda) \otimes (\Omega/2\pi)^{-1})', 0'\}']$$

$$= S\{\chi_\alpha(\lambda) \otimes A^{-1}(\lambda) f_e(\lambda) \otimes (\Omega/2\pi)^{-1}\} \qquad \forall \lambda \in [-\pi,\pi].$$

$$\Rightarrow L_{11} T_1(\lambda) = S T_2(\lambda) \quad \forall \lambda \in [-\pi,\pi] \qquad (5.45)$$

where $T_1(\lambda)$ and $T_2(\lambda)$ are defined in the statement of Theorem 2. Conversely, if (5.45) holds for some matrix S, then $Z_1 = SZ_2$ a.s. Therefore $\ddagger_\delta^{(AR)} = \ddagger_\delta^{(NP)}$ iff $L_{11} T_1(\lambda) = S T_2(\lambda)$ $\forall \lambda \in [-\pi,\pi]$, with S a real constant matrix. If $r' > 0$, then, in general, $G(z) = B(z)^{-1}$ is not a polynomial but has the infinite series expansion

$G(z) = \sum_{j=0}^{\infty} G_j z^j$ valid for $|z| \leq 1$.

The submatrices of $T_1(\lambda)$ are

$$\sum_{j=o}^{\infty} \sum_{k=o}^{t'} (G_j \otimes A_k') z^{j+k} - \sum_{j=o}^{t'} (B_o^{-1} A_j \otimes I_{(p)}) z^j, \quad z=e^{i\lambda} \tag{5.50}$$

and

$$\sum_{j=o}^{\infty} \sum_{k=o}^{t'} (G_j \otimes A_k') z^{j+\ell+k}, \quad 1 \leq \ell \leq r, \; z = e^{i\lambda} \tag{5.51}$$

The submatrices of $T_2(\lambda)$ are given by (5.32). If $r' > 0$, then it is clear from (5.32), (5.50) and (5.51) that there is usually no constant matrix S such that (5.45) holds. Therefore we will usually have $\ddagger_\delta^{(NP)} > \ddagger_\delta^{(AR)}$ if $r' > 0$. This completes the proof of (i) and (ii) of Theorem 2. (iii) is shown in the same way as for Theorem 1(iii), while (iv) follows from above. □

Proof of Corollary 2 If $G_j = 0$ for $j > g$, then (5.50) can be rewritten as

$$(G_o \otimes I_{(p)}) + \sum_{k=1}^{t'} (G_o \otimes A_k') z^k + \sum_{j=1}^{g} \sum_{k=o}^{t'} (G_j \otimes A_k') z^{j+k} -$$

$$- (B_o^{-1} \otimes I_{(p)}) - \sum_{k=1}^{t'} (B_o^{-1} A_k' \otimes I_{(p)}) z^j \tag{5.52}$$

$$= \sum_{k=1}^{t'} (G_o \otimes A_k') z^k + \sum_{j=1}^{g} \sum_{k=o}^{t'} (G_j \otimes A_k') z^{j+k} -$$

$$- \sum_{k=1}^{t'} (B_o^{-1} A_k' \otimes I_{(p)}) z^j, \quad z = e^{i\lambda},$$

where the first and third terms on the right in (5.52) cancel because $G_o = B_o^{-1}$. (5.51) becomes

$$\sum_{j=o}^{g} \sum_{k=o}^{t'} (G_j \otimes A_k') z^{j+k+\ell}, \quad 1 \leq \ell \leq r, \; z = e^{i\lambda} \tag{5.53}$$

If $(g+t'+r) \leq t$, then it is clear from (5.32), (5.52) and (5.53) that there exists a constant matrix S such that

$T_1(\lambda) = ST_2(\lambda)\ \forall\ \lambda \in [-\pi,\pi]$. An application of Theorem 2(iii) completes the proof. $\square$

BIBLIOGRAPHY

Aitchison, J. and Silvey, S.D. [1958]. Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math.Statist.*, 29, 813-828.

Akaike, H. [1973]. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60, 255-265.

———— [1974a]. Stochastic Theory of minimal realization. *IEEE Trans. Automat. Contr.*, 19, 667-674.

———— [1974b]. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Ann. Inst. Statist. Math.*, 26, 363-387.

———— [1975]. Markovian representation of stochastic processes by canonical variables. *SIAM J. Control*, 13, 162-173.

Anderson, T.W. [1971]. *The statistical analysis of time series*. John Wiley, New York.

Bard, Y. [1974]. *Nonlinear parametric estimation*. Academic Press, New York.

Bowden, R. [1973]. The theory of parametric identification. *Econometrica*, 41, 1069-1074.

Box, M.J., Davies, D. and Swann, W.H. [1969]. *Non-linear optimization techniques*. Oliver and Boyd, Edinburgh.

Brown, R.L., Durbin, J. and Evans, J.M. [1975]. Techniques for testing the constancy of regression relationships over time. *J.R.S.S.B.*, 37, 149-163.

Brown, B.M. and Eagleson, G.K. [1971]. Martingale convergence to infinitely divisible laws with finite variances. *Trans. Amer. Math.Soc.*, 162, 449-453.

Byron, R.P. and Pagan, A.R. [1975]. A synthetic approach to the estimation of models with autocorrelated disturbance terms. Mimeo. (To appear in Essays in Honour of A.W. Phillips, John Wiley, New York).

Deistler, M. [1975]. z- Transform and identification of linear Econometric models with autocorrelated errors. *Metrika*, 22, 13-25.

———— [1976]. The identifiability of linear Econometric models with autocorrelated errors. *Int. Ec. Rev.*, 17, 26-46.

Deistler, M., Dunsmuir, W. and Hannan, E.J. [1977]. Vector linear time series models: Corrections and extensions. Mimeo.

Dhrymes, P.J. [1973]. Restricted and unrestricted reduced forms: asymptotic distribution and relative efficiency. *Econometrica*, 41, 119-134.

Dhrymès, P.J. and Erlat, H. [1974]. Asymptotic properties of full information estimators in dynamic autoregressive simultaneous equation models. *J. Ec.*, 2, 247-259.

Dhrymes, P.J. and Taylor, J.B. [1976]. On an efficient two-step estimator for dynamic simultaneous equations models with autoregressive errors. *Int. Ec. Rev.* , 17, 362-376.

Doob, J.L. [1953]. *Stochastic Processes*. John Wiley, New York.

Dunsmuir, W. [1976]. Estimation for vector linear time series models. Unpublished Ph.D. thesis, A.N.U.

Dunsmuir, W. and Hannan, E.J. [1976]. Vector Linear Time Series Models. *Adv. Appl. Prob.*, 8, 339-364.

Engle, R.F. [1976]. Band spectrum regression methods with lagged dependent variables and stationary errors: an iterated instrumental variable approach. Mimeo.

Espasa, A. [1975]. The estimation of the multiple regression model with stationary errors and lagged endogenous variables. Mimeo, London School of Economics.

Espasa, A. and Sargan, J.D. [1975]. The spectral estimation of simultaneous equation systems with lagged endogenous variables. Paper presented at the World Congress of the Econometric Society, Toronto, August 1975.

Fisher, F.M. [1966]. *The identification problem in Econometrics*. McGraw-Hill, New York.

Führt, B.P. [1973]. Maximum likelihood identification of Astrom model by quasilinearization. *Proc. 3rd IFAC Symposium*, The Hague/Delft pt 2,pp.737-740, ed. P. Eykhoff.

Gertler, J. and Bányász, C. [1974]. A recursive (on-line) maximum likelihood identification method. *IEEE Trans. Automat. Contr.*, 19, 816-820.

Halmos, P.R. [1950]. *Measure Theory*. D. Van Nostrand, Princeton, New Jersey.

Hannan, E.J. [1969a]. The identification of vector mixed autoregressive-moving average systems. *Biometrika*, 56, 223-225.

——————— [1969b]. The estimation of mixed autoregressive moving average systems. *Biometrika*, 56, 579-593.

——————— [1970]. *Multiple Time Series*. John Wiley, New York.

Hannan, E.J. [1971a]. The identification problem for multiple equation systems with moving average errors. *Econometrica*, 39, 751-765.

———— [1971b]. Non-linear time series regression. *J. Appl. Prob.*, 8, 767-780.

———— [1973a]. Multivariate time series analysis. *J. Mult. Analysis*, 3, 395-407.

———— [1973b]. Central limit theorems for time series regression. *Z. Wahr. verw. Geb.*, 26, 157-170.

———— [1973c]. The asymptotic theory of linear time-series models. *J. Appl. Prob.*, 10, 130-145.

———— [1976a]. The identification and parameterization of ARMAX and state space forms. *Econometrica*, 44, 713-723.

———— [1976b]. Armax and state space systems and recursive calculations. Mimeo.

———— [1976c]. The convergence of some recursion. *Ann. of Statist.*, 4, 1258-1270.

Hannan, E.J. and Heyde, C.C. [1972]. On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.*, 43, 2058-2066.

Hannan, E.J. and Nicholls, D.F. [1972]. The estimation of mixed regression, autoregression, moving average, and distributed lag models. *Econometrica*, 40, 529-547.

Hannan, E.J. and Robinson, P.M. [1973]. Lagged regression with unknown lags. *J.R.S.S.B.*, 35, 252-267.

Hannan, E.J. and Tanaka, K. [1976]. ARMAX models and recursive calculations, Proc. Conf. Multivariable System Theory and Control of Dynamic Economic Systems, Springer Verlag.

Hannan, E.J. and Terrell, R.D. [1972]. Time series regression with linear constraints. *Int. Ec. Rev.*, 13, 189-200.

———— [1973]. Multiple equation systems with stationary errors. *Econometrica*, 41, 299-320.

Hatanaka, M. [1974]. An efficient two-step estimator for the dynamic adjustment model with autoregressive errors. *J. Ec.*, 2, 199-220.

———— [1975]. On the global identification of the dynamic simultaneous equations model with stationary disturbances. *Int.Ec. Rev.* 16, 545-554.

Hatanaka, M. [1976]. Several efficient two-step estimators for the dynamic simultaneous equations model with autoregressive disturbances. *J. Ec.*, 4, 189-204.

Hendry, D.F. [1976]. The structure of simultaneous equations estimators. *J. Ec.*, 4, 51-88.

Heyde, C.C. [1973]. An iterated logarithm result for martingales and its application in estimation theory for autoregressive processes. *J. Appl. Prob.*, 10, 146-157.

Hoerl, A.E. and Kennard, R.W. [1970a]. Ridge Regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-67.

———————— [1970b]. Ridge Regression: Applications to nonorthogonal problems. *Technometrics*, 12, 69-82.

Hurwicz, L. [1950]. Generalization of the concept of identification. Chapter IV in *Statistical inference in dynamic economic models*, Cowles Commission Monograph 10, T.C. Koopmans ed., John Wiley, New York.

Ibragimov, I.A. and Linnik, Yu. V. [1971]. *Independent and stationary sequences of random variables*. Wolters - Noordhoff, Groningen.

Jennrich, R.I. [1969]. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40, 633-643.

Kohn, R. [1977a]. Note concerning the Akaike and Hannan estimation procedures for an autoregressive moving average process. To appear in Biometrika.

———— [1977b]. Asymptotic properties of time domain Gaussian Estimators: I. Accepted for publication in J. Appl. Prob.

———— [1977c]. Reappraisal and extension of some previous identification results. Accepted for publication in J.Ec.

Koopmans, T.C., Rubin, H. and Leipnik, R.B. [1950]. Measuring the equation systems of dynamic economics. Chapter II in: *Statistical inference in dynamic economic models*, Cowles Commission Monograph 10, T.C. Koopmans ed., John Wiley, New York.

Kowalik, J. and Osborne, M.R. [1968]. *Methods for unconstrained optimization problems*. Elsevier, New York.

MacDuffee, C.C. [1956]. *The theory of matrices*. Chelsea, New York.

Neveu, J. [1975]. *Discrete-Parameter Martingales*. North-Holland, Amsterdam.

Nicholls, D.F. [1971]. Inference problems for vector linear time series models. Unpublished Ph.D. thesis, A.N.U.

————————— [1976]. The efficient estimation of vector linear time series models. *Biometrika*, 63, 381-390.

————————— [1977]. A comparison of estimation methods for vector linear time series models. *Biometrika*, 64, 85-90.

Nicholls, D.F., Pagan, A.R. and Terrell, R.D. [1975]. The estimation and use of models with moving average disturbance terms: A survey. *Inter. Eco. Rev.*, 16, 267-280.

Petrovski, I.G. [1966]. *Ordinary Differential Equations.* Prentice-Hall, N.J.

Phillips, A.W. [1966]. Estimation of systems of difference equations with moving average disturbances. Paper read to the Econometric Society Meeting, San Francisco.

Plackett, R.L. [1950]. Some theorems in least squares. *Biometrika*, 37, 149-157.

Rao, C.R. [1973]. *Linear Statistical Inference and Its Applications*, 2nd ed. John Wiley, New York.

Robinson, P.M. [1972a]. Non-linear regression for multiple time series. *J. Appl. Prob.*, 9, 758-768.

————————— [1972b]. The estimation of continuous time systems using discrete data. Unpublished Ph.D. thesis, A.N.U.

—————————[1975]. Continuous time regression with discrete data. *Ann. Statist.*, 3, 688-697.

—————————[1976a]. The estimation of linear differential equations with constant coefficients. *Econometrica*, 44, 751-764.

————————— [1976b]. Instrumental variable estimation of differential equations. *Econometrica*, 44, 765-776.

Rohde, C.A. [1965]. Generalized inverses of partitioned matrices. *SIAM J. Appl. Math.*, 13, 1033-1035.

Rosenbrock, H.H. [1970]. *State-space and multivariable theory.* Nelson, London.

Rothenberg, T.J. [1971]. Identification in parametric models. *Econometrica*, 39, 577-591.

Rudin, W. [1964]. *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York.

Sargan, J.D. [1975]. The identification and estimation of sets of simultaneous stochastic equations. Discussion paper No. A1, L.S.E.

Schmidt, P. [1976]. *Econometrics*, Marcel Dekker, New York.

Schönfeld, P. [1971]. A useful central limit theorem for m-dependent variables. *Metrika*, 17, 116-128.

Söderström, T., Ljung, L. and Gustavsson, I. [1974]. A comparative study of recursive identification methods. Report 7407, Lund Inst. of Tech., Dept. Aut. Control, Sweden.

Stout, W.F. [1974]. *Almost sure convergence.* Academic Press, New York.

Wald, A. [1949]. Note on the consistency of the maximum likelihood estimate. *Ann. Math.Statist.*, 20, 595-601.

——— [1950]. Note on the identification of economic relations. In *Statistical inference in dynamic economic models* (T.C. Koopmans, ed.)pp.238-244. Cowles Commission Monograph, No. 10, John Wiley, New York.

Wegge, L. [1965]. Identifiability criteria for a system of equations as a whole. *Aust. J. Statist.*, 7, 67-77.

Whittle, P. [1962]. Gaussian estimation in stationary time series. *Bull. Inter. Statist. Inst.*, 39, 105-129.

Wilkinson, J.H. [1971]. Linear systems, least squares and linear programming. In Linear Algebra (J.H. Wilkinson and C. Reinsch, Eds.), pp.1-8. Handbook for automatic computation. Springer-Verlag, New York.

Young, P.C. [1976]. Some observations on instrumental variable methods of time series analysis. *Int. J. Control*, 16, 593-612.

Zellner, A. and Palm, F. [1974]. Time Series analysis in simultaneous equation econometric models. *J. Ec.*, 2, 17-54.

Zygmund, A. [1959]. *Trigonometric series, Vol. I.* Cambridge University Press, Cambridge.