THE ESTIMATION OF CONTINUOUS-TIME SYSTEMS USING DISCRETE DATA

by

Peter M. Robinson

A thesis submitted to the
Australian National University
for the degree of
Doctor of Philosophy.
October, 1972.

PREFACE

The research for this thesis was carried out over a period of two years, commencing October, 1970. The multivariate treatment of Chapter 4 is my own but it extends work carried out with Professor E.J. Hannan, the results of Chapter 4 being included in modified form in Hannan and Robinson (1973). Also Theorems 2.5 and 2.8 are my own extensions, to wider circumstances, of results appearing in this paper. With these qualifications and unless otherwise stated, the results described in the thesis are my own. Many of the results of Chapter 2 and Chapter 3 (sections 1-3 and Appendix) will appear in Robinson (1972, 1973); the author is preparing for publication other results of Chapter 2 and the substance of Chapter 5 and Chapter 6 (sections 1 and 2); he hopes also to publish some of the results of Chapter 3 (section 4), Chapter 6 (section 3) and Chapter 7.

Peter M. Robinson

ACKNOWLEDGEMENTS

I am greatly indebted to Professor E.J. Hannan for his many helpful suggestions during the past two years; I feel that a number of my ideas were prompted by an idea of his for the approximation used in estimating the model of Chapter 4. I am very grateful to him for encouraging me to explore the remaining topics on my own and for his wise advice whenever it was sought.

I wish to thank Miss P.N. McCarter for her assistance with the computations of Chapter 4. The author carried out the computations of Chapters 5-7 on his own on the Australian National University's Univac 1108 computer. The data were supplied by Professor R.D. Terrell and Mr B.V. Hamon. Finally I am grateful to Mrs B.M. Geary for her excellent typing of the thesis.

CONTENTS

SUMMARY

This thesis is concerned with the estimation of parameters in continuous-time systems, when the available data consist of time series sampled at regular intervals of time.

Chapter 1 begins with a discussion of the circumstances in which the work may be relevant. We also describe useful results in matrix theory and in the spectral theory of continuous stationary processes.

In Chapter 2 the general continuous-time model is specified by a sequence of detailed assumptions. It may be regarded as the solution of a system that is linear in the endogenous and exogenous variables, the parameters possibly occuring in a non-linear fashion and satisfying non-linear constraints. The basic method of estimation involves the Fourier transformation of the model and the insertion of the discrete Fourier transforms to produce an approximate model that is of regression type and is thus relatively easy to handle, although the estimation must generally rely on numerical methods. We establish the strong consistency of the estimates and the asymptotic normality of the normed estimates with respect to the *true* model. We do not assume independence or normality for our processes but under our, much weaker, conditions the Fourier transformed residuals have these properties in large samples and it is then possible to choose estimates which are, in a sense, efficient.

The topic of Chapter 3 is the estimation of a regression matrix of less than full rank, a problem related to canonical correlation analysis. Asymptotic theory follows from the previous chapter since a non-linear regression approach is employed but the model is unlagged and much of this work was carried out before the author thought of the general model of Chapter 2. Possible computational procedures are suggested and the related problem of regression on an

unobservable variable is considered.

In Chapter 4 we deal with multivariate linear regressions involving lagged values of the exogenous variables, the lags being unknown and not necessarily multiples of the sampling interval. The identification and estimation problems are considered, and asymptotic theory is given for estimates of the lags and coefficients.

Chapter 5 is concerned with linear systems in which there are lagged endogenous variables with arbitrary real lags. The solution, stability and identification of single- and multiple-difference equation systems are discussed, and estimates of the lags and coefficients with good asymptotic properties are suggested.

In Chapter 6 systems of linear differential equations with constant coefficients are studied. Their solution, stability and identification are discussed and asymptotic theory is given for estimates of the coefficients. Also we propose consistent and efficient methods of estimation which do not require iteration.

The final topic, in Chapter 7, is the estimation of mixed difference-differential equation systems, which arise in many fields of study, and also integro-differential equations, involving a continuum of lagged endogenous variables.

In each of Chapters 4, 5, 6 and 7 we report the fitting of simple models to economic and oceanographic data.

CHAPTER 1

INTRODUCTORY DISCUSSION AND THEORY

1.1 Introduction

In the real world many natural and artificially-generated stochastic processes are of a continuous nature. Continuous observation of such processes may be unfeasible or uneconomical, however. In the social sciences it is out of the question. In some natural sciences it may be possible to obtain a continuous record by means of electrical or optical equipment, but even then only a discrete sample can be used in digital computations. We shall be concerned in the thesis with processes that are of an underlying continuous nature but that are sampled at equal intervals of time, or some other dimension. The work has relevance also to phenomena that are not continuous but that are much more frequently observable than in the available record. The sampling schedule envisaged is apparently a standard one, convenient for the purposes of both data collection and data analysis.

Scientific knowledge frequently suggests the existence of dynamic relationships between stochastic processes. When data are available one may attempt to gauge the extent to which theory agrees with empirical evidence. In order to achieve some generality the theoretical model will involve a number of unknown parameters to be estimated. Typically dynamic systems feature such linear operations as differentiation, integration and translation through time. When there is no information on the behaviour of the processes between the data points such operations will be difficult to handle. Of course a continuous-time system can be approximated by a discrete-time system but the accuracy of such approximations is often difficult to assess and it seems undesirable in any case that the specification of the model should be limited in this way by the nature of the sampling process. In the thesis we discuss parameter estimates, having desirable large-sample properties, for a class of dynamic systems relating endogenous and exogenous variables[1] that are stationary processes in continuous time. The class includes a number of interesting special cases that are believed to provide appropriate descriptions of many phenomena of the real world. In particular our attention is directed to lagged regressions and autoregressions involving unknown and possibly non-integral lags, linear differential equation systems and mixed difference-differential equation systems. The estimation problem for such models has received little previous consideration and it is thus difficult to predict in which fields our results are most likely to be applied. From time to time, nevertheless, we shall briefly mention real-life situations which our models may represent. We make no attempt to be comprehensive in this respect, however, concentrating on the presentation of a theory for the estimation of continuous-time systems that is motivated by the widespread practice of sampling at equal intervals. Time series analyses have frequently been carried out in the past for scientific or social scientific data of this type and, where our models are suitable, our methods may be useful in these fields.

Of all the possible areas of application the author's chief interest is in economics. This bias is reflected from time to time in the terminology and approach adopted. Models of the type we discuss are prominent in economic theory. Economic systems frequently involve leads or lags, expressing a direct or indirect dependence of economic

[1] By endogenous and exogenous variables we mean, respectively, dependent and independent variables. Here and elsewhere we make use of econometric jargon although our work may not always be of particular relevance to this field.

variables on their own past history and that of other variables, and the delay in reaction to changes elsewhere in the economy. Because there is often interest in the rate of change of economic variables and belief that the response of one variable to a change in another is not instantaneous, many economic models are formulated in terms of differential equations. Continuous observation of economic variables is not possible although the underlying processes are often continuous or almost continuous, reflecting the results of decisions taken at many points of time within the sampling interval. The general theory of Chapter 2 may also be relevant in that many economic models involve a number of both endogenous and exogenous variables[2] and their reduced forms are non-linear in the parameters. Furthermore because of limitations on the quantity of available data full use must be made of *a priori* knowledge, and we allow here for the presence of linear and non-linear constraints on the parameters. The methods also seem appropriate in an economic context because the application of ordinary least squares often produces inefficient estimates owing to the presence of serial correlation in the residuals. The spectral methods we adopt provide a flexible and powerful means of allowing for these effects. Fourier methods occupy a role of growing importance in economics and in other fields. Their use here in the estimation of continuous-time systems is a further illustration of their power.

However, some of the premises of our theoretical development are invalid for many economic processes. We assume our processes to be stationary whereas trends are an important feature of much economic data. Thus effects leading to non-stationarity must be eliminated beforehand. We also assume the spectra of our (mean-corrected) processes to be absolutely continuous, thus requiring the prior removal

---

[2] However, the theory is not relevant to "closed" systems, involving no exogenous variables.

of periodic and almost-periodic components. We do not allow for the fact that there may be errors in our observation of the exogenous variables. The most notable qualification arises from the fact that the properties of our estimates are demonstrably optimal only in an asymptotic sense. Moreover the efficient procedures suggested require a sample that is sufficiently large to enable the estimation of the residual spectral density at a number of frequencies.[3] In such fields as physics and the earth sciences very large samples, easily justifying the use of our distribution theory, are often forthcoming. Economists will not have access to series of comparable length and thus the application of our methods must proceed with care. However, they may not be misleading when applied to economic series of reasonable length, when other assumptions are appropriate. Useful information concerning the minimum acceptable sample size could perhaps be gained from systematic simulation studies, which are beyond the scope of the present work. Even when data are scarce, however, the procedures developed in the thesis have an intuitive appeal; they offer, in the author's opinion, a natural and unified approach to the estimation of continuous-time systems.

1.2 Matrix notation and definitions

Generally we deal with vector-valued parameters and variables and the use of matrix notation is indispensable to a succinct exposition. We shall denote matrices by bold-face capitals and vectors by bold-face lower-case characters. This policy is subject to two modifications: following convention we use f to denote spectral density matrices; also both null matrices and null vectors are denoted 0 , the dimensions depending on the context.

[3] Of course even if time-domain methods were used, and it seems that these are less appropriate here, the only theory we could give for non-linear models would be asymptotic.

We consider the complex $p \times q$ matrix $A = (a_{jk})$. The rank of A is denoted $\mathrm{rank}\{A\}$, the transpose A' , the complex conjugate, $\bar{A}$, and we write $A^* = \bar{A}'$. We denote by $\nu_j\{A\}$, $j = 1, \ldots, p$, the singular values of A , that is the square roots of the eigenvalues of AA^* ; if these are ordered so that $\nu_j \geq \nu_k$, $j < k$, we have Ky Fan's inequality[4]

$$\nu_{j+k-1}\{A+B\} \leq \nu_j\{A\} + \nu_k\{B\} , \tag{2.1}$$

where B is $p \times q$. We also have the Euclidean or Schur norm

$$\|A\| = [\mathrm{tr}\{AA^*\}]^{\frac{1}{2}} = \left[\sum \nu_j\{A\}^2\right]^{\frac{1}{2}} , \tag{2.2}$$

where the trace, $\mathrm{tr}\{\cdot\}$, is defined for square matrices.

We introduce $\mathrm{vec}\{A\}$, the $pq \times 1$ vector having a_{jk} as its $[(k-1)p+j]$th element and obtained by stacking the columns of A from left to right down a single column. For any matrix B the Kronecker product $A \otimes B = (a_{jk}B)$ is defined. Then useful relations are:

$$\mathrm{vec}\{ABC\} = (C' \otimes A)\mathrm{vec}\{B\} \tag{2.3}$$

$$\mathrm{tr}\{ABCD\} = \mathrm{vec}'\{C\}(D \otimes B')\mathrm{vec}\{A'\} \tag{2.4}$$

where $\mathrm{vec}'\{C\}$ is the transpose of $\mathrm{vec}\{C\}$ and A, B, C and D conform appropriately. For B also having q columns we introduce the extended product

$$A \odot B = (a_1 \otimes b_1 \vdots \ldots \vdots a_q \otimes b_q) , \tag{2.5}$$

a_j and b_j being respectively the jth columns of A and B .

We denote by A^+ the $q \times p$ unique Moore-Penrose pseudoinverse of A , defined by the relations

---

[4] We index equations which will be referred to later in the text by $(j \cdot k)$, j being the section number and k the equation number within the section. When referring to an equation that appears in a different chapter we add the chapter number as a prefix; thus (2.1) will be cited in later chapters as (1.2.1).

$$AA^+A = A \ , \quad A^+AA^+ = A^+ \ , \quad (AA^+)^* = AA^+ \ , \quad (A^+A)^* = A^+A \ . \qquad (2.6)$$

The remainder of the section deals exclusively with Hermitian[5] A . We may then write $A = U\Lambda U^*$, where Λ is the $p \times p$ diagonal matrix of (real) eigenvalues of A and U is unitary. Then if Λ^+ is derived from Λ by replacing all non-zero elements by their reciprocials, $A^+ = U\Lambda^+U^*$. When A contains null rows and columns and the matrix derived by deleting these is non-singular, as may often be the case in our work, A^+ may be derived by inverting the condensed matrix and reinstating the null rows and columns.

If A is positive-definite (p.d.) we write $A > 0$. For such matrices the determinant, $\det\{A\}$, is non-zero and the true inverse, A^{-1} , exists. If A is non-negative definite (n.n.d.) we write $A \geq 0$. If $A \geq 0$, $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$ and $A^{\frac{1}{2}} + A^{\frac{1}{2}} \geq 0$ then $A^{\frac{1}{2}}$ is the unique n.n.d. Hermitian square root of A , defined by $A^{\frac{1}{2}} = U\Lambda^{\frac{1}{2}}U^*$, where $\Lambda^{\frac{1}{2}}$ is the diagonal matrix of square roots of eigenvalues.

1.3 Stationary processes in continuous time

The theory of stationary processes is fundamental to our work. We introduce a measure space Ω and a probability measure P defined on a σ-field[6] F of Borel sets in Ω . Over the probability space represented by the triple (Ω, F, P) we define the $p \times 1$ vector process $\xi(t, \omega)$ [7], $t \in R$, $\omega \in \Omega$, having real-valued components

---

[5] These are $p \times p$ matrices for which $A^* = A$. Results are implied for real symmetric matrices (for which $A' = A$) which occur frequently in the thesis.

[6] A σ-field is a family of sets closed under the formation of complements and denumerable unions and intersections.

[7] Generally the symbol R^n denotes n-dimensional Euclidean space, but instead of R^1 we write simply R .

that are measurable functions of ω. We shall always be concerned with one particular realization of the process and thus we suppress reference to ω, writing $\xi(t) = (\xi_j(t))$ in place of $\xi(t, \omega)$. Now $\xi(t)$ will at most be observable at the N discrete equally-spaced points $n = \Delta$, $2\Delta, \ldots, N\Delta$, $\Delta > 0$ (when $\xi(t)$ is endogenous or exogenous), and may not be observable even at these points (when $\xi(t)$ is a residual). However, in either case we shall make assumptions about the underlying process $\xi(t)$, $t \in R$.

Though we shall not repeatedly emphasize this, every process that is involved in the thesis is assumed to be continuous in the mean square (m.s.) sense, so[8]

$$\underset{\tau \to t}{\text{l.i.m.}}\ \xi(\tau) = \xi(t) , \tag{3.1}$$

uniformly in $t \in R$. A necessary and sufficient condition for m.s. continuity is the continuity of $E\{\xi(t)\xi(\tau)'\}$ at all diagonal points $\tau = t$. (See Cramér and Leadbetter, 1967, p. 83.)

We shall always assume the means of our processes to be independent of t *and identically zero, so* $E\{\xi(t)\} \equiv 0$. *Moreover, we shall be concerned exclusively with processes that are strictly stationary,* for which the probability distribution of any finite collection of values

$$\xi_{m_1}(t_1+\tau), \ldots, \xi_{m_r}(t_r+\tau) , \tag{3.2}$$

$\tau, t_j \in R$, $m_j \in \{m \mid m = 1, \ldots, p\}$, $j = 1, \ldots, r$, $r = 1, 2, \ldots$, does not depend on τ. *In all that follows we shall refer to such*

---

[8] By the symbol l.i.m. we mean "limit in mean square". Thus (3.1) is

$$\lim_{\tau \to t} E\|\xi(\tau)-\xi(t)\|^2 = 0 ,$$

where, here and elsewhere, the expectation is taken for fixed t over all realizations. Of course, m.s. continuity does not imply the continuity of almost all realizations $\xi(t)$.

processes simply as stationary. In the m.s. sense the Cramér representation

$$\xi(t) = \int_R e^{it\lambda} d\chi(\lambda) \ , \quad t \in R \ , \tag{3.3}$$

exists, where the $p \times 1$ vector process $\chi(\lambda)$ has orthogonal increments and $\chi(-\infty) = 0$. (Cramér and Leadbetter, 1967, p. 129.) Now stationarity implies rth-order stationarity, $r = 2, 3, \ldots$, in which the expectation of the product of the values in (3.2) (taking $t_1 = 0$) depends only on $t_2, \ldots, t_r$ and not on τ . In particular for $r = 2$ m.s. continuity implies that $\xi(t)$ has finite second moments and thus the existence of the Bochner-Khinchine representation

$$E\{\xi(t)\xi(t+\tau)'\} = \int_R e^{i\tau\lambda} dF_{\xi\xi}(\lambda) \ , \quad \tau \in R \ ,$$

where the spectrum $F_{\xi\xi}$ is $p \times p$ with Hermitian n.n.d. increments, $dF_{\xi\xi}(\lambda) = E\{d\chi(\lambda)d\chi(\lambda)^*\}$.

We shall always assume that our processes have absolutely continuous spectra. In this case $\xi(t)$ has the representation (Ibragimov and Linnik, 1971, p. 298)

$$\xi(t) = \int_R \Gamma(\tau) d\eta(t-\tau) \ , \quad t \in R \ ,$$

where[9] $\Gamma(t) \in L^2$ and is $p \times p$ and $\eta(t)$ is a $p \times 1$ process with orthogonal increments and $E\{d\eta(t)d\eta(t)'\} = I_p dt$; moreover we have

$$dF_{\xi\xi}(\lambda) = f^c_{\xi\xi}(\lambda) d\lambda \ , \quad \lambda \in R \ ,$$

where $f^c_{\xi\xi}(\lambda)$ is the spectral density matrix of $\xi(t)$. We include the superscript c to indicate that the spectral density is that of

[9] By $\Gamma(t) \in L^q$ we mean $\Gamma(t)$ is measurable and $\int_R |\nu_1\{\Gamma(t)\}|^q dt < \infty$.

a continuous process. Correspondingly, we denote by $f^d_{\xi\xi}(\lambda)$ the spectral density matrix of the sequence $\xi(n)$, $n = 0, \pm\Delta, \pm 2\Delta, \ldots$, so that

$$f^d_{\xi\xi}(\lambda) = \sum_{j=-\infty}^{\infty} f^c_{\xi\xi}\left(\lambda + \frac{2\pi j}{\Delta}\right) , \quad -\frac{\pi}{\Delta} < \lambda \le \frac{\pi}{\Delta} .$$

For convenience we always take $\Delta = 1$. Then we have the Bochner-Herglotz representation

$$E\{\xi(t)\xi(t+j)'\} = \int_{-\pi}^{\pi} e^{ij\lambda} f^d_{\xi\xi}(\lambda)d\lambda \ , \quad j = 0, \pm 1, \ldots . \qquad (3.4)$$

At times in our work we shall be concerned with processes that are differentiable one or more times in the m.s. sense. The m.s. derivative is defined by

$$\underset{\tau\to 0}{\text{l.i.m.}} \frac{\xi(t+\tau)-\xi(t)}{\tau} = \frac{d}{dt}\,\xi(t) \ . \qquad (3.5)$$

A necessary and sufficient condition for (3.5) to hold uniformly in $t \in R$ is the existence of $\partial^2 E\{\xi(t)\xi(\tau)'\}/dt d\tau$ at all diagonal points $\tau = t$. The m.s. derivatives of higher order are likewise defined. Since $\xi(t)$ is stationary with absolutely continuous spectrum $d^n\xi(t)/dt^n$ exists in m.s. if and only if

$$\int_R \lambda^{2n} \mathrm{tr}\left\{f^c_{\xi\xi}(\lambda)\right\} d\lambda < \infty \ . \qquad (3.6)$$

(See Hannan, 1970, p. 55.) If $\xi(t)$ is m.s. differentiable it is almost surely (a.s.) continuous; if $\xi(t)$ is m.s. differentiable twice it is a.s. differentiable once, and so on.

We shall be very much concerned with the property of almost sure convergence, or convergence with probability one. In respect of our stationary processes we need the existence of certain ergodicity properties to ensure the convergence a.s. of statistics to their expected values. For a discussion of ergodicity see e.g. Cramér and

Leadbetter (1967, pp. 147-159). Briefly, however, we define a shift transform taking a set S (corresponding to a subset of the measure space) of variables $\xi(t)$ into a set S_τ of variables $\xi(t+\tau)$. A set S is called invariant if, for all fixed τ , S and S_τ differ at most by sets of probability measure zero. The invariant sets form a σ-field and include all sets of probability 0 or 1 . Then we call $\xi(t)$ ergodic if the σ-field of invariant sets contains *only* sets of probability 0 or 1 . If $\xi(t)$ is ergodic and $E\{\|\xi(t)\|\}$, $E\{\|\xi(t)\|^2\}$ are finite then

$$\lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} \xi(n) = E\{\xi(t)\} \text{ , a.s.,}$$

$$\lim_{N\to\infty} N^{-1} \sum_{1\le n, n+j\le N} \xi(n)\xi(n+j)' = E\{\xi(t)\xi(t+j)'\} \text{ , a.s.,}$$

the second property holding uniformly in fixed integral j such that $0 \le |j| < N$. In fact although we assume our underlying processes to be ergodic we use only these implied properties of the discrete sequences.

Under the assumptions of stationarity and ergodicity we shall prove strong laws of large numbers. Stricter conditions are needed, however, to prove central limit theorems for quantities such as $N^{-\frac{1}{2}} \sum_{n=1}^{N} \xi(n)$. While we do not assume the mutual independence of the $\xi(n)$ the introduction, to some extent, of the notion of independence seems necessary. A possible representation for $\xi(n)$ is the linear process

$$\xi(n) = \sum_{j=-\infty}^{\infty} A(j)\varepsilon(n-j) \text{ , } \sum_{j=-\infty}^{\infty} \nu_1\{A(j)\}^2 < \infty \text{ ,} \tag{3.7}$$

where the $p \times 1$ vectors $\varepsilon(n)$ are independent and identically distributed with null mean vector and covariance matrix I_p (i.i.d. $(0, I_p)$). Then (3.7) implies that $\xi(t)$ is stationary and ergodic. (See Hannan, 1970, p. 204). The specification is of some generality, including solutions of mixed autoregressive moving-averages, and widespread use of (3.7) has been made in establishing asymptotic theory. However when the underlying process is continuous it seems implausible to represent $\xi(n)$ as a moving-average of discrete random variables, and we shall make an alternative, more appropriate assumption.[10]

In recent years a series of "weak dependence" or "mixing" conditions, based on the proposition that the dependence between events decreases as the intervening time increases, has been made use of by such authors as Rosenblatt (1956), Rozanov (1967), Hannan (1970) and Ibragimov and Linnik (1971). It is doubtful whether central limit theorems can be proved under substantially weaker or intuitively more pleasing assumptions. We impose what we shall call the strong mixing condition, in order to make use of a basic theorem proved by Hannan (1970) under this condition[11] (together with a condition of Lyapounov type on the fourth cumulants). With $\xi(t)$, $t \in R$, we associate σ-fields F_a^b $(-\infty \leq a < b \leq \infty)$ generated by events depending upon $\xi(t)$ for $a \leq t \leq b$ only. Then $\xi(t)$ satisfies the strong mixing condition if there exists a positive function $\alpha(t)$, $0 \leq t < \infty$, such that $\alpha(t) \to 0$ as $t \to \infty$ and, for $\tau > t$,

$$\sup_{A \in F_{-\infty}^{t}, B \in F_{\tau}^{\infty}} |Pr(A \cap B) - Pr(A)Pr(B)| < \alpha(\tau - t) .$$

---

[10] Nevertheless the representation (3.7) satisfies the strong mixing condition about to be introduced.

[11] Incidentally Hannan (1970) and Rozanov (1967) refer to the condition as, respectively, uniform mixing and complete regularity. Ibragimov and Linnik (1971, p. 308) attach the description uniform mixing to a somewhat stricter condition.

CHAPTER 2

CONTINUOUS-TIME REGRESSIONS WITH DISCRETE DATA

2.1 Introduction

The present chapter provides the theoretical basis for our subsequent treatment of the estimation problem for a number of specific continuous-time systems. The general model we consider is

$$y(t) = B_0 \int_R \Gamma(\tau;\ \theta_0) z(t-\tau) d\tau + x(t) \ , \quad t \in R \ . \qquad (1.1)$$

Here,[12] $x(t)$ is a $p_1 \times 1$ vector residual variable; $y(t)$ is a $p_1 \times 1$ vector endogenous variable; $z(t)$ is a $p_2 \times 1$ vector exogenous variable; $B_0 = (\beta_{0jk})$ is a $p_1 \times p_3$ matrix of real parameters; $\theta_0 = (\theta_{0j})$ is a $1 \times q$ row vector of real parameters; $\Gamma(t;\ \theta)$ is a $p_3 \times p_2$ matrix of possibly non-linear functions or generalized functions of t and θ . *Throughout the thesis, when we are not necessarily referring to the true parameter values but to any admissible values satisfying the same conditions we omit the zero subscript;* in this case we would write simply θ, B . We shall not discuss particular examples of (1.1) at the present time. However, bearing in mind our discussions in later chapters, we make the following observation. Loosely speaking, (1.1) includes under suitable conditions the solutions of functional equations of the class[13]

$$\Lambda_1\{y(t)\} = \Lambda_2\{z(t)\} \ , \quad t \in R \ , \qquad (1.2)$$

where $y(t)$, $z(t) \in L^2$ and Λ_1, Λ_2 are linear combinations of differential, integral and translation operators. The equations (1.2) are similar to the class considered by Kitagawa (1937).

---

[12] We shall presently give conditions that ensure that (1.1) is well-defined as a m.s. limit.

[13] Omitting $x(t)$ and ignoring the stochastic nature of (1.1).

In the real world it may not be strictly reasonable to assume (1.1) to be valid over all $t \in R$. However, to restrict t to a given finite interval would make (1.1) more difficult to handle and it is convenient to suppose that our model continues to hold before and beyond the observation period. Alternatively we could confine ourselves to the half-line $t > 0$, say, but the Laplace transform techniques that are then indicated produce less elegant results than the Fourier methods we use. In the same way, we have taken R to be the range of integration, whereas $\Gamma(t;\, \theta) = 0$, $t < 0$, in a real life situation in which causation is implied. In any case, greater generality and flexibility will accrue if we allow for leads as well as lags.

We wish to estimate θ_0 and B_0. As previously indicated our estimation procedures are based not on a continuous record but on sequences $y(n)$, $z(n)$, $n = 1, \ldots, N$, with $y(t)$, $z(t)$ being sampled at equal, unit, intervals of time. Before we discuss the estimation of (1.1) it is convenient to indicate how it differs from and extends non-linear models of regression type that have recently been considered. We first introduce the Dirac delta function, $\delta(t)$, defined by

$$\delta(t) = \infty\,, \quad t = 0\,; \quad = 0\,, \quad t \neq 0\,; \quad \int_R \delta(t)dt = 1\,. \qquad (1.3)$$

By allowing Γ to involve delta functions we may derive a number of interesting models from (1.1). We prefer to deal explicitly with the generalized function $\delta(t)$, rather than express (1.1) as a Stieltje's integral. We first take $\Gamma(t;\, \theta) = \delta(t)\Gamma(\theta)$, say, where $\Gamma(\theta)$ is $p_3 \times p_2$. Then we have

$$y(t) = B_0\Gamma(\theta_0)z(t) + x(t)\,, \quad t \in R\,, \qquad (1.4)$$

a non-linear model of conventional type. Jennrich (1969) considered the scalar model

$$y_n = f_n(\theta_0) + x_n \ , \quad n = 1, \ldots, N \ , \tag{1.5}$$

where the observations do not necessarily form a time series. Assuming, in particular, that θ_0 lies in a compact subset of R^q and the x_n are i.i.d. $(0, \sigma^2)$, $\sigma^2 < \infty$, Jennrich proved a strong law of large numbers and a central limit theorem for the least squares (*LS*) estimate of θ_0 . Malinvaud (1970 a, b) has proved under somewhat different conditions weak consistency and a central limit theorem. The time series case of (1.5) was treated by Hannan (1971 a) under conditions similar to Jennrich's but allowing for stationary residuals. Robinson (1972), in a paper written during the research for the thesis, considered the vector model

$$y(n) = B_0 z(n; \theta_0) + x(n) \ , \quad n = 1, \ldots, N \ , \tag{1.6}$$

with non-linear constraints on θ_0 . Aitchison and Silvey (1958) earlier considered maximum-likelihood (*ML*) estimation under such constraints. Now because $z(n; \theta)$ may not admit a factorization $\Gamma(\theta)z(n)$, (1.1) and (1.4) exclude some cases of (1.5) and (1.6), for example the Cobb-Douglas production model

$$y(n) = \beta_0 \prod_{j=1}^{q} z_j(n)^{\beta_j} + x(n) \ , \quad n = 1, \ldots, N \ . \tag{1.7}$$

If we require, say, $\beta_j \geq 0$, $j = 1, \ldots, q$, $\sum_1^q \beta_j \leq 1$, thus constraining the β_j , $j \geq 1$, to a compact set, asymptotic theory for *LS* estimates follows from the papers we have cited above.[14] However the cases of (1.6) we discuss (in §3 and §6. Appendix) are

[14] Exponential type models such as (1.7) are more usually treated by replacing the additive residual by a multiplicative one and taking the log transform. For a comparison of the two approaches see Robinson (1969).

also of the form (1.4) and the extra generality does not seem worthwhile for its own sake. Furthermore the conditions needed on $z(n;\ \theta)$ are less perspicuous than the conditions we impose on $z(t)$.

For (1.4) - (1.7) it is of little consequence whether or not the underlying processes are regarded as continuous and these cannot account for the more interesting special cases of (1.1). No systematic attempt has apparently been made to estimate (1.1) or the systems of §§4-7 when only integer-indexed observations are available. By replacing t by n and choosing $B_0\Gamma(\tau;\ \theta_0)$ to be $\sum_{-\infty}^{\infty} B(j)\delta(\tau-j)$ we have the very popular distributed lag system

$$y(n) = \sum_{j=-\infty}^{\infty} B(j)z(n-j) + x(n)\ , \quad n = 1, \ldots, N\ . \tag{1.8}$$

The estimation of (1.8) is a relatively straightforward task although the parameter space must of course be of finite dimension; this is so if the sum in (1.8) is truncated or if (1.8) represents the solution of a system of linear difference equations. However, (1.8) is not an entirely satisfactory description of a relation between continuous processes. Sims (1971) has considered the validity of discrete approximations to continuous-time systems. If the process generated by (1.1) is sampled at unit intervals this sequence is also generated by a discrete model of the form (1.8); however (1.1) and (1.8) are not isomorphic because of the impossibility of interpolating a continuum on the basis of a discrete equal-spaced sequence. We shall consider the estimation of (1.1) rather than (1.8), but we shall need some way of replacing expressions involving the $z(t-\tau)$ by expressions involving the $z(n)$, and conditions to overcome an obstacle to identification, caused by aliasing. Some of the results of the present chapter appear in Robinson (1972) and other results in a paper the author is currently preparing.

The presence of a convolution integral in (1.1) leads us to consider its Fourier representation. In the m.s. sense (1.1) is equivalent to

$$\int_R e^{it\lambda}\{d\chi_y(\lambda)-B_0\tilde{\Gamma}(-\lambda;\ \theta_0)d\chi_z(\lambda)-d\chi_x(\lambda)\} = 0\ ,\quad t \in R\ ,$$

say, using (1.3.3) and with[15]

$$\tilde{\Gamma}(\lambda;\ \theta) = \int_R e^{i\lambda t}\Gamma(t;\ \theta)dt\ ,\quad \lambda \in R\ . \tag{1.9}$$

We propose to replace $d\chi_x$, $d\chi_y$ and $d\chi_z$ by the (normed and conjugated) discrete Fourier transforms, whose definition is typified by

$$w_x(s) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} x(n)\exp(in\lambda_s)\ , \tag{1.10}$$

$$\lambda_s = 2\pi s/N\ ,\quad -\tfrac{1}{2}N \le s \le [\tfrac{1}{2}N]\ .$$

For the purposes of estimation we thus consider the approximation to the discrete Fourier transform of (1.1)[16]

$$w_y(s) = B_0\tilde{\Gamma}(\lambda_s;\ \theta_0)w_z(s) + w_x(s)\ . \tag{1.11}$$

An explicit expression for $\tilde{\Gamma}$ must of course be available. In practice $\tilde{\Gamma}$ will often arise naturally from the structural form and it may indeed not be possible to actually evaluate the inverse transform to find Γ.[17] This is of no concern to us since our computing formulae involve $\tilde{\Gamma}$ rather than Γ. We may describe Γ and $\tilde{\Gamma}$ as respectively the kernel and transfer function of the

[15] We shall presently make formal assumptions concerning $x(t)$, $z(t)$ and (1.9).

[16] An alternative approach might involve fitting polynomials or trigonometric series to the $z_j(n)$ (using orthogonality properties to determine a suitable order), whereupon it may be possible to evaluate the integral in (1.1) and regress $y(t)$ on this function of t.

[17] Although of course a unique Fourier image must exist. Numerous transform pairs are tabulated in the Fourier transform literature.

filter. In (1.8) the parameter set may be reduced by taking the elements of the $B(j)$ to be ordinates of a polynomial or of a discrete frequency distribution. It may then be of interest to consider the "moments" of the "lag distribution". (See Dhrymes 1971 b.) We may do the same here: in the case $p_1 = p_2 = p_3 = 1$, for example, we have

$$E\{t^r\} = \frac{\partial^r}{\partial\lambda^r}\tilde{\Gamma}(\lambda;\ \theta_0)\Big|_{\lambda=0} \Big/ \tilde{\Gamma}(0;\ \theta_0)\ ,\quad r = 1, 2, \ldots .$$

Now whereas the sample provides information on the discrete sequences for the frequency band $(-\pi, \pi)$, we may not believe (1.1) to be a valid model for $y(n)$ over the whole of this band. Because of the assumptions we shall make we would wish to exclude low frequencies which we believe to correspond to trend, other frequencies corresponding to cyclic effects, and also frequencies for which the measuring device used produces an expecially low signal-to-noise ratio. Thus we restrict ourselves to a set $\mathcal{B} \subset (-\pi, \pi)$, composed of a finite number of disjoint open intervals that are symmetrical about $\lambda = 0$, so that if $\lambda \in \mathcal{B}$, $-\lambda \in \mathcal{B}$ also. Such a set $\mathcal{B}$ is considered also by Hannan and Robinson (1973). It will later be seen that the "smaller" $\mathcal{B}$ is the less stringent are the assumptions needed to overcome the aliasing problem.

We take $\hat{\theta}_N$, $\hat{B}_N$ to be the values, out of the set of all admissible θ, B, that absolutely minimize[18]

$$Q_N(\theta,\ B) = N^{-1} \sum_{\mathcal{B}} \left\| \Phi(\lambda_s)^{\frac{1}{2}} \{w_y(s) - B\tilde{\Gamma}(\lambda_s;\ \theta) w_z(s)\} \right\|^2 . \qquad (1.12)$$

Here and elsewhere, $\sum_{\mathcal{B}}$ is a sum over all $\lambda_s \in \mathcal{B}$ and

[18] The conditions we shall impose ensure that such values $\hat{\theta}_N$, $\hat{B}_N$ exist, and for N sufficiently large they will be uniquely defined a.s.

$\Phi(\lambda) = \Phi(\lambda)^{\frac{1}{2}}\Phi(\lambda)^{\frac{1}{2}*}$ is a n.n.d. $p_1 \times p_1$ Hermitian matrix with $\Phi(-\lambda) = \Phi(\lambda)'$, that is continuous in λ over $\mathcal{B}$. For convenience in our assumptions of §2.2 we shall always take $y(n)$ and $z(n)$ to be mean-corrected. In practice this may be accomplished by omitting from (1.12) the component for $s = 0$, which involves the deviation

$$w_y(0) - B\tilde{\Gamma}(0;\ \theta)w_z(0) = (N/2\pi)^{\frac{1}{2}}\{\overline{y} - B\tilde{\Gamma}(0;\ \theta)\overline{z}\}\ .$$

Now because of the symmetry of $\mathcal{B}$, $\hat{\theta}_N$ and $\hat{B}_N$ are ~~Q_N is~~ real.[19] Our problem has been reduced to one of non-linear regression. While considerable practical difficulties may be associated with the optimization of Q_N, the revolutionary development of high-speed computing facilities and the existence of sophisticated iterative procedures make many such tasks technically feasible.

The expression (1.12) is a Hermitian form in $w_x(s)$ with matrix consisting of diagonal blocks $\Phi(\lambda_s)$. It is motivated in the following way. We select some $\lambda \in \mathcal{B}$ and some integer $m << N$ and construct the frequencies $\lambda_{j_s} = 2\pi j_s/N$, $s = 1, \ldots, m$, the integers j_s being chosen so that the λ_{j_s} are clustered around λ. Then Hannan (1970) has shown that, under the assumptions we shall make for $x(n)$, the $w_x(j_s)$, $s = 1, \ldots, m$, converge to i.i.d. complex $N\left[0,\ f^d_{xx}(\lambda)\right]$ vectors, $f^d_{xx}(\lambda)$ being the spectral density matrix of $x(n)$, defined as in (1.3.4).[20] Thus although the parent

---

[19] It is implied that when the full band $(-\pi,\ \pi]$ is under consideration and N is even the component of frequency $\lambda_s = \pi$ is excluded, leaving only $N - 1$ summands. This is necessary because in general, and in the cases in which we are most interested,

$$Im\{\tilde{\Gamma}(\pi;\ \theta)\} \neq 0\ .$$

[20] There is a modification if $\lambda = 0, \pm\pi$; see Hannan (1970, Chapter IV, Theorems 13, 13').

$x(n)$ may be autocorrelated, for large N the summands in (1.12) are roughly independent and the problem of serial correlation may be eliminated. If Φ is replaced by ${f_{xx}^{d}}^{-1}$, when it exists, Q_N is a monotonic function of the likelihood based on i.i.d. normal $w_x(s)$; heuristically an optimal choice of Φ is thus suggested. Because f_{xx}^{d} is generally unknown we are led to discuss its estimation in §2.5. Our first task, however, is to introduce conditions that will be sufficient to give $\hat{\theta}_N$ and $\hat{B}_N$ certain desirable properties, as will be shown in §§2.3, 2.4. In discussing special cases of (1.1) in subsequent chapters we shall then be able to assert such properties with little or no proof and concentrate on other aspects.

2.2 Some assumptions

It is well known that asymptotic theory for non-linear regressions generally requires restrictions on the parameters. We confine the admissible θ to a set Θ defined as follows.[21]

(i) Θ *is the compact closure of a* $(q-r)$*-dimensional* C^k *manifold,*[22] $q > r \geq 0$, $k \geq 2$.

Requirements of compactness or boundedness have been used in similar circumstances by Jennrich (1969), Malinvaud (1970 a, b),

---

[21] Throughout, we index and refer to our assumptions by lower-case Roman numerals (i), (ii) etc.

[22] This may be defined as follows. A C^k map, $k \geq 0$, is a function f mapping an open set $A \subset R^{q-r}$ into a finite-dimensional vector space if f possesses continuous partial derivatives on A of all orders $\leq k$. Then the open set V is a $(q-r)$-dimensional C^k manifold if there exist a family of pairs (f_j, V_j) (where j is an index, V_j is an open set in V and f_j maps V_j into an open set in R^{q-r}) such that $\cup V_j = V$ and for all non-disjoint V_j, V_k, $f_k \circ f_j^{-1} : f_j(V_j \cap V_k) \to f_k(V_j \cap V_k)$ is a C^k map.

Hannan (1971 a) and Dhrymes (1971 a, b). The modification here and in Robinson (1972) is to allow for the presence of a number of given linear or non-linear equations in θ, which reduce the dimensionality of the manifold by the number of independent constraints. (If these equations do not themselves establish compactness the latter restriction is superimposed.) We have in mind circumstances in which, for example, θ is confined for the purpose of normalisation to the surface of a hypersphere or ellipsoid (both of which are compact C^{∞} manifolds) or θ consists of a number of mutually orthogonal vectors. The need to allow for such constraints is crucial in §3, and for one reason or another they may arise in connection with the other models we consider. We note that compactness, or at least boundedness, may enter in a fairly natural way into our considerations: first, to produce identification it may be necessary to "bound" θ since B is not constrained in this way; second to ensure the stability of certain functional equation systems (see §5). In any case when the choice of Θ is not determined on prior grounds it may be taken arbitrarily "large", subject to other restrictions, in which case no practical limitation is involved.

To allow for the incorporation of *a priori* information, to assist identification and also to facilitate the task of expressing a number of interesting models in the form (1.1) we allow B to satisfy the homogeneous linear restrictions

$$\beta C = 0 \, , \tag{2.1}$$

where C is a matrix of given constants, having $p_1 p_3$ rows and less than $p_1 p_3$ (possibly zero) linearly independent columns, and $\beta = \mathrm{vec}'\{B\}$.[23] The matrix

[23] We shall often use B and β interchangeably.

$$L = I_{p_1 p_3} - C(C'C)^{-1}C'$$

projects onto the orthogonal complement of $p_1 p_3$-dimensional vector space spanned by the rows of C and (2.1) is implied by the relations $\beta = \beta L$. (See Hannan, 1970, p. 416).

(ii) *The processes* $x(t)$, $z(t)$ *have null mean vectors, are mutually incoherent, are stationary and ergodic and have absolutely continuous spectra and continuous spectral density matrices. The sequences* $x(n)$, $z(n)$ *have continuous spectral density matrices* $f^d_{xx}(\lambda)$, $f^d_{zz}(\lambda)$ *respectively,* $|\lambda| \le \pi$, *defined as in* (1.3.4).

We have already explained the meaning of such conditions in §1.3.[24]

(iii) *Uniformly in* $\lambda \in R$, $\theta \in \Theta$, $\tilde{\Gamma}(\lambda;\, \theta)$ *exists and is given by* (1.9). *Uniformly in* $\theta \in \Theta$, $\tilde{\Gamma}(\lambda;\, \theta)$ *is continuous in* λ , *satisfying a Lipschitz condition of order greater than one half. Uniformly in* $\lambda \in B$, $\tilde{\Gamma}(\lambda;\, \theta)$ *is continuous in* θ .

The Lipschitz condition (see Zygmund, 1959, p. 43), needed for Theorem 2.6, is an unimportant restriction, every specific case of $\tilde{\Gamma}$ we consider being differentiable in λ and thus having modulus of continuity unity.

From (ii) and (iii) the relation

$$f^c_{yy}(\lambda) = B_0 \tilde{\Gamma}(\lambda;\, \theta_0) f^c_{zz}(\lambda) \tilde{\Gamma}(\lambda;\, \theta_0)^* B_0' + f^c_{xx}(\lambda) \qquad (2.2)$$

is integrable and thus (1.1) is well-defined in the m.s. sense. (*Cf.* Hannan, 1970, p. 8.)

---

[24] In fact continuity of the density matrix of the *continuous* process $x(t)$ is not necessary, but we have included the condition so as not to further complicate (ii).

Whereas (ii) is fundamental to our work stronger conditions on $z(t)$ will usually be needed because the attempt to estimate a continuous model from discrete data gives rise to an identification problem. Our estimation procedures are based on second-order functions of the data and we ignore any information provided by higher moments. Thus all we can attempt to estimate are the integer-lagged autocovariances between the elements of $[y(t)' \vdots z(t)']$, or equivalently the spectral density matrices

$$f_{yy}^{d}(\lambda) , \quad f_{yz}^{d}(\lambda) , \quad f_{zz}^{d}(\lambda) , \quad |\lambda| \le \pi .$$

Now from (2.2) we have, for $\lambda \in \mathcal{B}$,

$$\begin{aligned} f_{yy}^{d}(\lambda) &= \sum_{l=-\infty}^{\infty} f_{yy}^{c}(\lambda+2\pi l) \\ &= \sum_{-\infty}^{\infty} B_0 \tilde{\Gamma}(\lambda+2\pi l;\ \theta_0) f_{zz}^{c}(\lambda+2\pi l) \tilde{\Gamma}(\lambda+2\pi l;\ \theta_0)^{*} B_0' + f_{xx}^{d}(\lambda) . \qquad (2.3) \end{aligned}$$

Also we have

$$f_{yz}^{d}(\lambda) = B_0 \sum_{-\infty}^{\infty} \tilde{\Gamma}(\lambda+2\pi l;\ \theta_0) f_{zz}^{c}(\lambda+2\pi l) , \qquad (2.4)$$

$$f_{zz}^{d}(\lambda) = \sum_{-\infty}^{\infty} f_{zz}^{c}(\lambda+2\pi l) ,$$

$\lambda \in \mathcal{B}$. Now while we have information on $f_{zz}^{d}(\lambda)$ we have no means of isolating individual components $f_{zz}^{c}(\lambda+2\pi l)$. Thus there would no hope of identifying θ_0 and B_0 from (2.3) and (2.4) without the imposition of a further condition unless $\tilde{\Gamma}$ is periodic of period 2π ; $\tilde{\Gamma}$ has the latter form for the unlagged regression (1.4) $\{\tilde{\Gamma}(\lambda;\ \theta) \equiv \Gamma(\theta)\}$ and for the distributed lag system (1.8)

[25] More information could be gained by sampling in a less regular, even erratic, fashion. Such data would be much less susceptible to a neat mathematical treatment however, as well as being often difficult and expensive to record.

$\left(B\tilde{\Gamma}(\lambda;\ \theta) = \sum_{-\infty}^{\infty} B(j)\exp(ij\lambda)\right)$. Such models have formed the basis for most regression analyses of time series, but they will not be of concern to us. When $\tilde{\Gamma}$ is not periodic we need some way of choosing one out of the infinitely many $z(t)$ which generate the sequence $z(n)$.

(iv) *For* $\lambda = \mu + 2\pi l$ *,* $\mu \in B$ *,* $l = \pm 1, \pm 2, \ldots$ *and for* $|\lambda| > K$ *, some* $K < \infty$ *,*

$$f^{c}_{zz}(\lambda) = 0 \ .$$

Under this condition[26] there is no aliasing of the frequencies B , since $f^{c}_{zz}(\lambda) = f^{d}_{zz}(\lambda)$, $\lambda \in B$; any aliasing of $\overline{B} \cap (-\pi, \pi)$ is of no concern. (The extra requirement that f^{c}_{zz} be null everywhere outside an arbitrary finite interval is not practically restrictive but is needed in Theorems 2.5, 2.6.) When $B = (-\pi, \pi)$ we thus require $f^{c}_{zz}(\lambda) = 0$, $|\lambda| > \pi$, an assumption which is valid only if the sampling interval is sufficiently small to enable the detection of all frequency components of the continuous process. This "smoothness" assumption is strong, and one cannot adequately assess its validity without access also to data that are sampled at shorter intervals. Nevertheless in studying a continuous process it seems likely that one would sample with the intention of leaving a negligible spectral mass beyond the Nyquist frequency. For slow-moving economic series, for example, it may indeed be substantially concentrated in a much smaller band centred at the origin.

Henceforth, when no confusion is possible, we shall omit the superscript

[26] A slightly stronger condition, used by Hannan and Robinson (1973), would require the spectral mass to be null for $|\lambda| \geq 2\pi - \sup_{\lambda \in B} \lambda$.

d from the spectral densities of discrete processes. Then under (iv), for $\lambda \in \mathcal{B}$

$$f_{yy}(\lambda) = B_0\tilde{\Gamma}(\lambda;\ \theta_0)f_{zz}(\lambda)\tilde{\Gamma}(\lambda;\ \theta_0)^*B_0' + f_{xx}(\lambda)\ , \tag{2.5}$$

$$f_{yz}(\lambda) = B_0\tilde{\Gamma}(\lambda;\ \theta_0)f_{zz}(\lambda)\ . \tag{2.6}$$

Moreover, even if these relations do not serve to identify θ_0 and B_0 , $f_{xx}(\lambda)$ is identified as

$$f_{xx}(\lambda) = f_{yy}(\lambda) - f_{yz}(\lambda)f_{zz}(\lambda)^+f_{zy}(\lambda)\ ,$$

because from (1.2.6) we may replace f_{zz} in (2.5) by

$f_{zz}f_{zz}^+f_{zz}f_{zz}^+f_{zz}$ and then, from (2.6), $B_0\tilde{\Gamma}(\lambda;\ \theta_0)f_{zz}f_{zz}^+$ by $f_{yz}f_{zz}^+$.

Insofar as the identification problem stems from aliasing it is resolved by (iv), but this condition is by no means sufficient to ensure the identification of θ_0 and B_0 . Typically in systems whose reduced form is non-linear a number of somewhat arbitrary identification conditions are required, and these will be implied by the following statement.

(v) *Uniformly in admissible* $(\theta \vdots \beta) \neq (\theta_0 \vdots \beta_0)$

$$\mathrm{tr}\left\{(2\pi)^{-1}\int_{\mathcal{B}} \left[B_0\tilde{\Gamma}(\lambda;\ \theta_0)-B\tilde{\Gamma}(\lambda;\ \theta)\right]f_{zz}(\lambda)\left[B_0\tilde{\Gamma}(\lambda;\ \theta_0)\right.\right.$$
$$\left.\left. -B\tilde{\Gamma}(\lambda;\ \theta)\right]^*\Phi(\lambda)d\lambda\right\} > 0\ . \tag{2.7}$$

For an estimation procedure based on $Q_N(\theta,\ B)$, (v) is quite indispensable, for it asserts that the limit to which Q_N converges has a unique absolute minimum; corresponding conditions have been used by Jennrich (1969), Malinvauld (1970 a, b), Dhrymes (1971 a, b), Hannan (1971 a) and Robinson (1972). For reasonable Φ , f_{zz} the condition will hold when $B\tilde{\Gamma}(\lambda;\ \theta) \neq B_0\tilde{\Gamma}(\lambda;\ \theta_0)$ uniformly in admissible $(\theta \vdots \beta) \neq (\theta_0 \vdots \beta_0)$ almost everywhere in a non-degenerate subset of B ; since $\Phi \geq 0$ and $f_{zz} \geq 0$ uniformly, the left side of (2.7) is non-negative and will be positive if the matrix integral has a non-zero eigenvalue, uniformly in admissible $(\theta \vdots \beta) \neq (\theta_0 \vdots \beta_0)$. When we come to consider special cases of (1.1) we shall discuss the particular conditions which will produce identification. However it is convenient at this stage to comment generally on the effect on the possibility of identification of the presence of non-linear constraints, for which, as already indicated, we shall allow. Loosely speaking, when all constraints are linear the model is either (globally) identified or there exists a non-denumerable set of alternative $(\theta \vdots \beta)$ admissible in the sense (different from that above) that (2.7) is untrue. When θ satisfies a number of non-linear equations then although the model may not be globally identified it is possible that there exist only a finite or denumerable number of alternative admissible structures. In this case, when there exists a neighbourhood of the true parameter vector containing no other admissible vector, the model is said to be locally identified. (See for example Wald, 1950, Fisher, 1966, Chapter 5.) Then, at least in principle, Θ may be chosen to include one and only one of the admissible vectors and global identification is accomplished.

2.3 The strong law of large numbers

We prove first a theorem of fundamental importance to much of what follows.

THEOREM 2.1. *For* $k, l = 1, 2$ *define*

$$c_{kl}(j) = N^{-1} \sum_{1 \leq n, n+j \leq N} \xi_k(n)\xi_l(n+j) , \quad \gamma_{kl}(j) = \int_{-\pi}^{\pi} e^{ij\lambda} f_{kl}(\lambda) d\lambda ,$$

$$I_{kl}(s) = (2\pi N)^{-1} \sum_{m,m=1}^{N}\sum \xi_k(m)\xi_l(n)\exp\{i(m-n)\lambda_s\} \quad , \quad -\tfrac{1}{2}N < s \le [\tfrac{1}{2}N] \; ,$$

$$g_N(\theta;\, \varphi) = N^{-1} \sum_{B} I_{kl}(s)\varphi(\lambda_s;\, \theta) \;, \quad g(\theta;\, \varphi) = (2\pi)^{-1} \int_B f_{kl}(\lambda)\varphi(\lambda;\, \theta)d\lambda \,,$$

where $\varphi(\lambda;\, \theta)$ *is a complex-valued function that is continuous in* λ *, uniformly in* $\theta \in \Theta$ *, and continuous in* θ *, uniformly in* $\lambda \in B$ *. Let*

$$\lim_{N\to\infty} c_{kl}(j) = \gamma_{kl}(j) \;, \text{ a.s.,} \tag{3.1}$$

$k,\, l = 1,\, 2$ *, uniformly in fixed integral* j *,* $0 \le |j| < N$ *, and let* $f_{kl}(\lambda)$ *be continuous in* λ *. Then if condition* (i) *holds,*[27]

$$\lim_{N\to\infty} g_N(\theta;\, \varphi) = g(\theta;\, \varphi) \;, \text{ a.s.,}$$

uniformly in $\theta \in \Theta$ *.*

Proof. It is convenient to first prove pointwise convergence with B replaced by $(-\pi,\, \pi]$ and $\varphi(\lambda;\, \theta)$ taken to be continuous and periodic of period 2π . Then although the Fourier series of φ may not be absolutely convergent, since φ is continuous its Jth first-order Cesàro mean,

$$\varphi_J(\lambda;\, \theta) = (2\pi)^{-1} \int_{-\pi}^{\pi} \varphi(\mu;\, \theta)F_J(\lambda-\mu)d\mu \;, \tag{3.2}$$

converges uniformly in λ to φ , where by $F_J(\lambda)$ we shall always mean Fejér's kernel

$$F_J(\lambda) = J^{-1} \left|\sum_{j=0}^{J-1} e^{ij\lambda}\right|^2 = \sum_{|j|<J} \left(1-|j|J^{-1}\right)e^{ij\lambda} \;. \tag{3.3}$$

(See Zygmund, 1959, p. 89.) Then for given $\varepsilon > 0$

$$\sup_{\lambda} |\varphi(\lambda;\, \theta)-\varphi_J(\lambda;\, \theta)| < \varepsilon$$

for J sufficiently large. Thus

[27] Throughout the present section the compactness property only of (i) is needed, the differentiability properties being needed only for the central limit theorem.

$$|g_N(\theta;\ \varphi)-g_N(\theta;\ \varphi_J)| < \varepsilon N^{-1} \sum_s |I_{kl}(s)| \le \varepsilon N^{-1}\left[\prod_{k=1}^{2} \sum_s I_{kk}(s)\right]^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality, the sums being over $-\frac{1}{2}N < s \le [\frac{1}{2}N]$. The final expression is

$$(\varepsilon/2\pi)\left[\prod_{k=1}^{2} c_{kk}(0)\right]^{\frac{1}{2}} = O(\varepsilon) .$$

Since ε is arbitrary we may replace φ by φ_J . Now $g_N(\theta;\ \varphi_J)$ is

$$(2\pi N)^{-2} \sum_j \sum_m \sum_n \sum_s \int_{-\pi}^{\pi} \xi_k(m)\xi_l(n)\varphi(\lambda;\ \theta)\left(1-|j|J^{-1}\right)\exp\{i(m-n+j)\lambda_s - ij\lambda\}d\lambda$$

$$= (2\pi)^{-2} \sum_j \left[c_{kl}(j-N)+c_{kl}(j)+c_{kl}(j+N)\right]\left[\int_{-\pi}^{\pi} \varphi(\lambda;\ \theta)\left(1-|j|J^{-1}\right)e^{-ij\lambda}d\lambda\right]$$

for $N > J$ since

$$\sum_s \exp\left(in\lambda_s\right) = N\ ,\quad n = 0\ ,\quad \mathrm{mod}(N)\ ;\quad = 0\ ,\quad n \ne 0\ ,\quad \mathrm{mod}(N)\ .$$

However

$$c_{kl}(j-N) = 0\ ,\quad j \le 0\ ;\quad c_{kl}(j+N) = 0\ ,\quad j \ge 0\ .$$

Also, for $|j| \le J < N$, J fixed,

$$\lim_{N\to\infty} c_{kl}(j-N) = 0\ ,\quad j > 0\ ;\quad \lim_{N\to\infty} c_{kl}(j+N) = 0\ ,\quad j < 0\ ,$$

a.s. because the c_{kl} here involve sums over $N-j+1 \le n \le N$, $1 \le n \le -j$ respectively and the number of summands thus remains constant as $N \to \infty$. Then because of (3.1) and the boundedness of the Fourier coefficients of φ ,

$$\lim_{N\to\infty} g_N(\theta;\ \varphi_J) = g(\theta;\ \varphi_J) \quad \text{a.s.}$$

However

$$|g(\theta;\ \varphi_J)-g(\theta;\ \varphi)| < (\varepsilon/2\pi)\int_{-\pi}^{\pi}|f_{kl}(\lambda)|d\lambda$$

$$\leq (\varepsilon/2\pi)\left[\prod_{k=1}^{2}\int_{-\pi}^{\pi} f_{kk}(\lambda)d\lambda\right]^{\frac{1}{2}} = (\varepsilon/2\pi)\left[\prod_{k=1}^{2}\gamma_{kk}(0)\right]^{\frac{1}{2}} = O(\varepsilon)\ .$$

Thus pointwise convergence holds for our modified $g_N(\theta;\ \varphi)$.

We next show that pointwise convergence holds under the more general circumstances of the theorem. We introduce $\tilde{\varphi}$ defined as

$$\tilde{\varphi}(\lambda+2\pi l;\ \theta) = \varphi(\lambda;\ \theta)\ ,\quad \lambda \in B\ ;\quad = 0\ ,\quad \lambda \in \overline{B} \cap (-\pi,\ \pi)\ ,$$

$l = 0,\ \pm 1,\ \ldots$. Then $\tilde{\varphi}$ is periodic but has a finite number of discontinuities in $(-\pi,\ \pi]$ (in neighbourhoods of which the Cesàro mean fails to converge), namely the end-points of the intervals defining B and at π when $\varphi(\pi;\ \theta) \neq \varphi(-\pi;\ \theta)$. Let $C \subset (-\pi,\ \pi]$ be a union of neighbourhoods of these discontinuities and let $\int_C d\lambda = \eta < \infty$. (We follow a technique used by Hannan and Robinson, 1973.) Then

$$\left|N^{-1}\sum_{C} I_{kl}(s)\tilde{\varphi}(\lambda_s;\ \theta)\right| \leq \left\{\prod_{k=1}^{2} N^{-1}\sum_{C} I_{kk}(s)|\tilde{\varphi}(\lambda_s;\ \theta)|\right\}^{\frac{1}{2}}\ ,$$

where the sums are over $\lambda_s \in C$. The right side is bounded by

$$\left\{\prod_{k=1}^{2} N^{-1}\sum_{s} I_{kk}(s)\breve{\varphi}(\lambda_s;\ \theta)\right\}^{\frac{1}{2}}\ , \tag{3.4}$$

where $\breve{\varphi}$ is continuous and periodic and equal to $|\tilde{\varphi}|$ within C . However from what we have already proved (3.4) converges a.s. to

$$\left\{\prod_{k=1}^{2}(2\pi)^{-1}\int_{-\pi}^{\pi} f_{kk}(\lambda)\breve{\varphi}(\lambda;\ \theta)d\lambda\right\}^{\frac{1}{2}}\ . \tag{3.5}$$

If we choose $\breve{\varphi}$ to be small for $\lambda \notin C$ (3.5) is arbitrarily close to

$$(\eta/2\pi)\sup_{\lambda\in C}\breve{\varphi}(\lambda;\ \theta)\left\{\prod_{k=1}^{2}\sup_{\lambda\in C} f_{kk}(\lambda)\right\}^{\frac{1}{2}} = O(\eta)\ ,$$

since $\breve{\varphi}$, f_{kk} , $k = 1,\ 2$, are continuous. Thus the effect of the discontinuities is negligible and since $\tilde{\varphi}$ is continuous over $\overline{C} \cap (-\pi, \pi]$ its Cesàro mean converges uniformly to φ in $B \cap \overline{C}$ and zero in $\overline{B} \cap \overline{C} \cap (-\pi, \pi]$, and the pointwise convergence of g_N to g is established.

Finally we must show that the convergence is uniform in $\theta \in \Theta$; essentially this is implied by results of Jennrich (1969) but we give a proof for our different circumstances. Omitting the argument φ from g_N and g we have

$$|g_N(\theta)-g(\theta)| \le |g_N(\theta)-g_N(\theta_{\neq})| + |g_N(\theta_{\neq})-g(\theta_{\neq})| + |g(\theta_{\neq})-g(\theta)| , \quad (3.6)$$

for given $\theta, \theta_{\neq} \in \Theta$. For given $\varepsilon > 0$ there exists a neighbourhood N of $\theta_{\neq}$, $N \in \Theta$, such that

$$\sup_{\theta \in N} |g(\theta_{\neq})-g(\theta)| < \varepsilon , \quad (2\pi)^{-1} \int_B f_{kk}(\lambda)\sigma(\lambda;\theta_{\neq})d\lambda < \tfrac{1}{2}\varepsilon , \quad k = 1, 2 ,$$

where

$$\sigma(\lambda;\ \theta_{\neq}) = \sup_{\theta \in N} |\varphi(\lambda;\ \theta)-\varphi(\lambda;\ \theta_{\neq})| .$$

Also, for N sufficiently large,

$$|g_N(\theta_{\neq})-g(\theta_{\neq})| < \varepsilon , \text{ a.s.,}$$

$$\left|N^{-1} \sum_B I_{kk}(s)\sigma(\lambda_s;\ \theta_{\neq})-(2\pi)^{-1} \int_B f_{kk}(\lambda)\sigma(\lambda;\ \theta_{\neq})d\lambda\right| < \tfrac{1}{2}\varepsilon , \text{ a.s.,}$$

from what we have already proved and Lemma 1 of Jennrich (1969). Then uniformly in $\theta \in N$ and N sufficiently large

$$|g_N(\theta)-g_N(\theta_{\neq})| \le \left\{\prod_{k=1}^{2} N^{-1} \sum_B I_{kk}(s)\sigma(\lambda_s;\ \theta_{\neq})\right\}^{\frac{1}{2}}$$

$$< \left\{\prod_{k=1}^{2} (\tfrac{1}{2}\varepsilon+\tfrac{1}{2}\varepsilon)\right\}^{\frac{1}{2}} = \varepsilon , \text{ a.s.,}$$

using the Cauchy-Schwarz inequality, and thus

$$|g_N(\theta)-g(\theta)| < 3\varepsilon , \text{ a.s.,} \quad (3.7)$$

from (3.6). Now because Θ is compact every open cover of Θ contains a finite sub-cover. Thus (3.7) holds uniformly in $\theta \in \Theta$ and since ε is arbitrary the uniformity of convergence is established and the proof of the theorem concluded.

We shall use Theorem 2.1 principally to prove the consistency[28] of $\hat{\theta}_N$, $\hat{B}_N$. From (1.2.4) we have

$$\sum_{B} \mathrm{tr}\{B\tilde{\Gamma}(\lambda_s;\ \theta) I_{zy}(s)\Phi(\lambda_s)\} = \sum_{B} \mathrm{tr}\{I_{yz}(s)\tilde{\Gamma}(\lambda_s;\ \theta)^* B'\Phi(\lambda_s)\}$$

$$= N\beta h_N(\theta)\ ,$$

$$\sum_{B} \mathrm{tr}\{B\tilde{\Gamma}(\lambda_s;\ \theta) I_{zz}(s)\tilde{\Gamma}(\lambda_s;\ \theta)^* \Phi(\lambda_s)\} = N\beta H_N(\theta,\ \theta)\beta'\ ,$$

say, where $I_{yz}(s) = w_y(s)w_z(s)^*$ and I_{zy}, I_{zz} are analogously defined, and

$$h_N(\theta) = N^{-1} \sum_{B} \mathrm{vec}\{\Phi(\lambda_s) I_{yz}(s)\tilde{\Gamma}(\lambda_s;\ \theta)^*\}\ ,$$

$$H_N(\theta,\ \theta_{\neq}) = N^{-1} \sum_{B} \tilde{\Gamma}(-\lambda_s;\ \theta) I_{zz}(-s)\tilde{\Gamma}(\lambda_s;\ \theta_{\neq})' \otimes \Phi(\lambda_s)\ .$$

Then using (1.2.2), (1.12) may be re-written as

$$Q_N(\theta,\ B) = N^{-1} \sum_{B} \mathrm{tr}\{I_{yy}(s)\Phi(\lambda_s)\} - 2\beta h_N(\theta) + \beta H_N(\theta,\ \theta)\beta'\ . \qquad (3.8)$$

For a prescribed θ this is minimized by

$$\hat{\beta}_N(\theta) = h_N(\theta)' [LH_N(\theta,\ \theta)L]^+\ , \qquad (3.9)$$

where for L defined in §2.2 we use the fact that $\beta = \beta L$ and $L(LAL)^+ = (LAL)^+$ and we assume[29]

$$\mathrm{rank}\{LH_N(\theta,\ \theta)L\} = \mathrm{rank}\{L\}\ . \qquad (3.10)$$

Now because ~~as earlier noted, Q_N is real~~ $\mathcal{B}$ is symmetric it follows that h_N, H_N

---

[28] Whenever we speak of consistency we shall always mean *strong* consistency.

[29] It will later be apparent that, with $\theta = \hat{\theta}_N$, (3.10) is true a.s. for N sufficiently large.

and $\hat{\beta}_N(\theta)$ are ~~also~~ real. On inserting $\hat{\beta}_N(\theta)$ in place of β in (3.8) we deduce that, instead of minimizing Q_N simultaneously with respect to θ and B, we may find $\hat{\theta}_N$ by maximizing the concentrated function

$$R_N(\theta) = h_N(\theta)'\left[LH_N(\theta,\ \theta)L\right]^+ h_N(\theta)\ ,$$

and then compute $\hat{\beta}_N = \hat{\beta}_N\left(\hat{\theta}_N\right)$ from (3.9). Because numerical methods will generally be needed R_N may be the more manageable function since it involves fewer parameters. In any case, because our compactness assumptions relate to θ but not to β we shall establish the consistency of $\hat{\theta}_N$, and thence that of $\hat{\beta}_N$, by considering R_N.

THEOREM 2.2. *Let* (1.1) *and conditions* (i)-(iv) *hold. Then*

$$\lim_{N\to\infty} R_N(\theta) = R(\theta)\ ,\ \text{a.s.},$$

uniformly in $\theta \in \Theta$, *where*

$$R(\theta) = \beta_0 H\left(\theta_0,\ \theta\right)[LH(\theta,\ \theta)L]^+ H\left(\theta,\ \theta_0\right)\beta_0'\ ,$$

with

$$H\left(\theta,\ \theta_{\neq}\right) = (2\pi)^{-1}\int_B \tilde{\Gamma}(-\lambda;\ \theta)f_{zz}(-\lambda)\tilde{\Gamma}\left(\lambda;\ \theta_{\neq}\right)' \otimes \Phi(\lambda)d\lambda\ .$$

Proof. We must prove the convergence of $H_N(\theta,\ \theta)$, $h_N(\theta)$ to $H(\theta,\ \theta)$, $H\left(\theta,\ \theta_0\right)\beta_0'$ respectively. Now the elements of these involve products of elements of $\tilde{\Gamma}$ and Φ which, under (iii), satisfy the conditions on φ in Theorem 2.1. Moreover the elements of H_N, h_N and H are linear combinations of a finite number of terms g_N and g respectively because $y(t)$ and $z(t)$ are stationary and ergodic. Thus a.s. and uniformly in $\theta \in \Theta$,

$$\lim_{N\to\infty} \mathrm{H}_N(\theta,\ \theta) = \mathrm{H}(\theta,\ \theta)\ ,$$

$$\begin{aligned}\lim_{N\to\infty} \mathrm{h}_N(\theta) &= (2\pi)^{-1}\int_B \operatorname{vec}\left\{\Phi(\lambda)\mathrm{f}^d_{yz}(\lambda)\tilde{\Gamma}(\lambda;\ \theta)^*\right\}d\lambda \\ &= \sum_{-\infty}^{\infty}(2\pi)^{-1}\int_B \operatorname{vec}\left\{\Phi(\lambda)\mathrm{B}_0\tilde{\Gamma}\left(\lambda+2\pi l;\ \theta_0\right)\mathrm{f}^c_{zz}(\lambda+2\pi l)\tilde{\Gamma}(\lambda;\ \theta)^*\right\}d\lambda \\ &= \mathrm{H}\left(\theta,\ \theta_0\right)\beta_0'\ ,\end{aligned}$$

under (iv), using (1.2.3). The proof of the theorem is completed.

LEMMA 2.1. *If condition* (v) *holds then*

$$R\left(\theta_0\right) > R(\theta)\ ,$$

uniformly in admissible $\theta \neq \theta_0$.

Proof. Using (1.2.4),(2.7) may be re-written as

$$\beta_0\mathrm{H}\left(\theta_0,\ \theta_0\right)\beta_0' - 2\beta_0\mathrm{H}\left(\theta_0,\ \theta\right)\beta' + \beta\mathrm{H}(\theta,\ \theta)\beta' > 0\ . \qquad (3.11)$$

On replacing β by $\beta_0\mathrm{H}\left(\theta_0,\ \theta\right)[\mathrm{LH}(\theta,\ \theta)\mathrm{L}]^+$ (which is the limit to which $\hat{\beta}_N(\theta)$ converges) and using (1.2.6) the lemma is proved.

LEMMA 2.2. *If condition* (v) *holds then*

$$\operatorname{rank}\{\mathrm{LH}\left(\theta_0,\ \theta_0\right)\mathrm{L}\} = \operatorname{rank}\{\mathrm{L}\}\ .$$

Proof. On putting $\theta = \theta_0$ in (3.11) we have

$$0 < \left(\beta_0-\beta\right)\mathrm{H}\left(\theta_0,\ \theta_0\right)\left(\beta_0-\beta\right)' = \left(\beta_0-\beta\right)\mathrm{L}\left[\mathrm{LH}\left(\theta_0,\ \theta_0\right)\mathrm{L}\right]\mathrm{L}\left(\beta_0-\beta\right)'\ ,$$

uniformly in $\beta \neq \beta_0$, β , β_0 satisfying (2.1). The result of the lemma follows immediately.

THEOREM 2.3. *Denote by* $\left(\hat{\theta}_N \vdots \hat{\beta}_N\right)$ *the vector minimizing* $Q_N(\theta,\ \mathrm{B})$ *over admissible* $(\theta \vdots \beta)$, $\theta \in \Theta$. *Let* $\theta_0 \in \Theta$ *and let* (1.1) *and conditions* (i)-(v) *hold. Then*

$$\lim_{N\to\infty}\left(\hat{\theta}_N \vdots \hat{\beta}_N\right) = \left(\theta_0 \vdots \beta_0\right)\ ,\ \text{a.s.}$$

Proof. The proof of consistency of $\hat{\theta}_N$ is similar to the proof of Theorem 6 in Jennrich (1969). In the first place the existence of some $\hat{\theta}_N$ maximizing $R_N(\theta)$ over Θ follows from Jennrich's Lemma 2. (It will be apparent that if Θ is unbounded or bounded but not compact such a $\hat{\theta}_N$ may not exist.) Now suppose the sequence $\{\hat{\theta}_N\}$ has a sub-sequence $\{\hat{\theta}_n\}$ converging to $\theta_{\neq} \neq \theta_0$. Then from Theorem 2.2 and the continuity of $R(\theta)$, $R_n(\hat{\theta}_n) - R_n(\theta_0) > 0$ converges to $R(\theta_{\neq}) - R(\theta_0)$, which contradicts Lemma 2.1. Thus $\hat{\theta}_N$ is consistent. Because of this, the continuity in θ of $\hat{\beta}_N(\theta)$ and the uniform convergence of H_N, h_N ,

$$\hat{\beta}_N(\hat{\theta}_N) \to \beta_0 \mathrm{H}(\theta_0, \theta_0)[\mathrm{LH}(\theta_0, \theta_0)\mathrm{L}]^+ \text{ , a.s.}$$

However the multiplier of β_0 is

$$\mathrm{LH}(\theta_0, \theta_0)\mathrm{L}[\mathrm{LH}(\theta_0, \theta_0)\mathrm{L}]^+ = \mathrm{L} \text{ ,}$$

from Lemma 2.2. Thus $\hat{\beta}_N \to \beta_0 \mathrm{L} = \beta_0$ and the theorem is established.

2.4 The central limit theorem

Our next task is to establish the asymptotic normality of $N^{\frac{1}{2}}(\hat{\theta}_N - \theta_0 \vdots \hat{\beta}_N - \beta_0)$. As a preliminary the behaviour of more basic quantities must be examined. We need to supplement the assumptions of §2.2, in particular adopting a mixing condition of the type introduced in §1.3 to formally weaken the dependence within $\mathrm{x}(t)$.

(vi) $\mathrm{x}(t)$ *satisfies the strong mixing condition and, uniformly in* $k_j \in \{k \mid k = 1, \ldots, p_1\}$, $j = 0, 1, 2, 3$,

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=0}^{\infty} \left|\kappa_{k_0 k_1 k_2 k_3}(0, n_1, n_2, n_3)\right| < \infty \text{ ,}$$

the summand being the absolute value of the fourth cumulant of

$$x_{k_0}(n)\ ,\quad x_{k_1}(n+n_1)\ ,\quad x_{k_2}(n+n_2)\quad and\quad x_{k_3}(n+n_3)\ .$$

THEOREM 2.4. *Denote by* $T(\lambda)$ *a* p_1p_2*-rowed matrix that is continuous in* λ *over* B *and such that*

$$T(-\lambda) = Re\{T(\lambda)\} - i\ Im\{T(\lambda)\}\ .$$

Let conditions (ii) *and* (vi) *hold. Then as* $N \to \infty$

$$v_N(T) = N^{-\frac{1}{2}} \sum_B \{w_z(s)' \otimes w_x(s)^*\}T(\lambda_s) \xrightarrow{\mathcal{D}} N[0,\ \Sigma]\ ,$$

where

$$\Sigma = (2\pi)^{-1} \int_B T(\lambda)^*\{f_{zz}(-\lambda) \otimes f_{xx}(\lambda)\}T(\lambda)d\lambda\ .$$

Proof. As in Theorem 2.1 it is convenient to first replace B by $(-\pi,\ \pi]$ and take T to be continuous and periodic of period 2π. Then the Jth Cesàro mean, $T_J(\lambda)$, (*cf.* (3.2)) converges uniformly. We shall show that the effect of replacing T by T_J in $v_N(T)$ is asymptotically negligible in m.s. Putting

$$A_s = (2\pi)^{-1}\left[w_z(s)' \otimes I_{p_1}\right]\left[T(\lambda_s)-T_J(\lambda_s)\right]$$

we write the conditional expectation of $\|v_N(T-T_J)\|^2$, over $x(t)$ only, as

$$N^{-2}E \sum_m \sum_n \sum_s \sum_t \mathrm{tr}\{A_s^* x(m)x(n)'A_t\}\exp\{i(m\lambda_s - n\lambda_t)\}$$

$$= N^{-2} \int_{-\pi}^{\pi} \mathrm{tr}\left\{\left[\sum_n \sum_s A_s^*\exp\{in(\lambda_s-\lambda)\}\right] f_{xx}(\lambda) \left[\sum_n \sum_s A_s^*\exp\{in(\lambda_s-\lambda)\}\right]^*\right\}d\lambda$$

$$= N^{-2} \int_{-\pi}^{\pi} \left\|\sum_n \sum_s A_s^*\exp\{in(\lambda_s-\lambda)\} f_{xx}(\lambda)^{\frac{1}{2}}\right\|^2 d\lambda$$

$$\le N^{-2} \int_{-\pi}^{\pi} \left\|\sum_n \sum_s A_s^*\exp\{in(\lambda_s-\lambda)\}\right\|^2 \mathrm{tr}\{f_{xx}(\lambda)\}d\lambda\ , \tag{4.1}$$

by (1.2.2) and the Cauchy-Schwarz inequality. Since f_{xx} is continuous

$\sup \operatorname{tr}\{f_{xx}(\lambda)\} = c < \infty$ and (4.1) is not greater than

$$cN^{-2} \int_{-\pi}^{\pi} \sum_m \sum_n \sum_s \sum_t \operatorname{tr}\{A_s^* A_t\} \exp\{i[m(\lambda_s - \lambda) - n(\lambda_t - \lambda)]\} d\lambda$$

$$= 2\pi c N^{-1} \sum_s \|A_s\|^2 \le 2\pi c N^{-1} \sum_s \|T(\lambda_s) - T_J(\lambda_s)\|^2 \|w_z(s)\|^2 \ , \qquad (4.2)$$

on integrating. For given $\varepsilon > 0$ a sufficiently large J may be chosen for (4.2) to be bounded by

$$c\varepsilon^2 \operatorname{tr}\left\{N^{-1} \sum_n z(n) z(n)'\right\} = O(\varepsilon) \ .$$

Thus we may replace $v_N(T)$ by $v_N(T_J)$, which reduces to

$$(2\pi)^{-2} \sum_j N^{\frac{1}{2}} \operatorname{vec}'\{C_{xz}(j-N) + C_{xz}(j) + C_{xz}(j+N)\}\left\{\int_{-\pi}^{\pi} T(\lambda)\left(1 - |j| J^{-1}\right) e^{ij\lambda} d\lambda\right\}$$

for $N > J$, using (1.2.3) and with

$$C_{xz}(j) = N^{-1} \sum_n x(n) z(n+j)' \ .$$

Now (*cf.* the proof of Theorem 2.1) for $|j| < J$ either $N^{\frac{1}{2}} C_{xz}(j \pm N) = 0$ or $N^{\frac{1}{2}} C_{xz}(j \pm N) \to 0$. For fixed finite j the asymptotic normality of $N^{\frac{1}{2}} \operatorname{vec}\{C_{xz}(j)\}$ under (ii) and (vi) follows from Hannan (1970, Chapter IV, Theorem 10'), in which the properties of linear regression coefficients are considered. There are two differences that require comment. First, Hannan allows for certain non-stationarity effects, assuming, loosely speaking, that sample autocorrelations of $z(t)$ rather than autocovariances converge, and our stricter conditions are implied by these. Second, our $z(t)$ is lagged a finite number, j , units and this merely affects the covariance structure. We have

$$E\left[N \operatorname{vec}\{C_{xz}(j_1)\} \operatorname{vec}'\{C_{xz}(j_2)\}\right] = N^{-1} E \sum_m \sum_n z(m+j_1) z(n+j_2)' \otimes x(m) x(n)' \ ,$$

where the sums are over $1 \le m, m + j_1 \le N$, $1 \le n, n + j_2 \le N$.

This is

$$N^{-1}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathrm{f}_{zz}(\mu)\otimes \mathrm{f}_{xx}(\lambda)\exp\{i(j_2-j_1)\mu\}$$

$$\times \sum_m\sum_n \exp\{i(n-m)(\mu+\lambda)\}d\mu d\lambda\ . \quad (4.3)$$

However

$$N^{-1}\sum_m\sum_n e^{i(n-m)\lambda} = F_N(\lambda) - N^{-1}\left(\sum_m{}' \sum_{n=1}^{N} + \sum_{m=1}^{N}\sum_n{}'' - \sum_m{}'\sum_n{}''\right)e^{i(n-m)\lambda} \quad (4.4)$$

where $\sum'$, $\sum''$ are sums over a finite number, j_1, j_2 respectively, of terms. Then as $N\to\infty$, $F_N(\lambda)\to 2\pi\delta(\lambda)$, dominating (4.4) in a neighbourhood of $\lambda = 0$, beyond which all terms converge to zero. Thus (4.3) converges to

$$2\pi\int_{-\pi}^{\pi} \mathrm{f}_{zz}(-\lambda)\otimes \mathrm{f}_{xx}(\lambda)\exp\{i(j_2-j_1)\lambda\}d\lambda\ .$$

It follows that as $N\to\infty$

$$E\{\mathrm{v}_N(\mathrm{T}_J)'\mathrm{v}_N(\mathrm{T}_J)\}\to(2\pi)^{-1}\int_{-\pi}^{\pi}\mathrm{T}_J(\lambda)^*\{\mathrm{f}_{zz}(-\lambda)\otimes \mathrm{f}_{xx}(\lambda)\}\mathrm{T}_J(\lambda)d\lambda\ ,$$

which differs by arbitrarily little, as J is increased, from Σ. Thus the theorem is proved for the modified v_N. Under the more general circumstances of the theorem we define

$$\tilde{\mathrm{T}}(\lambda+2\pi l) = \mathrm{T}(\lambda)\ ,\quad \lambda\in B\ ;\quad = 0\ ,\quad \lambda\in\overline{B}\cap(-\pi,\pi]\ ,$$

$l = 0, \pm1, \ldots$. Now $\tilde{\mathrm{T}}$ is periodic but, within $(-\pi, \pi]$, discontinuous at a finite number of points. Let us, as before, denote the union of neighbourhoods of these C, with $\int_C d\lambda = \eta < \infty$. We choose $\tilde{\mathrm{T}}_\eta(\lambda)$ such that $\tilde{\mathrm{T}}(\lambda)+\tilde{\mathrm{T}}_\eta(\lambda)$ is continuous, $\lambda\in(-\pi,\pi]$, and equal to $\tilde{\mathrm{T}}(\lambda)$, $\lambda\in\overline{C}\cap(-\pi,\pi]$. From what has already been shown,

$$\mathrm{v}_N(\tilde{\mathrm{T}}+\tilde{\mathrm{T}}_\eta)\xrightarrow{\mathcal{D}} N[0,\Sigma_\eta]\xrightarrow{\eta\to0} N[0,\Sigma]\ .$$

Moreover $E\|\mathrm{v}_N(\tilde{\mathrm{T}}_\eta)\|^2$ is bounded by a quantity which converges to

$$(2\pi)^{-1}\int_{-\pi}^{\pi}\|\breve{T}(\lambda)\|^2\mathrm{tr}\{f_{zz}(\lambda)\}\mathrm{tr}\{f_{xx}(\lambda)\}d\lambda\ ,$$

where $\breve{T}$ is continuous over $(-\pi, \pi]$, equal to $\tilde{T}_\eta$, $\lambda \in C$, and such that $\|\breve{T}_\eta(\lambda)\|$ is small for $\lambda \in \overline{C} \cap (-\pi, \pi]$. Thus as η is arbitrary $v_N(\tilde{T}_\eta) \to 0$, in m.s. Then the asymptotic normality of $v_N(T)$ follows from a lemma of Bernstein (1926-7, p. 12) because

$$v_N(T) = v_N(\tilde{T}) = v_N(\tilde{T}+\tilde{T}_\eta) - v_N(\tilde{T}_\eta)\ ,$$

and the proof of the theorem is completed.

The use of an approximation for estimation purposes did not hinder us in establishing the consistency of the estimates with respect to the true model. To prove a relatively neat central limit theorem, however, the approximation seems to make necessary the imposition of a stronger (but not very restrictive) condition on $z(t)$ and the proving of a somewhat complicated theorem.

(vii) *Uniformly in* $t_j \in R$ *and* $k_j \in \{k \mid k = 1, \ldots, p_2\}$, $j = 1, \ldots, 4$, *the cumulant function of* $z_{k_1}(t_1)$, $z_{k_2}(t_2)$, $z_{k_3}(t_3)$ *and* $z_{k_4}(t_4)$ *is finite and given by*

$$\underset{\Sigma\lambda_j=0}{\int\int\int\int} f_{k_1k_2k_3k_4}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\exp\{i\Sigma t_j\lambda_j\}\Pi d\lambda_j\ ,$$

where $f_{k_1k_2k_3k_4}$ *is continuous over* $\Sigma\lambda_j = 0$.

THEOREM 2.5. *Define*

$$I(s;\ \tau) = (2\pi N)^{-1}\sum_{n,m=1}^{N}\sum z_k(m-\tau)z_l(n)\exp\{i(m-n)\lambda_s\}\ ,\quad \lambda_s \in B\ ,$$

$$q_N = N^{-\frac{1}{2}}\sum_B \tilde{a}(\lambda_s)\left\{\int_R b(\tau)I(s;\ \tau)d\tau - \tilde{b}(\lambda_s)I(s;\ 0)\right\}\ ,$$

where $\tilde{a}(\lambda)$ *is continuous in* λ *over* B *and*

$$\tilde{b}(\lambda) = \int_R b(t)e^{i\lambda t}dt$$

satisfies a Lipschitz condition of order greater than one half over $[-\pi, \pi]$. *Then if* $\mathbf{z}(t) = \{z_k(t)\}$ *is second-order stationary with absolutely continuous spectrum and continuous spectral density and conditions* (iv) *and* (vii) *hold,*

$$\underset{N\to\infty}{\text{l.i.m.}}\ q_N = 0 .$$

Proof. Using the abbreviation

$$\sum{}' = \sum_B \sum_B \sum_{n_1} \sum_{n_2} \sum_{n_3} \sum_{n_4}$$

we write $\mathsf{E}|q_N|^2$ as

$$(4\pi^2 N^3)^{-1} \sum{}' \ \tilde{a}(\lambda_s)\tilde{a}(-\lambda_t)\mathsf{E}\Big\{\Big[\int_R \int_R b(\tau_1)b(\tau_2)z_k(n_1-\tau_1)z_k(n_2-\tau_2)d\tau_1 d\tau_2$$

$$- \int_R b(\tau)\tilde{b}(-\lambda_t)z_k(n_1-\tau)z_k(n_2)d\tau - \tilde{b}(\lambda_s) \int_R b(\tau)z_k(n_1)z_k(n_2-\tau)d\tau$$

$$+\tilde{b}(\lambda_s)\tilde{b}(-\lambda_t)z_k(n_1)z_k(n_2)\Big]z_k(n_3)z_l(n_4)\Big\}\exp\{i[(n_1-n_3)\lambda_s-(n_2-n_4)\lambda_t]\} ,$$

that is,

$$(4\pi^2 N^3)^{-1} \sum{}' \int\int_{\Sigma\lambda_j=0}\int\int \tilde{a}(\lambda_s)\tilde{a}(-\lambda_t)[\tilde{b}(-\lambda_1)-\tilde{b}(\lambda_s)][\tilde{b}(-\lambda_2)-\tilde{b}(-\lambda_t)]$$

$$\times [f^c_{kl}(\lambda_3)f^c_{kl}(\lambda_4)\delta(\lambda_1+\lambda_3)\delta(\lambda_2+\lambda_4)+f^c_{kl}(\lambda_4)f^c_{kl}(\lambda_3)\delta(\lambda_1+\lambda_4)\delta(\lambda_2+\lambda_3)$$

$$+f^c_{kk}(\lambda_1)f^c_{ll}(\lambda_3)\delta(\lambda_1+\lambda_2)\delta(\lambda_3+\lambda_4)+f(\lambda_1, \lambda_2, \lambda_3, \lambda_4)]$$

$$\times \exp\{i[\Sigma n_j\lambda_j+(n_1-n_3)\lambda_s-(n_2-n_4)\lambda_t]\}\Pi d\lambda_j .$$

Here f^c_{kk}, f^c_{ll} and f^c_{kl} are spectral and cross-spectral densities of z_k, z_l , and $f = f_{kkll}$ in the notation of (vii). Let us consider first the term in $f^c_{kl}(\lambda_3)f^c_{kl}(\lambda_4)$, which corresponds to $|\mathsf{E}\{q_N\}|^2$. We must show that

$$N^{-\frac{1}{2}} \sum_B \tilde{a}(\lambda_s) \int_R [\tilde{b}(\lambda)-\tilde{b}(\lambda_s)]f^c_{kl}(\lambda)F_N(\lambda-\lambda_s)d\lambda \to 0 \qquad (4.5)$$

as $N \to \infty$. However,

$$N^{\frac{1}{2}} \int_R \left[\tilde{b}(\lambda)-\tilde{b}(\lambda_s)\right] f_{kl}^{c}(\lambda) F_N(\lambda-\lambda_s)\, d\lambda$$

$$= N^{\frac{1}{2}} \int_B \left[\tilde{b}(\lambda)-\tilde{b}(\lambda_s)\right] f_{kl}^{c}(\lambda) F_N(\lambda-\lambda_s)\, d\lambda$$

$$+ N^{\frac{1}{2}} \sum_{1 \le |j| \le J} \int_{(-\pi,\pi)\cap\overline{B}} \left[\tilde{b}(\lambda+2\pi j)-\tilde{b}(\lambda_s)\right] f_{kl}^{c}(\lambda+2\pi j) F_N(\lambda+2\pi j-\lambda_s)\, d\lambda ,$$

$J < \infty$, because of (iv). The first term is not greater than

$$N^{\frac{1}{2}} \sup_{\lambda \in B} |f_{kl}^{c}(\lambda)| \int_B |\tilde{b}(\lambda)-\tilde{b}(\lambda_s)| F_N(\lambda-\lambda_s)\, d\lambda = O\left(N^{\frac{1}{2}-\alpha}\right) \qquad (4.6)$$

if $\tilde{b}(\lambda)$ satisfies a Lipschitz condition of order α (Zygmund, 1959, p. 91); since $\alpha > \frac{1}{2}$, (4.6) $\to 0$. For $\lambda_s \in B$ and B open, $0 < \Delta \le |\lambda+2\pi(j+m)-\lambda_s| \le \pi$, all $\lambda \in \overline{B} \cap (-\pi, \pi)$ for some m; thus the second term converges to zero since $F_N(\lambda) = O\left(N^{-1}\operatorname{cosec}^2\frac{1}{2}\Delta\right)$, $0 < \Delta \le |\lambda| \le \pi$, F_N is periodic of period 2π and the number of summands is finite. Then because $\tilde{a}$ is continuous (4.5) is true.

The remaining terms in $E|q_N|^2$ are more difficult to handle. We consider the term involving $f(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. We shall show first that

$$r_N = \left(4\pi^2 N^3\right)^{-1} \iint\limits_{\Sigma\lambda_j=0} \iint \sum{}' \tilde{c}(\lambda_s)\tilde{d}(-\lambda_t) g(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$

$$\times \exp\{i\left[\Sigma n_j\lambda_j+(n_1-n_3)\lambda_s-(n_2-n_4)\lambda_t\right]\}\Pi d\lambda_j$$

converges as $N \to \infty$ to

$$(2\pi)^{-1} \int_B \int_B \tilde{c}(\mu)\tilde{d}(-\lambda) g(-\mu,\lambda, \mu,-\lambda)\, d\mu\, d\lambda , \qquad (4.7)$$

where $\tilde{c}$ and $\tilde{d}$ are continuous and

$$\sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} \sum_{j_3=-\infty}^{\infty} \sum_{j_4=-\infty}^{\infty} g(\lambda_1+2\pi j_1, \lambda_2+2\pi j_2, \lambda_3+2\pi j_3, \lambda_4+2\pi j_4) \qquad (4.8)$$

is continuous over $\Sigma\lambda_j = 0$, $|\lambda_j| < \pi$, $j = 1, \ldots, 4$, the summands in (4.8) being zero when any argument $\lambda_m + 2\pi j_m$ is such that $\lambda_m \in B$ and $|j_m| \geq 1$. (We note that f satisfies the conditions on g because of (iv).) Now since (4.8) occurs in r_N only on the sub-manifold $\Sigma\lambda_j = 0$ we may make the transformation $\mu_1 = \lambda_1$, $\mu_2 = \lambda_1 + \lambda_3$, $\mu_3 = \lambda_1 + \lambda_2 + \lambda_3$ and express it in terms of the μ_j , denoting it $h(\mu_1, \mu_2, \mu_3)$. Then replacing integration over $\Sigma\lambda_j = 0$ by integration over $-\pi < \mu_1, \mu_2, \mu_3 < \pi$ and noting the periodicity of the exponential term we write r_N as

$$r_N(h) = (4\pi^2 N^3)^{-1} \sum{}' \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \tilde{c}(\lambda_s)\tilde{d}(-\lambda_t)h(\mu_1, \mu_2, \mu_3)$$
$$\times \exp\{i[(n_1-n_3)(\lambda_s+\mu_1)-(n_2-n_4)(\lambda_t-\mu_3)-(n_2-n_3)\mu_2]\}\Pi d\mu_j \ .$$

Apart from neighbourhoods of $\mu_j = \pm\pi$, $j = 1, 2, 3$, at which h is not continuous, the Jth Cesàro mean of the Fourier series of h , denoted h_J , converges uniformly in $(-\pi, \pi)$. The discontinuities of h are of no concern for their neighbourhoods may be taken sufficiently small for their contribution to be negligible. Now for given $\varepsilon > 0$ a sufficiently large J may be chosen so that $|r_N(h-h_J)|$ is less than

$$\varepsilon(4\pi^2 N^3)^{-1} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\sum{}' \ \tilde{c}(\lambda_s)\tilde{d}(-\lambda_t)\exp\{i[(n_1-n_3)(\lambda_s+\mu_1)$$
$$-(n_2-n_4)(\lambda_t-\mu_3)-n_2\mu_2+n_3\mu_2]\}|\Pi d\mu_j$$

$$\leq \varepsilon(4\pi^2 N^3)^{-1}\left[\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left|\sum_B \sum_{n_1} \sum_{n_3} \tilde{c}(\lambda_s)\exp\{i[(n_1-n_3)(\lambda_s+\mu_1)+n_3\mu_2]\}\right|^2 \Pi d\mu_j\right.$$

$$\left.\times \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left|\sum_B \sum_{n_2} \sum_{n_4} \tilde{d}(-\lambda_t)\exp\{i[(n_4-n_2)(\lambda_t-\mu_3)-n_2\mu_2]\}\right|^2 \Pi d\mu_j\right]^{\frac{1}{2}} ,$$

by the Cauchy-Schwarz inequality. This expression is

$$\varepsilon\left(4\pi^2N^3\right)^{-1}\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum{}'\ \tilde{c}(\lambda_s)\tilde{c}(-\lambda_t)\exp\{i\left[(n_1-n_3)\lambda_s-(n_2-n_4)\lambda_t\right.\right.$$

$$\left.+(n_1-n_2-n_3+n_4)\mu_1+(n_3-n_4)\mu_2\right]\}\Pi d\mu_j \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum{}'\tilde{d}(\lambda_s)\tilde{d}(-\lambda_t)$$

$$\left.\times\ \exp\{i\left[(n_1-n_3)\lambda_s-(n_2-n_4)\lambda_t+(n_1-n_2)\mu_2-(n_1-n_2-n_3+n_4)\mu_3\right]\}\Pi d\mu_j\right]^{\frac{1}{2}} .$$

On integrating this reduces to

$$2\pi\varepsilon N^{-1}\left[\sum_B\sum_B \tilde{c}(\lambda_s)\tilde{c}(-\lambda_t)F_N(\lambda_s-\lambda_t)\ \sum_B\sum_B \tilde{d}(\lambda_s)\tilde{d}(-\lambda_t)F_N(\lambda_s-\lambda_t)\right]^{\frac{1}{2}}$$

$$= 2\pi\varepsilon N^{-1}\left[\sum_B |\tilde{c}(\lambda_s)|^2 \sum_B |\tilde{d}(\lambda_s)|^2\right]^{\frac{1}{2}} = O(\varepsilon) .$$

Thus we may consider $r_N(h_J)$, which is

$$\left(32\pi^5N^3\right)^{-1}\sum{}'\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\tilde{c}(\lambda_s)\tilde{d}(-\lambda_t)\left[\sum_{j_1}\sum_{j_2}\sum_{j_3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}h(\lambda_1,\ \lambda_2,\ \lambda_3)\right.$$

$$\left.\times\ \left(1-|j_1|J^{-1}\right)\left(1-|j_2|J^{-1}\right)\left(1-|j_3|J^{-1}\right)\exp\left\{i\sum_m j_m(\lambda_m-\mu_m)\right\}\right]\Pi d\lambda_m$$

$$\times\ \exp\{i\left[(n_1-n_3)(\lambda_s+\mu_1)-(n_2-n_4)(\lambda_t-\mu_3)-(n_2-n_3)\mu_2\right]\}\Pi d\mu_m$$

$$=(2\pi N)^{-2}\sum_B\sum_B\sum_{j_1}\sum_{j_2}\sum_{j_3}\tilde{c}(\lambda_s)\tilde{d}(-\lambda_t)\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}h(\lambda_1,\ \lambda_2,\ \lambda_3)$$

$$\times\ \left(1-|j_1|J^{-1}\right)\left(1-|j_2|J^{-1}\right)\left(1-|j_3|J^{-1}\right)\exp\left\{i\left(\sum j_m\lambda_m+j_1\lambda_s-j_3\lambda_t\right)\right\}\Pi d\lambda_m$$

$$= 2\pi N^{-2}\sum_B\sum_B \tilde{c}(\lambda_s)\tilde{d}(-\lambda_t)h_J(-\lambda_s,\ 0,\ \lambda_t)$$

$$\to (2\pi)^{-1}\int_B\int_B \tilde{c}(\mu)\tilde{d}(-\lambda)h_J(-\mu,\ 0,\ \lambda)d\mu d\lambda ,$$

as $N\to\infty$. As J increases the last expression is arbitrarily close to

$$(2\pi)^{-1}\int_B\int_B \tilde{c}(\mu)\tilde{d}(-\lambda)h(-\mu,\ 0,\ \lambda)d\mu d\lambda .$$

But

$$h(-\mu, 0, \lambda) = g(-\mu, \lambda, \mu, -\lambda) \ , \quad \mu, \lambda \in B \ ,$$

since

$$g\left(-\mu-2\pi j_1, \lambda+2\pi j_2, \mu+2\pi j_3, -\lambda-2\pi j_4\right) = 0 \ , \quad \mu, \lambda \in B \ ,$$

$j_m = \pm 1, \pm 2, \ldots$, $m = 1, \ldots, 4$. Thus r_N converges to (4.7). Now on choosing $\tilde{c} = \tilde{d} = \tilde{a}\tilde{b}$, $g = f$, or $\tilde{c} = \tilde{a}\tilde{b}$, $\tilde{d} = \tilde{a}$, $g = \tilde{b}(-\lambda_2)f$, or $\tilde{c} = \tilde{a}$, $\tilde{d} = \tilde{a}\tilde{b}$, $g = \tilde{b}(-\lambda_1)f$, or $\tilde{c} = \tilde{d} = \tilde{a}$, $g = \tilde{b}(-\lambda_1)\tilde{b}(-\lambda_2)f$ we find that each of the four terms in f in $E|q_N|^2$ converges to

$$(2\pi)^{-1} \int_B \int_B \tilde{a}(\mu)\tilde{a}(-\lambda)\tilde{b}(\mu)\tilde{b}(-\lambda)f(-\mu, \lambda, \mu, -\lambda)d\mu d\lambda \ .$$

Thus the complete term in f converges to zero. In a manner that is so similar as to make a separate proof unnecessary it may be shown that each of the four terms in $f^c_{kl}(\lambda_4)f^c_{kl}(\lambda_3)$ converges to

$$(2\pi)^{-1} \int_B \left[\tilde{a}(\lambda)\tilde{b}(\lambda)f^c_{kl}(\lambda)\right]^2 d\lambda \ ,$$

and each of the four terms in $f^c_{kk}(\lambda_1)f^c_{ll}(\lambda_3)$ converges to

$$(2\pi)^{-1} \int_B |\tilde{a}(\lambda)\tilde{b}(\lambda)|^2 f^c_{kk}(\lambda)f^c_{ll}(\lambda)d\lambda \ .$$

The proof of the theorem is completed.

We are now equipped to establish the major result of the present section. As indicated in §2.2 our attention is directed principally to situations in which $\boldsymbol{\theta}$ satisfies a vector of r $(0 \leq r < q)$ possibly non-linear equations

$$\boldsymbol{\xi}(\boldsymbol{\theta}) = \left[\xi_1(\boldsymbol{\theta}) \ \vdots \ \ldots \ \vdots \ \xi_r(\boldsymbol{\theta})\right] = \mathbf{0} \ .$$

Correspondingly we introduce a $1 \times r$ Lagrange coefficient vector, $\boldsymbol{\gamma}$. We assume:

(viii) *Uniformly in* $\theta \in \Theta$ *the derivatives*

$$\tfrac{1}{2}(\partial/\partial\theta \,\vdots\, \partial/\partial\beta)\xi(\theta)' = \Xi$$

exist and are continuous; θ_0 *is a regular point*[30] *of* Ξ *and* $\mathrm{rank}\{\Xi_0\} = r$.

(ix) *Uniformly in* $\lambda \in R$, $\theta \in \Theta$ *the derivatives*

$$(\partial/\partial\theta_k)\tilde{\Gamma}(\lambda;\,\theta)\ ,\quad \left(\partial^2/\partial\theta_k\partial\theta_l\right)\tilde{\Gamma}(\lambda;\,\theta)\ ,\quad k,\,l = 1,\,\ldots,\,q\ ,$$

exist, being continuous in λ *, uniformly in* $\theta \in \Theta$ *, and continuous in* θ *, uniformly in* $\lambda \in B$ *;* $(\theta_0 \,\vdots\, \beta_0)$ *is a regular point of* $[\Psi(\Phi) \,\vdots\, \Xi']$ *, where*

$$\Psi(\Phi) = (2\pi)^{-1}\int_B T(\lambda;\,\theta,\,B)^*\{f_{zz}(-\lambda) \otimes \Phi(\lambda)\}T(\lambda;\,\theta,\,B)d\lambda\ ,$$

with

$$T(\lambda;\,\theta,\,B) = \left[(\partial/\partial\theta)\mathrm{vec}\{B\tilde{\Gamma}(\lambda;\,\theta)\} \,\vdots\, \left\{\tilde{\Gamma}(\lambda;\,\theta)' \otimes I_{p_1}\right\}L\right] .$$

In (viii), Ξ_0 is Ξ evaluated at θ_0 ; likewise $\Psi_0(\Phi)$ will denote $\Psi(\Phi)$ evaluated at $(\theta_0 \,\vdots\, \beta_0)$.

THEOREM 2.6. *Denote by* $(\hat{\theta}_N \,\vdots\, \hat{\beta}_N \,\vdots\, \hat{\gamma}_N)$ *the vector minimizing* $Q_N(\theta,\,B) + \xi(\theta)\gamma'$ *over admissible* $(\theta \,\vdots\, \beta \,\vdots\, \gamma)$, $\theta \in \Theta$. *Let* θ_0 *be an interior point of* Θ *and let* (1.1) *and conditions* (i)-(ix) *hold. Then as* $N \to \infty$

$$N^{\frac{1}{2}}(\hat{\theta}_N - \theta_0 \,\vdots\, \hat{\beta}_N - \beta_0 \,\vdots\, \hat{\gamma}_N) \xrightarrow{\mathcal{D}} N\left[0,\ \Psi_2^+\Psi_1\Psi_2^+\right] ,$$

where

[30] For a matrix $A(\theta)$ that is continuous in θ , θ_0 is a regular point of $A(\theta)$ if there exists a neighbourhood of θ_0 in which $\mathrm{rank}\{A(\theta)\}$ is constant. If the elements of $A(\theta)$ are analytic the set of irregular points has measure zero. (Fisher, 1966, p. 167.)

$$\Psi_1 = \begin{bmatrix} \Psi_0(\Phi f_{xx}\Phi) & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix}, \quad \Psi_2 = \begin{bmatrix} \Psi_0(\Phi) & \vdots & \Xi_0' \\ \cdots\cdots & \vdots & \cdots \\ \Xi_0 & \vdots & 0 \end{bmatrix},$$

$\Psi_0(\Phi f_{xx}\Phi)$ *being derived from* $\Psi_0(\Phi)$ *by replacing* $\Phi(\lambda)$ *by* $\Phi(\lambda)f_{xx}(\lambda)\Phi(\lambda)$.

Proof. For brevity we write $\alpha = (\theta \vdots \beta)$. Because the interior of Θ is an open set in a differentiable manifold there exists a neighbourhood N of α_0 which is diffeomorphic[31] to an open set in Euclidean space, and any such diffeomorphism establishes a co-ordinate system on N . For N sufficiently large $\hat{\alpha}_N \in N$ a.s., since $\hat{\alpha}_N \to \alpha_0$ a.s. Then there exists a sequence $\tilde{\alpha}_N \in N$, all N , which is tail equivalent to $\hat{\alpha}_N$ and (from Lemma 3 and Theorem 7 of Jennrich, 1969) for each equation a measurable function $\bar{\alpha}_N$ such that

$\|\bar{\alpha}_N-\alpha_0\| \leq \|\tilde{\alpha}_N-\alpha_0\|$ and

$$(\partial/\partial\tilde{\alpha}_j)S_N(\tilde{\alpha}_N,\hat{\gamma}_N) - (\partial/\partial\alpha_{0j})S_N(\alpha_0, \hat{\gamma}_N)$$
$$= (\tilde{\alpha}_N-\alpha_0)\left(\partial^2/\partial\bar{\alpha}_N'\partial\bar{\alpha}_N\right)S_N(\bar{\alpha}_N, \hat{\gamma}_N), \quad j = 1, \ldots, q+p_1p_3$$

$$0 = \xi_j(\tilde{\theta}_N) - \xi_j(\theta_0) = (\tilde{\theta}_N-\theta_0)(\partial/\partial\bar{\theta}_N')\xi_j(\bar{\theta}_N), \quad j = 1, \ldots, r,$$

applying the mean-value theorem, with $S_N(\alpha, \gamma) = Q_N(\theta, B) + \xi(\theta)\gamma'$. The $\bar{\alpha}_N$, $\bar{\theta}_N$ vary from equation to equation, of course. Then for N sufficiently large $\tilde{\alpha}_N = \hat{\alpha}_N$ and (since $(\partial/\partial\hat{\alpha}_N)S_N(\hat{\alpha}_N, \hat{\gamma}_N) = 0$)

$$-\tfrac{1}{2}\left[N^{\frac{1}{2}}(\partial/\partial\alpha_0)Q_N(\theta_0, B_0) \vdots 0\right] = N^{\frac{1}{2}}(\hat{\alpha}_N-\alpha_0 \vdots \hat{\gamma}_N)\bar{\Psi}_2 ,$$

where

---

[31] The C^k manifolds V, V' are diffeomorphic if there exists a homeomorphism $f : V \to V'$ such that f, f^{-1} are C^k mappings. Then f is said to be a diffeomorphism between V and V' .

$$\overline{\Psi}_2 = \frac{1}{2}\left[\begin{array}{c:c} \left(\partial^2/\partial\overline{\alpha}_N'\partial\overline{\alpha}_N\right)Q_N\left(\overline{\theta}_N,\overline{B}_N\right) & \left(\partial/\partial\overline{\alpha}_N'\right)\xi\left(\overline{\theta}_N\right) \\ \hdashline \left(\partial/\partial\alpha_0\right)\xi\left(\theta_0\right)' & 0 \end{array}\right],$$

our use of $\overline{\alpha}_N$ being symbolic, a different $\overline{\alpha}_N$ occurring in each column of $\overline{\Psi}_2$. Now

$$-\tfrac{1}{2}N^{\frac{1}{2}}\left(\partial/\partial\alpha_0\right)Q_N\left(\theta_0,\ B_0\right) = N^{\frac{1}{2}}\left(\partial/\partial\alpha_0\right)\left[\beta_0 h_N\left(\theta_0\right)-\tfrac{1}{2}\beta_0 H_N\left(\theta_0,\ \theta_0\right)\beta_0'\right] .$$

On substituting for $y(n)$ from (1.1) we write this as the sum of two terms. The first is

$$N^{-\frac{1}{2}}\sum_{B}\{w_z(s)'\otimes w_x(s)^*\Phi\left(\lambda_s\right)\}T\left(\lambda_s;\ \theta_0,\ B_0\right), \qquad (4.10)$$

whose limiting distribution, under (ii), (iii), (vi) and (ix), follows from Theorem 2.4. The second term is

$$N^{-\frac{1}{2}}\beta_0\sum_{B}\left\{\left[\int_R \Gamma\left(\tau;\ \theta_0\right)I(-s;\ \tau)d\tau-\tilde{\Gamma}\left(-\lambda_s;\ \theta_0\right)I(-s;\ 0)\right]\right.$$
$$\left.\otimes\ \Phi\left(\lambda_s\right)\right\}T\left(\lambda_s;\ \theta_0,\ B_0\right), \qquad (4.11)$$

where

$$I(s;\ \tau) = (2\pi N)^{-1}\sum_{n_1,n_2=1}^{N}\sum z\left(n_1-\tau\right)z\left(n_2\right)'\exp\{i\left(n_1-n_2\right)\lambda_s\} .$$

The expression (4.11) arises from the use of an approximation to estimate (1.1); in general it is non-null unless $\Gamma\left(t;\ \theta_0\right) \propto \delta(t)$. However each element is seen to have the form q_N of Theorem 2.5 and thus (4.11) converges to 0 in m.s. under (ii), (iii), (iv), (vii) and (ix). Then from (4.9), as $N \to \infty$

$$N^{\frac{1}{2}}(\hat{\alpha}_N-\alpha_0\ \vdots\ \hat{\gamma}_N)\overline{\Psi}_2 \xrightarrow{\mathcal{D}} N\left[0,\ \Psi_1\right] .$$

Now it is straightforward to show that $\overline{\Psi}_2 \to \Psi_2$ a.s. since $\hat{\alpha}_N \to \alpha_0$ and (i)-(iii) and (ix) hold. Because of the local identification implied by (v) and because α_0 is a regular point of both

$[\Psi(\Phi) \vdots \Xi']$ and Ξ , it follows as in Rothenberg (1971, p. 581)[32] that the former, evaluated at α_0 , has rank $q + \text{rank}\{L\}$. Since $\Psi_0(\Phi)$ has the form $A'A$, say, and the null spaces of A and $A'A$ are identical, using Khatri (1968, p. 269)

$$\text{rank}\{\Psi_2\} = \text{rank}\{A' \vdots \Xi_0'\} + \text{rank}\{\Xi_0\} = q + \text{rank}\{L\} + r ,$$

under (viii) and (ix). Thus

$$(\hat{\alpha}_N-\alpha_0 \vdots \hat{\gamma}_N)\Psi_2\Psi_2^+ = (\hat{\alpha}_N-\alpha_0 \vdots \hat{\gamma}_N)\begin{bmatrix} I_q & \vdots & 0 & \vdots & 0 \\ \cdots & & \cdots & & \cdots \\ 0 & \vdots & L & \vdots & 0 \\ \cdots & & \cdots & & \cdots \\ 0 & \vdots & 0 & \vdots & I_r \end{bmatrix} = (\hat{\alpha}_N-\alpha_0 \vdots \hat{\gamma}_N) .$$

The proof of the theorem is completed.

COROLLARY 2.6. *If* $r = 0$ *the covariance matrix in the limiting distribution of* $N^{\frac{1}{2}}(\hat{\theta}_N-\theta_0 \vdots \hat{\beta}_N-\beta_0)$ *is*

$$\Psi_0(\Phi)^+\Psi_0(\Phi f_{xx}\Phi)\Psi_0(\Phi)^+ . \qquad (4.12)$$

We note in conclusion that our limiting multivariate normal distribution in the corollary is of course singular when $\text{rank}\{L\} < p_1p_3$, whereas that in the theorem is singular, for $r \geq 1$, even when $\text{rank}\{L\} = p_1p_3$, having the same rank as Ψ_0 . Moreover in subsequent chapters it will be implicitly assumed that the matrix corresponding to Ψ_2 has the desired rank, and we shall omit explicit reference to the connection with regularity and local identification.

2.5 Efficient estimation

The efficiency of $(\hat{\theta}_N \vdots \hat{\beta}_N)$ depends on the choice of Φ . To

---

[32] In Robinson (1972) the following discussion is omitted, it merely being assumed that the matrix corresponding to Ψ_2 has the desired rank, since the author had not at that time read Rothenberg's article.

establish the optimal choice we prove the following theorem, under the assumption:

(x) $f_{XX}(\lambda)$ *is non-singular, uniformly in* $\lambda \in B$.

THEOREM 2.7. *Denote* $\Psi_2^+\Psi_1\Psi_2^+$ *by*

$$\left[\begin{array}{c:c} \Psi_{11} & \Psi_{12} \\ \hdashline \Psi_{12}' & \Psi_{22} \end{array}\right],$$

where the partition of rows and columns is $q + p_1p_3 : r$. *Then*

$$\Psi_{11} - \tilde{\Psi}\left(f_{XX}^{-1}\right) \geq 0 , \tag{5.1}$$

where

$$\tilde{\Psi}(\Phi) = \Psi_0(\Phi)^+\left\{I_{q+p_1p_3} - \Xi_0'\left[\Xi_0\Psi_0(\Phi)^+\Xi_0'\right]^+\Xi_0\Psi_0(\Phi)^+\right\} .$$

If condition (x) *holds and* $\Phi(\lambda) = f_{XX}(\lambda)^{-1}$, $\lambda \in B$, *the equality in* (5.1) *is attained and*

$$\Psi_{12} = 0 , \quad \Psi_{22} = \left[\Xi_0\Psi_0\left(f_{XX}^{-1}\right)^+\Xi_0'\right]^+ .$$

Proof. For convenience we write $\Psi_0\left(\Phi f_{XX}\Phi\right) = A$, $\Psi_0(\Phi) = B$ and $\Psi_0\left(f_{XX}^{-1}\right) = C$. Using a result of Rohde (1965) for a generalized inverse of a partitioned matrix,

$$\Psi_{11} = \tilde{\Psi}(\Phi)A\tilde{\Psi}(\Phi) , \quad \Psi_{12} = \tilde{\Psi}(\Phi)AB^+\Xi_0'\left[\Xi_0B^+\Xi_0\right]^+ ,$$

$$\Psi_{22} = \left[\Xi_0B^+\Xi_0'\right]^+\Xi_0B^+AB^+\Xi_0'\left[\Xi_0B^+\Xi_0'\right]^+ .$$

On taking $\Phi = f_{XX}^{-1}$ the desired values are easily obtained. To show that the choice is optimal, the left side of (5.1) is

$$\left[\tilde{\Psi}(\Phi) \;\vdots\; -\tilde{\Psi}\left(f_{XX}^{-1}\right)\right]\left[\begin{array}{c:c} A & B \\ \hdashline B & C \end{array}\right]\left[\tilde{\Psi}(\Phi) \;\vdots\; -\tilde{\Psi}\left(f_{XX}^{-1}\right)\right]' , \tag{5.2}$$

since

$$\tilde{\Psi}(\Phi)B\tilde{\Psi}\left(f_{xx}^{-1}\right) = \tilde{\Psi}\left(f_{xx}^{-1}\right)B\tilde{\Psi}(\Phi) = \tilde{\Psi}\left(f_{xx}^{-1}\right)C\tilde{\Psi}\left(f_{xx}^{-1}\right) = \tilde{\Psi}\left(f_{xx}^{-1}\right) .$$

Now (5.2) is n.n.d. if the central matrix of (5.2) is n.n.d. The last event occurs if a factorization of the integrand of the form DD^* is possible. Since $f_{zz} \geq 0$ and

$$\begin{bmatrix} \Phi f_{xx}\Phi & \Phi \\ \Phi & f_{xx}^{-1} \end{bmatrix} = \begin{bmatrix} \Phi f_{xx}^{\frac{1}{2}} \\ f_{xx}^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi f_{xx}^{\frac{1}{2}} \\ f_{xx}^{-\frac{1}{2}} \end{bmatrix}^*$$

such a factorization is available and the theorem is established.

COROLLARY 2.7. *The matrix* (4.12) *achieves its minimum,* $\Psi_0\left(f_{xx}^{-1}\right)^+$, *when* $\Phi = f_{xx}^{-1}$.

Unfortunately f_{xx} will generally be unknown and an estimate will be needed. The computational effort involved in the estimation of θ_0, β_0, f_{xx} and $\Psi_0\left(f_{xx}^{-1}\right)$ is likely to be considerable. However even when N is very large fast Fourier transform techniques enable the rapid calculation of spectral quantities, which are likely to be of interest in any case in a serious examination of time series data. The major problems will be encountered in attempting to find the optimal parameter estimates. We have the choice of minimizing $Q_N(\theta, B)$ or maximizing $R_N(\theta)$, both of which may be highly non-linear in θ . These functions may possess a large number of stationary points and a preliminary search procedure will usually be necessary to ensure the convergence of the iterative procedure to the optimal vector. While the estimation of some simple scalar models (reported in §§4-7) was not a troublesome task, when the parameter space is large even with modern computing facilities this process of finding suitable initial values may be extremely onerous and expensive, unless prior notions of plausibility substantially limit the region to be scanned. For the iterative stage numerous algorithms are available but when non-linear constraints are present Aitchison and Silvey's (1958) modification of Newton's method may be useful. *The fact that we shall not dwell further here on these computational aspects is no attempt to minimize their importance.* However it seems that the precise method of approach will be dictated in part by the nature of the model and it may perhaps be possible in the future to develop techniques that are especially suitable for some of our models. In any case we are more interested in the properties of our statistics then in their derivation. It seems worthwhile, nevertheless, to suggest estimates of f_{xx} and to discuss the extent to which their insertion leaves our asymptotic theory intact and provides efficient estimation of θ_0, β_0.

Let us suppose we have available initial consistent estimates $\overline{\theta}_N$, $\overline{B}_N$ possibly obtained by minimizing Q_N with $\Phi(\lambda) \equiv I_{p_1}$. Then naturally we would seek to estimate f_{xx} by means of the residuals that such estimates generate. A possible procedure, as in Hannan (1971 a), would be to assert that f_{xx} is rational of prescribed degree, and estimate its parameters. However we shall adopt procedures which avoid prior assumptions of this kind and are analogous to the broad-band efficient methods of Hannan (1963, 1970) for estimating linear regressions. For an integer $M << N$ we introduce the $2M$ sets

$$S_m = \{\lambda \mid \omega_m - \pi/2M < \lambda \le \omega_m + \pi/2M;\ \omega_m = \pi m/M;\ \lambda \in \mathcal{B}\}\ ,\quad |m| \le M-1\ ,$$

$$S_M = \{\lambda \mid -\pi < \lambda \le -\pi(1-\tfrac{1}{2M})\ ,\quad \pi(1-\tfrac{1}{2M}) < \lambda \le \pi;\ \lambda \in \mathcal{B}\}\ .$$

Then we estimate $f_{xx}(\lambda)$ by

$$\hat{f}_M(\lambda) = 2MN^{-1} \sum_{s(m)} \left[w_y(s) - \overline{B}_N\tilde{\Gamma}(\lambda_s;\ \overline{\theta}_N)w_z(s)\right]\left[w_y(s) - \overline{B}_N\tilde{\Gamma}(\lambda_s;\ \overline{\theta}_N)w_z(s)\right]^*, \quad (5.3)$$

all $\lambda \in S_m$, where the sum is over $\lambda_s \in S_m$. It will always be implied that the number of summands in (5.3) is not less than p_1. (Otherwise the $\hat{f}_M$ would certainly be singular.) Now the greater the value of N available the greater the value of M we would wish to take, in order to estimate f_{xx} at a finer grid of frequencies and improve the efficiency of our estimates of θ_0, B_0, bearing in mind Theorem 2.7.

THEOREM 2.8. *Denote by* $(\tilde{\theta}_N \vdots \tilde{\beta}_N)$ *the vector minimizing*

$$\tilde{Q}_N(\theta,\ B) = N^{-1} \sum_m \sum_{s(m)} \left\|\hat{f}_M(\omega_m)^{-\frac{1}{2}}\{w_y(s) - B\tilde{\Gamma}(\lambda_s;\ \theta)w_z(s)\}\right\|^2$$

over admissible $(\theta \vdots \beta)$, *where the first sum is over* $\omega_m \in B$. *Then under the conditions of Theorems* 2.3, 2.6 *and* (x) *there exists a sequence* M *increasing with* N *in such a way that*[33] *the conclusions of these theorems are valid for* $(\tilde{\theta}_N \vdots \tilde{\beta}_N)$, *the covariance matrix in the limiting distribution of* $N^{\frac{1}{2}}(\tilde{\theta}_N - \theta_0 \vdots \tilde{\beta}_N - \beta_0)$ *being* $\tilde{\Psi}\left(f_{xx}^{-1}\right)$, *and* $\Psi_0\left(f_{xx}^{-1}\right)$ *is consistently estimated by*

$$N^{-1} \sum_m \sum_{s(m)} T\left(\lambda_s;\ \tilde{\theta}_N,\ \tilde{B}_N\right)^*\left\{I_{zz}(-s) \otimes \hat{f}_M\left(\omega_m\right)^{-1}\right\}T\left(\lambda_s;\ \tilde{\theta}_N,\ \tilde{B}_N\right). \quad (5.4)$$

Proof. We prove first the existence of a sequence $M = M(N)$ increasing with N in such a way that

[33] We could introduce conditions to establish our results under a prescribed maximum rate of increase of M with N (as in Hannan, 1963, 1970) but the conditions needed are unverifiable and the exercise would have little practical value.

$$\lim_{M,N\to\infty} \sup_{\lambda\in B} \|\hat{f}_{M(N)}(\lambda)-f_{XX}(\lambda)\| = 0 \text{ , a.s.} \tag{5.5}$$

Using Theorem 2.1 and the consistency of $\overline{\theta}_N$, $\overline{B}_N$,

$$\lim_{N\to\infty} \hat{f}_M(\lambda) = (M/\pi) \int_{S_m} f_{XX}(\lambda)d\lambda = f_M(\lambda) \equiv f_M(\omega_m) \text{ ,} \tag{5.6}$$

say, a.s. and uniformly in $\lambda \in S_m \subset B$ and (fixed) M . Then

$$\lim_{M\to\infty} \sup_{\lambda\in B} \|f_M(\lambda)-f_{XX}(\lambda)\| = 0$$

as S_m degenerates. Now since

$$\hat{f}_M(\lambda) - f_{XX}(\lambda) = [\hat{f}_M(\lambda)-f_M(\lambda)] + [f_M(\lambda)-f_{XX}(\lambda)] \text{ ,}$$

(5.5) will follow if there exists an $M(N)$ such that

$$d_{M(N)} = \sup_{\lambda\in B} \|\hat{f}_M(\lambda)-f_M(\lambda)\| \to 0 \text{ , a.s.,}$$

as $N \to \infty$; that is, for all $\varepsilon > 0$, $\eta > 0$ there exists some $N(\varepsilon, \eta)$ such that

$$Pr\{|d_{M(N)}| \le \eta, \text{ uniformly in } N \ge N(\varepsilon, \eta)\} \ge 1 - \varepsilon \text{ .} \tag{5.7}$$

Now using (5.6) we introduce a monotonically increasing function $g(M)$, defined on integral M , such that

$$Pr\{|d_M| \le \eta, \text{ uniformly in } N \ge g(M)\} \ge 1 - \varepsilon_M \text{ ,}$$

$\varepsilon_M > 0$, $\sum \varepsilon_M < \infty$, where d_M is a function of M also. We choose $N(\varepsilon, \eta)$ such that $\sum_G \varepsilon_M \le \varepsilon$, G being the set of all M for which $g(M) \ge N(\varepsilon, \eta)$, and choose $M(N)$ to be $g^{-1}(N)$, the function inverse of $g(M)$. Then the left side of (5.7) is

$$1-Pr\{|d_{M(N)}| > \eta, \text{ some } N \ge N(\varepsilon, \eta)\} \ge 1-Pr\{\bigcup_G[|d_M| > \eta, \text{ some } N \ge g(M)]\}$$

$$\ge 1 - \sum_G Pr\{|d_M| > \eta, \text{ some } N \ge g(M)\} \ge 1 - \sum \varepsilon_M \ge 1 - \varepsilon \text{ .}$$

Thus we have proved (5.5). Then since $\det\{\cdot\}$ is continuous in its elements, a.s. $\det\{\hat{f}_M\} \to \det\{f_{XX}\} \ne 0$ under (x), for some sequence

$M = M(N)$. So, for this $M(N)$, for N sufficiently large $\hat{f}_M > 0$ a.s. and

$$\sup\left\|\hat{f}_M^{-1}-f_{XX}^{-1}\right\| \le \sup\left\|f_{XX}^{-1}\right\|\left\|f_{XX}-\hat{f}_M\right\|\left\|\hat{f}_M^{-1}\right\| < \varepsilon ,$$

a.s. for arbitrary $\varepsilon > 0$. Putting $\Phi = \hat{f}_M^{-1} - f_{XX}^{-1}$ in $H_N(\theta, \theta)$ and $h_N(\theta)$ their norms are found to bounded respectively by

$$(\varepsilon/2\pi)\sup\|\tilde{\Gamma}(\lambda;\ \theta)\|^2 \mathrm{tr}\left\{N^{-1} \sum z(n)z(n)'\right\} = O(\varepsilon) ,$$

$$(\varepsilon/2\pi)\sup\|\tilde{\Gamma}(\lambda;\ \theta)\| \left[\mathrm{tr}\left\{N^{-1} \sum y(n)y(n)'\right\}\mathrm{tr}\left\{N^{-1} \sum z(n)z(n)'\right\}\right]^{\frac{1}{2}} = O(\varepsilon) ,$$

a.s. for N sufficiently large and some $M(N)$. Then because Theorem 2.3 holds with $\Phi = f_{XX}^{-1}$ it must hold for some $M(N)$ with $\Phi = \hat{f}_M^{-1}$. We next show that Theorem 2.6 holds. Since $f_M \to f_{XX}$, $f_M > 0$ for M sufficiently large. Then (4.10) with $\Phi = \hat{f}_M^{-1} - f_M^{-1}$ is the sum of a finite number (for fixed M) of terms of the form

$$\mathrm{vec}'\left\{I_{p_2} \otimes \left[\hat{f}_M(\omega_m)^{-1}-f_M(\omega_m)^{-1}\right]\right\}N^{-\frac{1}{2}} \sum_{s(m)} w_z(s) \otimes w_x(-s) \otimes T(\lambda_s;\ \theta_0,\ B_0)$$

using (1.2.3). However for fixed M the first factor converges to 0 a.s. and uniformly and the expectation of the squared norm of the second factor converges to

$$(2\pi)^{-1} \int_{S_m} \|T(\lambda;\ \theta_0,\ B_0)\|^2 \mathrm{tr}\{f_{zz}(\lambda)\}\mathrm{tr}\{f_{XX}(\lambda)\}d\lambda = O(1) .$$

Thus with $\Phi = \hat{f}_M^{-1} - f_M^{-1}$, (4.10) $\to 0$ as similarly does (4.11). It follows that, for fixed M , $N^{\frac{1}{2}}(\tilde{\theta}_N-\theta_0 \vdots \tilde{\beta}_N-\beta_0)$ has the limiting distribution of Theorem 2.6, with $\Phi = f_M^{-1}$ in the covariance matrix. Then, increasing M , there exists a sequence $M(N)$ such that Theorem 2.6 holds with covariance matrix given by Theorem 2.7. Thus we have sequences $M_1(N)$, $M_2(N)$, say, for which Theorems 2.3, 2.6

respectively hold. Then the sequence $\min\{M_1(N), M_2(N)\}$ allows them to hold simultaneously. The consistency of (5.4) follows, and the proof of the theorem is completed.

When the full band $(-\pi, \pi)$ is under consideration an alternative estimator to (5.3) is

$$\tilde{f}_M(\lambda) = N^{-1} \sum_s \left[w_y(s) - B_N\tilde{\Gamma}(\lambda_s; \overline{\theta}_N)w_z(s)\right]\left[w_y(s) - \overline{B}_N\tilde{\Gamma}(\lambda_s; \overline{\theta}_N)w_z(s)\right]^* F_M(\lambda_s - \omega_m) \, , \qquad (5.8)$$

all $\lambda \in S_m$, the sum being over $\lambda_s \in (-\pi, \pi)$ with F_M given by (3.3). Then for fixed M,

$$\lim_{N\to\infty} \tilde{f}_M(\lambda) = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{xx}(\mu)F_M(\lambda-\mu)d\mu = f_M(\lambda) \, ,$$

say, a.s. and uniformly. Then because f_{xx} is continuous the Cesàro mean $f_M(\lambda) \to f_{xx}(\lambda)$ uniformly as $M \to \infty$. We could thus prove Theorem 2.7 in exactly the same way with $\tilde{f}_M$ in place of $\hat{f}_M$.

Finally we consider the sense in which our estimates may be described as efficient. We have already established in Theorem 2.7 the optimal estimates of the class minimizing $Q_N(\theta, B)$. In addition, if we suppose the asymptotic properties of the $w_x(s)$, mentioned in §2.1, to be valid for given M, N we are led to set up an approximate likelihood, taking the $w_x(s)$ to be i.i.d. complex $N[0, f_{xx}(\omega_m)]$, all $\lambda_s \in S_m$. Then let $\breve{\theta}_N, \breve{\beta}_N, \breve{f}_M(\omega_m)$ be the ML estimates. It is easily shown that, for fixed M, their information matrix is block diagonal with the $N^{\frac{1}{2}}[\breve{f}_M(\omega_m) - f_M(\omega_m)]$ asymptotically independent of $N^{\frac{1}{2}}(\breve{\theta}_N - \theta_0 \vdots \breve{\beta}_N - \beta_0)$, whose limiting covariance matrix is that of $N^{\frac{1}{2}}(\tilde{\theta}_N - \theta_0 \vdots \tilde{\beta}_N - \beta_0)$. As $M \to \infty$ both the latter have covariance matrix $\tilde{\Psi}\left(f_{XX}^{-1}\right)$. *Thus* $\left(\tilde{\theta}_N \vdots \tilde{\beta}_N\right)$ *is efficient in the sense that the normed estimates have the same limiting distribution, under our assumptions, as have the normed ML estimates, under Gaussian assumptions.*

The convergence of the $w_X(s)$ to normality under our assumptions suggests that $\left(\tilde{\theta}_N \vdots \tilde{\beta}_N\right)$ will be actually "close" to *ML* estimates. The discrepancy between $\left(\tilde{\theta}_N \vdots \tilde{\beta}_N\right)$ and $\left(\breve{\theta}_N \vdots \breve{\beta}_N\right)$ may be remedied, to the extent desired, by iterating. We may use $\left(\tilde{\theta}_N \vdots \tilde{\beta}_N\right)$ to re-estimate the $f_{XX}\left(\omega_m\right)$, which in turn are used to re-estimate $\left(\theta_0 \vdots \beta_0\right)$, continuing this process indefinitely and judging whether a satisfactory degree of convergence has been attained by, possibly, comparing the values of $\tilde{Q}_N$ on successive steps.

CHAPTER 3

THE ESTIMATION OF A REGRESSION MATRIX OF LESS THAN FULL RANK

3.1 Introduction

In the present chapter we consider a problem related to classical canonical correlation analysis, namely that of estimating a time series regression matrix that is constrained to have less than full rank. The asymptotic theory of our estimates will follow in a fairly direct way from the results of §2. However our work here differs from that of other chapters in that our model is unlagged, being of the special form (2.1.4); thus no approximation is needed for the purposes of estimation. However we make particular use of our provision in §2 for non-linear constraints. Many of the results of §3.1-3, Appendix appear in Robinson (1973); for the most part the work was carried out before the author thought of the model (2.1.1) and was based on the simpler discrete non-linear model considered in Robinson (1972). We suppose our underlying processes to be continuous here for the sake of conformity, while in fact our results are of equal relevance to discrete processes.

Later in the present section we shall discuss the relationship between our problem and that of ordinary canonical correlation analysis. Making an appropriate parameterization of our model we present in §3.2 asymptotic theory for our estimates, which may be efficient even when the residuals are serially correlated. The estimates must in general be obtained by iteration, and we leave to §3.Appendix a discussion of possible procedures. Some estimates that are closely related to ordinary canonical correlations and vectors, and obtained by solving an eigenvalue problem, are consistent when the residuals are stationary, but inefficient unless the residuals are independent. We always assume the rank of the regression matrix to be known *a priori*, but in §3.3 we present a test for determining rank. In §3.4 we consider a related problem that arises in econometrics, that of regression on an unobservable variable.

The theory of canonical correlation analysis (CCA), introduced by Hotelling (1936), has become established as a powerful exploratory tool in the analysis and summarisation of multivariate data, in such fields as economics, psychology, the biological sciences and geology. It enables the relationship between two sets of variables to be condensed into a relatively small number of equations involving orthonormal linear combinations of the variables. Let

$$Y = [y(1) \vdots \ldots \vdots y(N)] \ , \quad Z = [z(1) \vdots \ldots \vdots z(N)]$$

and consider a pair of vectors lying respectively in the p_1-space spanned by the rows of Y and the p_2-space spanned by the rows of Z. If the angle between them is smaller than the angle between any other such pair the cosine of this angle is designated as the first canonical correlation and the vectors themselves, uniquely defined up to simultaneous reversal of direction, as the first pair of canonical vectors. The second pair are the vectors orthogonal to the first pair that are separated by the smallest angle, and so on. The $r \le \min(p_1, p_2)$ greatest squared canonical correlations between Y and Z are denoted ρ^2_{aj}, $j = 1, \ldots, r$, and are the r greatest zeros ρ^2 of

$$\det\left\{\rho^2\hat{\Omega}_{yy}-\hat{\Omega}_{yz}\hat{\Omega}^{-1}_{zz}\hat{\Omega}_{zy}\right\} , \tag{1.1}$$

where, for example, $\hat{\Omega}_{yz} = N^{-1}YZ'$. Then the canonical variates are $\mu'_{aj}y(n)$, $\nu'_{aj}z(n)$, $j = 1, \ldots, r$, having correlations ρ_{aj}, where the μ_{aj}, ν_{aj} are the corresponding eigenvectors of

$\hat{\Omega}_{yz}\hat{\Omega}_{zz}^{-1}\hat{\Omega}_{zy}$ in the metric of $\hat{\Omega}_{yy}$ and $\hat{\Omega}_{zy}\hat{\Omega}_{yy}^{-1}\hat{\Omega}_{yz}$ in the metric of $\hat{\Omega}_{zz}$, respectively. The μ_{aj}, ν_{aj} obey the orthonormality conditions

$$\mu'_{aj}\hat{\Omega}_{yy}\mu_{ak} = \nu'_{aj}\hat{\Omega}_{zz}\nu_{ak} = \delta_{jk} \ , \quad j, k = 1, \ldots, r \ , \qquad (1.2)$$

where δ_{jk} is Kronecker's delta.

Distribution theory for canonical correlations and vectors appears until now to have been based on independence and normality assumptions for both $y(n)$ and $z(n)$. (See Anderson, 1951, 1958, Bartlett, 1947 a, b, Constantine and James, 1958, Hsu, 1939, 1941, James, 1964.) The use of normality assumptions is understandable in that we are concerned with correlations. The characteristic equation determining canonical correlations is symmetrical in Y and Z and thus it is not necessary to regard one particular vector variable as dependent on another, as in regression, but merely to think of Y and Z as being associated. Correspondingly, reference to a residual random variable has usually been suppressed. However we shall approach the problem from a regression point of view and will see that this enables us to derive asymptotic properties of what we may regard as generalized canonical estimates under alternative assumptions that do not involve normality or independence and appear to be relatively weak in other respects.

We consider first the model

$$y(n) = B_0 z(n) + x(n) \ , \quad n = 1, \ldots, N \ , \qquad (1.3)$$

where $x(n)$ is, for the time being, a sequence of i.i.d. $(0, \Omega_{xx})$ $p_1 \times 1$ vector variables and B_0 is a $p_1 \times p_2$ matrix of regression coefficients. It is believed that B_0 has rank $r < \min(p_1, p_2)$, a situation resulting from the existence of a number of (we assume unknown) linear or non-linear relations obtaining among the elements of B_0 . An appealing physical interpretation, offered by Brillinger (1969), is to regard B_0 as being capable of providing only r channels for the transmission of the information derived from $\mathbf{z}(n)$; one introduces the $p_1 \times r$ and $r \times p_2$ matrices B_1, B_2 , transmits the $r \times 1$ vector $B_2 z(n)$ and on receipt forms the $p_1 \times 1$ vector $B_1 B_2 z(n)$. The following theorem provides us with an estimator of B_0 .

THEOREM 3.1. *The function*

$$N^{-1}\mathrm{tr}\left\{(Y-BZ)(Y-BZ)'\hat{\Omega}_{xx}^{-1}\right\} , \tag{1.4}$$

where

$$\hat{\Omega}_{xx} = \hat{\Omega}_{yy} - \hat{\Omega}_{yz}\hat{\Omega}_{zz}^{-1}\hat{\Omega}_{zy} ,$$

is minimized with respect to B *of rank* $r \le \min(p_1, p_2)$ *by*

$$B_b = \sum_{j=1}^{r} \rho_{bj}\mu_{bj}\nu'_{bj} = \sum_{j=1}^{r} \mu_{bj}\mu'_{bj}\hat{\Omega}_{yz}\hat{\Omega}_{zz}^{-1} ,$$

where the ρ_{bj} *are the square roots of the* r *greatest zeros* ρ^2 *of*

$$\det\left\{\rho^2\hat{\Omega}_{xx}-\hat{\Omega}_{yz}\hat{\Omega}_{zz}^{-1}\hat{\Omega}_{zy}\right\} \tag{1.5}$$

and the μ_{bj}, ν_{bj} *are respectively the corresponding eigenvectors of* $\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{yz}\hat{\Omega}_{zz}^{-1}\hat{\Omega}_{zy}\hat{\Omega}_{xx}^{-1}$ *in the metric of* $\hat{\Omega}_{xx}^{-1}$ *and* $\hat{\Omega}_{zy}\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{yz}$ *in the metric of* $\hat{\Omega}_{zz}$.

The theorem is easily proved either by using a result reported in Rao (1965) for finding the best approximation to a matrix by one of lower rank, or by using Ky Fan's inequality (1.2.1), or by reparameterizing and differentiating, using the constraints

$$\mu'_j\hat{\Omega}_{xx}^{-1}\mu_k = \nu'_j\hat{\Omega}_{zz}\nu_k = \delta_{jk} , \quad j, k = 1, \ldots, r . \tag{1.6}$$

In the sense of §2.5 our estimation procedure is efficient;[34] moreover it is equivalent to CCA. For if we reparameterize B as $\sum_{j=1}^{r} \rho_j \mu_j \nu_j'$, take $\mu_j' \Omega_{XX}^{-1} \mu_k = \delta_{jk}$ and pre-multiply (1.3) by $\mu_{0j}' \Omega_{XX}^{-1}$ we get

$$\mu_{0j}' \Omega_{XX}^{-1} y(n) = \rho_{0j} \nu_{0j}' z(n) + u_j(n) \ , \quad j = 1, \ldots, r \ ,$$

where the $u_j(n)$ are i.i.d. (0, 1) . Likewise, replacing B_0 by B_b in (1.3) and pre-multiplying by $\mu_{bj}' \hat{\Omega}_{XX}^{-1}$, we have the estimated equations

$$\mu_{bj}' \hat{\Omega}_{XX}^{-1} y(n) = \rho_{bj} \nu_{bj}' z(n) \ , \quad j = 1, \ldots, r \ , \tag{1.7}$$

because of (1.6). However, by comparing (1.1) and (1.5) we find that $\rho_{bj} = \rho_{aj} \left(1-\rho_{aj}^2\right)^{-\frac{1}{2}}$. Also we have the relations $\mu_{aj}' \hat{\Omega}_{XX} \mu_{aj} = 1 - \rho_{aj}^2$ (from (1.1), (1.2)) and $\mu_{bj}' \hat{\Omega}_{XX}^{-1} \hat{\Omega}_{XX} \hat{\Omega}_{XX}^{-1} \mu_{bj} = 1$, so $\mu_{bj} = \left(1-\rho_{aj}^2\right)^{-\frac{1}{2}} \hat{\Omega}_{XX} \mu_{aj}$. Finally $\nu_{bj} = \nu_{aj}$ and substituting in (1.7) we have the CCA relations

$$\mu_{aj}' y(n) = \rho_{aj} \nu_{aj}' z(n) \ , \quad j = 1, \ldots, r \ .$$

3.2 Identification, estimation and asymptotic theory

We turn now to a more general class of estimates. We consider

$$y(t) = U_0 R_0 V_0' z(t) + x(t) \ , \quad t \in R \ . \tag{2.1}$$

Here, $U = (\mu_1 \vdots \ldots \vdots \mu_r)$, $V = (\nu_1 \vdots \ldots \vdots \nu_r)$ and R is the $r \times r$ diagonal matrix with jth diagonal element ρ_j ; thus

[34] Although $\hat{\Omega}_{XX}$, the second moment matrix of *LS* residuals, is not the *ML* estimate of Ω_{XX} when B_0 is constrained to have less than full rank, under our conditions it is consistent and B_b will be close to, and have the same limiting properties as, the *ML* estimate.

$URV' = \sum_{j=1}^{r} \rho_j \mu_j \nu_j'$. We impose the orthonormality conditions

$$U'\Pi_\mu U = V'\Pi_\nu V = I_r \ , \tag{2.2}$$

Π_μ, Π_ν being arbitrary symmetric p.d. matrices. An alternative way of seeking a B of rank r would be to require all square submatrices of B of order greater than r to have zero determinants. Such constraints would seem in general to be of a very complicated nature, however, and we would not be able to form a set of r simultaneous equations as we may from (2.1) and (2.2)[35].

In terms of (2.1.1) we get (2.1) if we take $B = \rho \otimes I_{p_1}$, $\Gamma(t;\ \theta)' = (\nu_1\mu_1' \vdots \ldots \vdots \nu_r\mu_r')\delta(t)$, with $\rho = (\rho_1 \vdots \ldots \vdots \rho_r)$ and $\delta(t)$ given by (2.1.3). In §2 we allowed B to satisfy linear constraints of the type implied here. Moreover (2.2) constrains the μ_j, ν_j to a compact subset of a differentiable manifold because, for example, μ_j is a point on the manifold generated by the intersection in $p_1 r$-dimensional Euclidean space of the (p_1-1)-dimensional ellipsoid $\mu_j'\Pi_\mu\mu_j = 1$ and the bilinear functions $\mu_j'\Pi_\mu\mu_k = 0$, $k \neq j$.

Because (2.1) is unlagged we may estimate the exact discrete Fourier transform and no approximation is involved. Correspondingly the identification problem is simplified because (iv) is not needed, since

$$f_{yz}^d(\lambda) = U_0 R_0 V_0' f_{zz}^d(\lambda) \ , \quad \lambda \in B \ .$$

However the parameterization in (2.1) poses problems, requiring somewhat

---

[35] We may form $U_0'\Pi_\mu y(t) \doteq R_0 V_0' z(t) + u(t)$. In this respect our work has some connection with that elsewhere in the thesis concerned with the estimation of linear functional equations. However our system here is singular when $r < p_1$ and the identification conditions differ.

arbitrary prescriptions to distinguish the r individual components $\rho_{0j}\mu_{0j}\nu'_{0j}$ and then to identify ρ_{0j}, μ_{0j} and ν_{0j} separately. Using (1.2.4) we may write the present version of the left side of (2.2.7) as

$$\mathrm{vec}'\{U_0R_0V_0'-URV'\}\Psi(\Phi)\mathrm{vec}\{U_0R_0V_0'-URV'\} , \tag{2.3}$$

where

$$\Psi(\Phi) = (2\pi)^{-1}\int_B f_{zz}(-\lambda) \otimes \Phi(\lambda)d\lambda . \tag{2.4}$$

Now (2.3) will be positive when $\Psi(\Phi) > 0$ and $URV' \neq U_0R_0V_0$ uniformly in admissible $(U' \vdots V' \vdots R) \neq (U_0' \vdots V_0' \vdots R_0)$. The last condition will be achieved if we impose (2.2) on U_0 and V_0 (which will imply that $\mu_{0j} \neq \mu_{0k}$, $\nu_{0j} \neq \nu_{0k}$, $j \neq k$), a sign constraint on one element of each row of $(U_0' \vdots V_0')$ and an ordering such as

$$\rho_{01} > \rho_{02} > \ldots > \rho_{0r} > 0 .$$

In fact *local* identification follows from the constraints (2.2), the fact that the ρ_{0j} are distinct and non-zero and the condition on $\Psi(\Phi)$; *global* identification is completed by the ordering of the ρ_{0j} (enabling the choice of one out of $r!$ permutations of the $\rho_{0j}\mu_{0j}\nu'_{0j}$) and the sign constraints.

In deriving the covariance matrix for the limiting distribution of our estimates we must include the components corresponding to the constraints (2.2). For an $m \times n$ matrix A we have

$$\partial\mathrm{vec}\{A\}/\partial\mathrm{vec}'\{A'\} = \tilde{I}_n^m , \tag{2.5}$$

where $\tilde{I}_n^m$ is $mn \times mn$ and null except for the $[(k-1)m+j, (j-1)n+k]$th elements, $j = 1, \ldots, m$, $k = 1, \ldots, n$, all of which are unity. Now from (1.2.3):

$$\text{vec}\{U'\Pi_\mu U\} = \left(U'\Pi_\mu \otimes I_r\right)\text{vec}\{U'\} = \left(I_r \otimes U'\Pi_\mu\right)\text{vec}\{U\} ,$$

$$\text{vec}\{V'\Pi_\nu V\} = \left(V'\Pi_\nu \otimes I_r\right)\text{vec}\{V'\} = \left(I_r \otimes V'\Pi_\nu\right)\text{vec}\{V\} .$$

Then from (2.4) it follows that

$$\partial\text{vec}\{U'\Pi_\mu U\}/\partial\text{vec}'\{U\} = \left(U'\Pi_\mu \otimes I_r\right)\tilde{I}_r^{p_1} + \left(I_r \otimes U'\Pi_\mu\right) = A_1 ,$$

$$\partial\text{vec}\{V'\Pi_\nu V\}/\partial\text{vec}'\{V'\} = \left(V'\Pi_\nu \otimes I_r\right) + \left(I_r \otimes V'\Pi_\nu\right)\tilde{I}_{p_2}^{r} = A_2 ,$$

say[36]. We may now compose the matrix

$$\frac{\partial\text{vec}\{U'\Pi_\mu U - I_r \vdots V'\Pi_\nu V - I_r\}}{\partial(\text{vec}'\{U\} \vdots \text{vec}'\{V'\} \vdots \rho)} = \left[\begin{array}{c:c:c} A_1 & 0 & 0 \\ \hdashline 0 & A_2 & 0 \end{array}\right] .$$

From this we form the $r(r+1) \times r\left(p_1+p_2+1\right)$ matrix 2Ξ by deleting one member of each of the $r(r-1)$ pairs of identical rows, corresponding to the duplication of the off-diagonal constraints in (2.2).

THEOREM 3.2. *Denote by* $\left(\hat{U}_N' \vdots \hat{V}_N' \vdots \hat{R}_N\right)$ *the matrix minimizing*

$$Q_N = N^{-1} \sum_{B} \|\Phi\left(\lambda_s\right)^{\frac{1}{2}}\{w_y(s) - URV'w_z(s)\}\|^2 \qquad (2.6)$$

over admissible $(U' \vdots V' \vdots R)$ *such that*

$$\rho_1 > \rho_2 > \ldots > \rho_r > 0$$

and $(U' \vdots V') \in \Theta$ *, where*

$$\Theta \subset \left\{(U' \vdots V') \mid U'\Pi_\mu U = I_r;\ V'\Pi_\nu V = I_r;\ \mu_{j_1} > 0,\ j = 1, \ldots, r\right\} ,$$

μ_{j_1} *being the first element of* μ_j *. Let* $\left(U_0' \vdots V_0' \vdots R_0\right)$ *satisfy these conditions, let* $\Psi(\Phi)$ *be non-singular and let* (2.1) *and conditions* (i) *and* (ii) *hold. Then*

$$\lim_{N\to\infty} \left(\hat{U}_N' \vdots \hat{V}_N' \vdots \hat{R}_N\right) = \left(U_0' \vdots V_0' \vdots R_0\right) \text{, a.s.,}$$

---

[36] The reason we use the inconsistent orderings vec{U} , vec{V'} is to enable us to present the covariance matrix in Theorem 3.2 below in a relatively concise form.

$$\lim_{N\to\infty} \hat{B}_N = \lim_{N\to\infty} \hat{U}_N \hat{R}_N \hat{V}'_N = B_0 \text{ , a.s.}$$

Moreover if $\left(U'_0 \vdots V'_0\right)$ *is an interior point of* Θ *and condition* (vi) *holds, then as* $N \to \infty$

$$N^{\frac{1}{2}}\left(\text{vec}'\{\hat{U}_N - U_0\} \vdots \text{vec}'\{\hat{V}'_N - V'_0\} \vdots \hat{\rho}_N - \rho_0\right) \xrightarrow{\mathcal{D}} N[0, \Sigma] ,$$

where Σ *is the* $r(p_1+p_2+1)$*-order leading square submatrix of the* $r(p_1+p_2+r+2)$*-order matrix*

$$\begin{bmatrix} T'_0\Psi(\Phi)T_0 & \vdots & \Xi'_0 \\ \cdots\cdots & \vdots & \cdots \\ \Xi_0 & \vdots & 0 \end{bmatrix}^{-1} \begin{bmatrix} T'_0\Psi(\Phi f_{xx}\Phi)T_0 & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{bmatrix} T'_0\Psi(\Phi)T_0 & \vdots & \Xi'_0 \\ \cdots\cdots & \vdots & \cdots \\ \Xi_0 & \vdots & 0 \end{bmatrix}^{-1} , \quad (2.7)$$

with

$$T = \left(VR \otimes I_{p_1} \vdots I_{p_2} \otimes UR \vdots V \odot U\right) .$$

Proof. Substantially the proof follows from that of Theorems 2.3, 2.6; a few comments are relevant however. The sign constraints[37] on the μ_{j_1} destroy compactness so, formally, a compact subset must be chosen. (For example we might take $\mu_{j_1} \geq \Delta$ for some $\Delta > 0$, $j = 1, \ldots, r$.) The matrix T_0 has column rank $r\left(p_1+p_2-r\right) < r\left(p_1+p_2+1\right)$ because of the existence of $r(r+1)$ linearly independent relations between the columns. Then because $\left(p_1-r\right)\left(p_2-r\right) \geq 0$ it follows that $r\left(p_1+p_2-r\right) \leq p_1p_2 = \text{rank}\{\Psi(\Phi)\}$, and thus $T'_0\Psi(\Phi)T_0$ is singular of rank $r\left(p_1+p_2-r\right)$. However the true inverses in (2.7) exist since the augmented matrix has full rank, $T'_0\Psi(\Phi)T_0$ and Ξ_0 having non-intersecting null spaces. The form of

---

[37] Of course the sign constraint on μ_{j_1} could be transferred to any other element of $\left(\mu'_j \vdots \nu'_j\right)$.

T follows from the derivatives[38]

$$\partial \mathrm{vec}\{B\}/\partial \mathrm{vec}'\{U\} = \left(VR \otimes I_{p_1}\right)\partial \mathrm{vec}\{U\}/\partial \mathrm{vec}'\{U\} = VR \otimes I_{p_1} \ ,$$

$$\partial \mathrm{vec}\{B\}/\partial \mathrm{vec}'\{V'\} = \left(I_{p_2} \otimes UR\right)\partial \mathrm{vec}\{V'\}/\partial \mathrm{vec}'\{V'\} = I_{p_2} \otimes UR \ ,$$

$$\partial \mathrm{vec}\{B\}/\partial \rho = (\partial/\partial \rho) \sum_{j=1}^{r} \rho_j \nu_j \otimes \mu_j = V \odot U \ .$$

In Robinson (1973) a similar theorem is given, omitting the assumption that $z(n)$ has absolutely continuous spectrum (thus allowing for almost-periodic components), and with $x(n)$ satisfying (1.3.7).

The matrices and scalars

$$(2\pi)^{-1}\ \rho_{0j}^2 \int_B \nu_{0j}' f_{zz}(\lambda)\nu_{0j}\Phi(\lambda)d\lambda \ , \quad (2\pi)^{-1}\ \rho_{0j}^2 \int_B f_{zz}(\lambda)\mu_{0j}'\Phi(\lambda)\mu_{0j}d\lambda \ ,$$

$$(2\pi)^{-1} \int_B \nu_{0j}' f_{zz}(\lambda)\nu_{0j}\mu_{0j}'\Phi(\lambda)\mu_{0j}d\lambda \ ,$$

$j = 1, \ldots, r$, are non-singular and non-zero respectively under the identification conditions. Because they lie along the main diagonal of $T_0'\Psi(\Phi)T_0$ we could partition this matrix and base an evaluation of (2.7) on their inverses. In general explicit expressions for variances and covariances will be very complicated but it is feasible to derive them for the not unimportant case $r = 1$. We consider the efficient situation with $\Phi = f_{xx}^{-1}$. Then

$$T_0'\Psi\left\{f_{xx}^{-1}\right\}T_0 = \left[\begin{array}{c:c:c} \Psi_1 & \Psi_3 & \psi_1 \\ \hdashline & \Psi_2 & \psi_2 \\ \hdashline & & \psi \end{array}\right] \ , \qquad (2.8)$$

[38] Alternatively, following Theorem 2.6, the covariance matrix may be expressed in terms of pseudoinverses, with

$$\beta = \rho \otimes \mathrm{vec}'\left\{I_{p_1}\right\} \ , \quad L = p_1^{-1} I_r \otimes \mathrm{vec}\left\{I_{p_1}\right\}\mathrm{vec}'\left\{I_{p_1}\right\} \ ,$$

and then rows and columns that are null or repetitions may be deleted.

say, where the partition of rows and columns is $p_1 : p_2 : 1$. Now (2.8) has rank $p_1 + p_2 - 1$ because of the relations

$$\rho_{01}^{-1}\mu_{01}'(\Psi_1 \vdots \Psi_3 \vdots \psi_1) = \rho_{01}^{-1}\nu_{01}'(\Psi_3' \vdots \Psi_2 \vdots \psi_2) = (\psi_1' \vdots \psi_2' \vdots \psi) .$$

However we may invert the augmented matrix, which includes

$$\Xi_0 = \begin{bmatrix} \mu_{01}'\Pi_\mu & 0 & 0 \\ 0 & \nu_{01}'\Pi_\nu & 0 \end{bmatrix},$$

because of the non-singularity of Ψ_1, Ψ_2, ψ.

COROLLARY 3.2. *With* $r = 1$, $\Phi = f_{xx}^{-1}$ *in* (2.1), (2.6) *and* $\Pi_\mu = I_{p_1}$, $\Pi_\nu = I_{p_2}$ *the covariance matrix in the limiting distribution of* $N^{\frac{1}{2}}(\hat{\mu}_{N1}' - \mu_{01}' \vdots \hat{\nu}_{N1}' - \nu_{01}' \vdots \hat{\rho}_{N1} - \rho_{01})$ *is*

$$\begin{bmatrix} \Sigma_1 & \Sigma_3 & \sigma_1 \\ & \Sigma_2 & \sigma_2 \\ & & \sigma \end{bmatrix}, \tag{2.9}$$

where

$$\Sigma_1 = \left(\Psi_1 - \Psi_4\Psi_3\Psi_2^{-1}\Psi_3'\right)^{-1}\Psi_4;\ \Sigma_2 = \left(\Psi_2 - \Psi_5\Psi_3'\Psi_1^{-1}\Psi_3\right)^{-1}\Psi_5$$

$$\Sigma_3 = -\Sigma_1\Psi_3\Psi_2^{-1}\left(I_{p_2} - \nu_{01}\nu_{01}'\right);\quad \sigma_1 = -\rho_{01}\Sigma_1\Psi_3\Psi_2^{-1}\nu_{01};$$

$$\sigma_2 = -\rho_{01}\Sigma_2\Psi_3'\Psi_1^{-1}\mu_{01};\quad \sigma = \psi^{-1}\left(1 - \psi_1'\sigma_1 - \psi_2'\sigma_2\right);$$

in which

$$\Psi_4 = I_{p_1} - \mu_{01}\left(\mu_{01}'\Psi_1^{-1}\mu_{01}\right)^{-1}\mu_{01}'\Psi_1^{-1};\ \Psi_5 = I_{p_2} - \nu_{01}\left(\nu_{01}'\Psi_2^{-1}\nu_{01}\right)^{-1}\nu_{01}'\Psi_2^{-1}.$$

We shall not give the proof, which is of an excessively tedious nature. It is not difficult to verify the result, however. The Lagrange multipliers corresponding to the constraints

$\mu_1'\mu_1 = \nu_1'\nu_1 = 1$, namely $\hat{\gamma}_{N1}, \hat{\gamma}_{N2}$, may easily be found to be identically zero, and thus $N^{\frac{1}{2}}\hat{\gamma}_{N1}, N^{\frac{1}{2}}\hat{\gamma}_{N2}$ each have the degenerate distribution $\delta(t)$. The covariance matrix of $N^{\frac{1}{2}}(\hat{\mu}_{N1}'-\mu_{01}' \vdots \hat{\nu}_{N1}'-\nu_{01}' \vdots \hat{\rho}_{N1}-\rho_{01} \vdots \hat{\gamma}_{N1} \vdots \hat{\gamma}_{N2})$ thus has the form

$$\begin{bmatrix} \Sigma_4 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} .$$

That Σ_4 is given by (2.9) may then be verified from the relation

$$T_0'\Psi\left\{f_{xx}^{-1}\right\}T_0\Sigma_4T_0'\Psi\left\{f_{xx}^{-1}\right\}T_0 = T_0'\Psi\left\{f_{xx}^{-1}\right\}T_0 .$$

We note that $\operatorname{rank}\{\Sigma_1\} = p_1 - 1$, $\operatorname{rank}\{\Sigma_2\} = p_2 - 1$.

When Φ varies over $\mathcal{B}$ the efficient estimation of parameters requires iterative procedures and is likely to present a difficult and lengthy task. (Correspondingly no simple and satisfying geometrical interpretation seems possible.) We have relegated to an appendix our discussion of the computation of efficient estimates in the presence of serial correlation, since our suggestions are not illustrated by a numerical example. When Φ is constant, however, the estimation may reduce to the solution of an eigenvalue problem, and it is of some interest to consider the asymptotic properties of some such estimates, namely those of Theorem 3.1. *In all that follows in the present section we replace* $\mathcal{B}$ *by* $(-\pi, \pi]$.

THEOREM 3.3. *Denote the estimates of Theorem* 3.1 U_b, V_b, R_b , *with the usual ordering and the direction of* μ_{bj} *being determined by taking its first element to be positive,* $j = 1, \ldots, r$. *Let* $\rho_{01} > \rho_{02} > \ldots > \rho_{0r} > 0$ *and* $(U_0' \vdots V_0') \in \Theta$, *where* Θ *is as in Theorem* 3.2 *with* $\Pi_\mu = \Omega_{xx}^{-1}$ *and* $\Pi_\nu = E\{z(t)z(t)'\} = \Omega_{zz}$, Ω_{xx} *and*

Ω_{ZZ} *being non-singular. Let* (2.1) *and conditions* (i), (ii) *and* (vi) *hold. Then the asymptotic properties of* $(U_b' \vdots V_b' \vdots R_b)$ *are the same as those of* $(\hat{U}_N' \vdots \hat{V}_N' \vdots \hat{R}_N)$ *, except that in* (2.7) $\Phi \equiv \Omega_{XX}^{-1}$ *, and then*

$$\Psi\left(\Omega_{XX}^{-1}\right) = \Omega_{ZZ} \otimes \Omega_{XX}^{-1} \; , \quad \Psi\left(\Omega_{XX}^{-1} f_{XX} \Omega_{XX}^{-1}\right) = \left(I_{p_2} \otimes \Omega_{XX}^{-1}\right) \Psi\left(f_{XX}\right) \left(I_{p_2} \otimes \Omega_{XX}^{-1}\right) .$$

Proof. The theorem would follow immediately from Theorem 3.2 for estimates minimizing

$$Q_N = N^{-1} \mathrm{tr}\left\{(Y-BZ)(Y-BZ)'\Omega_{XX}^{-1}\right\} \tag{2.10}$$

subject to

$$U'\Omega_{XX}^{-1}U = V'\Omega_{ZZ}V = I_r \; , \tag{2.11}$$

for (2.10) is just (2.6) with $\Phi \equiv \Omega_{XX}^{-1}$. In the theorem, however, we are minimizing (1.4) subject to (1.6), the constraints varying with N . Nevertheless the region defined by (1.6) "converges to" that defined by (2.11) in the sense that each point of (1.6) is eventually arbitrarily close to a point of (2.11) since a.s. $\hat{\Omega}_{ZZ} \to \Omega_{ZZ}$ and

$$\lim_{N\to\infty} \hat{\Omega}_{XX} = \lim_{N\to\infty} N^{-1}\left(B_0Z+X\right)\left(I_N - Z'(ZZ')^{-1}Z\right)\left(B_0Z+X\right)' = \Omega_{XX} \; . \tag{2.12}$$

Then the strong law hold for estimates minimizing (2.10) subject to (1.6) because these estimates converge to values satisfying (2.11), and (2.10) is continuous in B and (v) is implied by our conditions. However the modulus of the difference between (1.4) and (2.10) is bounded by

$$\left\|\hat{\Omega}_{XX}^{-1} - \Omega_{XX}^{-1}\right\| \mathrm{tr}\{\hat{\Omega}_{yy} - B\hat{\Omega}_{zy} - \Omega_{yz}B' + B\hat{\Omega}_{ZZ}B'\} \; . \tag{2.13}$$

The first factor of (2.13) is not greater than

$$\left\|\Omega_{XX}^{-1}\right\| \left\|\Omega_{XX} - \hat{\Omega}_{XX}\right\| \left\|\hat{\Omega}_{XX}^{-1}\right\| < \varepsilon \; ,$$

a.s., N sufficiently large, for arbitrary $\varepsilon > 0$ because of (2.12).

Since the second factor in (2.13) is $O(1)$ the strong law holds for $\left(\mathrm{U}_b' \vdots \mathrm{V}_b' \vdots \mathrm{R}_b\right)$. For the central limit theorem we consider

$$Q_N'' = Q_N' + \mathrm{tr}\left\{\left(\mathrm{U}'\Omega_{\mathrm{XX}}^{-1}\mathrm{U}-\mathrm{I}_r\right)\hat{\Gamma}_\mu + \left(\mathrm{V}'\Omega_{\mathrm{ZZ}}\mathrm{V}-\mathrm{I}_r\right)\hat{\Gamma}_\nu\right\},$$

where $\hat{\Gamma}_\mu$, $\hat{\Gamma}_\nu$ are $r \times r$ symmetric matrices of Lagrange coefficients corresponding to the constraints (2.11). Now

$$N^{\frac{1}{2}}\partial Q_N''/\partial \mathrm{U}\Big|_0 = -2N^{-\frac{1}{2}}\Omega_{\mathrm{XX}}^{-1}\mathrm{XZ}'\mathrm{V}_0\mathrm{R}_0 + 2\Omega_{\mathrm{XX}}^{-1}\mathrm{U}_0 N^{\frac{1}{2}}\hat{\Gamma}_\mu, \qquad (2.14)$$

where we evaluate the derivative at $\left(\mathrm{U}_0' \vdots \mathrm{V}_0' \vdots \mathrm{R}_0\right)$. From Theorem 2.6 we note that the central limit theorem for estimates minimizing Q_N'' stems from the asymptotic normality of (2.14). However

$$\underset{N\to\infty}{\mathrm{l.i.m.}} \left(\hat{\Omega}_{\mathrm{XX}}^{-1}-\Omega_{\mathrm{XX}}^{-1}\right)N^{-\frac{1}{2}}\mathrm{XZ}'\mathrm{V}_0\mathrm{R}_0 = 0$$

because the first factor $\to 0$ a.s. and the expectation of the squared norm of the second is

$$N^{-1}\mathrm{Etr}\left\{\mathrm{R}_0\mathrm{V}_0' \sum_{m,n=1}^{N}\sum \mathrm{z}(m)\mathrm{x}(m)'\mathrm{x}(n)\mathrm{z}(n)'\mathrm{V}_0\mathrm{R}_0\right\}$$

$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathrm{f}_{\mathrm{XX}}(\mu)\}\mathrm{tr}\{\mathrm{R}_0\mathrm{V}_0'\mathrm{f}_{\mathrm{ZZ}}(\lambda)\mathrm{V}_0\mathrm{R}_0\}F_N(\mu-\lambda)d\mu d\lambda$$

$$< c\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F_N(\mu-\lambda)d\mu d\lambda = 4\pi^2 c = O(1),$$

using (ii). Now from Theorem 2.6, $N^{\frac{1}{2}}\hat{\Gamma}_\mu$ converges in distribution so

$$\underset{N\to\infty}{\mathrm{l.i.m.}} \left(\hat{\Omega}_{\mathrm{XX}}^{-1}-\Omega_{\mathrm{XX}}^{-1}\right)\mathrm{U}_0 N^{\frac{1}{2}}\hat{\Gamma}_\mu = 0 .$$

Thus (2.14) with Ω_{XX} replaced by $\hat{\Omega}_{\mathrm{XX}}$ is asymptotically normal also. In an entirely similar way the remaining derivatives of Q_N'' may be modified to complete the proof of the theorem.

The following theorem deals with the situation when $\left(\mathrm{U}_b' \vdots \mathrm{V}_b' \vdots \mathrm{R}_b\right)$ is efficient.

THEOREM 3.4. *Let the conditions of the previous theorem hold and moreover let* $x(n)$ *be i.i.d.* $(0, \Omega_{xx})$ *. Then as* $N \to \infty$ *the variates* $N^{\frac{1}{2}}(\rho_{bj}-\rho_{0j})$ *,* $j = 1, \ldots, r$*, have independent unit normal distributions and they are also independent of the vector variates* $N^{\frac{1}{2}}(\mu_{bj}-\mu_{0j})$ *,* $N^{\frac{1}{2}}(\nu_{bj}-\nu_{0j})$ *,* $j = 1, \ldots, r$ *, which have a limiting joint multivariate normal distribution with null mean vector and the following variances and covariances.*

$$\left.\begin{aligned}
\lim_{N\to\infty} E\{N(\mu_{bj}-\mu_{0j})(\mu_{bk}-\mu_{0k})'\} &= \sum_{l}{}' \frac{\left(\rho_{0j}^2+\rho_{0l}^2\right)}{\left(\rho_{0j}^2-\rho_{0l}^2\right)^2} \mu_{0l}\mu_{0l}' \;, \quad j = k \;, \\
&= -\frac{\left(\rho_{0j}^2+\rho_{0k}^2\right)}{\left(\rho_{0j}^2-\rho_{0k}^2\right)^2} \mu_{0k}\mu_{0j}' \;, \quad j \neq k \;; \\
\lim_{N\to\infty} E\{N(\nu_{bj}-\nu_{0j})(\nu_{bk}-\nu_{0k})'\} &= \sum_{l}{}' \frac{\left(\rho_{0j}^2+\rho_{0l}^2\right)}{\left(\rho_{0j}^2-\rho_{0l}^2\right)^2} \nu_{0l}\nu_{0l}' \;, \quad j = k \;, \\
&= -\frac{\left(\rho_{0j}^2+\rho_{0k}^2\right)}{\left(\rho_{0j}^2-\rho_{0k}^2\right)^2} \nu_{0k}\nu_{0j}' \;, \quad j \neq k \;; \\
\lim_{N\to\infty} E\{N(\mu_{bj}-\mu_{0j})(\nu_{bk}-\nu_{0k})'\} &= 2\sum_{l}{}' \frac{\rho_{0j}\rho_{0l}}{\left(\rho_{0j}^2-\rho_{0l}^2\right)^2} \mu_{0l}\nu_{0l}' \;, \quad j = k \;, \\
&= -2 \frac{\rho_{0j}\rho_{0k}}{\left(\rho_{0j}^2-\rho_{0k}^2\right)^2} \mu_{0k}\nu_{0j}' \;, \quad j \neq k \;;
\end{aligned}\right\} \quad (2.15)$$

where the sums are over l *such that* $1 \le l \le \min(p_1, p_2)$ *,* $l \neq j$ *and we put* $\rho_{0l} = 0$ *,* $l > r$ *. Consistent estimates of the variances and covariances are obtained by replacing* $\mu_{0j}, \nu_{0j}, \rho_{0j}$ *, by* $\mu_{bj}, \nu_{bj}, \rho_{bj}$ *respectively,* $j = 1, \ldots, r$ *.*

Proof. The expressions (2.11) may be more easily derived by means of perturbation expansions (see Wilkinson, 1965, Chapter 2) for eigenvalues and eigenvectors than by directly evaluating

$$\left[\begin{array}{c:c} T_0'\Psi\left(\Omega_{XX}^{-1}\right)T_0 & \Xi_0' \\ \hdashline \Xi_0 & 0 \end{array}\right]^{-1} \left[\begin{array}{c:c} T_0'\Psi\left(\Omega_{XX}^{-1}\right)T_0 & 0 \\ \hdashline 0 & 0 \end{array}\right] \left[\begin{array}{c:c} T_0'\Psi\left(\Omega_{XX}^{-1}\right)T_0 & \Xi_0' \\ \hdashline \Xi_0 & 0 \end{array}\right]^{-1} .$$

We expand the eigenvectors μ_{bj} of

$$\hat{\Sigma}_N = \hat{\Omega}_{XX}^{-1}\hat{\Omega}_{yz}\hat{\Omega}_{zz}^{-1}\hat{\Omega}_{zy}\hat{\Omega}_{XX}^{-1}$$

about their limits, the eigenvectors μ_{0j} of

$$\Sigma_0 = \Omega_{XX}^{-1}B_0\Omega_{ZZ}B_0'\Omega_{XX}^{-1} ,$$

giving

$$\mu_{bj} \doteqdot \mu_{0j} + \sum_l{}' \left(\rho_{0j}^2-\rho_{0l}^2\right)^{-1}\mu_{0l}\mu_{0l}'\left(\hat{\Sigma}_N-\Sigma_0\right)\mu_{0j}$$

$$\doteqdot \mu_{0j} + \sum_l{}' \left(\rho_{0j}^2-\rho_{0l}^2\right)^{-1}\mu_{0l}\mu_{0l}'\Omega_{XX}^{-1}\left(B_0\hat{\Omega}_{ZX}+\hat{\Omega}_{XZ}B_0'\right)\Omega_{XX}^{-1}\mu_{0j}$$

$$= \mu_{0j} + \sum_l{}' \left(\rho_{0j}^2-\rho_{0l}^2\right)^{-1}\mu_{0l}\left(\rho_{0l}\nu_{0l}'\hat{\Omega}_{ZX}\Omega_{XX}^{-1}\mu_{0j}+\rho_{0j}\mu_{0l}'\Omega_{XX}^{-1}\hat{\Omega}_{XZ}\nu_{0j}\right) ,$$

where $\hat{\Omega}_{ZX} = N^{-1}ZX'$. Thus $N^{\frac{1}{2}}\left(\mu_{bj}-\mu_{0j}\right)$ has the same limiting distribution as

$$\sum_l{}' \left(\rho_{0j}^2-\rho_{0l}^2\right)^{-1}\mu_{0l}\left[\left(\rho_{0l}\nu_{0l}' \otimes \mu_{0j}'\right)+\left(\rho_{0j}\nu_{0j}' \otimes \mu_{0l}'\right)\right]v , \tag{2.16}$$

where

$$v = N^{-\frac{1}{2}}\mathrm{vec}\left\{\Omega_{XX}^{-1}\hat{\Omega}_{XZ}\right\} \xrightarrow{\mathcal{D}} N\left[0,\ \Psi\left(\Omega_{XX}^{-1}\right)\right] . \tag{2.17}$$

Similarly $N^{\frac{1}{2}}\left(\nu_{bj}-\nu_{0j}\right)$ is equivalent to

$$\sum_l{}' \left(\rho_{0j}^2-\rho_{0l}^2\right)^{-1}\nu_{0l}\left[\left(\rho_{0l}\nu_{0j}' \otimes \mu_{0l}'\right)+\left(\rho_{0j}\nu_{0l}' \otimes \mu_{0j}'\right)\right]v . \tag{2.18}$$

Finally

$$\rho_{bj}^2 \doteq \rho_{0j}^2 + \mu_{0j}'\Omega_{XX}^{-1}\left(\hat{\Omega}_{XZ}B_0'+B_0\hat{\Omega}_{ZX}\right)\Omega_{XX}^{-1}\mu_{0j} = \rho_{0j}^2\left[1+2\rho_{0j}^{-1}\mu_{0j}'\Omega_{XX}^{-1}\hat{\Omega}_{XZ}\nu_{0j}\right] .$$

Then from the square root expansion,

$$N^{\frac{1}{2}}\left(\rho_{bj}-\rho_{0j}\right) \doteq \left(\nu_{0j}' \otimes \mu_{0j}'\right)\mathrm{v} .$$

Thus, heuristically, (2.15) may be constructed from (2.16)-(2.18). We may verify the result by forming the $r(p_1+p_2+1)$-order square matrix Σ_5 from (2.15), appropriately ordered, and showing that

$$T_0'\left[\Omega_{ZZ} \otimes \Omega_{XX}^{-1}\right]T_0\Sigma_5T_0'\left[\Omega_{ZZ} \otimes \Omega_{XX}^{-1}\right]T_0 = T_0'\left[\Omega_{ZZ} \otimes \Omega_{XX}^{-1}\right]T_0 ,$$

since the Lagrange multipliers are identically zero.

COROLLARY 3.4. *If* $r = 1$ *in* (2.1) *and* (1.4) *then* $N^{\frac{1}{2}}\left(\mu_{b1}-\mu_{01}\right)$, $N^{\frac{1}{2}}\left(\nu_{b1}-\nu_{01}\right)$, $N^{\frac{1}{2}}\left(\rho_{b1}-\rho_{01}\right)$ *are asymptotically independent with covariance matrices and variance respectively*

$$\rho_{01}^{-2}\left(\Omega_{XX}-\mu_{01}\mu_{01}'\right) , \quad \rho_{01}^{-2}\left[\Omega_{ZZ}^{-1}-\nu_{01}\nu_{01}'\right] , \quad 1 .$$

3.3 A test for rank

The distribution theory we have given may be used to test hypotheses about the μ_{0j}, ν_{0j}, ρ_{0j} : for example we could test whether certain components of $y(n)$, $z(n)$ may be omitted from any of the simultaneous equations. Of course the forming of $\hat{B}_N$ submerges the identities of the individual $\hat{\mu}_{Nj}$, $\hat{\nu}_{Nj}$, $\hat{\rho}_{Nj}$, and these cease to be of more than intrinsic interest. The question of most importance concerns the rank of B_0, that is the number of simultaneous equations needed to adequately describe the data. Unfortunately we cannot fit a matrix of suitably high rank r and then use our distribution theory to test $\text{rank}\{B_0\} = s < r$ versus $\text{rank}\{B_0\} = r$ by putting $\rho_{0j} = 0$, $s + 1 \leq j \leq r$ or by setting up $s - r$ equalities between the $(\mu'_{0j} \vdots \nu'_{0j})$, for to do so would contravene the identification conditions. If we are to attempt to find efficient estimates by means of iteration a test for rank is needed that can be carried out before this lengthy procedure. We propose a rough likelihood-ratio test. From our discussions of §2.5 we are led to set up the approximate log-likelihood

$$-(N/4M) \sum_m \log \det\{\mathsf{f}_{\mathsf{xx}}(\omega_m)\} - \tfrac{1}{2}Q_N , \tag{3.1}$$

where in Q_N we replace Φ by $\mathsf{f}_{\mathsf{xx}}^{-1}$, taking $\mathsf{f}_{\mathsf{xx}}(\lambda_s) \equiv \mathsf{f}_{\mathsf{xx}}(\omega_m)$, $\lambda_s \in S_m$. We denote by $\hat{\mathsf{f}}_r(\omega_m)$, $\hat{\mathsf{f}}_f(\omega_m)$ estimates of the form

$$2MN^{-1} \sum_{s(m)} [\mathsf{w}_{\mathsf{y}}(s)-\mathsf{B}_b\mathsf{w}_{\mathsf{z}}(s)][\mathsf{w}_{\mathsf{y}}(s)-\mathsf{B}_b\mathsf{w}_{\mathsf{z}}(s)]^* ,$$

using $\mathsf{B}_b = \mathsf{U}_b\mathsf{R}_b\mathsf{V}'_b$, $\mathsf{B}_b = \mathsf{YZ}'(\mathsf{ZZ}')^{-1}$ respectively. These easily computed B_b are consistent when the true rank is r , $\min(p_1, p_2)$ respectively, and thus so also are $\hat{\mathsf{f}}_r$, $\hat{\mathsf{f}}_f$, under our conditions. Now the estimation of a B_0 of rank r effectively disposes of

$$\sum_{j=1}^{r} (p_1-j+p_2-j+1) = p_1p_2 - (p_1-r)(p_2-r) \tag{3.2}$$

degrees of freedom[39], taking account of (2.2). Then from (3.1) and (3.2) the hypothesis $\operatorname{rank}\{\mathsf{B}_0\} = r$ would be accepted if the log likelihood-ratio

$$2MN^{-1} \left[\log \det\left\{\prod_m \hat{\mathsf{f}}_r(\omega_m)\right\} - \log \det\left\{\prod_m \hat{\mathsf{f}}_f(\omega_m)\right\}\right] \tag{3.3}$$

is less than a suitable percentage point of $\chi^2_{(p_1-r)(p_2-r)}$.

[39] If $r = \min(p_1, p_2)$, (3.2) is

$$p_1p_2 - [p_1-\min(p_1, p_2)][p_2-\min(p_1, p_2)] = p_1p_2 ,$$

as we should expect.

The test reduces under certain circumstances to one that is commonly used for the testing of canonical correlations. If $\mathcal{B}$ is replaced by $(-\pi, \pi]$ and if the $\mathsf{X}(n)$ are believed to be independent we would, for all m, $-M+1 \le m \le M$, replace $\hat{\mathsf{f}}_f(\omega_m)$ by $(2\pi)^{-1}\hat{\Omega}_{\mathsf{XX}}$ and $\hat{\mathsf{f}}_r(\omega_m)$ by

$$(2\pi)^{-1}\hat{\Omega}_r = (2\pi N)^{-1}\left(\mathsf{Y}-\mathsf{U}_b\mathsf{R}_b\mathsf{V}_b'\mathsf{Z}\right)\left(\mathsf{Y}-\mathsf{U}_b\mathsf{R}_b\mathsf{V}_b'\mathsf{Z}\right)'$$

$$= (2\pi)^{-1}\left[\hat{\Omega}_{\mathsf{XX}}+\sum{}' \rho_{bj}^2\mu_{bj}\nu_{bj}'\right] ,$$

the sum being over $r+1 \le j \le \min(p_1,p_2)$. Then (3.3) reduces to

$$N \log \det\left\{\hat{\Omega}_{\mathsf{XX}}^{-1}\hat{\Omega}_r\right\} = N \log \det\left\{I_{p_1}+\hat{\Omega}_{\mathsf{XX}}^{-1} \sum{}' \rho_{bj}^2\mu_{bj}\nu_{bj}'\right\}$$

$$= N \sum{}' \log\left[1+\rho_{bj}^2\right]$$

$$= -N \sum{}' \log\left[1-\rho_{aj}^2\right] .$$

Thus (3.3) generalizes the familiar test for canonical correlations.

3.4 Regression on an unobservable variable

Economic models are sometimes formulated in terms of variables that are themselves unobservable but believed to be related to observable variables. We shall consider the model

$$\mathsf{y}(t) = \beta_0\tilde{z}(t) + \mathsf{x}(t) , \quad t \in R , \tag{4.1}$$

where β_0 is a $p_1 \times 1$ vector and $\tilde{z}(t)$ is an unobservable scalar satisfying

$$\tilde{z}(t) = \alpha_0'\mathsf{z}(t) , \quad t \in R , \tag{4.2}$$

α_0 being a $p_2 \times 1$ vector. It is perhaps more plausible to append a residual to (4.2) but to do so would raise problems of identification. (See Goldberger, 1972.) Our model generalizes one treated by Zellner

(1970), who takes $p_1 = 2$, $\beta_0' = (\beta_0 \vdots 1)$ and gives explicit expressions for the *LS* estimates of β_0 and α_0 . Goldberger (1972) gives an iterative *ML* procedure for estimating (4.1), (4.2), based on normality and independence assumptions. No attention seems to have been paid to distribution theory, under these and more general conditions. We shall give asymptotic theory for a general class of estimates of α_0 and β_0 .

To begin with, on substituting for $\tilde{z}(t)$ in (4.1), the reduced form is

$$y(t) = \beta_0\alpha_0'z(t) + x(t) \ , \quad t \in R \ .$$

For identification purposes we require

$$\alpha_0 \in \Theta = \{\alpha \mid \alpha'\alpha = 1; \ \alpha_1 \geq \Delta > 0\} \ ,$$

α_1 being the first element of α and Δ arbitrarily small. The normalisation and sign constraint on α are thus modified to make Θ compact. Alternative prescriptions are of course possible. We take $\hat{\alpha}_N$, $\hat{\beta}_N$ to be the values minimizing

$$Q_N = N^{-1} \sum_B \left\| \Phi(\lambda_s)^{\frac{1}{2}}\{w_y(s) - \beta\alpha' w_z(s)\} \right\|^2$$

over $\alpha \in \Theta$. Then without bothering to state a formal theorem, under the conditions of Theorem 3.2

$$\lim_{N\to\infty} (\hat{\alpha}_N' \vdots \hat{\beta}_N') = (\alpha_0' \vdots \beta_0') \ , \text{ a.s.},$$

$$N^{\frac{1}{2}}(\hat{\alpha}_N' - \alpha_0' \vdots \hat{\beta}_N' - \beta_0') \xrightarrow{\mathcal{D}} N[0, \Sigma] \ ,$$

where Σ is derived by deleting the final row and column from (2.7) and taking

$$T = \left(\beta \otimes I_{p_1} \vdots I_{p_2} \otimes \alpha\right) \ , \quad \Xi = (\alpha \vdots 0) \ .$$

In the efficient case $\Phi = f_{xx}^{-1}$ we make the partition

$$T_0'\Psi\left(f_{xx}^{-1}\right)T_0 = \begin{bmatrix} \Psi_1 & \vdots & \Psi_3 \\ \cdots & \cdot & \cdots \\ \Psi_3' & \vdots & \Psi_2 \end{bmatrix}$$

and then

$$\lim_{N\to\infty} E\{N(\hat{\alpha}_N-\alpha_0)(\hat{\alpha}_N-\alpha_0)'\} = \left(\Psi_1-\Psi_4\Psi_3\Psi_2^{-1}\Psi_3'\right)^{-1}\Psi_4 \ ,$$

$$\lim_{N\to\infty} E\{N(\hat{\alpha}_N-\alpha_0)(\hat{\beta}_N-\beta_0)'\} = -\Psi_2^{-1}\Psi_3'\left(\Psi_1-\Psi_4\Psi_3\Psi_2^{-1}\Psi_3'\right)^{-1}\Psi_4 \ , \qquad (4.3)$$

$$\lim_{N\to\infty} E\{N(\hat{\beta}_N-\beta_0)(\hat{\beta}_N-\beta_0)'\} = \Psi_2^{-1}\left[I_{p_1}+\Psi_3'\left(\Psi_1-\Psi_4\Psi_3\Psi_2^{-1}\Psi_3'\right)^{-1}\Psi_4\Psi_3'\Psi_2^{-1}\right] \ ,$$

where

$$\Psi_4 = I_{p_2} - \alpha_0\left(\alpha_0'\Psi_1^{-1}\alpha_0\right)^{-1}\alpha_0'\Psi_1^{-1} \ .$$

We shall not discuss efficient estimation in the presence of serial correlation, although possible techniques are implied in §3. Appendix. When Φ is constant, however, the estimates result from the solution of an eigenvalue problem. Taking $\Phi \equiv \hat{\Omega}_{xx}^{-1}$, from Theorem 3.1 we may call $\tilde{\alpha}_N$ the eigenvector of $\hat{\Omega}_{zy}\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{yz}$ in the metric of $\hat{\Omega}_{zz}$ corresponding to the greatest eigenvalue, with $\tilde{\beta}_N = \hat{\Omega}_{yz}\tilde{\alpha}_N$; here we now use the normalisation rule $\alpha'\hat{\Omega}_{zz}\alpha = 1$. When in fact the $x(n)$ are i.i.d. $(0, \Omega_{xx})$ it follows from (4.3) that the vectors $N^{\frac{1}{2}}(\tilde{\alpha}_N-\alpha_0)$, $N^{\frac{1}{2}}(\tilde{\beta}_N-\beta_0)$ are asymptotically independent, with covariance matrices respectively

$$\left(\beta_0'\Omega_{xx}^{-1}\beta_0\right)^{-1}(\Omega_{zz}-\alpha_0\alpha_0') \ , \quad \Omega_{xx} \ .$$

3. Appendix. The computation of efficient estimates

The source of our difficulties in efficiently estimating $\left(U_0' \vdots V_0' \vdots R_0\right)$ lies principally in the simultaneous variation[40] of f_{xx} and f_{zz} over B ; if either of these are taken as constant a neat solution results. This is also the case when $x(n)$ is generated by the mixed autoregressive moving-average

$$\sum_{j=0}^{m} x(n-j)\alpha(j) = \sum_{j=0}^{n} \varepsilon(n-j)\beta(j) ,$$

where $\varepsilon(n)$ is i.i.d. $(0, \Omega)$ *and the* $\alpha(j)$, $\beta(j)$ *are scalars.* (The final condition renders the specification of little practical value.) We have

$$f_{xx}(\lambda) = (2\pi)^{-1} \left| \sum_{j=0}^{n} \beta(j)e^{ij\lambda} \Big/ \sum_{j=0}^{m} \alpha(j)e^{ij\lambda} \right|^2 \Omega .$$

Thus, assuming such a representation, our spectral estimates will be such that $f_{xx}(\lambda_s)^{-1}$ has the form $d(s)D$; replacing Φ in (2.6) we have

$$Q_N = N^{-1}\mathrm{tr}\left\{\left[\sum I_{yy}d - B \sum I_{zy}d - \sum I_{yz}dB' + B \sum I_{zz}dB'\right]D\right\} ,$$

because the scalars d commute with the matrices. Here and elsewhere in the appendix we omit the argument s and we mean the sum to include all $\lambda_s \in B$. It now follows exactly in the manner of Theorem 3.1 that the optimal choice is $B_c = \sum_1^r \rho_{cj}\mu_{cj}\nu_{cj}'$, the ρ_{cj} being the square roots of the greatest eigenvalues of $D \sum I_{yz}d\left(\sum I_{zz}d\right)^{-1} \sum I_{zy}dD$ in the metric of D , the μ_{cj} being the corresponding eigenvectors and $\nu_{cj} = \left(\sum I_{zz}d\right)^{-1} \sum I_{zy}d\mu_{cj}$, $j = 1, \ldots, r$.

[40] Hannan (1970, p. 299), Brillinger (1969) consider canonical correlation over a narrow frequency band, in which spectra are constant, whence the results are analogous to the classical results reported in §3.1.

For more general situations iterative procedures seem necessary, and we turn now to a discussion of these. Let us suppose that we have estimates $\hat{f}_{xx}(\lambda_s)$, possibly derived in the manner of §2.5. Then we consider minimizing

$$\tilde{Q}_N = N^{-1}\sum \left\| \hat{f}_{xx}(\lambda_s)^{-\frac{1}{2}}\{w_y(s)-URV'w_z(s)\}\right\|^2$$
$$+ \operatorname{tr}\left\{\left(U'\hat{\Omega}_{xx}^{-1}U-I_r\right)\Gamma_\mu + \left(V'\hat{\Omega}_{zz}V-I_r\right)\Gamma_\nu\right\},$$

where $\Gamma_\mu = \left(\gamma_{\mu jk}\right)$ and $\Gamma_\nu = \left(\gamma_{\nu jk}\right)$ are $r \times r$ symmetric matrices of Lagrange coefficients. We have chosen $\Pi_\mu = \hat{\Omega}_{xx}^{-1}$, $\Pi_\nu = \hat{\Omega}_{zz}$ in order that our estimates will be directly comparable with $\left(U_b' \vdots V_b' \vdots R_b\right)$. It will be convenient to consider first the case $r = 1$, since the procedure we propose for $r > 1$ is somewhat different. We have

$$\tfrac{1}{2}\partial\tilde{Q}_N/\partial\mu = -\rho \sum GI_{yz}\nu + \rho^2 \sum G\mu\nu' I_{zz}\nu + N\hat{\Omega}_{xx}^{-1}\mu\gamma_{\mu 11}, \qquad (A.1)$$

$$\tfrac{1}{2}\partial\tilde{Q}_N/\partial\nu = -\rho \sum I_{zy}G\mu + \rho^2 \sum \mu' G\mu I_{zz}\nu + N\hat{\Omega}_{zz}\nu\gamma_{\nu 11}, \qquad (A.2)$$

$$\tfrac{1}{2}\partial\tilde{Q}_N/\partial\rho = - \sum \mu' GI_{yz}\nu + \rho \sum \mu' G\mu\nu' I_{zz}\nu . \qquad (A.3)$$

Here we have omitted the subscript "1" from our parameters and G denotes $\hat{f}_{xx}^{-1}$. We equate (A.1)-(A.3) to zero, replacing μ, ν, ρ, $\gamma_{\mu 11}$, $\gamma_{\nu 11}$ by corresponding hatted quantities which are any values obeying the first-order optimization conditions; then pre-multiplying (A.1) by $\hat{\mu}'$ and (A.2) by $\hat{\nu}'$ gives $\hat{\gamma}_{\mu 11} = \hat{\gamma}_{\nu 11} = 0$, from (A.3). We commence with initial estimates $\hat{\mu}_0$, $\hat{\nu}_0$. On the $(k+1)$th iterative step, $k = 0, 1, \ldots$ we compute

$$\hat{\rho}_{k+1} = \left[\sum \hat{\mu}_k' G \hat{\mu}_k \hat{\nu}_k' I_{zz} \hat{\nu}_k\right]^{-1} \sum \hat{\mu}_k' G I_{yz} \hat{\nu}_k \ , \tag{A.4}$$

$$\hat{\mu}_{k+1} = \left[\hat{\rho}_{k+1} \sum G \hat{\nu}_k' I_{zz} \hat{\nu}_k\right]^{-1} \sum G I_{yz} \hat{\nu}_k \ , \tag{A.5}$$

$$\hat{\nu}_{k+1} = \left[\hat{\rho}_{k+1} \sum \hat{\mu}_{k+1}' G \hat{\mu}_{k+1} I_{zz}\right]^{-1} \sum I_{zy} G \hat{\mu}_{k+1} \ , \tag{A.6}$$

and continue until convergence is hopefully achieved.

When $r > 1$ we have

$$\tfrac{1}{2}\partial\tilde{Q}_N/\partial U = -\sum G I_{yz} V R + \sum G B I_{zz} V R + N\hat{\Omega}_{xx}^{-1} U \Gamma_\mu \ , \tag{A.7}$$

$$\tfrac{1}{2}\partial\tilde{Q}_N/\partial V = -\sum I_{zy} G U R + \sum I_{zz} B' G U R + N\hat{\Omega}_{zz} V \Gamma_\nu \ , \tag{A.8}$$

$$\tfrac{1}{2}\partial\tilde{Q}_N/\partial\rho_j = -\sum \mu_j' G I_{yz} \nu_j + \sum \mu_j' G B I_{zz} \nu_j \ , \quad j = 1, \ldots, r \ . \tag{A.9}$$

On pre-multiplying the first-order conditions derived from (A.7) and (A.8) by $\hat{U}'$ and $\hat{V}'$ respectively we get

$$\hat{\Gamma}_\mu = N^{-1}\hat{U}' \sum G(I_{yz} - \hat{B} I_{zz})\hat{V}\hat{R} \ ,$$

$$\hat{\Gamma}_\nu = N^{-1}\hat{V}' \sum (I_{zy} - I_{zz}\hat{B}')G\hat{U}\hat{R} \ .$$

We observe that $\hat{\Gamma}_\mu$ and $\hat{\Gamma}_\nu$ have zero diagonal elements (from (A.9)) but, in general, non-zero off-diagonal elements. (This is the reason we have distinguished between the cases $r = 1$ and $r > 1$.) We shall obtain equations determining an iterative step by substituting for $\hat{\Gamma}_\mu$, $\hat{\Gamma}_\nu$ in the first-order conditions derived from (A.7) and (A.8). We begin with the initial estimate $(\hat{U}_0' \,\vdots\, \hat{V}_0' \,\vdots\, \hat{R}_0)$; on the $(k+1)$th step, $k = 0, 1, \ldots$, we compute

$$\hat{U}_{k+1} = \left\{\hat{U}_k \hat{U}_k' \sum G I_{yz} - \hat{\Omega}_{xx} \sum G(I_{yz} - \hat{B}_k I_{zz})\right\}\hat{V}_k\left\{\sum \hat{U}_k' G \hat{B}_k I_{zz} \hat{V}_k\right\}^{-1} \ ,$$

$$\hat{V}_{k+1} = \left\{\hat{V}_k \hat{V}_k' \sum I_{zy} G - \hat{\Omega}_{zz}^{-1} \sum (I_{zy} - I_{zz}\hat{B}_k')G\right\}\hat{U}_k\left\{\sum \hat{V}_k' I_{zz} \hat{B}' G \hat{U}_k\right\}^{-1} \ ,$$

$$\hat{\rho}_{k+1}' = \left[(\hat{V}_k \odot \hat{U}_k)' \sum (I_{zz}(\hat{V}_k \odot G\hat{U}_k)\right]^{-1} (\hat{V}_k \odot \hat{U}_k)' \sum \mathrm{vec}\{G I_{yz}\} \ ,$$

adapting (A.7)-(A.9). Here $\hat{B}_k = \hat{U}_k\hat{R}_k\hat{V}_k'$, $\hat{R}_k$ having diagonal $\hat{\rho}_k$, and the last expression is derived using (1.2.3) and (1.2.5). We shall not discuss conditions for, or speed of, convergence of this procedure although it is hoped to examine its performance at a later date. The method suggested by Aitchison and Silvey (1958) may converge more rapidly but it would require the inversion of a much larger matrix than those involved above. The obvious choice of initial values is $(U_b' \vdots V_b' \vdots R_b)$, although this will not necessarily be sufficiently close to the efficient estimate to provide convergence to the *absolute* minimum of $\tilde{Q}_N$.

In CCA search and iterative procedures may be dispensed with because the solution of the eigenvalue problem automatically provides the optimal solution out of all those satisfying the first-order conditions. To illustrate the difference between CCA and the present situation we shall briefly describe an alternative procedure for $r = 1$, requiring the solution of an eigenvalue problem at each step. Putting (A.3) = 0 and substituting for $\sum \mu' G I_{yz} \nu$ in $\tilde{Q}_N$ it follows that we must maximize

$$c = \rho^2 \sum \mu' G \mu \nu' I_{zz} \nu \ .$$

Writing some of the $\hat{\mu}, \hat{\nu}$'s as $\hat{\mu}_k, \hat{\nu}_k$ and others as $\hat{\mu}_{k+1}, \hat{\nu}_{k+1}$ we have from (A.1)-(A.3),

$$\left[\begin{array}{c:c} \hat{\rho}_{k+1} \sum G\hat{\nu}_k' I_{zz}\hat{\nu}_k & -\sum G I_{yz} \\ \hdashline -\sum I_{zy} G & \hat{\rho}_{k+1} \sum \hat{\mu}_k' G \hat{\mu}_k I_{zz} \end{array}\right] \left[\begin{array}{c} \hat{\mu}_{k+1} \\ \hdashline \hat{\nu}_{k+1} \end{array}\right] = 0 \ .$$

We put $d_k = \sum \hat{\mu}_k' G \hat{\mu}_k \hat{\nu}_k' I_{zz} \hat{\nu}_k$ with $c_{k+1} = \hat{\rho}_{k+1}^2 d_k$ and we take as c_{k+1} the greatest zero c of

$$\det\left\{c \sum G\hat{\nu}_k' I_{zz}\hat{\nu}_k - d_k \sum G I_{yz}\left[\sum \hat{\mu}_k' G\hat{\mu}_k I_{zz}\right]^{-1} \sum I_{zy} G\right\} . \qquad \text{(A.11)}$$

We start by computing d_0 from initial values $\hat{\mu}_0, \hat{\nu}_0$. From (A.11) we find c_1, derive $\hat{\rho}_1 = (c_1/d_0)^{\frac{1}{2}}$, solve (A.10) for $\hat{\mu}_1, \hat{\nu}_1$ obeying the same normalisation constraints as $\hat{\mu}_0, \hat{\nu}_0$, compute d_1, and so on. The volume of computation may be greater than that generated by (A.4)-(A.6), although the greatest eigenvalue of a matrix can be found rapidly by means of the power method. Moreover (A.11) could be used to check the optimality of a solution derived by an alternative method: if the estimates are $\tilde{\rho}, \tilde{\mu}, \tilde{\nu}$ and we replace $\hat{\mu}_k, \hat{\nu}_k$ in (A.11) by $\tilde{\mu}, \tilde{\nu}$ the greatest zero should equal $\tilde{\rho}^2 \sum \tilde{\mu}' G\tilde{\mu}\tilde{\nu}' I_{zz}\tilde{\nu}$.

The techniques we have described are computationally onerous and it seems that one would often be content with the estimates U_b, V_b, R_b. It was shown in Theorem 3.3 that these have desirable properties even in the presence of serial correlation and indeed it is possible to measure their efficiency without actually computing efficient estimates. The matrix (2.7) (with $\Phi = f_{xx}^{-1}$) and the covariance matrix of Theorem 3.3 may be estimated by replacing sums by integrals, $(U_0' \vdots V_0' \vdots R_0)$ by $(U_b' \vdots V_b' \vdots R_b)$, f_{xx} by $\hat{f}_{xx}$, Π_μ by $\hat{\Omega}_{xx}^{-1}$ and Π_ν by $\hat{\Omega}_{zz}$. Then measures based on the non-zero eigenvalues of the leading $r(p_1+p_2+1)$-order square submatrices may be compared.

CHAPTER 4

TIME SERIES REGRESSION WITH UNKNOWN LAGS

4.1 Introduction

Linear regressions are usually fitted to time series data on the supposition that the endogenous variables depend on the past as well as the present behaviour of the exogenous variables. Thus we are led to formulate the model

$$y(t) = \sum_{j=1}^{r} B_0(j) z\left(t-\zeta_{0j}\right) + x(t) \ , \quad t \in R \ . \tag{1.1}$$

Here r is given, the ζ_{0j} are unknown and possibly non-integral lags[41] and the $B_0(j)$ are matrices of regression coefficients. We shall presently discuss the specification of (1.1) in greater detail. A more plausible model than (1.1) might relate $y(t)$ to a denumerable or a non-denumerable infinity of lagged values of $z(t)$. One might regard (1.1) as an approximation to, or truncation of, such a model to keep the dimension of the parameter space finite. (There are alternative ways of accomplishing the latter aim, which we discuss in subsequent chapters.) In the following section we consider the identification and estimation of the ζ_{0j}, $B_0(j)$ on the basis of a sample $y(n)$, $z(n)$, $n = 1, \ldots, N$, and present asymptotic theory. In §4.3 we describe the application of an elementary scalar model,

$$y(t) = \beta_0 z\left(t-\zeta_0\right) + x(t) \ , \quad t \in R \ , \tag{1.2}$$

to economic and oceanographic data.

Regression models that relate one time series to lagged values

[41] If $\zeta_{0j} < 0$, some j, ζ_{0j} represents a lead rather than a lag. As in §2 we allow for such a phenomenon but for brevity we shall always refer to the ζ_{0j} as lags.

of a second series have long been of interest in a variety of practical situations. However attention has focussed almost exclusively on models which include only lags that are multiples of the sampling interval, for example

$$y(n) = \sum_{j=-q}^{r} B(j)z(n-j) + x(n) \ , \quad n = 1, \ldots, N \ . \tag{1.3}$$

For the estimation of (1.3) Hannan (1963, 1967) proposes an efficient procedure using broad-band spectral techniques and also some estimates that are not in general efficient but possess a desirable property analogous to that of orthogonal polynomials: the $B(j)$ need not be re-estimated when q or r are increased. Notwithstanding the practical convenience of such procedures, when $y(n)$, $z(n)$ are sampled from a continuous process there is no reason in general why the underlying dependence should be confined to integral lags and there may be a need for more precise information about the lag structure. The results of the present chapter, for the most part, appear in modified form in Hannan and Robinson (1973), in which asymptotic theory for (1.2) is proved but that for (1.1) is merely stated. We shall follow the opposite course here because results for the multivariate case follow also from §2, and we include results for (1.2) as a corollary.

We confirm that (1.1) is of the form (2.1.1) by choosing

$$B = [B(1) \vdots \ldots \vdots B(r)] \ ,$$

$$\Gamma(t;\ \theta) = \left[\delta\left(t-\zeta_1\right) \vdots \ldots \vdots \delta\left(t-\zeta_r\right)\right]' \otimes I_{p_2} \ , \tag{1.4}$$

whence we have $p_3 = rp_1$. If $p_1 = p_2 = r = 1$ we get (1.2). More generally with $p_1 = p_2 = 1$ but $r > 1$ we have a univariate distributed lag model. However (1.1) deserves additional comment for when either p_1 or p_2 or both are greater then unity we must allow for the fact that we shall often wish to take a number of elements of the $B_0(j)$ to be zero. Indeed, if prior considerations suggest no duplications of ζ_{0j}, for some j, either between or within equations we would constrain every element but one of $B_0(j)$ to be zero. When there is no reason for believing any two pairs of components of $y(t)$, $z(t)$ to exhibit identical lags all $B_0(j)$ will have this form. Alternatively if a column, say the lth, of $B_0(j)$ possesses more than one non-zero element then $z_l(t-\zeta_{0j})$ appears in more than one equation. Again, the event of $B_0(j)$ having more than one non-null column implies that ζ_{0j} appears with several $z_l(t)$. The last two specifications may be unrealistic in view of the availability of a continuum of lags from which to choose, but if a rough equality between several lags seems likely we may be willing to sacrifice accuracy for computational convenience by acting as if they were in fact equal and keeping r to a minimum. In order to incorporate zero[42], and more general homogeneous linear restrictions, we take, from (1.4), $\beta = \mathrm{vec}'\{B\}$ and, following §2.2, introduce the projection matrix L so that $\beta = \beta L$ embodies the constraints on β.

[42] When all constraints are of the exclusion kind our results could be expressed in an alternative form which replaces the pseudoinverses in the theorem below by true inverses. Suppose there are altogether m elements, $r \leq m \leq p_1p_2r$, of B *not* constrained to be zero. We define E to be the $p_1p_2r \times m$ matrix derived from $I_{p_1p_2r}$ by omitting columns such that $\beta = \mathrm{vec}'\{B\}E$ is the $1 \times m$ vector of non-zero elements of B. In general post-multiplication by E of a p_1p_2r-columned matrix (or pre-multiplication by E' of a p_1p_2r-rowed matrix) has the effect of eliminating columns (or rows) corresponding to zero elements of B.

4.2 Identification, estimation and asymptotic theory

The possibility of the ζ_{0j} being non-integral produces problems of identification that do not exist for the integer-lagged system (1.3). Corresponding to the relations (2.2.3), (2.2.4), for $\lambda \in B$,

$$f_{yy}^{d}(\lambda) = \sum_{-\infty}^{\infty}\sum_{1}^{r}\sum_{1}^{r} B_0(j)f_{zz}^{c}(\lambda+2\pi l)B_0(k)'\exp\{i(\zeta_{0j}-\zeta_{0k})(\lambda+2\pi l)\} + f_{xx}^{d}(\lambda) ,$$

$$f_{yz}^{d}(\lambda) = \sum_{-\infty}^{\infty}\sum_{1}^{r} B_0(j)\exp\{i\zeta_{0j}(\lambda+2\pi l)\}f_{zz}^{c}(\lambda+2\pi l) .$$

Without being able to deduce the f_{zz}^{c} from f_{zz}^{d} we cannot possibly identify the ζ_{0j} , $B_0(j)$ from knowledge of f_{yy}^{d}, f_{yz}^{d} unless the ζ_{0j} are integers. Thus we impose (iv) whereupon, for example,

$$f_{yz}(\lambda) = \sum_{j=1}^{r} B_0(j)e^{i\zeta_{0j}\lambda}f_{zz}(\lambda) , \quad \lambda \in B ,$$

omitting the superscript d . However, further identification conditions are required. We take $\zeta_{0j} \neq \zeta_{0k}$, $j \neq k$, $j, k = 1, \ldots, r$, for otherwise there would be a term $[B_0(j)+B_0(k)]z(t-\zeta_{0j})$ and $B_0(j)$, $B_0(k)$ would not be identified.[43] To identify ζ_{0j} we must have $B_0(j) \neq 0$, all j, (for (1.2) we need $\beta_0 \neq 0$) and an ordering such as $\zeta_{01} < \zeta_{02} < \ldots < \zeta_{0r}$. Because an almost-periodic function is uniquely defined by its exponents and coefficients the conditions we have imposed ensure that $\sum_{j=1}^{r} B(j)\exp(i\zeta_j\lambda) \neq \sum_{j=1}^{r} B_0(j)\exp(i\zeta_{0k}\lambda)$ almost everywhere in B when $(\zeta \vdots \beta) \neq (\zeta_0 \vdots \beta_0)$, where $\zeta = (\zeta_1 \vdots \ldots \vdots \zeta_r)$. Then for

[43] Trivially, we could of course use our provision for linear restrictions to express the model in an uneconomical but identified fashion with $\zeta_{0j} = \zeta_{0j}$, some $j \neq k$.

suitable Φ , f_{zz} , (v) will hold. We cannot use the distribution theory below to test the hypotheses $B_0(j) = 0$, for any j , or $\zeta_{0j} = \zeta_{0k}$, for $j \neq k$, because were these true the identification conditions would be contravened.

We consider the approximation

$$w_y(s) = \sum_{j=1}^{r} B_0(j)\exp(i\zeta_{0j}\lambda_s)w_z(s) + w_x(s) \;, \quad \lambda_s \in B \;.$$

THEOREM 4.1. *Denote by* $(\hat{\zeta}_N \vdots \hat{\beta}_N)$ *the vector minimizing*

$$Q_N = N^{-1} \sum_{B} \left\| \Phi(\lambda_s)^{\frac{1}{2}} \left\{ w_y(s) - \sum_{j=1}^{r} B(j)\exp(i\zeta_j\lambda_s)\, w_z(s) \right\} \right\|^2$$

over admissible $(\zeta \vdots \beta)$ *such that, for given*[44] K_1, K_2 ,

$$-\tfrac{1}{2}N < K_1 \leq \zeta_1 < \ldots < \zeta_r \leq K_2 \leq \tfrac{1}{2}N \;.$$

Let ζ_0 *satisfy these constraints and let* (1.1) *and conditions* (ii), (iv) *and* (v) *hold. Then*

$$\lim_{N\to\infty} (\hat{\zeta}_N \vdots \hat{\beta}_N) = (\zeta_0 \vdots \beta_0) \;, \text{ a.s.}$$

Moreover if $\zeta_{01} > K_1$, $\zeta_{0r} < K_2$ *and conditions* (vi) *and* (vii) *also hold, then as* $N \to \infty$

$$N^{\frac{1}{2}}(\hat{\zeta}_N - \zeta_0 \vdots \hat{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} N\left[0,\; \Psi_0(\Phi)^{+}\Psi_0(\Phi f_{xx}\Phi)\Psi_0(\Phi)^{+}\right] \;,$$

where

$$\Psi(\Phi) = (2\pi)^{-1} \int_B T(\lambda)^*\{f_{zz}(-\lambda) \otimes \Phi(\lambda)\}T(\lambda)d\lambda \;,$$

with

$$T(\lambda) = \left[i\lambda e(i\zeta\lambda) \odot \left(\text{vec}\{B(1)\} \vdots \ldots \vdots \text{vec}\{B(r)\}\right) \vdots e(i\zeta\lambda) \otimes I_{p_1 p_2} L \right],$$

---

[44] However because of the particular way the ζ_j are involved, Hannan's (1973 b) proof of the uniform convergence of autocovariances seems to imply that the assumption of compactness is unnecessary for theoretical purposes here.

in which

$$e(\zeta\omega) = \left[\exp(\zeta_1\omega) \vdots \dots \vdots \exp(\zeta_r\omega)\right] .$$

Proof. The proof follows from Theorems 2.3, 2.6. The interval $[K_1, K_2]$ is taken to be a subset of $(-\frac{1}{2}N, \frac{1}{2}N]$ because Q_N is periodic of period N in the ζ_j. Because $e(i\zeta\lambda)$ and its derivatives are continuous in λ and ζ, (iii) and (ix) are implied. The form of Ψ follows by considering derivatives of

$$\mathrm{vec}\left\{\sum_{j=1}^{r} B(j)\exp(i\zeta_j\lambda)\right\} = \mathrm{vec}\left\{[B(1) \vdots \dots \vdots B(r)]\left[e(i\zeta\lambda)' \otimes I_{p_2}\right]\right\}$$

$$= \left\{e(i\zeta\lambda) \otimes I_{p_1p_2}\right\}\beta' .$$

The derivative in β follows immediately; the derivative in ζ is

$$i\lambda\left[\exp(i\zeta_1\lambda)\mathrm{vec}\{B(1)\} \vdots \dots \vdots \exp(i\zeta_r\lambda)\mathrm{vec}\{B(r)\}\right]$$

$$= i\lambda e(i\zeta\lambda) \odot \left(\mathrm{vec}\{B(1)\} \vdots \dots \vdots \mathrm{vec}\{B(r)\}\right) .$$

On re-writing Q_N in the manner of (2.3.6) we have for given ζ the estimate

$$\hat{\beta}_N(\zeta) = h_N(\zeta)'\left[LH_N(\zeta, \zeta)L\right]^+ , \tag{2.1}$$

where

$$h_N(\zeta) = N^{-1} \sum_B e(i\zeta\lambda_s)^* \otimes \mathrm{vec}\{\Phi(\lambda_s)I_{yz}(s)\} ,$$

$$H_N(\zeta, \zeta) = N^{-1} \sum_B e(i\zeta\lambda_s)^* e(i\zeta\lambda_s) \otimes I_{zz}(-s) \otimes \Phi(\lambda_s) .$$

We could of course assume the value of ζ_0 and find an explicit linear estimate of β_0 from (2.1); our theorem implies desirable properties for such an estimate. However when we also wish to estimate ζ_0 we may proceed by maximizing $h_N(\zeta)'\left[LH_N(\zeta, \zeta)L\right]^+h_N(\zeta)$ and then substituting in (2.1).

COROLLARY 4.1. *Denote by* $(\hat{\zeta}_N : \hat{\beta}_N)$ *the vector minimizing*

$$Q_N = N^{-1} \sum_{B} \left|w_y(s) - \beta\exp\left(i\zeta\lambda_s\right)w_z(s)\right|^2 \varphi\left(\lambda_s\right) ,$$

where φ *is continuous and even. Let* (1.2) *hold with* $\beta_0 \neq 0$. *Then the variates* $N^{\frac{1}{2}}\left(\hat{\zeta}_N - \zeta_0\right)$, $N^{\frac{1}{2}}\left(\hat{\beta}_N - \beta_0\right)$ *are asymptotically independent with variances*

$$(2\pi)^{-1} \int_B \lambda^2 f_{zz}(\lambda) f_{xx}(\lambda) \varphi^2(\lambda) d\lambda \bigg/ \left\{\beta_0 (2\pi)^{-1} \int_B \lambda^2 f_{zz}(\lambda)\varphi(\lambda) d\lambda\right\}^2 ,$$

$$(2\pi)^{-1} \int_B f_{zz}(\lambda) f_{xx}(\lambda) \varphi^2(\lambda) d\lambda \bigg/ \left\{(2\pi)^{-1} \int_B f_{zz}(\lambda)\varphi(\lambda) d\lambda\right\}^2 ,$$

respectively. If moreover, for all $\lambda \in B = (-\pi, \pi)$,

$$f_{zz}(\lambda) = \rho f_{xx}(\lambda) > 0 , \quad \varphi(\lambda) = f_{xx}(\lambda)^{-1}$$

the variances are respectively

$$3/\rho\left(\beta_0 \pi\right)^2 , \quad 1/\rho .$$

The corollary easily follows from the theorem. We note that the variances are independent of ζ_0 and that the precision of $\hat{\zeta}_N$ is directly related to the strength of the relationship between $y(t)$ and $z\left(t-\zeta_0\right)$, that is on the magnitude of $|\beta_0|$. The second part of the corollary deals with the efficient case with constant signal-to noise ratio.

Because we have not incorporated lagged endogenous variables in (1.1) the results of the present chapter are perhaps of less relevance to economics than to certain branches of the physical sciences. Our models may be suitable in circumstances in which a common signal is received by two recorders, arriving at one recorder with a lag relative to the other, the lag being determined by the speed of propagation and direction of the signal and the distance separating the recorders. Hannan and Thomson (1972) envisage such a situation, but over a very narrow frequency band. Our broad band model may be less plausible in this case because it seems that different frequency components of the signal are likely to give rise to different lags. We note also that if, as often in the natural sciences, it is possible to choose a very small sampling interval one may be content to consider integral lags, the available extra precision being of no value. However the number of integers to be scanned may be very large and the computational effort involved in minimizing over integral ζ_j may be comparable to that of minimizing over all $\zeta_j \in R$. Furthermore an asymptotic theory for such estimates may be difficult to develop.

4.3 Numerical Examples

The performance on two sets of data of the elementary scalar model (1.2) was examined. The first data used consisted of a series of monthly measurements of U.S. Money Stock (not seasonally adjusted) and a series of monthly indices of U.S. Industrial Production, both for the period 1955-1970, each series comprising 192 observations. There is foundation for arguing that either of these series leads the other, an observed lag being the result of averaging leads as well as lags; it is thus perhaps unreasonable to regard either variable as purely exogenous. A more plausible model would be of the multiple type considered in §4.2 or in subsequent chapters, possibly treating both money stock and production as endogenous and introducing also variables which are more acceptable as exogenous. Nevertheless, $y(t)$ was taken as money stock and $z(t)$ as production.

We deduce from §4.2 that, after mean-correcting the data, $\hat{\zeta}_N$ is the value that maximizes

$$\left\{N^{-1} \sum_{B} I_{zy}(s)\exp\left(i\zeta\lambda_s\right)\varphi\left(\lambda_s\right)\right\}^2 . \tag{3.1}$$

Then

$$\hat{\beta}_N = \sum_{B} I_{zy}(s)\exp\left(i\hat{\zeta}_N\lambda_s\right)\varphi\left(\lambda_s\right) \Big/ \sum_{B} I_{zz}(s)\varphi\left(\lambda_s\right) . \tag{3.2}$$

In the first place we took $B = (-\pi, \pi)$ and $\varphi \equiv 1$. (In this case, incidentally, the maximization of (3.1) is very similar to the maximization of squared autocovariances because, for integral $\zeta > 0$, (3.1) is

$$\left\{(2\pi)^{-1}\left[C_{zy}(\zeta)+C_{zy}(\zeta-N)\right]\right\}^2 ,$$

$|\zeta| < N$.) Secondly these estimates were used to estimate $f_{xx}(\lambda)$ from

$$\hat{f}_{xx}\left(\omega_m\right) = 2MN^{-1} \sum_{s(m)} \left|w_y(s)-\hat{\beta}_N\exp\left(i\hat{\zeta}_N\lambda_s\right)w_z(s)\right|^2 ,$$

for $\omega_m = 0, \pm \frac{1}{8}\pi, \ldots, \pm \frac{7}{8}\pi, \pi$. Thus $M = 8$. Then efficient estimates were derived by inserting the $\hat{f}_{xx}^{-1}$ in place of φ in (3.1) and (3.2). Because independent examination of the coherence between $y(n)$ and $z(n)$ indicated little mass beyond a relatively small band centred at the origin, the same procedures were carried out with $B = (-\frac{3}{16}\pi, \frac{3}{16}\pi)$. The results are as follows. (For brevity we put f for $\hat{f}_{xx}^{-1}$.)

Table 4.1					
Economic Data					
B	φ	$\hat{\zeta}_N$	s.e. $\hat{\zeta}_N$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
$(-\pi, \pi)$	1	0.342	0.080	0.741	0.061
"	f	0.027	0.024	0.873	0.038
$(-\frac{3}{16}\pi, \frac{3}{16}\pi)$	1	1.477	0.616	0.749	0.056
"	f	0.489	0.352	0.817	0.052

In fact (3.1) was in each case maximized over $|\zeta| \leq 20$. The estimates using $\hat{f}_{xx}$ are the result of carrying out the cycle of estimating ζ_0 , β_0 and f_{xx} four times although a satisfactory degree of convergence had been attained after the second iteration. The standard errors were computed from the formulae given in the first part of Corollary 4.1, replacing integrals by sums, spectral densities by periodograms and β_0 by $\hat{\beta}_N$. The difference in $\hat{\zeta}_N$ between lines 1 and 2 of Table 4.1 suggests that the model is unsuitable over $(-\pi, \pi)$. It seems that the relative over-weighting of low frequencies in line 1 , where possibly a large average lag is present but the coherence is still low, gives rise to the high value of $\hat{\zeta}_N$. It was felt that the lag might more reasonably be regarded as constant over $(-\frac{3}{16}\pi, \frac{3}{16}\pi)$, and in fact the proportional discrepancy between lines 3 and 4 is less. We emphasize that inspection of the data reveals substantial trend and seasonal effects: money stock shows little variation during the first half of the period apart from an annual cycle, and then shows a gradual upward trend. Thus apart from the model being invalid our stationarity assumptions are contravened. Because of the likely absence of high frequencies, however, (iv) is probably reasonable, as it seems to be in respect of the second set of data. This consisted of measurements of sea level, recorded twice daily at two localities approximately 270 miles apart on the Australian East Coast, namely Coff's Harbour and Fort Denison. A shelf wave is known to travel from south to north and so Coff's Harbour, which is the most northerly, was chosen as $y(t)$ and Fort Denison as $z(t)$. Four sets of estimates were again computed, in the same way as before (with $M = 8$ still) but taking (on the basis of prior information) B as

$(-\frac{5}{16}\pi, \frac{5}{16}\pi)$ in the second place.

Table 4.2					
Oceanographic Data					
B	φ	$\hat{\zeta}_N$	s.e. $\hat{\zeta}_N$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
$(-\pi, \pi)$	1	1.483	0.133	0.925	0.054
"	f	3.271	0.039	0.488	0.036
$(-\frac{5}{16}\pi, \frac{5}{16}\pi)$	1	1.418	0.175	0.935	0.054
"	f	1.629	0.154	0.894	0.046

Because of the combination of low coherence and small $\hat{f}_{xx}$ at high frequencies the inverse weighting by $\hat{f}_{xx}$ distorts $\hat{\zeta}_N$ in line 2. The situation is rectified, however, by the exclusion of high frequencies, the results in lines 1, 3 and 4 agreeing reasonably with prior expectations.

CHAPTER 5

DIFFERENCE EQUATIONS WITH UNKNOWN DIFFERENCES

5.1 Introduction

There are often good reasons for including in a linear model lagged values of the endogenous variables, relating the latter to their own past history besides that of other variables. Such systems, which have long interested economists, tend to imply that the effect of $z(t)$ on the future behaviour of $y(t)$ is long-term as well as short-term. We consider the system

$$y(t) + \sum_{j=0}^{q} A_0(j)y(t-\eta_{0j}) = \sum_{j=1}^{r} B_0(j)z(t-\zeta_{0j}) + u(t) \ , \quad t \in R \ . \qquad (1.1)$$

Here q and r are given integers, the η_{0j}, ζ_{0j} are unknown and possibly non-integral lags and the $A_0(j)$, $B_0(j)$ are respectively $p_1 \times p_1$ and $p_1 \times p_2$ coefficient matrices. We denote the residual process $u(t)$ rather than $x(t)$ for a reason that will presently be indicated.

We wish to estimate the η_{0j}, ζ_{0j}, $A_0(j)$, $B_0(j)$ by means of N observations on $y(n)$, $z(n)$ sampled at unit intervals. There has apparently been no previous discussion of the identification and estimation of models such as (1.1) when the η_{0j}, ζ_{0j} are not known integers. The inclusion of the lagged $y(t)$ introduces a complication but the type of estimates we suggest follows from §2. To express (1.1) in the form (2.1.1) we must consider its solution or reduced form, whose nature depends on whether or not the system is stable. We describe the general system (2.1.2) as stable if and only if all zeros ω of $\det\{\Lambda_1[\exp(\omega t)]\}$ have negative real parts, where we replace each element of $y(t)$ by $\exp(\omega t)$. Then if (1.1) is stable we shall see that its stationary solution expresses $y(t)$ in terms of $z(\tau)$ for $\tau \leq t$ only, unless $\zeta_{0j} < 0$, some j , in which case values of $z(\tau)$ over a finite future period are also involved. Unfortunately the mathematical problems of listing possible solutions of (1.1) and laying down perspicuous stability conditions in terms of the parameter values are more difficult than for autoregressive systems involving only integral differences. The treatment of linear difference equations with non-integral differences in the mathematical literature is of little use to us here. More interest has been shown, in any case, in the mixed difference-differential equation systems considered in §7 which, it must be said, appear to offer a more plausible description of many continuous-time relations than does (1.1).

Because of the relative lack of previous work in the field, it seems desirable to postpone our discussion of (1.1) by dealing first, in §5.2, with the solution and stability of simpler scalar models. In §5.3 the identification and estimation problems for (1.1) are considered, and asymptotic theory presented. Finally in §5.4 we describe the application of a simple model to our economic and oceanographic data. The results of the present chapter are the subject of an article the author is preparing for publication.

5.2 Scalar difference equations

We consider the functional equation

$$\sum_{j=0}^{q} \alpha(j) y\left(t-\eta_j\right) = w(t) \ , \quad t \in R \ , \tag{2.1}$$

with $q \geq 1$. Without loss of generality we impose the normalisation constraint $\alpha(0) = 1$, and take $\alpha(j) \neq 0$ all j and

$0 = \eta_0 < \eta_1 < \ldots < \eta_q$. A more common form than (2.1) in the mathematical literature is

$$\sum_{j=0}^{q} \gamma(j)y\left(x+\theta_j\right) = v(x) \ , \tag{2.2}$$

where $0 = \theta_0 < \theta_1 < \ldots < \theta_q$, the θ_j being referred to not only as differences or lags but as spans or retardations. If we take $x = t - \eta_q$, $\theta_j = \eta_q - \eta_{q-j}$, $\gamma(j) = \alpha(q-j)$; $j = 0, 1, \ldots, q$, $v\left(t-\eta_q\right) = w(t)$ then (2.2) becomes (2.1). For real x , θ_j and complex $\gamma(j)$, (2.2) has been studied by Bochner (1931), Raclis (1930), Martin (1938); for complex x , θ_j by Pincherle (1926), Carmichael (1933), Strodt (1948). However there are fundamental differences between our approach and motivation and that of these authors. They take $v(x)$ to be a member of some class of functions (e.g. a function of exponential type or an entire function) or alternatively consider the homogeneous case $v(x) \equiv 0$. In (2.1), on the other hand, all quantities are real and $w(t)$ is not a deterministic function but a stationary stochastic process whose history is indexed by t . Moreover (2.2) is generally taken to be valid only over a proper subset of the complex plane or R , in which case boundary conditions are needed. In the common situation in which (2.2) is valid for (real) $x > 0$, $y(x)$ is assumed for $-\theta_q < x \le 0$. The general solution is then not strictly of the form (2.1.1), involving an extra term dependent on these conditions.[45] For (2.1) with $t \in R$, however, such a term will not arise, although conditions on $y(t)$ as $|t| \to \infty$ are implied. Finally the study of non-integral difference equations has centred around the question

[45] In the event of stability, however, this term decays to zero as $x \to \infty$. If $y(x) = 0$, $-\theta_q < x \le 0$, the term disappears uniformly in $x > 0$.

of the existence and uniqueness of solutions of (2.2), partly as an intermediary to the study of more advanced equations in which the $\gamma(j)$ are functions of x, in particular when they are asymptotically constant. Our concerns are more mundane. In the belief that $y(t)$ is causally dependent on only past $w(\tau)$ we wish to know whether given η_j, $\alpha(j)$ imply stability and to derive the solution. We shall suggest a simple sufficient condition for stability.

In the m.s. sense (2.1) is

$$\int_R e^{it\lambda}\{\tilde{\alpha}(-i\lambda)d\chi_y(\lambda)-d\chi_w(\lambda)\} = 0 \ , \quad t \in R \ ,$$

where $\chi_y(\lambda)$, $\chi_w(\lambda)$ have orthogonal increments (taking $y(t)$, $w(t)$ to be stationary) and

$$\tilde{\alpha}(\omega) = \sum_{j=0}^{q} \alpha(j)\exp\left(\eta_j\omega\right) \ . \tag{2.3}$$

We shall always assume that $\tilde{\alpha}(i\lambda)$ *is bounded away from zero,*[46] *uniformly in* $\lambda \in R$. The formal solution is then

$$y(t) = \int_R e^{it\lambda}\tilde{\alpha}(-i\lambda)^{-1}d\chi_w(\lambda) \ , \quad t \in R \ . \tag{2.4}$$

The form of (2.4) in terms of lagged $w(t)$ depends upon the distribution of the zeros of $\tilde{\alpha}(-\omega)$. Now the almost-periodic function $\tilde{\alpha}(-\omega)$ has infinitely many zeros, all of which lie in a strip[47] $|\sigma| < K < \infty$, where σ will always denote $Re\{\omega\}$. As in Bochner (1931), Martin (1938), the set of real parts of all zeros of $\tilde{\alpha}(-\omega)$, with limit points annexed, has a complementary set on R consisting of at most denumerably many open intervals, which include two half-

[46] If $i\lambda$ is bounded away from the zeros of $\tilde{\alpha}(-\omega)$, uniformly in $\lambda \in R$, then $\tilde{\alpha}(i\lambda)$ is bounded away from zero, uniformly in $\lambda \in R$. (Bellman and Cooke, 1963, p. 403.)

[47] This stems from the dominance of the first term, $\alpha(0)$, over the remaining terms in (2.3) for $\sigma > K$ and the dominance of $\alpha(q)\exp\{-\eta_q\omega\}$ for $\sigma < -K$.

lines, the right being denoted J_0 and the left, J_1. Now the set $\{\eta_j\}$ forms a basis for the numbers $\sum_{j=0}^{q} n_j\eta_j$ for all $n_j \in \{n \mid n = 0, \pm 1, \dots\}$, $j = 0, \dots, q$. We denote by μ_0 the number for which $n_j = 0$, all j, (so $\mu_0 = 0$) and denote the remaining distinct numbers $\mu_1, \mu_2, \dots$, in some arbitrary order. Then for $\sigma \in J_m$, $\tilde{\alpha}(-\omega)^{-1}$ is an analytic almost-periodic function with generalized Dirichlet expansion

$$\tilde{\alpha}(-\omega)^{-1} = \sum_{n=0}^{\infty} \kappa_m(\mu_n)\exp(\mu_n\omega) \;, \tag{2.5}$$

$m = 0, 1, \dots$. The real coefficients $\kappa_m(\mu_n)$ are uniquely determined and (2.5) converges absolutely, $\sigma \in J_m$. Now if $\{0\} \in J_m$ we can re-write (2.4) as

$$\begin{aligned} y(t) &= \int_B \sum_{n=0}^{\infty} \kappa_m(\mu_n)\exp\{i(t+\mu_n)\lambda\}d\chi_w(\lambda) \\ &= \sum_{n=0}^{\infty} \kappa_m(\mu_n)w(t+\mu_n) \;, \quad t \in R \;. \end{aligned} \tag{2.6}$$

It is not difficult to check the validity of (2.6) as a solution of (2.1) by means of the relations

$$\sum_{j=0}^{q} \alpha(j)\kappa_m(\mu_n+\eta_j) = 1 \;, \quad n = 0; \; = 0, \; n \neq 0 \;,$$

which follow from the identity

$$1 = \tilde{\alpha}(-\omega)\tilde{\alpha}(-\omega)^{-1} = \sum_{n=0}^{\infty}\sum_{j=0}^{q} \alpha(j)\kappa_m(\mu_n)\exp\{(\mu_n-\eta_j)\omega\} \;,$$

on collecting terms with like exponents.[48]

Now $\kappa_m(\mu_n) = 0$, all $\mu_n > 0$ when $m = 0$, for otherwise $\sum \kappa_0(\mu_n)\exp(\mu_n\omega)$ would not converge uniformly in $\sigma \geq 0$. If, conversely, $\tilde{\alpha}(-\omega) = 0$, some $\sigma > 0$, the Dirichlet expansion of $\tilde{\alpha}(-\omega)^{-1}$ (for $\sigma = 0$) will involve non-zero $\kappa_m(\mu_n)$ for some $\mu_n > 0$, for otherwise the expansion would converge uniformly in $\sigma \geq 0$. Thus a necessary and sufficient condition for (2.6) to involve $w(\tau)$ only for $\tau \leq t$ is that all zeros of $\tilde{\alpha}(-\omega)$ have negative real parts. Now for $\sigma \in J_0$,

$$\tilde{\alpha}(-\omega)^{-1} = \sum_{k=0}^{\infty} \left[- \sum_{j=1}^{q} \alpha(j)\exp(-\eta_j\omega)\right]^k$$

$$= 1 + \sum_{k=1}^{\infty} (-1)^k \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \prod_{l=1}^{k} \alpha(j_l)\exp\left(-\eta_{j_l}\omega\right) .$$

Thus when (2.1) is stable its solution, under our conditions, is

$$y(t) = w(t) + \sum_{k=1}^{\infty} (-1)^k \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \alpha(j_1) \cdots$$

$$\times \alpha(j_k)w\left(t - \eta_{j_1} - \cdots - \eta_{j_k}\right) , \quad t \in R . \qquad (2.7)$$

Stability conditions for given η_j , $\alpha(j)$ are of course

---

[48] If (2.1) is valid for $t > 0$ only, with $y(t)$ given for $-\eta_q < t \leq 0$, the series in (2.6) is truncated by the omission of terms in which w has non-positive argument. Moreover a further term is included and it is not hard to show that this has the form

$$-\sum_1^q\alpha(j)\sum_0^\infty\kappa_m(\mu_n)\int_{-\eta_j}^{0} y(\tau)\delta(t-\tau+\mu_n-\eta_j)d\tau = -\sum_1^q\alpha(j) \sum_{n(j)}{}' \kappa_m(\mu_n)y(t+\mu_n-\eta_j) ,$$

where the second sum is over n for which $0 < t+\mu_n < \eta_j$. This term, in the initial values of $y(t)$, may alternatively be expressed as $\sum p_j(t)\exp\{\omega_j t\}$, the sum being over distinct zeros ω_j of $\tilde{\alpha}(-\omega)$ and $p_j(t)$ being a polynomial of order less then the multiplicity of ω_j . The term decays to zero as $t \to \infty$ if $Re\{\omega_j\} < 0$ uniformly.

required, the actual computation of even a small number of zeros of $\tilde{\alpha}(-\omega)$ generally being out of the question. Unfortunately no such conditions appear to have been developed for situations in which the η_j are not of a very special form. However because $\tilde{\alpha}(\omega)$ is continuous a sufficient condition for $\{0\} \in J_0$ is

$$\left|\sum_{j=1}^{q} \alpha(j)\exp\left(-\eta_j\omega\right)\right| < 1 \ , \quad \sigma \geq 0 \ .$$

Since the left side is bounded, uniformly in $\sigma \geq 0$, by $\sum_{j=1}^{q} |\alpha(j)|$, a sufficient condition for stability is

$$\sum_{j=1}^{q} |\alpha(j)| < 1 \ . \tag{2.8}$$

Thus *uniformly in* $\eta_j > 0$, $j = 1, \ldots, q$, (2.1) is stable whenever $[\alpha(1) \vdots \ldots \vdots \alpha(q)]$ is an interior point of the hypercube having vertices at $(\pm 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, \pm 1)$. Our condition lacks refinement in that it uses only the fact that the η_j are positive and distinct so for given $\eta_1, \ldots, \eta_q$ it may fail to detect the presence of stability. For the second-order integral-difference equation

$$y(t) + \alpha(1)y(t-1) + \alpha(2)y(t-2) = w(t) \ , \quad t \in R \ , \tag{2.9}$$

necessary and sufficient conditions for stability are

$$\alpha(2) < 1 \ , \quad \alpha(1) - \alpha(2) - 1 > 0 \ , \quad \alpha(1) + \alpha(2) + 1 > 0 \ . \tag{2.10}$$

The region (2.8) represents only one half of (2.10). We note that (2.10) provides a *necessary* condition for (2.1) (with $q = 2$) to be stable *uniformly in* $\left(\eta_1 \vdots \eta_2\right)$. A further necessary condition when $q \geq 2$ is obtained by considering the extreme unstable case $\{0\} \in J_1$, for which the convergent solution is found to be

$$y(t) = \alpha(q)^{-1}\Big\{w\big(t+\eta_q\big) + \sum_{k=1}^{\infty} [-\alpha(q)]^k \sum_{j_1} \dots \sum_{j_k} \alpha\big(j_1\big) \dots$$

$$\times\, \alpha\big(j_k\big) w\Big(t+(k+1)\eta_q - \eta_{j_1} - \dots - \eta_{j_k}\Big)\Big\} , \quad t \in R .$$

Then a (very mild) necessary condition for (2.1) to be stable uniformly in $\eta_j > 0$, $j = 1, \dots, q$, is found to be

$$|\alpha(q)| \le 1 + \sum_{j=1}^{q-1} |\alpha(j)| .$$

(Necessary conditions will not be useful in determining an admissible parameter set for estimation purposes.)

We turn now to a consideration of some special cases of (2.1) for which the precise assessment of stability is easier and which involve a generalization of conventional integral-difference equations that is more limited but may have some practical value. We commence with the first-order equation, with $\eta > 0$,

$$y(t) + \alpha y(t-\eta) = w(t) , \quad t \in R . \tag{2.11}$$

The zeros of $\tilde{\alpha}(-\omega) = 1 + \alpha\exp\{-\eta\omega\}$ are

$$\omega_k = \eta^{-1}\log|\alpha| + i\eta^{-1}[2\pi k+\arg(-\alpha)] , \quad k = 0, \pm 1, \dots .$$

The zeros are thus spaced $2\pi/\eta$ units apart along the line $\sigma = \eta^{-1}\log|\alpha|$. Then we have stability if and only if $|\alpha| < 1$, whence the solution is[49]

$$y(t) = \sum_{j=0}^{\infty} (-\alpha)^j w(t-j\eta) , \quad t \in R . \tag{2.12}$$

We observe that (2.12) is a generalization of the geometric distributed lag model in which $\eta = 1$, which has been much discussed in an

[49] When (2.11) is valid only for $t > 0$ and $y(t)$ is given for $-\eta < t \le 0$ the series in (2.12) is truncated by the exclusion of terms in which w has non-positive argument and the term $y(\tau)(-\alpha)^{(t-\tau)/\eta}$ is included, τ being uniquely defined by the requirement that $-\eta < \tau \le 0$ and $(t-\tau)/\eta$ is integral. This term decays to zero as $t \to \infty$ if $|\alpha| < 1$.

econometric context.

If the η_j in (2.1) have the form $\eta_j = \eta j$, where $\eta > 0$ and the j are non-negative integers, the η_j are said to be *commensurable*, in the terminology of Langer (1931). (We now allow $\alpha(j) = 0$ for any j for which the lag ηj is not to appear.) In particular when $\eta = 1$ we have an ordinary integral-difference equation, for the estimation of which a number of methods already exist. The introduction of the parameter η , which may be estimated, may be worthwhile when it is practically undesirable to allow for incommensurable differences but unreasonable to take the fundamental difference to be the sampling interval. One might consider extending the various distributed lag structures for integral lags that have been suggested (see for example Jorgenson, 1966, Dhrymes, 1971 b) to allow for general η . The zeros of $\tilde{\alpha}(-\omega)$ are obtainable from the zeros ν_l of the polynomial $\psi(\nu) = 1 + \sum_{j=1}^{q} \alpha(j)\nu^j$ and are

$$\omega_{kl} = \eta^{-1}\log|\nu_l| + i\eta^{-1}[2\pi k+\arg(\nu_l)] , \quad k = 0, \pm 1, \ldots, l = 1, \ldots, q .$$

The ω_{kl} form a finite number of chains, not necessarily distinct, each consisting of a denumerable number of zeros on the lines $\sigma = \eta^{-1}\log|\nu_l|$. To examine stability we may thus consider $\psi(\nu)$. Marden (1949, Chapter X) discusses conditions for $|\nu_l| > 1$, all l , the actual computation of the ν_l often being unfeasible. (Dhrymes, 1971 b, p. 332, considers a stability test for integral-difference equations which is relevant here; Hannan, 1973 a, has developed asymptotic theory for the estimated ν_l .) The stable solution is easily deduced from (2.7) by grouping terms appropriately.[50] We note finally that the stability problem continues to be relatively tractable if we assume that (2.1) gives rise to a characteristic function of the form

$$\tilde{\alpha}(-\omega) = \prod_{l=1}^{m} \left\{1 + \sum_{j=1}^{q_l} \alpha_l(j)\exp(-j\eta_l\omega)\right\} .$$

For each of the $\bar{m} > 1$ factors the components in the sum are commensurable and stability is achieved when each of the polynomials

$1 + \sum_{j=1}^{q_l} \alpha_l(j)\nu^j$, $l = 1, \ldots, m$, has all of its zeros outside the unit circle.

5.3 Identification, estimation and asymptotic theory

We consider first the identification of (1.1) and its scalar analogue. We introduce the generating functions

$$\tilde{A}(\omega) = I_{p_1} + \sum_{j=0}^{q} A(j)\exp(\eta_j\omega) \quad , \quad \tilde{B}(\omega) = \sum_{j=1}^{r} B(j)\exp(\zeta_j\omega) \quad ,$$

defined for all admissible parameter values and complex ω . We assume throughout that $\det\{\tilde{A}_0(i\lambda)\}$ is bounded away from zero, uniformly in $\lambda \in R$. Then for $\lambda \in B$,

$$f^d_{yy}(\lambda) = \sum_{-\infty}^{\infty} \tilde{A}_0(i\lambda_l)^{-1}\left\{\tilde{B}_0'(i\lambda_l)f^c_{zz}(\lambda_l)\tilde{B}_0(i\lambda_l)^* + f^c_{uu}(\lambda_l)\right\}\tilde{A}_0(i\lambda_l)^{-1*} \quad (3.1)$$

$$f^d_{yz}(\lambda) = \sum_{-\infty}^{\infty} \tilde{A}_0(i\lambda_l)^{-1}\tilde{B}_0(i\lambda_l)f^c_{zz}(\lambda_l) \quad , \quad (3.2)$$

where $\lambda_l = \lambda + 2\pi l$ and $f^c_{uu}(\lambda)$, $\lambda \in R$, is the spectral density matrix of the m.s. continuous stationary process $u(t)$, which has absolutely continuous spectrum. For non-integral η_{0j}, ζ_{0j} we must,

---

[50] In this case the Dirichlet development of $\tilde{\alpha}(-\omega)^{-1}$ in vertical strips may be replaced by the Laurent developement of $\psi(\nu)^{-1}$ in annular rings.

for identification purposes, impose (iv) whereupon we have, from (3.2),

$$f_{yz}(\lambda) = \tilde{A}_0(i\lambda)^{-1}\tilde{B}_0(i\lambda)f_{zz}(\lambda) \ , \quad \lambda \in \mathcal{B} \ , \tag{3.3}$$

omitting superscripts. We also assume

(xi) *For* $\lambda \in \mathcal{B}$, $l = \pm 1, \pm 2, \ldots$,

$$f^c_{uu}(\lambda + 2\pi l) = 0 \ .$$

Then under (xi) we may invert (3.1), using (3.3), to get

$$f_{uu}(\lambda) = \tilde{A}_0(i\lambda)\left\{f_{yy}(\lambda) - f_{yz}(\lambda)f_{zz}(\lambda)^+ f_{zy}(\lambda)\right\}\tilde{A}_0(i\lambda)^* \ , \tag{3.4}$$

$\lambda \in \mathcal{B}$. Thus if $\tilde{A}_0(i\lambda)$ is identified by (3.3), $f_{uu}(\lambda)$ is identified by (3.4).

A number of extra conditions are however needed to identify $\tilde{A}_0$ and $\tilde{B}_0$. We consider these first of all in relation to the scalar model

$$\sum_{j=0}^{q} \alpha_0(j)y\left(t - \eta_{0j}\right) = \sum_{j=1}^{r} \beta_0(j)z\left(t - \zeta_{0j}\right) + u(t) \ , \quad t \in \mathbb{R} \ . \tag{3.5}$$

The equation will not be fixed in time and the $\alpha_0(j)$ will not be identified without requiring, for example, that $0 = \eta_{00} < \eta_{01} < \ldots < \eta_{0q}$ and $\alpha_0(0) = 1$. (An alternative normalisation and sign constraint to $\alpha_0(0) = 1$ is $\sum \alpha_0(j)^2 = 1$, $\alpha_0(0) > 0$.) We clearly need $\alpha_0(j) \neq 0$, all j, to avoid annihilating any η_{0j}. Likewise we need $\zeta_{01} < \zeta_{02} < \ldots < \zeta_{0r}$, $\beta_0(j) \neq 0$, $j = 1, \ldots, r$. Unfortunately further, less perspicious, conditions are required because of the necessity of identifying the parameters from $\tilde{\alpha}_0(i\lambda)^{-1}\tilde{\beta}_0(i\lambda)$, where

$$\tilde{\alpha}(\omega) = \sum_{j=0}^{q} \alpha(j)\exp(\eta_j\omega) \ , \quad \tilde{\beta}(\omega) = \sum_{j=1}^{r} \beta(j)\exp(\zeta_j\omega) \ .$$

Now Ritt (1927) has shown that all exponential sums $\tilde{\alpha}(\omega)$ (with $\alpha(0) = 1$, $\eta_0 = 0$) have a unique factorization

$$\tilde{\alpha}(\omega) = \Pi\tilde{\alpha}_j^S(\omega)\Pi\tilde{\alpha}_j^I(\omega) \ .$$

Here the $\tilde{\alpha}_j^S$ are "simple" factors, being exponential sums with commensurable exponents. If the fundamental lag in $\tilde{\alpha}_j^S$ is denoted $\eta^{(j)}$ we take $\eta^{(j)}/\eta^{(k)}$ to be irrational, $j \neq k$. (Otherwise, if $\eta^{(j)}/\eta^{(k)} = n/m$, say, n, m integral, $\tilde{\alpha}_j^S\tilde{\alpha}_k^S$ is itself simple with fundamental lag $\eta^{(j)}/n = \eta^{(k)}/m$.) The factors $\tilde{\alpha}_j^I$ are termed "irreducible" by Ritt; each is an exponential sum divisible by no exponential sum but itself. Now suppose $\tilde{\alpha}_0$ and $\tilde{\beta}_0$ have one or more factors in common. Then it may be possible to replace these common factors by alternative common factors to get a function $\tilde{\alpha}_1(\omega)^{-1}\tilde{\beta}_1(\omega)$, involving parameters η_{1j}, ζ_{1j}, $\alpha_1(j)$, $\beta_1(j)$ that are admissible but are not all identical to the corresponding η_{0j}, ζ_{0j}, $\alpha_0(j)$, $\beta_0(j)$. Then since $\tilde{\alpha}_1(\omega)^{-1}\tilde{\beta}_1(\omega) \equiv \tilde{\alpha}_0(\omega)^{-1}\tilde{\beta}_0(\omega)$, (3.5) is not identified.[51] We assume that no such common factors are present. A sufficient condition is that $\tilde{\alpha}_0$ and $\tilde{\beta}_0$ are irreducible and $\tilde{\alpha}_0 \not\equiv K\tilde{\beta}_0$, for any constant K. Ritt (1927) has discussed the problem of extracting the factors of $\tilde{\alpha}$ and $\tilde{\beta}$ when they are not irreducible. If $\tilde{\alpha}$ is divisible by an exponential sum the exponents of the latter are linear combinations of the η_j with rational coefficients. Moreover there exist parameters μ_j, $j = 1, \ldots, n$, say, that are linearly independent (in the sense that there exists no set of rational coefficients $q_j \not\equiv 0$ for which $\sum_{j=1}^{n} q_j\mu_j = 0$) and such that each η_j is expressible as a linear combination of the μ_j with positive integral coefficients. Then we may write $x_j = \exp(\mu_j\omega)$ and express $\tilde{\alpha}(\omega)$ as a "generalized polynomial" in the x_j. Ritt demonstrates how the simple factors of $\tilde{\alpha}$ correspond to the factors of the polynomial consisting of two terms, and how the irreducible factors are derived from the factors of

[51] For example suppose $\tilde{\alpha}_0(\omega) = 1 + \alpha_0\exp(\eta_0\omega)$, $\tilde{\beta}_0(\omega) = \tilde{\alpha}_0(\omega)[\beta_0+\gamma_0\exp(\zeta_0\omega)]$, with $\alpha_0, \gamma_0 \neq 0$, $0 < \eta_0 < \zeta_0$. Then $\tilde{\alpha}_0(\omega)$ may be replaced by $\tilde{\alpha}_1(\omega) = 1 + \alpha_1\exp(\eta_1\omega)$, all $\alpha_1 \neq 0$, $0 < \eta_1 < \zeta_0$. However suppose we know also that $\gamma_0 = -\alpha_0\beta_0$, $\zeta_0 = \eta_0$ and $\beta_0 \neq 0$. Then

$$\tilde{\beta}_0(\omega) = \beta_0\tilde{\alpha}_0(\omega)[1-\alpha_0\exp(\eta_0\omega)] = \beta_0[1-\alpha_0^2\exp(2\eta_0\omega)] .$$

Then there exists no $\tilde{\alpha}_1(\omega)$, $(\eta_1 \vdots \alpha_1) \neq (\eta_0 \vdots \alpha_0)$, such that $\tilde{\alpha}_1(\omega)[1-\alpha_0\exp(\eta_0\omega)]$ has the same form as $\tilde{\beta}_0(\omega)$. Thus the equation

$$y(t) + \alpha_0 y(t-\eta_0) = \beta_0[z(t)+\gamma_0 z(t-\zeta_0)] + u(t)$$

is identified when $\eta_0, \zeta_0, \alpha_0, \beta_0, \gamma_0 \neq 0$ and either $\eta_0 \neq \zeta_0$ or else $\eta_0 = \zeta_0$ but $\alpha_0 \neq \gamma_0$.

more than two terms. Of course in practice we do not know $\tilde{\alpha}_0$, $\tilde{\beta}_0$ so we cannot determine whether common factors are present, and all one can do is guard against redundancies in specification.

For (1.1) we again take

$$0 = \eta_{00} < \eta_{01} < \ldots < \eta_{0q} \ , \quad \zeta_{01} < \zeta_{02} < \ldots < \zeta_{0r} \ .$$

When $I_{p_1} + A_0(0)$ possesses no null columns, that is when a different $y_j(t)$ is led in each equation (for example when $I_{p_1} + A_0(0)$ is non-singular) we order the rows and normalise in such a way that the diagonal elements of $A_0(0)$ are all zero; we have not amalgamated I_{p_1} into $A_0(0)$ in order that this constraint be homogeneous. Thus a normalisation and sign constraint has been imposed on each equation; when $I_{p_1} + A_0(0)$ has one or more null columns alternative elements must be constrained. We take $A_0(j) \neq 0$, $j = 1, \ldots, q$, $B_0(j) \neq 0$, $j = 1, \ldots, r$, although possibly $A_0(0) = 0$. The multiple system differs from the scalar case in that conditions are needed to ensure that individual equations may be distinguished from one another, and no equation may be represented as a linear combination of the remainder. We write

$$A_0 = [A_0(0) \vdots \ldots \vdots A_0(q)] \ , \quad B_0 = [B_0(1) \vdots \ldots \vdots B_0(r)] \ .$$

Let us take p elements, $0 < p < p_1(q+1) + p_2 r$, of the first row of $(A_0 \vdots B_0)$ to be zero. We denote by C the $(p_1-1) \times p$ matrix consisting of the columns beneath these zero elements. Then if $p < p_1 - 1$ or if $p \geq p_1 - 1$ but $\text{rank}\{C\} < p_1 - 1$ it is possible to find a linear combination of rows of C that is a null vector, in which case the first equation is not identified. We therefore prescribe at least $p_1 - 1$ elements of each row of $(A_0 \vdots B_0)$ to be zero, and assume the rank condition is satisfied for each row.[52] As in §4 it is likely that we shall wish to take a number of coefficients to be zero in order to avoid certain lag-variable-equation combinations, so the situation may in fact be one of over-identification. We require always of course that our parameters be uniquely defined by knowledge of $\tilde{A}_0(\omega)^{-1}\tilde{B}_0(\omega)$ in such a way that (v) holds. Thus we must consider the possibility of $\tilde{A}_0$ and $\tilde{B}_0$ possessing a common non-singular $p_1 \times p_1$ left-hand divisor having elements that are exponential sums. Such a matrix may correspond to redundancies in specification, as in the scalar case. However the multivariate case is more difficult to deal with because the common divisor may have constant determinant in which case the problem cannot be resolved by eliminating redundancies since such matrices are units of the ring of matrices $\tilde{A}_0$ and divide every element of the ring. Hannan (1971 b) has discussed the identification of systems such as (1.1) when the η_{0j}, ζ_{0j} are all known integers; for just-identification, for example, Hannan shows that redundancies must be eliminated and the null spaces of $A_0(q)'$, $B_0(r)'$ must have null intersection . We shall not attempt a correspondingly precise analysis for our more general situation, merely asserting that (v) holds.

We consider now the solution of (1.1). We may expand $\det\{\tilde{A}_0(\omega)\}$ as

[52] The rank and order conditions may be fulfilled by the presence of more general linear restrictions but for simplicity we confine our attention to zero constraints.

$$\det\{\tilde{A}_0(\omega)\} = \sum_{j_1=0}^{q} \cdots \sum_{j_{p_1}=0}^{q} \det\left\{A\left(j_1, \ldots, j_{p_1}\right)\right\} \exp\left\{\sum_{k=1}^{p_1} \eta_{j_k}\omega\right\},$$

where $A\left(j_1, \ldots, j_{p_1}\right)$ is the $p_1 \times p_1$ matrix whose kth row is the kth row of $I_{p_1} + A_0(0)$, if $j_k = 0$, or $A_0\left(j_k\right)$, if $j_k \geq 1$. For brevity we put $\bar{A} = I_{p_1} + A_0(0)$. Then under the conditions on the η_{0j} the only constant term in the sum is $\det\{\bar{A}\}$. If $\det\{\bar{A}\} \neq 0$ then $\det\{\tilde{A}_0(-\omega)\} = O(1)$, $\sigma \geq 0$, but if $\det\{\bar{A}\} = 0$, $\det\{\tilde{A}_0(-\omega)\} = O\left(\exp(-\theta\omega)\right)$, say, where $\theta \geq \eta_{01}$. In the latter case, unless $\zeta_{01} \geq \theta$, terms in $z(\tau)$ for $\tau > t$ will be present in the solution, although if all zeros of $\exp(\theta\omega)\det\{\tilde{A}_0(-\omega)\}$ have negative real parts there are only a finite number of such terms and the system is stable in our sense. *We shall confine our attention to systems in which all zeros of* $\det\{\tilde{A}_0(-\omega)\}$ *have negative real parts and the matrix* $\bar{A}$ *is non-singular.* In this case we have the Dirichlet expansion

$$\tilde{A}_0(-\omega)^{-1} = \bar{A}^{-1} \sum_{k=0}^{\infty} \left\{- \sum_{j=1}^{q} A_0(j)\bar{A}^{-1}\exp\left(-\eta_{0j}\omega\right)\right\}^k, \quad \sigma \geq 0, \quad (3.6)$$

and the solution

$$y(t) = \bar{A}^{-1} \sum_{j=1}^{r} \left\{B_0(j)z\left(t-\zeta_{0j}\right) + \sum_{k=1}^{\infty} (-1)^k \sum_{j_1} \cdots \sum_{j_k} A_0\left(j_1\right)\bar{A}^{-1} \cdots \right.$$
$$\left. \times A_0\left(j_k\right)\bar{A}^{-1}B_0(j)z\left(t - \eta_{0j_1} - \cdots - \eta_{0j_k} - \zeta_{0j}\right)\right\} + x(t), \quad t \in R,$$

where

$$x(t) = \bar{A}^{-1}\left\{u(t) + \sum_{k=1}^{\infty} (-1)^k \sum_{j_1} \cdots \sum_{j_k} A_0(j_1)\bar{A}^{-1} \cdots \right.$$

$$\left. \times A_0(j_k)\bar{A}^{-1}u\left(t - \eta_{0j_1} - \cdots - \eta_{0j_k}\right)\right\} .$$

Of course even when there are zeros with positive real parts there exists a solution of (1.1) of the form

$$y(t) = \sum_{n=0}^{\infty} K(\mu_n) \sum_{j=1}^{r} B_0(j)z(t+\mu_n-\zeta_{0j}) + x(t) , \qquad (3.7)$$

with

$$x(t) = \sum_{n=0}^{\infty} K(\mu_n)u(t+\mu_n) .$$

By successive use of (1.2,3), (3.7) may be expressed in the form (2.1.1) as

$$y(t) = \left(\beta_0 \otimes I_{p_1}\right) \int_R \left[\sum_{n=0}^{\infty} \delta_n(\tau)' \otimes I_{p_2} \otimes \mathrm{vec}\{K(\mu_n)\}\right] z(t-\tau)d\tau + x(t) ,$$

where

$$\beta = \mathrm{vec}'\{B\} , \quad \delta_n(t) = \left[\delta(t+\mu_n-\zeta_{01}) \vdots \cdots \vdots \delta(t+\mu_n-\zeta_{0r})\right] .$$

However we would naturally wish our estimated system to have a stationary solution of the same form as the true system and while it may be difficult to restrict the admissible parameter set to the stable domain it seems that it will usually be very much harder to derive estimates satisfying alternative pre-specified conditions on the zeros. In any case, of course, the stable case is by far the most important in practice.

We shall now give asymptotic theory for some parameter estimates obtained by transforming and approximating the reduced form of (1.1), following §2.1. We define

$$\eta = (\eta_1 \vdots \cdots \vdots \eta_q) , \quad \zeta = (\zeta_1 \vdots \cdots \vdots \zeta_r)$$

and we allow for restrictions of the form (2.2.1) on $\alpha = \mathrm{vec}'\{A\}$, expressed by $\alpha = \alpha L_1$, and on β, expressed by $\beta = \beta L_2$, where L_1 and L_2 are the appropriate projection matrices.

THEOREM 5.1. *Denote by* $(\hat{\eta}_N \vdots \hat{\zeta}_N \vdots \hat{\alpha}_N \vdots \hat{\beta}_N)$ *the vector minimizing*

$$Q_N = N^{-1} \sum_{B} \left\| \Phi(\lambda_s)^{\frac{1}{2}} \left\{ w_y(s) - \tilde{A}(i\lambda_s)^{-1} \tilde{B}(i\lambda_s) w_z(s) \right\} \right\|^2$$

over admissible $(\eta \vdots \zeta \vdots \alpha \vdots \beta)$ *such that*

$$0 < K_1 \le \eta_1 < \dots < \eta_q \le K_2 \le N, \quad -\tfrac{1}{2}N < K_3 \le \zeta_1 < \dots < \zeta_r \le K_4 \le \tfrac{1}{2}N,$$

$$\alpha \in \Theta \subset \left\{ \alpha \mid |\det\{\tilde{A}(-\omega)\}| \ge K_5 > 0,\ \sigma \ge 0;\ \det\left\{ I_{p_1} + A(0) \right\} \ne 0 \right\}.$$

Let $(\eta_0 \vdots \zeta_0 \vdots \alpha_0)$ *obey these constraints and let* (1.1) *and conditions* (i), (ii), (iv) *and* (v) *hold. Then*

$$\lim_{N\to\infty} (\hat{\eta}_N \vdots \hat{\zeta}_N \vdots \hat{\alpha}_N \vdots \hat{\beta}_N) = (\eta_0 \vdots \zeta_0 \vdots \alpha_0 \vdots \beta_0), \text{ a.s.}$$

Moreover if $\eta_{01} > K_1$, $\eta_{0q} < K_2$, $\zeta_{01} > K_3$, $\zeta_{0r} < K_4$, *if* α_0 *is an interior point of* Θ *and if conditions* (vi) *and* (vii) *also hold, then as* $N \to \infty$

$$N^{\frac{1}{2}} (\hat{\eta}_N - \eta_0 \vdots \hat{\zeta}_N - \zeta_0 \vdots \hat{\alpha}_N - \alpha_0 \vdots \hat{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} N\left[0,\ \Psi_0(\Phi)^{+} \Psi_0(\Phi f_{xx} \Phi) \Psi_0(\Phi)^{+}\right],$$

where

$$\Psi(\Phi) = (2\pi)^{-1} \int_B T(\lambda)^* \left\{ f_{zz}(-\lambda) \otimes \tilde{A}(i\lambda)^{-1*} \Phi(\lambda) \tilde{A}(i\lambda)^{-1} \right\} T(\lambda)\, d\lambda,$$

with

$$T(\lambda) = \left[-\tilde{B}(i\lambda)' \tilde{A}(i\lambda)^{-1'} \otimes I_{p_1} \vdots I_{p_1 p_2} \right] \left[\begin{array}{c:c:c:c} C_1 & 0 & C_3 L_1 & 0 \\ \hdashline 0 & C_2 & 0 & C_4 L_2 \end{array} \right],$$

in which

$$C_1 = i\lambda e(i\eta\lambda) \odot [\mathrm{vec}\{A(1)\} \vdots \dots \vdots \mathrm{vec}\{A(q)\}],$$

$$C_2 = i\lambda e(i\zeta\lambda) \odot [\mathrm{vec}\{B(1)\} \vdots \dots \vdots \mathrm{vec}\{B(r)\}],$$

$$C_3 = [1 \,\vdots\, e(i\eta\lambda)] \otimes I_{p_1^2} ,$$

$$C_4 = e(i\zeta\lambda) \otimes I_{p_1 p_2} .$$

Proof. While the proof substantially follows from that of earlier asymptotic theorems a few comments are in order. In choosing $(\hat{\eta}_N \,\vdots\, \hat{\zeta}_N)$ we take account of the fact that Q_N is periodic of period N in the η_j, ζ_j. The restriction on $\boldsymbol{\alpha}$ to ensure compactness and the stability of the estimated system may be chosen as follows. For the scalar model (3.5), for example, we modify (2.8) by taking Θ to be

$$\left\{[\alpha(1) \,\vdots\, \dots \,\vdots\, \alpha(q)] \in R^q \,\middle|\, \sum_{j=1}^{q} |\alpha(j)| \le 1 - \Delta;\ |\alpha(j)| \ge \Delta,\ j = 1, \dots, q\right\}$$

for arbitrarily small $\Delta > 0$. For the system (1.1) we note from Wilkinson (1965, p. 59) that a sufficient condition for the convergence of the expansion (3.6) is

$$\nu_1\left\{\sum_{j=1}^{q} A_0(j)\bar{A}^{-1}\exp\left(-\eta_{0j}\omega\right)\right\} < 1 , \quad \sigma \ge 0 .$$

By successive use of (1.2.1) with $j = k = 1$ the left side is not less than

$$\sum_{j=1}^{q} \nu_1\left\{A_0(j)\bar{A}^{-1}\exp\left(-\eta_{0j}\omega\right)\right\} \le \sum_{j=1}^{q} \nu_1\left\{A_0(j)\bar{A}^{-1}\right\} , \quad \sigma \ge 0 .$$

Thus a sufficient condition for the stability of (1.1), uniformly in (positive) $\eta_{01}, \dots, \eta_{0q}$ is

$$\sum_{j=1}^{q} \nu_1\left\{A_0(j)\bar{A}^{-1}\right\} < 1 .$$

Then we may form a compact set Θ as before although, of course, the system may be stable with $\boldsymbol{\alpha}_0 \notin \Theta$.

The strong mixing condition is imposed on the reduced form residuals $x(t)$ rather than on the structural form residuals $u(t)$. In general for a representation of the form

$$x(t) = \int_R \Gamma(\tau)u(t-\tau)d\tau \ , \quad t \in R \ ,$$

the fact that $u(t)$ is strong mixing does *not* imply that $x(t)$ is strong mixing, and vice versa. (See Ibragimov and Linnik, 1971, p. 352.) However under (xi) and the usual conditions

$$\underset{N\to\infty}{\text{l.i.m.}} \left[N^{-\frac{1}{2}} \sum_B \text{vec}\{\Phi(\lambda_s)I_{xz}(s)\} - N^{-\frac{1}{2}} \sum_B \text{vec}\{\Phi(\lambda_s)\tilde{\Gamma}(\lambda_s)I_{uz}(s)\}\right] = 0 \ ,$$

where $\tilde{\Gamma}$ is defined as in (2.1.9) and $I_{uz}(s) = w_u(s)w_z(s)^*$, with $w_u(s)$ defined as in (2.1.10). Now if $u(t)$ is strong mixing, and suitable ancillary conditions hold, the second term in the square brackets is asymptotically normal and then so also is the first.

The fact that the covariance matrix is as stated follows by considering derivatives of $\text{vec}\{\tilde{A}(i\lambda)^{-1}\tilde{B}(i\lambda)\}$. The derivatives in ζ and β are easily deduced from §4.2. Also

$$\text{vec}\left\{I_{p_1}\right\} = \text{vec}\{\tilde{A}(i\lambda)^{-1}\tilde{A}(i\lambda)\}$$

$$= \left\{\tilde{A}(i\lambda)' \otimes I_{p_1}\right\}\text{vec}\{\tilde{A}(i\lambda)^{-1}\} \qquad (3.8)$$

$$= \left\{I_{p_1} \otimes \tilde{A}(i\lambda)^{-1}\right\}\text{vec}\{\tilde{A}(i\lambda)\} \ . \qquad (3.9)$$

Then using the rule for product differentiation

$$0 = (\partial/\partial\eta)\text{vec}\left\{I_{p_1}\right\} = \left\{\tilde{A}(i\lambda)' \otimes I_{p_1}\right\}(\partial/\partial\eta)\text{vec}\{\tilde{A}(i\lambda)^{-1}\}$$

$$+ \left\{I_{p_1} \otimes \tilde{A}(i\lambda)^{-1}\right\}i\lambda\left[\exp\left(i\eta_1\lambda\right)\text{vec}\{A(1)\} \vdots \ldots \vdots \exp\left(i\eta_q\lambda\right)\text{vec}\{A(q)\}\right] \ .$$

Rearranging and inserting $\tilde{B}$ we have

$$(\partial/\partial\eta)\text{vec}\{\tilde{A}(i\lambda)^{-1}\tilde{B}(i\lambda)\} = -i\lambda\{\tilde{B}(i\lambda)'\tilde{A}(i\lambda)^{-1'} \otimes \tilde{A}(i\lambda)^{-1}\}C_1 \ .$$

Moreover (3.9) may be written also as

$$\text{vec}\{\tilde{A}(i\lambda)^{-1}\} + \left\{[1 \vdots e(i\eta\lambda)] \otimes I_{p_1} \otimes \tilde{A}(i\lambda)^{-1}\right\}\alpha'$$

whence, using also (3.8),

$$(\partial/\partial\boldsymbol{\alpha})\text{vec}\{\tilde{A}(i\lambda)^{-1}\tilde{B}(i\lambda)\} = -\{\tilde{B}(i\lambda)'\tilde{A}(i\lambda)^{-1'} \otimes \tilde{A}(i\lambda)^{-1}\}C_3L_1 \ .$$

The estimation procedure may proceed in the following way. We can re-write Q_N in the manner of (2.3.6) whence we have, for given $(\eta \vdots \zeta \vdots \alpha)$, the estimate

$$\hat{\beta}_N(\eta,\ \zeta,\ \alpha) = h_N'[L_2H_NL_2]^+ \ , \tag{3.10}$$

where

$$h_N = N^{-1} \sum_B e(i\zeta\lambda_s)^* \otimes \text{vec}\left\{\tilde{A}(i\lambda_s)^{-1*}\Phi(\lambda_s)I_{yz}(s)\right\} ,$$

$$H_N = N^{-1} \sum_B e(i\zeta\lambda_s)^* e(i\zeta\lambda_s) \otimes I_{zz}(-s) \otimes \tilde{A}(i\lambda_s)^{-1*}\Phi(\lambda_s)\tilde{A}(i\lambda_s)^{-1} \ .$$

Thus we would attempt to estimate $(\eta_0 \vdots \zeta_0 \vdots \alpha_0)$ by maximizing $h_N'[L_2H_NL_2]^+h_N$ (which may well be a formidable task) before using (3.10). The results of the theorem remain true, incidentally, if there are homogeneous linear constraints linking A and B in which case the final $p_1^2(q+1) + p_1p_2r$ columns of the second factor in $T(\lambda)$ are

$$\begin{bmatrix} C_3 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & C_4 \end{bmatrix}\begin{bmatrix} L_1 & \vdots & L_3 \\ \cdots & \cdots & \cdots \\ L_3' & \vdots & L_2 \end{bmatrix} ,$$

L_3 being the projection matrix for these constraints. However the two-stage estimation procedure described above needs modification when $L_3 \neq 0$.

When $\Phi = f_{xx}^{-1}$ (assuming (x) holds) we have

$$\tilde{A}_0(i\lambda)^{-1*}\Phi(\lambda)\tilde{A}_0(i\lambda)^{-1} = \tilde{A}_0(i\lambda)^{-1*}\Phi(\lambda)f_{xx}(\lambda)\Phi(\lambda)\tilde{A}_0(i\lambda)^{-1} = f_{uu}(\lambda)^{-1} ,$$

under (xi), since in this case

$$f_{xx}(\lambda) = \tilde{A}_0(i\lambda)^{-1} f_{uu}(\lambda) \tilde{A}_0(i\lambda)^{-1*} .$$

Then the covariance matrix in the limiting distribution is

$$\left[(2\pi)^{-1} \int_B T_0(\lambda)^* \left\{ f_{zz}(-\lambda) \otimes f_{uu}(\lambda)^{-1} \right\} T_0(\lambda) d\lambda \right]^+ .$$

Finally we give a corollary for a simple scalar model, including the cases when the spectral densities of $z(t)$ and $u(t)$ have the same shape and when the true lag is an integer. We shall not give the proof, which is elementary and tedious.

COROLLARY 5.1. *Denote by* $\left(\hat{\eta}_N \vdots \hat{\alpha}_N \vdots \hat{\beta}_N\right)$ *the vector minimizing*

$$Q_N = N^{-1} \sum_B \left| w_y(s) - \beta \left[1 + \alpha \exp\left(i\eta\lambda_s\right)\right]^{-1} w_z(s) \right|^2 \varphi\left(\lambda_s\right) .$$

Let

$$y(t) + \alpha_0 y\left(t - \eta_0\right) = \beta_0 z(t) + u(t) , \quad t \in R , \qquad (3.11)$$

where $\eta_0 > 0$, $0 < |\alpha_0| < 1$ *and* $\beta_0 \neq 0$. *Then the covariance matrix in the limiting distribution of* $N^{\frac{1}{2}}\left(\hat{\eta}_N - \eta_0 \vdots \hat{\alpha}_N - \alpha_0 \vdots \hat{\beta}_N - \beta_0\right)$ *is* $\Psi_0(\varphi)^{-1} \Psi_0\left(f_{xx}\varphi^2\right) \Psi_0(\varphi)^{-1}$ *where*

$$\Psi(\varphi) = \frac{1}{2\pi} \int_B \left[\begin{array}{c:c:c} (\alpha\beta\lambda)^2 & 0 & \alpha\beta\lambda\sin\eta\lambda \\ \hdashline & \beta^2 & -\beta(\alpha+\cos\eta\lambda) \\ \hdashline & & 1+\alpha^2+2\alpha\cos\eta\lambda \end{array} \right] \frac{f_{zz}(\lambda)\varphi(\lambda)d\lambda}{\left(1+\alpha^2+2\alpha\cos\eta\lambda\right)^2} .$$

If moreover condition (xi) *holds and for all* $\lambda \in B = (-\pi, \pi)$

$$f_{zz}(\lambda) = \rho f_{uu}(\lambda) > 0 , \quad \varphi(\lambda) = f_{xx}(\lambda)^{-1} ,$$

the covariance matrix is the inverse of $\Psi_0 = \left(\rho\varphi_{0jk}\right)$, *where*

$$\varphi_{11} = \alpha^2\beta^2\left(1-\alpha^2\right)^{-1}\left\{\pi^2\left(\tfrac{1}{3}+\kappa\right) + 4 \sum_{j=1}^{\infty} (-\alpha)^j \left(j^{-2}\eta^{-2}\cos j\eta\pi - j^{-3}\eta^{-3}\sin j\eta\pi\right)\right\} ;$$

$$\varphi_{12} = 0 ; \quad \varphi_{13} = (\alpha\beta/\eta)\left\{\log\left|1+\alpha e^{i\eta\pi}\right| + \pi^{-1}\sum_{j=1}^{\infty}(-\alpha)^j j^{-2}\eta^{-2}\sin j\eta\pi\right\} ;$$

$$\varphi_{22} = \beta(1-\alpha^2)^{-1}(1+\kappa/\pi) \;;\quad \varphi_{23} = \beta\kappa/2\pi\alpha \;;\quad \varphi_{33} = 1 \;;$$

with

$$\kappa = -i\eta^{-1}\log\{(1+\alpha e^{-i\eta\pi})/(1+\alpha e^{i\eta\pi})\} \;.$$

If η *is an even integer these expressions reduce to*

$$\varphi_{11} = \alpha^2\beta^2(1-\alpha^2)^{-1}\left\{\tfrac{1}{3}\pi^2 + 4\sum_{j=1}^{\infty}(-\alpha)^j j^{-2}\eta^{-2}\right\} \;;\quad \varphi_{12} = 0 \;;$$

$$\varphi_{13} = (\alpha\beta/\eta)\log(1+\alpha);\quad \varphi_{22} = \beta^2(1-\alpha^2)^{-1} \;;\quad \varphi_{23} = 0 \;;\quad \varphi_{33} = 1 \;.$$

If η *is an odd integer we have instead*

$$\varphi_{11} = \alpha^2\beta^2(1-\alpha^2)^{-1}\left\{\tfrac{1}{3}\pi^2 + 4\sum_{j=1}^{\infty}\alpha^j j^{-2}\eta^{-2}\right\} \;;\quad \varphi_{13} = (\alpha\beta/\eta)\log(1-\alpha) \;.$$

5.4 Numerical examples

The scalar model (3.11) was fitted to both our economic and oceanographic data; $y(t)$ and $z(t)$ were assigned in the same way as in §4.3 and will continue to be so assigned in our practical examples of §6.4, §7.5. The statistic

$$\left\{\sum_s I_{zy}(s)\left[1+\alpha\exp(i\eta\lambda_s)\right]^{-1}\varphi(\lambda_s)\right\}^2 \Big/ \sum_s I_{zz}(s)\left|1+\alpha\exp(i\eta\lambda_s)\right|^{-2}\varphi(\lambda_s)$$

was maximized over a subset of the stable domain, namely

$$0.01 \le \eta \le 20.0 \;,\quad 0.01 \le |\alpha| \le 0.99 \;,$$

to give $(\hat{\eta}_N \,\vdots\, \hat{\alpha}_N)$. Again we took $\varphi \equiv 1$ and then, iterating, $\varphi = \hat{f}_{xx}^{-1}$, with $M = 8$ as before. (The estimates for $\hat{f}_{xx}^{-1}$ in the tables below are in fact third iterates in this sense.) We emphasize that the sums for both sets of data, here and in §6.4, §7.5, are over $\lambda_s \in (-\pi, \pi)$. The results were as follows.

Table 5.1						
Economic Data						
φ	$\hat{\eta}_N$	s.e. $\hat{\eta}_N$	$\hat{\alpha}_N$	s.e. $\hat{\alpha}_N$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
1	2.104	0.213	-0.398	0.138	0.456	0.102
f	2.063	0.198	-0.113	0.052	0.783	0.044

Table 5.2						
Oceanographic Data						
φ	$\hat{\eta}_N$	s.e. $\hat{\eta}_N$	$\hat{\alpha}_N$	s.e. $\hat{\alpha}_N$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
1	0.064	0.616	-0.956	0.393	0.038	0.377
f	0.103	0.256	-0.957	0.098	0.037	0.092

The entire covariance matrix was estimated using the first part of Corollary 5.1, but we reproduce only standard errors. The estimated difference for the economic data corresponds to about[53] two months. A difference or differential type of model is perhaps a more appropriate description of the economic data than is (4.1.2) because the introduction of differences or rates of change of money stock to a model with production as exogenous produces a relationship between flow variables. The smallness of the estimated difference (one hour) for the oceanographic data suggests that a differential equation might be more suitable, and we now turn to the problem of estimating such equations.

[53] One disadvantage of our models as compared with regressions and autoregressions involving only integral lags lies in the fact that they cannot be precisely used for prediction purposes on the basis of observed and conjectured future discrete values of the exogenous variables.

CHAPTER 6

LINEAR DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS

6.1 Introduction

In many areas of research there is reason to believe that differential equations provide a suitable description of dynamic relationships between continuous processes. When considering the response of one continuous-time variable to changes in another it seems natural to make explicit reference to rates of change, as well as to the sample paths of the variables themselves. A number of open and closed economic systems have been formulated in terms of differential equations. (See Bergstrom, 1967.) The dynamic Leontief model (Leontief, 1953), for example, relates the supply and demand for commodities. In the theory of servomechanisms differential equations control power by comparing input with output. (See for example James, Nicholls and Phillips, 1947.) Tustin (1957) has provided an interesting comparison between economic and physical differential equation systems. In the biological sciences differential equations may represent the interaction of organisms and in demography, changes in population. The applications in these and other fields are too many and varied to be adequately summarised here, and in any case it is not clear in which fields the results of the present chapter will be most useful. We shall confine our attention here to systems of linear differential equations that relate endogenous variables to exogenous variables, or outputs to inputs in control theory terminology. We discuss the solution and stability of such systems in the present section. In §6.2 we consider their identification and estimation in the presence of data consisting of time series recorded at regular intervals, and give asymptotic theory. In §6.3 we present an alternative "three-stage" method of estimation, using instrumental variables, which does not require iteration and produces consistent estimates on the first stage and efficient estimates on the third. We illustrate these methods by applying a simple model to our data. The author is preparing the results of the present chapter in a form suitable for publication.

In view of the profusion of differential equation models in scientific and social scientific disciplines it is somewhat surprising that the estimation problem has been largely neglected. Of course in control theory, where primarily a practical rather than a descriptive role is being fulfilled, it may be appropriate to use any values which assure a stable solution and fulfil certain other conditions. Also, in scientific fields there may be strong *a priori* information available or data of the type we assume may be unavailable. In the past the estimation problem has been dealt with by substituting an appropriate difference equation with integral differences: Phillips (1959), Durbin (1961) and Bergstrom (1966) have adopted various approaches of this kind towards the identification and estimation of closed systems. Sargan (1968) has considered the estimation of the open first-order system

$$\mathsf{y}(t) - \mathsf{A}\frac{d}{dt}\mathsf{y}(t) = \mathsf{B}\mathsf{z}(t) + \mathsf{u}(t) \ . \tag{1.1}$$

Here A and B are respectively $p_1 \times p_1$ and $p_1 \times p_2$ matrices and $\mathsf{u}(t)$ is a residual vector. In the corresponding discrete-time system the parameters are involved in a highly complicated fashion so Sargan considers an approximate difference system which is easier to handle. Sargan (1968) and Wymer (1972) have discussed the asymptotic properties of estimates of A and B derived from this approximation. The technique we shall adopt here involves a different type of approximation, one which follows naturally from §2.

We consider the system[54]

$$\left[I_{p_1} + A_0(0)\right] y(t) + \sum_{j=1}^{m} A_0(j) \frac{d^j}{dt^j} y(t)$$

$$= B_0(0) z(t) + \sum_{j=1}^{n} B_0(j) \frac{d^j}{dt^j} z(t) + u(t) \ , \quad t \in R \ . \qquad (1.2)$$

The derivatives are assumed to exist in the m.s. sense; thus from (1.3.6) we require that

$$\int_R \lambda^{2m} \mathrm{tr}\{f_{yy}(\lambda)\} d\lambda \ , \quad \int_R \lambda^{2n} \mathrm{tr}\{f_{zz}(\lambda)\} d\lambda \qquad (1.3)$$

be finite. The $A_0(j)$ and $B_0(j)$ are respectively $p_1 \times p_1$ and $p_1 \times p_2$ coefficient matrices and m and n are given. We introduce the operator $D = d/dt$ and the matrix polynomials[55]

$$\tilde{A}(\omega) = I_{p_1} + \sum_{j=0}^{m} A(j)(-\omega)^j \ , \quad \tilde{B}(\omega) = \sum_{j=0}^{n} B(j)(-\omega)^j \ .$$

Then we re-write (1.2) as

$$\tilde{A}_0(-D) y(t) = \tilde{B}_0(-D) z(t) + u(t) \ , \quad t \in R \ .$$

We propose to estimate the $A_0(j)$, $B_0(j)$ subject to various restrictions. In particular we would not, for identification and prior reasons, include all derivatives up to the orders m, n for each $y_k(t)$, $z_k(t)$; we can most easily allow for exclusions by retaining (1.2) but introducing linear constraints as in previous chapters. While we shall concentrate on multiple systems our implied results are new even in respect of simple scalar models such as

$$y(t) + \alpha_0 \frac{d}{dt} y(t) = \beta_0 z(t) + u(t) \ , \quad t \in R \ . \qquad (1.4)$$

[54] The reason for the seeming redundancy in not absorbing I_{p_1} into $A_0(0)$ is the same as in §5.

[55] It is hoped that no confusion will arise from using identical symbols in §§5, 6, 7 for different generating functions.

We shall now show that under suitable conditions the solution of (1.2) has the form (2.1.1). In the m.s. sense, from (1.3.3),

$$D^n y(t) = \int_R e^{it\lambda}(i\lambda)^n d\chi_y(\lambda) \ , \quad t \in R \ .$$

Then in this sense (1.2) is equivalent to

$$\int_R e^{it\lambda}\{\tilde{A}_0(-i\lambda)d\chi_y(\lambda)-\tilde{B}_0(-i\lambda)d\chi_z(\lambda)-d\chi_u(\lambda)\} = 0 \ , \quad t \in R \ .$$

We shall always assume that $\det\{\tilde{A}_0(-\omega)\}$ *has no zeros on the imaginary axis*; in this case the formal solution is

$$y(t) = \int_R e^{it\lambda}\tilde{A}_0(-i\lambda)^{-1}\{\tilde{B}_0(-i\lambda)d\chi_z(\lambda)+d\chi_u(\lambda)\} \ , \quad t \in R \ . \tag{1.5}$$

We confine our attention to circumstances in which the numerator of each element of $\tilde{A}_0(\omega)^{-1}\tilde{B}_0(\omega)$ is of not greater degree than its denominator. This case seems plausible for most practical purposes, excluding the possibility that the output involves derivatives of the input.[56] Then typical elements of both $\tilde{A}_0(-\omega)^{-1}$ and $\tilde{A}_0(-\omega)^{-1}\tilde{B}_0(-\omega)$ have the form

$$\gamma_0 + \sum_{j=1}^{q}\sum_{k=1}^{r_j} \gamma_{jk}\left(\omega-\omega_j\right)^{-k} \ , \tag{1.6}$$

where γ_0 and the γ_{jk} are constants,[57] q is the number of poles ω_j ($Re\{\omega_j\} \neq 0$) and r_j is the multiplicity of ω_j . Now it is easily established by integration by parts that the inverse Fourier transform of $\left(i\lambda-\omega_j\right)^{-k}$ for $Re\{\omega_j\} < 0$ is

$$t^{k-1}\exp\left(\omega_j t\right)/(k-1)! \ , \quad t \geq 0 \ ; \quad 0 \ , \quad t < 0 \ , \tag{1.7}$$

---

[56] In this event, incidentally, the solution of (1.2) is well-defined in the m.s. sense under (ii), the conditions on (1.3) being necessary only for the existence of the structural form.

[57] We allow for an instantaneous response, occurring when $\gamma_0 \neq 0$.

and for $Re\{\omega_j\} > 0$ it is

$$-t^{k-1}\exp(\omega_j t)/(k-1)! \ , \quad t \leq 0 \ ; \quad 0 \ , \quad t > 0 \ . \qquad (1.8)$$

Then (1.6), evaluated at $i\lambda$, is

$$\gamma_0 + \left\{\sum_j{}' \int_0^\infty - \sum_j{}'' \int_{-\infty}^0\right\} \sum_{k=1}^{r_j} t^{k-1}\exp\{(\omega_j - i\lambda)t\}dt/(k-1)! \ , \qquad (1.9)$$

where $\sum'$, $\sum''$ are sums over ω_j having respectively negative and positive real parts. Then evidently we may construct matrices $\Gamma_1(t)$, $\Gamma_2(t)$, whose elements are linear combination of terms (1.7), (1.8), such that (1.5) is

$$y(t) = \int_R \int_R e^{i(t-\tau)\lambda}\{\Gamma_1(\lambda)d\chi_z(\lambda)+\Gamma_2(\lambda)d\chi_u(\lambda)\}d\tau$$

$$= \int_R \Gamma_1(\tau)z(t-\tau)d\tau + x(t) \ , \quad t \in R \ ,$$

where

$$x(t) = \int_R \Gamma_2(\tau)u(t-\tau)d\tau \ , \quad t \in R \ .$$

It is clear from (1.9) that, as we should expect, if all zeros of $\det\{\tilde{A}_0(-\omega)\}$ have negative real parts[58] then $\Gamma_1(t) = 0$, $t < 0$ and $y(t)$ is unrelated to $z(\tau)$, $\tau > t$. If we knew the parameter values we would naturally seek to determine whether a stable system is implied. For (1.4) the condition[59] $\alpha_0 > 0$ is in fact necessary and sufficient for the solution

---

[58] Stability will still be achieved if $\det\{\tilde{A}_0(-\omega)\}$ has a zero with positive real part if this zero is common to the numerator of each element of $\tilde{A}_0(-\omega)^{-1}\tilde{B}_0(-\omega)$. The control theory interpretation of this eventuality provided by James, Nicholls and Phillips (1947) is that "the network has an undamped natural mode but the input terminals are so connected to the filter that no input can excite this type of response". Of course the presence of such a zero constitutes a redundancy in specification and is likely to cause loss of identification.

[59] When $\alpha_0 = 0$, (1.4) is stable in a trivial sense.

$$y(t) = \alpha_0^{-1}\beta_0 \int_0^\infty \exp\left(-\tau/\alpha_0\right)z(t-\tau)d\tau + x(t) \ , \quad t \in R \ .$$

When $\alpha_0 < 0$ the integral is undefined as a m.s. limit for stationary $z(t)$ and (1.4) is described as *resonant*, the stationary solution being

$$y(t) = -\alpha_0^{-1}\beta_0 \int_0^\infty \exp\left(\tau/\alpha_0\right)z(t+\tau)d\tau + x(t) \ , \quad t \in R \ .$$

For (1.1) the necessary and sufficient condition for stability is that all eigenvalues of A should fall to the left of the imaginary axis. For higher-order systems such as (1.2) the actual derivation of the zeros may not be feasible. However a number of techniques for testing stability have been developed, for example the Hurwitz-Routh criterion, the Nyquist criterion and the root-locus method, many descriptions of which are available in the literature. (See e.g. Kaplan, 1962, Marden, 1949.) We merely mention that a necessary condition is that the coefficients of each power of ω in $\det\{\tilde{A}_0(-\omega)\}$ have the same sign. (Kaplan, 1962, p. 403.) Thus in the scalar case each of the coefficients of $y(t)$ and its derivatives must have the same sign. If we believe the true system to be stable then, in practice, by testing possible estimates we may restrict our admissible parameter set to those values which imply stability.

6.2 Identification, estimation and asymptotic theory

The fact that a discrete record only is available again presents an aliasing problem. We shall have relations of the form (5.3.1) and (5.3.2), with $\tilde{A}_0$ and $\tilde{B}_0$ redefined. Under (iv) and (xi), however, we get relations corresponding to (5.3.3) and (5.3.4), which may enable us to identify $\tilde{A}_0$, $\tilde{B}_0$ and f_{uu}. We note that if the spectral mass is in fact restricted to a finite interval, $(-K, K)$ say, the expressions (1.3) are finite under (ii), being bounded by

$$\sup_\lambda \operatorname{tr}\{f_{yy}(\lambda)\}2K^{2m+1}(2m+1)^{-1}\ ,\quad \sup_\lambda \operatorname{tr}\{f_{zz}(\lambda)\}2K^{2n+1}(2n+1)^{-1}$$

respectively, and thus (1.2) is well-defined as a m.s. limit.

Of course after imposing (iv) further conditions will still be needed to identify (1.2). In the first place, we prevent the multiplication of the equations by constant factors by taking the diagonal of $A_0(0)$ to be null. (As in §5.3 this may not be possible, in which case alternative elements are constrained.) We must have $B_0(j) \neq 0$, some j : we cannot consider closed systems in the present circumstances. (For (1.4) we take $\beta_0 \neq 0$.) To ensure the uniqueness, one from the other, of the equations we impose the rank and order conditions, described in §5.3, on $(A_0 \vdots B_0)$, where this matrix is defined by the relations

$$\tilde{A}(\omega) - I_{p_1} = A\left\{1_m(\omega)' \otimes I_{p_1}\right\}\ ,\quad \tilde{B}(\omega) = B\left\{1_n(\omega)' \otimes I_{p_2}\right\}\ ,$$

with

$$1_m(\omega) = \left[1 \vdots -\omega \vdots \dots \vdots (-\omega)^m\right]\ ,\quad 1_n(\omega) = \left[1 \vdots -\omega \vdots \dots \vdots (-\omega)^n\right]\ .$$

As in §5.3 we allow for linear homogeneous constraints expressed by the relations $\alpha = \alpha L_1$, $\beta = \beta L_2$, where $\alpha = \operatorname{vec}'\{A\}$, $\beta = \operatorname{vec}'\{B\}$. We note that from §2 our asymptotic results are valid under suitable conditions when there are non-linear constraints on the parameters. In §6.Appendix we briefly discuss the presence of such constraints on an ordinary system of simultaneous equations, these considerations readily extending to (1.2).

Finally in our treatment of the identification problem we must contend with the possibility of factorization by asserting that except for at most a finite number of points in $\mathcal{B}$,

$\tilde{A}(i\lambda)^{-1}\tilde{B}(i\lambda) \neq \tilde{A}_0(i\lambda)^{-1}\tilde{B}_0(i\lambda)$ uniformly in admissible $(\alpha \vdots \beta) \neq (\alpha_0 \vdots \beta_0)$, such that (v) holds. The identification problem is analogous to that for mixed autoregressive moving-averages with exogenous variables, rigorously discussed by Hannan (1971 b). (See §5.3.)

The first estimates we suggest are essentially those of §2.1.

THEOREM 6.1. *Denote by* $(\hat{\alpha}_N \vdots \hat{\beta}_N)$ *the vector minimizing*

$$Q_N = N^{-1} \sum_B \left\| \Phi(\lambda_s)^{\frac{1}{2}} \left\{ w_y(s) - \tilde{A}(i\lambda_s)^{-1}\tilde{B}(i\lambda_s) w_z(s) \right\} \right\|^2$$

over admissible $(\alpha \vdots \beta)$ *such that*

$$\alpha \in \Theta \subset \{\alpha \mid \det\{\tilde{A}(-\omega)\} \neq 0 \ , \ Re\{\omega\} \geq 0\} .$$

Let $\alpha_0 \in \Theta$ *and let* (1.2) *and conditions* (i), (ii), (iv) *and* (v) *hold. Then*

$$\lim_{N\to\infty} (\hat{\alpha}_N \vdots \hat{\beta}_N) = (\alpha_0 \vdots \beta_0) \ , \text{ a.s.}$$

Moreover if α_0 *is an interior point of* Θ *and if conditions* (vi) *and* (vii) *hold, then as* $N \to \infty$

$$N^{\frac{1}{2}}(\hat{\alpha}_N - \alpha_0 \vdots \hat{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} N\left[0, \ \Psi_0(\Phi)^{+}\Psi_0(\Phi f_{xx}\Phi)\Psi_0(\Phi)^{+}\right] ,$$

where

$$\Psi(\Phi) = (2\pi)^{-1} \int_B T(\lambda)^* \left\{ f_{zz}(-\lambda) \otimes \tilde{A}(i\lambda)^{-1*}\Phi(\lambda)\tilde{A}(i\lambda)^{-1} \right\} T(\lambda) d\lambda \ ,$$

with

$$T(\lambda) = \left[-\left\{ 1_m(i\lambda) \otimes \tilde{B}(i\lambda)'\tilde{A}(i\lambda)^{-1'} \right\} L_1 \vdots \left\{ 1_n(i\lambda) \otimes I_{p_2} \right\} L_2 \right] .$$

Proof. The proof follows directly from Theorems 2.3, 2.6. Unlike in §5.3, stability restrictions do not naturally produce a bounded set that can be compactified. In a somewhat more artificial way a compact subset of the stable region must be isolated: for

(1.4) we may take $0 \le \alpha \le K < \infty$. A two-stage estimation procedure entirely analogous to that described in previous chapters may be devised, estimating α_0 numerically and then β_0 explicitly. It would be tedious to give the details, particularly in view of the fact that in §6.3 we discuss wholly explicit methods of estimation, using the linearity of the structural form.

COROLLARY 6.1. *Denote by* $(\hat{\alpha}_N \vdots \hat{\beta}_N)$ *the vector minimizing*

$$Q_N = N^{-1} \sum_{B} \left| w_y(s) - \beta\left(1-\alpha i\lambda_s\right)^{-1} w_z(s)\right|^2 \varphi(\lambda_s) .$$

Let (1.4) *hold, with* $\alpha_0 \ge 0$ *and* $\beta_0 \ne 0$. *Then the covariance matrix in the limiting distribution of* $N^{\frac{1}{2}}(\hat{\alpha}_N - \alpha_0 \vdots \hat{\beta}_N - \beta_0)$ *is* $\Psi_0(\varphi)^{-1}\Psi_0\left(f_{xx}\varphi^2\right)\Psi_0(\varphi)^{-1}$, *where*

$$\Psi(\varphi) = \frac{1}{2\pi}\int_B \begin{bmatrix} \beta^2 & -\alpha\beta \\ -\alpha\beta & \lambda^{-2}+\alpha^2 \end{bmatrix} \left(\frac{\lambda}{1+\alpha^2\lambda^2}\right)^2 f_{zz}(\lambda)\varphi(\lambda)d\lambda .$$

When $\alpha_0 = 0$, $N^{\frac{1}{2}}\hat{\alpha}_N$ *and* $N^{\frac{1}{2}}(\hat{\beta}_N - \beta_0)$ *are asymptotically independent with the same variances as* $N^{\frac{1}{2}}(\hat{\zeta}_N - \zeta_0)$, $N^{\frac{1}{2}}(\hat{\beta}_N - \beta_0)$, *respectively, of Corollary* 4.1. *On the other hand if condition* (xi) *holds and for all* $\lambda \in B = (-\pi, \pi)$

$$f_{zz}(\lambda) = \rho f_{uu}(\lambda) > 0 , \quad \varphi(\lambda) = f_{xx}(\lambda)^{-1} ,$$

the covariance matrix is

$$\frac{1}{\rho\kappa}\begin{bmatrix} \alpha_0^2/\beta_0^2(1-\kappa) & \alpha_0/\beta_0 \\ \alpha_0/\beta_0 & 1 \end{bmatrix} ,$$

where

$$\kappa = \left(2\pi\alpha_0\right)^{-1}\tan^{-1}\left\{2\pi\alpha_0/\left(1-\alpha_0^2\pi^2\right)\right\} .$$

We conclude the present section by noting that our results are easily extended to circumstances in which the parameters are involved in a non-linear fashion in the structural form. For example, for the model

$$1 + \alpha_0 Dy(t) + \dots + \alpha_0^m D^m y(t) = \beta_0 z(t) + u(t) \ , \quad t \in R \ ,$$

we would estimate

$$w_y(s) = \left\{1-\left(-\alpha_0 i\lambda_s\right)^{m+1}\right\}^{-1}\left(1+\alpha_0 i\lambda_s\right)w_z(s) + w_x(s) \ .$$

The zeros of the characteristic equation are $\alpha_0^{-1}\exp\{2\pi i j/(m+1)\}$, $j = 1, \dots, m$. For $m = 1, 2$ the model is stable when $\alpha_0 > 0$. However, the cases $m = 3$, mod (4) are inadmissible since two zeros are purely imaginary and for $m > 3$ there are always zeros with positive real parts. The model

$$\left(1+\alpha_0 D\right)^m y(t) = \beta_0 z(t) + u(t) \ , \quad t \in R \ ,$$

on the contrary, is stable uniformly in $\alpha_0 \geq 0$ and $m \geq 1$.

6.3 Instrumental variables estimation

Because the structural form (1.2) is linear it is natural to examine the possibility of deriving explicit estimates with good asymptotic properties. For simplicity of exposition, and because we shall use data to estimate (1.4), we shall examine the problem principally with respect to the elementary scalar model (1.4). We shall, however, state analogous results for (1.2).

We approximate the discrete Fourier transform of (1.4) by

$$w_y(s) = \alpha_0 i\lambda_s w_y(s) + \beta_0 w_z(s) + w_u(s) \ . \qquad (3.1)$$

The most obvious estimation procedure is ordinary LS , which produces the estimates

$$\left(\bar{\alpha}_N \;\vdots\; \bar{\beta}_N\right) = \left[0 \;\vdots\; \sum_B I_{yz}\right]\left\{\sum_B \begin{bmatrix} \lambda_s^2 I_{yy} & \vdots & i\lambda_s I_{yz} \\ \cdots\cdots & \vdots & \cdots\cdots \\ -i\lambda_s I_{zy} & \vdots & I_{zz}\end{bmatrix}\right\}^{-1},$$

omitting arguments s . Defining

$$g = \left[(2\pi)^{-1}\int_B \lambda^2 f_{xx}(\lambda)d\lambda \;\vdots\; 0\right],$$

$$G = \frac{1}{2\pi}\int_B \begin{bmatrix} \beta_0^2 & \vdots & -\alpha_0\beta_0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ -\alpha_0\beta_0 & \vdots & \lambda^{-2}+\alpha_0^2\end{bmatrix} \frac{\lambda^2 f_{zz}(\lambda)}{1+\alpha_0^2\lambda^2}\,d\lambda + \begin{bmatrix} g \\ \cdots \\ 0\end{bmatrix},$$

then under (ii) and (iv), a.s.,

$$\lim_{N\to\infty}\left(\bar{\alpha}_N \;\vdots\; \bar{\beta}_N\right) = \left[0 \;\vdots\; \beta_0(2\pi)^{-1}\int_B f_{zz}(\lambda)\left(1-\alpha_0 i\lambda\right)^{-1}d\lambda\right]G^{-1}$$

$$= \left[(\alpha_0 \;\vdots\; \beta_0)G-\alpha_0 g\right]G^{-1}$$

$$= \left(\alpha_0 \;\vdots\; \beta_0\right) - \alpha_0 gG^{-1}.$$

As we should expect, $\left(\bar{\alpha}_N \;\vdots\; \bar{\beta}_N\right)$ is not consistent[60] (unless $\alpha_0 = 0$). In general, by way of analogy with ordinary simultaneous equations systems, estimates of the parameters in the systems of §§5-7 derived by minimizing squared deviations from the approximate *structural* form will not be consistent. In the present case α_0 is easily shown to be (asymptotically) under- or over-estimated depending upon whether $\alpha_0 > 0$ or $\alpha_0 < 0$ respectively, and (unless $\alpha_0 = 0$) β_0 is under- or over-estimated depending upon whether $\beta_0 > 0$ or $\beta_0 < 0$ respectively.

[60] *Cf.* ordinary *LS* estimates for the corresponding difference equation,

$$y(n) + \alpha_0 y(n-1) = \beta_0 z(n) + u(n),$$

which are well known to be consistent only when the $u(n)$ are serially independent.

Consistent estimation thus requires a more sophisticated approach. Let us suppose we have available N observations $v(n)$, $n = 1, \ldots, N$, sampled from a process $v(t)$, $t \in R$. *We require the instrumental variable* $v(t)$ *to be incoherent of* $u(\tau)$, *all* $t, \tau \in R$ *and to satisfy conditions* (ii) *and* (iv) *on* $z(t)$. Indeed $z(t)$ could itself serve as such an instrumental variable, although for reasons of efficiency we would like $v(t)$ to be strongly coherent with $y(t)$. Now post-multiplying (3.1) by

$$\left[-i\lambda_s w_v(-s) \ \vdots \ w_z(-s)\right]$$

(with w_v defined by (2.1.10)) and summing, we have the estimates

$$(\tilde{\alpha}_N \ \vdots \ \tilde{\beta}_N) = \tilde{g}_N \tilde{G}_N^{-1}, \qquad (3.2)$$

where

$$\tilde{g}_N = N^{-1} \sum_B \left[-i\lambda_s I_{yv} \ \vdots \ I_{yz}\right], \quad \tilde{G}_N = N^{-1} \sum_B \begin{bmatrix} \lambda_s^2 I_{yv} & i\lambda_s I_{yz} \\ -i\lambda_s I_{zv} & I_{zz} \end{bmatrix}$$

with $I_{yv} = w_u \overline{w}_v$, $I_{zv} = w_z \overline{w}_v$. (There is no point in including a weight function φ, since we merely require our estimates to be consistent.) Then

$$(\tilde{\alpha}_N - \alpha_0 \ \vdots \ \tilde{\beta}_N - \beta_0) = \left[\tilde{g}_N - (\alpha_0 \ \vdots \ \beta_0)\tilde{G}_N\right]\tilde{G}_N^{-1}.$$

The first factor is

$$N^{-1} \sum_B \left[w_y(s) - \alpha_0 i\lambda_s w_y(s) - \beta_0 w_z(s)\right]\left[-i\lambda_s w_v(-s) \ \vdots \ w_z(-s)\right]. \qquad (3.3)$$

Since the first factor of the summand is close to $w_u(s)$ and $v(t)$, $z(t)$ are each incoherent of $u(t)$, (3.3) vanishes a.s. as $N \to \infty$. Then since $\tilde{G}_N = O(1)$

$$\lim_{N\to\infty} (\tilde{\alpha}_N \ \vdots \ \tilde{\beta}_N) = (\alpha_0 \ \vdots \ \beta_0), \text{ a.s.},$$

under the usual conditions.[61] Now after multiplying (3.3) by $N^{\frac{1}{2}}$ it follows from Theorem 2.5 that the effect of replacing the first factor of the summand by $w_u(s)$ is asymptotically negligible in m.s. Then if $u(t)$ obeys also condition (vi) on $x(t)$,

$$N^{\frac{1}{2}}\left[\tilde{g}_N-(\alpha_0 \vdots \beta_0)\tilde{G}_N\right] \xrightarrow{\mathcal{D}} N[0, \Psi_1] ,$$

where

$$\Psi_1 = \frac{1}{2\pi}\int_B \begin{bmatrix} \lambda^2 f_{vv}(\lambda) & \vdots & i\lambda f_{vz}(\lambda) \\ \cdots\cdots & \vdots & \cdots\cdots \\ -i\lambda f_{zv}(\lambda) & \vdots & f_{zz}(\lambda) \end{bmatrix} f_{uu}(\lambda)d\lambda . \qquad (3.4)$$

Then

$$N^{\frac{1}{2}}(\tilde{\alpha}_N-\alpha_0 \vdots \tilde{\beta}_N-\beta_0) \xrightarrow{\mathcal{D}} N\left[0, \Psi_2^{-1}\Psi_1\Psi_2^{-1}\right] ,$$

providing

$$\Psi_2 = \lim_{N\to\infty} \tilde{G}_N = \frac{1}{2\pi}\int_B \begin{bmatrix} i\lambda(1-\alpha_0 i\lambda)^{-1} \\ \cdots\cdots \\ 1 \end{bmatrix}\left[-i\lambda f_{zv}(\lambda) \vdots f_{zz}(\lambda)\right]d\lambda \quad (3.5)$$

is non-singular.

While $(\tilde{\alpha}_N \vdots \tilde{\beta}_N)$ meets the prime requirement of consistency we should like estimates that are more efficient. To find these we must first estimate $f_{uu}(\lambda)$. We consider

$$f_{uu}(\lambda) = 2MN^{-1} \sum_{s(m)} |w_y(s)-\tilde{\alpha}_N i\lambda_s w_y(s)-\tilde{\beta}_N w_z(s)|^2 , \qquad (3.6)$$

$\lambda \in S_m$, $\omega_m \in B$, in the notation of §2.5. Then in much the same way as in §2.5, for some sequence $M = M(N)$ increasing with N , $\hat{f}_{uu}(\lambda) \to f_{uu}(\lambda)$ a.s. and uniformly. We now post-multiply (3.1) by $\hat{h}_N(\omega_m)w_z(-s)$, where

$$\hat{h}_N(\omega_m) = \left[-\tilde{\beta}_N i\omega_m(1+\tilde{\alpha}_N i\omega_m)^{-1} \vdots 1\right]\hat{f}_{uu}(\omega_m)^{-1} .$$

---

[61] However no formal requirements of compactness are needed in the present section.

Then we consider the estimates

$$(\hat{\alpha}_N \vdots \hat{\beta}_N) = \hat{g}_N \hat{G}_N^{-1} , \tag{3.7}$$

where

$$\hat{g}_N = N^{-1} \sum_m \sum_{s(m)} I_{yz}(s) \hat{h}_N(\omega_m) , \quad \hat{G}_N = N^{-1} \sum_m \sum_{s(m)} \begin{bmatrix} i\lambda_s I_{yz}(s) \\ \cdots\cdots\cdots \\ I_{zz}(s) \end{bmatrix} \hat{h}_N(\omega_m) .$$

We have

$$(\hat{\alpha}_N - \alpha_0 \vdots \hat{\beta}_N - \beta_0) = [\hat{g}_N - (\alpha_0 \vdots \beta_0)\hat{G}_N]\hat{G}_N^{-1} .$$

Then there exists an increasing sequence $M(N)$ such that the first factor (which is similar to (3.3)) vanishes asymptotically a.s. and

$$\lim_{N\to\infty} (\hat{\alpha}_N \vdots \hat{\beta}_N) = (\alpha_0 \vdots \beta_0) , \text{ a.s.}$$

Now $N^{\frac{1}{2}}[\hat{g}_N - (\alpha_0 \vdots \beta_0)\hat{G}_N]$ may be expressed as the sum of a finite number (for fixed M) of the terms

$$\left\{N^{-\frac{1}{2}} \sum_{s(m)} [w_y(s) - \alpha_0 i\lambda_s w_y(s) - \beta_0 w_z(s)] w_z(-s)\right\} \hat{h}_N(\omega_m) .$$

Then as before the effect of replacing the term in square braces by $w_u(s)$ is asymptotically negligible in m.s. Also, keeping M fixed,

$$\lim_{N\to\infty} \hat{h}_N(\omega_m) = \left[-\beta_0 i\omega_m(1+\alpha_0 i\omega_m)^{-1} \vdots 1\right] f_M(\omega_m)^{-1} = h_M(\omega_m) ,$$

say. Then for fixed M we may replace $\hat{h}_N$ by h_M and

$$N^{\frac{1}{2}}[\hat{g}_N - (\alpha_0 \vdots \beta_0)\hat{G}_N] \xrightarrow{\mathcal{D}} N\left[0, \sum_m h_M(\omega_m)^* (2\pi)^{-1} \int_{S_m} f_{zz}(\lambda) f_{uu}(\lambda) d\lambda h_M(\omega_m)\right] .$$

Then on increasing M it follows that there exists some sequence $M(N)$ such that

$$N^{\frac{1}{2}}(\hat{\alpha}_N - \alpha_0 \vdots \hat{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} N[0, \Psi^{-1}] ,$$

where, for $f_{uu}(\lambda) > 0$,

$$\Psi = \frac{1}{2\pi} \int_B \left[\begin{array}{c:c} \beta_0^2 & -\alpha_0\beta_0 \\ \hdashline -\alpha_0\beta_0 & \lambda^{-2}+\alpha_0^2 \end{array}\right] \left(\frac{\lambda^2}{1+\alpha_0^2\lambda^2}\right) \frac{f_{zz}(\lambda)}{f_{uu}(\lambda)}\, d\lambda \ . \tag{3.8}$$

We have thus established a three-stage efficient estimation procedure for (1.4) that does not require iteration. Corresponding methods are of course available for the multiple system (1.2). We shall present the estimates and their asymptotic properties without further proof since the proof is essentially the same as that above but more tedious. For some $p_1 \times 1$ vector instrumental variable process $\mathsf{v}(t)$, incoherent of $\mathsf{u}(\tau)$ for all $t, \tau \in R$, we have the estimates

$$(\tilde{\alpha}_N \vdots \tilde{\beta}_N) = \sum_B [g_{1s} \vdots g_{2s}] \left\{ L \left(\sum_B \left[\begin{array}{c:c} G_{11s} & G_{12s} \\ \hdashline G_{21s} & G_{22s} \end{array}\right] \otimes I_{p_1} \right) L \right\}^+ ,$$

where

$$g_{1s} = -\, 1_m(i\lambda_s) \otimes \mathrm{vec}'\{I_{yv}(-s)\} \ , \qquad g_{2s} = 1_n(i\lambda_s) \otimes \mathrm{vec}'\{I_{yz}(-s)\} \ ,$$

$$G_{11s} = 1_m(i\lambda_s)^* 1_m(i\lambda_s) \otimes I_{yv}(-s) \ , \quad G_{12s} = -1_m(i\lambda_s)^* 1_n(i\lambda_s) \otimes I_{yz}(-s) \ ,$$

$$G_{21s} = -\, 1_n(i\lambda_s)^* 1_m(i\lambda_s) \otimes I_{zv}(-s) \ , \quad G_{22s} = 1_n(i\lambda_s)^* 1_n(i\lambda_s) \otimes I_{zz}(-s) \ .$$

Then

$$N^{\frac{1}{2}}(\tilde{\alpha}_N - \alpha_0 \vdots \tilde{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} N\left[0, \ (L\Psi_2 L)^+ \Psi_1 (L\Psi_2 L)^+\right] ,$$

where

$$\Psi_1 = \frac{1}{2\pi} \int_B \left[\begin{array}{c:c} G_{11}(\lambda) & G_{12}(\lambda) \\ \hdashline G_{21}(\lambda) & G_{22}(\lambda) \end{array}\right] \otimes f_{uu}(\lambda) d\lambda \ ,$$

with

$$G_{11}(\lambda) = 1_m(i\lambda)^* 1_m(i\lambda) \otimes f_{vv}(-\lambda) \ , \quad G_{12}(\lambda) = -1_m(i\lambda)^* 1_n(i\lambda) \otimes f_{vz}(-\lambda) \ ,$$

$$G_{21}(\lambda) = -1_n(i\lambda)^* 1_m(i\lambda) \otimes f_{zv}(-\lambda) \ , \quad G_{22}(\lambda) = 1_n(i\lambda)^* 1_n(i\lambda) \otimes f_{zz}(-\lambda) \ ,$$

and where

$$\Psi_2 = (2\pi)^{-1} \int_B \left\{G_1(\lambda)^* G_2(\lambda) \otimes I_{p_1}\right\} d\lambda \ ,$$

with

$$G_1(\lambda) = \left[-1_m(i\lambda) \otimes \tilde{B}_0(i\lambda)'\tilde{A}_0(i\lambda)^{-1'} \ \vdots \ 1_n(i\lambda) \otimes I_{p_2}\right] ,$$

$$G_2(\lambda) = \left[-1_m(i\lambda) \otimes f_{zv}(-\lambda) \ \vdots \ 1_n(i\lambda) \otimes f_{zz}(-\lambda)\right] .$$

Next we form

$$\hat{f}_{uu}(\omega_m) = 2MN^{-1} \sum_{s(m)} \left[\tilde{A}_N(i\lambda_s)w_y(s) - \tilde{B}_N(i\lambda_s)w_z(s)\right]\left[\tilde{A}_N(i\lambda_s)w_y(s) - \tilde{B}_N(i\lambda_s)w_z(s)\right]^* ,$$

$\omega_m \in B$, where $\tilde{A}_N$ and $\tilde{B}_N$ are generating functions formed by replacing $(\alpha_0 \ \vdots \ \beta_0)$ by $(\tilde{\alpha}_N \ \vdots \ \tilde{\beta}_N)$ in $\tilde{A}_0$ and $\tilde{B}_0$ respectively. Then we have the estimates

$$(\hat{\alpha}_N \ \vdots \ \hat{\beta}_N) = \sum_m \sum_{s(m)} \mathrm{vec}'\{I_{yz}(-s)\}H_N(\omega_m)\left[L \sum_m \sum_{s(m)} G_N(\lambda_s)^* H_N(\omega_m) L\right]^+ ,$$

where

$$G_N(\lambda_s) = \left[-1_m(i\lambda_s) \otimes I_{zy}(-s) \ \vdots \ 1_n(i\lambda_s) \otimes I_{zz}(-s)\right] \otimes I_{p_1} ,$$

$$H_N(\lambda) = \left[-1_m(i\lambda) \otimes \tilde{B}_N(i\lambda)'\tilde{A}_N(i\lambda)^{-1'} \ \vdots \ -1_n(i\lambda) \otimes I_{p_2}\right] \otimes \hat{f}_{uu}(\lambda)^{-1} .$$

Then for some sequence $M = M(N)$,

$$N^{\frac{1}{2}}(\hat{\alpha}_N - \alpha_0 \ \vdots \ \hat{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} N\left[0, \left\{(2\pi)^{-1} \int_B \underset{\circ}{T}(\lambda)^* \left\{f_{zz}(-\lambda) \otimes f_{uu}(\lambda)^{-1}\right\} \underset{\circ}{T}(\lambda) d\lambda\right\}^+\right]$$

in the notation of §6.2.

6.4 Numerical examples

The alternative methods of estimation suggested in the present chapter were compared in estimating (1.4) using our economic and oceanographic data. The procedures of §6.2 based on the reduced form involved finding $\hat{\alpha}_N$ to maximize

$$\left\{\sum_s I_{zy}(s)\left(1-\alpha i\lambda_s\right)^{-1}\varphi\left(\lambda_s\right)\right\}^2 \Big/ \sum_s I_{zz}(s)\left(1+\alpha^2\lambda_s^2\right)^{-1}\varphi\left(\lambda_s\right) , \quad (4.1)$$

with $\hat{\beta}_N$ given by

$$\hat{\beta}_N = \sum_s I_{zy}(s)\left(1-\hat{\alpha}_N i\lambda_s\right)^{-1}\varphi\left(\lambda_s\right) \Big/ \sum_s I_{zz}(s)\left[1+\hat{\alpha}_N^2\lambda_s^2\right]^{-1}\varphi\left(\lambda_s\right) . \quad (4.2)$$

The optimization of (4.1) was carried out over the stable region $0 \le \alpha \le 5.0$. The estimates, for both sets of data, were derived using $\varphi \equiv 1$ and $\varphi = f_{xx}^{-1}$ as before, and appear in lines (a) and (b) respectively in Tables 6.1, 6.2 below, along with their standard errors, computed using the formulae given in the first part of Corollary 6.1.

Next the methods of §6.3 were used, with $z(n)$ taken as the instrumental variable. The initial, consistent, estimates (3.2) are then given by

$$\hat{\alpha}_N = -\sum_s i\lambda_s I_{yz}(s) \Big/ \sum_s \lambda_s^2 I_{yz}(s)$$

$$\hat{\beta}_N = \left\{\left[\sum_s i\lambda_s I_{yz}(s)\right]^2 + \sum_s I_{yz}(s) \sum_s \lambda_s^2 I_{yz}(s)\right\} \Big/ \sum_s \lambda_s^2 I_{yz}(s) \sum_s I_{zz}(s) .$$

These estimates are presented in line (c) in our tables, together with their standard errors which were computed using (3.4) and (3.5). Then $f_{uu}(\lambda)$ was estimated from (3.6), with $M = 8$. The efficient estimates (3.7) were finally computed exactly in the manner described in §6.3. Although the estimates are explicit iteration was carried out in respect of the estimation of $f_{uu}(\lambda)$, and we reproduce estimates and standard errors (derived from (3.8)) for the first and fourth iterates in lines (d) and (e) below.

Table 6.1				
Economic Data				
	$\hat{\alpha}_N$	s.e. $\hat{\alpha}_N$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
(a)	1.035	0.443	0.751	0.059
(b)	0.009	0.022	0.863	0.037
(c)	0.295	0.091	0.743	0.066
(d)	0.136	0.055	0.772	0.044
(e)	0.047	0.014	0.811	0.039

Table 6.2				
Oceanographic Data				
	$\hat{\alpha}_N$	s.e. $\hat{\alpha}_N$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
(a)	1.509	0.286	0.957	0.059
(b)	2.443	0.345	0.965	0.058
(c)	1.746	0.429	0.953	0.018
(d)	2.673	0.426	0.928	0.017
(e)	2.504	0.419	0.938	0.018

There is substantial discrepancy between the estimates of lines (c) and (e) and for the economic data at least the efficiency of the instrumental variables estimates of α_0 is poor. Although lines (d) and (e) agree well in Table 6.2 the difference between the corresponding lines in Table 6.1 indicates that it is wise to iterate with respect to the estimation of f_{uu}, to overcome the effect of a possibly inaccurate first-stage estimate. As we would wish, there is reasonable agreement between lines (b) and (e); this would have been improved by further iterations although even after iteration these procedures will still in general produce different estimates. For the simple model (1.4) the methods of §6.3 may not save very much computational labour but for systems with more parameters these explicit estimates will be very much faster than a multi-dimensional search and iteration.

6. Appendix. Linear systems with non-linear constraints

Whereas non-linear parameter constraints were allowed for in §2 and employed in §3 we have since ignored the possibility of their presence. It seems that non-linear restrictions will arise with much less frequency in connection with linear models than will linear restrictions. However, Malinvaud (1970 a, b) has considered a class of non-linear models that may be re-defined as linear models having non-linear constraints. Moreover Fisher (1966) suggests two ways in which such constraints may arise: namely when the linear model is a Taylor series approximation to a non-linear one and when a particular structure is imposed on a distributed lag model. The model of §3 may be formulated as a linear regression with non-linear constraints (on the cofactors). Also an alternative normalisation condition is sometimes used for linear systems, restricting the parameter point to, say, the surface of the unit hypersphere. Finally non-linear restrictions may arise from *a priori* knowledge. We shall briefly consider the effect of such constraints on the familiar econometric system

$$A_0 y(t) + B_0 z(t) = u(t) , \quad t \in R , \tag{A.1}$$

where A_0 and B_0 are respectively $p_1 \times p_1$ and $p_1 \times p_2$ matrices. It will be seen that much of our discussion is relevant to the more elaborate linear systems considered in the thesis.

When the only restrictions on A and B are linear (A.1) is either identified or there is a continuum of alternative structures, reflecting the nature of solutions of linear simultaneous equations. However, as noted in §2.2, when non-linear restrictions are involved it is possible that, whereas these may be insufficient to provide *global* identification, the system is *locally* identified. Then, by restricting attention to a sufficiently small neighbourhood Θ of one of the alternatives, global identification is achieved. Such a Θ could conceivable arise from inequalities on the parameters, which are common in economics. Wegge (1965) and Rothenberg (1971) have considered the local identification of A_0 and B_0[62]. Writing $\alpha = \text{vec}'\{A\}$, $\beta = \text{vec}'\{B\}$ we introduce the row vector of constraints, $\xi(\alpha, \beta) = 0$. Then[63] if $(\alpha_0 \,\vdots\, \beta_0)$ is a regular point of

$$(\partial/\partial\alpha)\xi(\alpha, \beta)'\left(I_{p_1} \otimes A'\right) + (\partial/\partial\beta)\xi(\alpha, \beta)'\left(I_{p_1} \otimes B'\right) ,$$

a necessary and sufficient condition for local identifiability is that this matrix have rank p_1^2 at $(\alpha_0 \,\vdots\, \beta_0)$.

We consider now some estimates of $(A_0 \,\vdots\, B_0)$ and their asymptotic properties. Frequently economic models involve identities in the variables, and it seems worthwhile to allow for such contingencies here. (See Malinvaud, 1970 b, pp. 646-648.) We re-write (A.1) as

$$\begin{bmatrix} A_{01} & \vdots & A_{02} \\ \cdots & \cdots & \cdots \\ A_{03} & \vdots & A_{04} \end{bmatrix} \begin{bmatrix} y_1(t) \\ \cdots \\ y_2(t) \end{bmatrix} + \begin{bmatrix} B_{01} \\ \cdots \\ B_{02} \end{bmatrix} z(t) = \begin{bmatrix} u_1(t) \\ \cdots \\ 0 \end{bmatrix} , \quad t \in R .$$

Here all partitioning is $p_4 : p_1 - p_4$, $p_4 > 0$. The identities are expressed by the lower partition, with $[A_{03} \,\vdots\, A_{04} \,\vdots\, B_{02}]$ a known matrix. We assume the identities to be linearly independent and the elements of $y(t)$ to have been alloted to $y_1(t)$ and $y_2(t)$

---

[62] These authors also consider the identification of $E\{u(t)u(t)'\}$.

[63] This is an extension of results of Fisher (1959), (1966) for the identification of a single equation in the system.

in such a way that A_{04} is non-singular. On substituting for $y_2(t)$ in the upper partition we have

$$\bar{A}_0 y_1(t) + \bar{B}_0 z(t) = u_1(t) ,$$

where

$$\bar{A} = A_1 - A_2 A_{04}^{-1} A_{03} , \quad \bar{B} = B_1 - A_2\, A_{04}^{-1} B_{02} .$$

We write

$$\alpha_1 = \mathrm{vec}'\{A_1\} , \quad \alpha_2 = \mathrm{vec}'\{A_2\} , \quad \beta = \mathrm{vec}'\{B_1\} ,$$

and introduce the $1 \times r$ constraint vector

$$\xi(\alpha_1, \alpha_2, \beta) = 0 , \tag{A.2}$$

requiring its first derivatives,

$$(\partial/\partial\alpha_1 \vdots \partial/\partial\alpha_2 \vdots \partial/\partial\beta)\xi(\alpha_1, \alpha_2, \beta)' = \Xi ,$$

to be continuous and linearly independent. We take $(\hat{\alpha}_{N1} \vdots \hat{\alpha}_{N2} \vdots \hat{\beta}_N)$ to minimize[64]

$$Q_N = N^{-1} \sum_{B} \left\| \Phi(\lambda_s)^{\frac{1}{2}} \left\{ w_1(s) - \bar{A}^{-1}\bar{B} w_z(s) \right\} \right\|^2$$

over $(\alpha_1 \vdots \alpha_2 \vdots \beta) \in \Theta$, a compact subset of the admissible set satisfying (A.2). When all restrictions are linear an efficient method of estimation for the time series case is given by Hannan and Terrell (1973). Of course simpler and longer established methods exist for finding estimates that are at least consistent. (See for example Malinvaud, 1970 b, Chapter 19.) When non-linear restrictions are present such methods require modification, and iteration will certainly be needed. The modification of Newton's method suggested by Aitchison and Silvey (1958) may be useful.

It would be repetitive to state a formal theorem but from

---

[64] $w_1(s)$ is the discrete Fourier transform of $y_1(n)$ and Φ is now $p_4 \times p_4$.

Theorems 2.3 and 2.6, under the usual conditions[65], our estimates are consistent and as $N \to \infty$,

$$N^{\frac{1}{2}}\left(\hat{\alpha}_{N1}-\alpha_{01} \vdots \hat{\alpha}_{N2}-\alpha_{02} \vdots \hat{\beta}_{N}-\beta_{0}\right) \xrightarrow{\mathcal{D}} N[0, \Sigma] ,$$

where Σ is derived by deleting the final r rows and columns from a matrix of the form (3.2.7), with $\Psi(\Phi)$ given by (3.2.4) and

$$T = \left[-\bar{B}'\bar{A}^{-1\prime} \vdots \left(A_{04}^{-1}A_{03}\bar{A}^{-1}\bar{B}\right)' \vdots I_{p_2}\right] \otimes \bar{A}^{-1} .$$

[65] We do not require (iv) and (vii), however.

CHAPTER 7

DIFFERENCE-DIFFERENTIAL AND INTEGRO-DIFFERENTIAL EQUATIONS

7.1 Introduction

The previous two chapters have been devoted to the estimation of simple systems of linear functional equations that can be solved by the Fourier transform. Difference and differential equations apparently provide reasonable descriptions of a wide variety of dynamic situations in the real world. However prior evidence may call for a system of greater sophistication. In particular it may well be the case that the rate of change of a process depends not only on its present state but also on its past history. Thus we are led to examine differential equations that involve lagged variables and their derivatives and which are most commonly known as difference-differential equations. Bernoulli (1728) was apparently the first to consider such an equation, in connection with a problem in mechanics. Subsequently difference-differential equations have frequently arisen in such fields as the physical and biological sciences, engineering and economics, sometimes as approximations to partial differential equations and other functional equations. In recent years there has accumulated a vast literature concerning the solution and stability of difference-differential equations.

The theory of §2 allows us to estimate linear difference-differential equations having constant coefficients. In addition it allows us to estimate difference-kernel integral equations involving a continuum of lagged endogenous variables. We shall consider the estimation of a composite linear difference-differential and integro-differential equation system of a type similar to those discussed in different contexts by, for example, Kitagawa (1937) and Pinney (1958), involving the linear operations of differentiation, integration and translation through time. As far as is known, no previous attention has been directed to the estimation of such systems, in the presence of a discrete sample at least. Before we describe our estimates and their asymptotic properties we shall attempt to provide some background and motivation by considering the solution and stability of some simple difference-differential and integral equations that occur in practice. The fact that generally solutions are of a recondite nature, and the assessment of stability of a given system is very difficult, will limit the usefulness of our results.

7.2 Difference-differential equations

The functional equation we consider is

$$\sum_{j=0}^{m} \sum_{k=0}^{q} \alpha(j,\, k) D^j y\left(t-\eta_k\right) = w(t) \ , \quad t \in R \ . \tag{2.1}$$

We take $\alpha(0,\, 0) = 1$ and $0 = \eta_0 < \eta_1 < \ldots < \eta_q$. If $\alpha(j,\, k) = 0$, all $j > 0$, (2.1) is a pure difference equation; if $\alpha(j,\, k) = 0$, all $k > 0$, it is a pure differential equation. Thus we take $\alpha(j,\, k) \neq 0$ for some j , all $k > 0$, and for some $(j,\, k)$, $j > 0$. Under the usual assumptions, the formal solution in the m.s. sense is

$$y(t) = \int_R e^{it\lambda} \tilde{\alpha}(-i\lambda)^{-1} d\chi_w(\lambda) \ , \quad t \in R \ ,$$

where we assume

$$\tilde{\alpha}(i\lambda) = \sum_{j=0}^{m} \sum_{k=0}^{q} \alpha(j,\, k)(-i\lambda)^j \exp\left(i\eta_k\lambda\right)$$

is bounded away from zero, $\lambda \in R$. Naturally we are interested in determining whether all zeros of $\tilde{\alpha}(-\omega)$ have negative real parts. Now these zeros may have arbitrarily large modulus and they tend to lie along "chains", of which there are essentially three types. A chain is said to be retarded, neutral or advanced depending upon whether, as $|\omega| \to \infty$, $Re\{\omega\} \to -\infty$, $Re\{\omega\}$ is bounded or $Re\{\omega\} \to \infty$, respectively. A necessary condition for stability is thus the absence of advanced chains. However the problem of establishing perspicuous stability conditions in terms of the $\alpha(j, k)$, η_k appears to be intractable in general, although specific cases have been discussed by such authors as Cebotarev and Meiman (1949, Chapter VII), Bellman and Cooke (1963) and Pinney (1958). When the η_k are commensurable the situation is simplified. We introduce the two-dimensional polynomial

$$\psi(\omega, \sigma) = \sum_{j=0}^{m} \sum_{k=0}^{q} \alpha(j, k)\omega^j\sigma^{q-k} .$$

Given that $\alpha(j, 0) \neq 0$, some j , a principal term $\alpha(l, 0)\omega^l\sigma^q$ of $\psi(\omega, \sigma)$ exists if and only if $\alpha(l, 0) \neq 0$ and for all (j, k) such that $\alpha(j, k) \neq 0$, $j \leq l$ and $k < q$. Pontryagin (1955) has shown that if $\psi(\omega, \sigma)$ has no principal term the function

$$\tilde{\alpha}(-\omega) = e^{-q\omega}\psi(\omega, e^{\omega})$$

has an unbounded number of zeros with arbitrarily large real part. Given that such a term is present, however, Pontryagin presents sufficient conditions for stability.[66] Pontryagin's results have implications for a number of difference-differential equations that have arisen in practice and seem worthy of mention.

The homogeneous equation

$$y(t) + \alpha_1 y(t-\eta) + \alpha_2 Dy(t) = 0 , \qquad (2.2)$$

with $\eta > 0$, has been considered by Kalecki (1935) in the study of business cycles. The solution and stability of (2.2) have been discussed by, for example, Frisch and Holme (1935), James and Brelz

[66] Pontryagin takes all differences to be integer-valued. An elementary transformation makes his results applicable to equations with an arbitrary fundamental difference.

(1936), Hayes (1950) and Burger (1956). For our parameterization (2.2) is stable if and only if, for $\eta > 0$,

$$\frac{1}{\alpha_2} + \frac{1}{\eta} > 0 \ , \quad -\frac{1}{\alpha_2} < \frac{\alpha_1}{\alpha_2} < \left\{\left(\frac{\omega}{\eta}\right)^2 + \frac{1}{\alpha_2^2}\right\}^{\frac{1}{2}} \ , \tag{2.3}$$

where ω is the zero of $\alpha_2\omega + \eta\tan\omega$ in $(0, \pi)$. Tinbergen (1935) suggests an (unstable) equation for profit margin, namely

$$y(t) + \alpha_1 y(t-\eta) + \alpha_2 y(t-2\eta) + \alpha_3 Dy(t-2\eta) = 0 \ ,$$

and Pinney (1958) has considered its solution when $\alpha_2 = 0$. The equation

$$y(t) + \alpha_1 Dy(t) + \alpha_2 D^2 y(t) + \alpha_3 D^n y(t-\eta) = w(t) \ ,$$

where it is necessary that $n = 1$ or 2 for stability, is of importance in control theory. (See Ansoff and Krumhansl, 1948.) Bellman and Cooke (1963) reproduce stability conditions for the last equation, derived using Pontryagin's results. Sargan (1958), observing that the Leontief dynamic system is unrealistic in its assumption that supply is at all times equal to demand, considers the effect of superimposing a difference-differential structure on the system, involving small time lags. Further applications in economics are discussed by Kalecki (1943).

7.3 Integral equations

We consider first the difference-kernel equation of the first kind, namely

$$\int_a^b \gamma_1(\tau) y(t-\tau) d\tau = w(t) \ , \quad t \in R \ , \tag{3.1}$$

where possibly $a = -\infty$, $b = \infty$. Under the usual conditions the formal solution, in the m.s. sense, is

$$y(t) = \int_R e^{it\lambda} \tilde{\gamma}_1(-\lambda)^{-1} d\chi_w(\lambda) \ , \quad t \in R \ ,$$

where we assume

$$\tilde{\gamma}_1(\lambda) = \int_a^b e^{i\lambda t}\gamma_1(t)dt$$

is bounded away from zero, $\lambda \in R$. Then a solution of the form

$$y(t) = \int_R \kappa(\tau)w(t-\tau)d\tau \ , \quad t \in R \ ,$$

exists. While the general nature of (3.1) makes precise discussion difficult, simple stability conditions for such equations are available. Langer (1931) presents a result of G. Pólya stating that[67] if $a = 0$ and $b = 1$ and γ_1 is positive and non-decreasing on $(0, 1)$, (3.1) is stable.

We examine next the integral equation of the second kind,

$$y(t) - \int_a^b \gamma_2(\tau)y(t-\tau)d\tau = w(t) \ , \quad t \in R \ .$$

Then it is known that a unique solution satisfying our conditions exists, namely

$$y(t) = \int_R e^{it\lambda}\left(1-\tilde{\gamma}_2(-\lambda)\right)^{-1}d\chi_w(\lambda) \ ,$$

where

$$\tilde{\gamma}_2(\lambda) = \int_a^b e^{i\lambda t}\gamma_2(t)dt \ ,$$

providing $|\tilde{\gamma}_2(\lambda)| < 1$, $\lambda \in R$. In this case the solution may be written as

$$y(t) = w(t) + \int_a^b \gamma_2(\tau)w(t-\tau)d\tau + \ldots$$

$$+ \int_a^b \cdots \int_a^b \gamma_2(\tau_1) \ldots \gamma_2(\tau_k)w(t - \tau_1 - \ldots - \tau_k)d\tau_1 \ldots d\tau_k + \ldots \ .$$

[67] Providing $\gamma(t)$ is not an "exceptional" function, that is a step function with discontinuites finite in number and occurring only at rational points of $(0, 1)$.

Thus if $b > a \geq 0$, $y(t)$ is unrelated to $w(\tau)$ for $\tau > t$.

7.4 Identification, estimation and asymptotic theory

We shall consider the estimation of the general multiple system

$$y(t) + \sum_{j=0}^{m} \left\{ \sum_{k=0}^{q} A_0(j, k) D^j y(t-\eta_{0k}) + \int_R \Gamma_j(\tau; \theta_0) D^j y(t-\tau) d\tau \right\}$$

$$= \sum_{j=0}^{n} \left\{ \sum_{k=1}^{r} B_0(j, k) D^j z(t-\zeta_{0k}) + \int_R \Delta_j(\tau; \theta_0) D^j z(t-\tau) d\tau \right\} + u(t), \quad (4.1)$$

$t \in R$. Here the $A_0(j, k)$ and $B_0(j, k)$ are respectively $p_1 \times p_1$ and $p_1 \times p_2$ matrices of coefficients and the $\Gamma_j(t; \theta)$ and $\Delta_j(t; \theta)$ are respectively $p_1 \times p_1$ and $p_1 \times p_2$ matrices of, in general, non-linear functions of θ . We wish to estimate the η_{0j}, ζ_{0j}, $A_0(j, k)$ $B_0(j, k)$ and θ_0 given the sample $y(n)$, $z(n)$, $n = 1, \ldots, N$.

For the identification of (4.1), after imposing (iv) , it is necessary that the parameters can be determined from knowledge of $\tilde{A}_0(i\lambda)^{-1}\tilde{B}_0(i\lambda)$, in such a way that (v) holds, where

$$\tilde{A}(i\lambda) = I_{p_1} + \sum_{j=0}^{m} (-i\lambda)^j \left\{ \sum_{k=0}^{q} A(j, k) \exp(i\eta_k\lambda) + \tilde{\Gamma}_j(\lambda; \theta) \right\} ,$$

$$\tilde{B}(i\lambda) = \sum_{j=0}^{n} (-i\lambda)^j \left\{ \sum_{k=1}^{r} B(j, k) \exp(i\zeta_k\lambda) + \tilde{\Delta}_j(\lambda; \theta) \right\} ,$$

with

$$\tilde{\Gamma}_j(\lambda; \theta) = \int_R e^{i\lambda t} \Gamma_j(t; \theta) dt , \quad \tilde{\Delta}_j(\lambda; \theta) = \int_R e^{i\lambda t} \Delta_j(t; \theta) dt ,$$

and $\det\{\tilde{A}_0(i\lambda)\}$ is taken to be bounded away from zero, $\lambda \in R$. The presence of the terms involving the Γ_j and Δ_j complicates the identification problem and prevents further precise analysis.

We consider, however, the pure difference-differential[68] case $\Gamma_j(t;\ \theta) = 0$, all j , $\Delta_j(t;\ \theta) = 0$, all j . As usual we take

$$0 = \eta_{00} < \eta_{01} < \ldots < \eta_{0q}\ ,\quad \zeta_{01} < \zeta_{02} < \ldots < \zeta_{0r}\ .$$

Then we normalise the system by taking the diagonal of $A_0(0,\ 0)$ to be null, when this is possible (see §5.3). We require $A_0(j,\ k) \neq 0$, some j , $k = 1, \ldots, q$ and $B_0(j,\ k) \neq 0$, some j , $k = 1, \ldots, r$ to identify the η_{0j}, ζ_{0j} . Also the rank and order conditions of §5.3 must be imposed on the $p_1 \times [(m+1)(q+1)p_1+(n+1)rp_2]$ matrix

$$[A_0(0,\ 0) \vdots \ldots \vdots A_0(m,\ q) \vdots B_0(0,\ 1) \vdots \ldots \vdots B_0(n,\ r)]\ .$$

For the general system (4.1) we take $(\hat{\eta}_N \vdots \hat{\zeta}_N \vdots \hat{\alpha}_N \vdots \hat{\beta}_N \vdots \hat{\theta}_N)$ to be the admissible $(\eta \vdots \zeta \vdots \alpha \vdots \beta \vdots \theta)$ minimizing

$$Q_N = N^{-1} \sum_{B} \left\| \Phi(\lambda_s)^{\frac{1}{2}}\left\{w_y(s)-\tilde{A}(i\lambda_s)^{-1}\tilde{B}(i\lambda_s)w_z(s)\right\}\right\|^2\ ,$$

where now $\alpha = \text{vec}'\{A\}$, $\beta = \text{vec}'\{B\}$, with

$$A = \begin{bmatrix} A(0,\ 0) & \vdots & \quad & \vdots & A(m,\ 0) \\ \cdots\cdots & \vdots & \cdots & \vdots & \cdots\cdots \\ & \vdots & & \vdots & \\ \cdots\cdots & \vdots & \cdots & \vdots & \cdots\cdots \\ A(0,\ q) & \vdots & & \vdots & A(m,\ q) \end{bmatrix}\ ,\quad B = \begin{bmatrix} B(0,\ 1) & \vdots & \quad & \vdots & B(n,\ 1) \\ \cdots\cdots & \vdots & \cdots & \vdots & \cdots\cdots \\ & \vdots & & \vdots & \\ \cdots\cdots & \vdots & \cdots & \vdots & \cdots\cdots \\ B(0,\ r) & \vdots & & \vdots & B(n,\ r) \end{bmatrix}\ .$$

Of course the $\tilde{\Gamma}_j(\lambda;\ \theta)$, $\tilde{\Delta}_j(\lambda;\ \theta)$ must be known functions of λ and θ . It would be tedious to state a formal theorem since all we are doing is amalgamating previous results. However under the usual conditions our estimates are consistent and

[68] The system (4.1) includes such special cases as $y(t) + \alpha Dy(t) = \beta z(t-\zeta) + u(t)$. This is not a difference-differential equation but a differential equation in which there is a period ζ of pure delay before a given change in $z(t)$ produces a response in $y(t)$. The quantity ζ is sometimes known as "dead time". Such equations are of substantial practical importance but we do not consider them in isolation because of course their solution follows from §6 and their estimation from the present section, taking $q = 0$.

$$N^{\frac{1}{2}}\left(\hat{\eta}_N-\eta_0 \vdots \hat{\zeta}_N-\zeta_0 \vdots \hat{\alpha}_N-\alpha_0 \vdots \hat{\beta}_N-\beta_0 \vdots \hat{\theta}_N-\theta_0\right)$$

$$\xrightarrow{\mathcal{D}} N\left[0,\ \Psi_0(\Phi)^{+}\Psi_0\left(\Phi f_{xx}\Phi\right)\Psi_0(\Phi)^{+}\right] ,$$

where

$$\Psi(\Phi) = (2\pi)^{-1}\int_B T(\lambda)^*\left\{f_{zz}(-\lambda) \otimes \tilde{A}(i\lambda)^{-1*}\Phi(\lambda)\tilde{A}(i\lambda)^{-1}\right\}T(\lambda)d\lambda ,$$

with

$$T(\lambda) = \left[-\tilde{B}(i\lambda)'\tilde{A}(i\lambda)^{-1'} \otimes I_{p_1} \vdots I_{p_1p_2}\right]\left[C_1 \vdots C_2 \vdots C_3L \vdots C_4\right] ,$$

in which

$$C_1 = \begin{bmatrix} i\lambda e(i\eta\lambda) \odot \sum_{j=0}^{m} \left(\mathrm{vec}\{A(j,1)\} \vdots \cdots \vdots \mathrm{vec}\{A(j,q)\}\right) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 0 \end{bmatrix} ,$$

$$C_2 = \begin{bmatrix} 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ i\lambda e(i\zeta\lambda) \odot \sum_{j=0}^{n} \left(\mathrm{vec}\{B(j,1)\} \vdots \cdots \vdots \mathrm{vec}\{B(j,r)\}\right) \end{bmatrix} ,$$

$$C_3 = \begin{bmatrix} 1_m(i\lambda) \otimes I_{p_1} \otimes [1 \vdots e(i\eta\lambda)] \otimes I_{p_1} & \vdots & 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots & \vdots & \cdots\cdots\cdots\cdots\cdots\cdots \\ 0 & \vdots & 1_n(i\lambda) \otimes I_{p_2} \otimes e(i\zeta\lambda) \otimes I_{p_1} \end{bmatrix} ,$$

$$C_4 = \begin{bmatrix} \left\{1_m(i\lambda) \otimes I_{p_1^2}\right\}(\partial/\partial\theta)\mathrm{vec}\{\tilde{\Gamma}_0(\lambda;\theta) \vdots \cdots \vdots \tilde{\Gamma}_m(\lambda;\theta)\} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \left\{1_n(i\lambda) \otimes I_{p_1p_2}\right\}(\partial/\partial\theta)\mathrm{vec}\{\tilde{\Delta}_0(\lambda;\theta) \vdots \cdots \vdots \tilde{\Delta}_n(\lambda;\theta)\} \end{bmatrix} .$$

7.5 Numerical examples

Our sequence of numerical examples is completed by the fitting of the simple difference-differential equation

$$y(t) + \alpha_{01}y\left(t-\eta_0\right) + \alpha_{02}Dy(t) = \beta_0 z(t) + u(t) ,\quad t \in R ,$$

to our data. The estimate $\left(\hat{\eta}_N \vdots \hat{\alpha}_{N1} \vdots \hat{\alpha}_{N2}\right)$ was found by maximizing

$$\frac{\left\{\sum_s I_{zy}(s)\left[1+\alpha_1\exp\left(i\eta\lambda_s\right)-\alpha_2 i\lambda_s\right]^{-1}\varphi\left(\lambda_s\right)\right\}^2}{\sum_s I_{zz}(s)\left|1+\alpha_1\exp\left(i\eta\lambda_s\right)-\alpha_2 i\lambda_s\right|^{-2}\varphi\left(\lambda_s\right)} \tag{5.1}$$

over the region

$$0.01 \le \eta \le 20.00\ ,\quad 0.01 \le \alpha_1 \le 0.99\ ,\quad 0 \le \alpha_2 \le 5.00 \tag{5.2}$$

which is properly contained by the stable domain (2.3). (A portion of the region $\alpha_2 < 0$ is also admissible but this was not scanned.)

Then we have

$$\hat{\beta}_N = \frac{\sum_s I_{zy}(s)\left[1+\hat{\alpha}_{N1}\exp\left(i\hat{\eta}_N\lambda_s\right)-\hat{\alpha}_{N2}i\lambda_s\right]^{-1}\varphi\left(\lambda_s\right)}{\sum_s I_{zz}(s)\left|1+\hat{\alpha}_{N1}\exp\left(i\hat{\eta}_N\lambda_s\right)-\hat{\alpha}_{N2}i\lambda_s\right|^{-2}\varphi\left(\lambda_s\right)}\ .$$

Now

$$N^{\frac{1}{2}}\left(\hat{\eta}_N-\eta_0 \,\vdots\, \hat{\alpha}_{N1}-\alpha_{01} \,\vdots\, \hat{\alpha}_{N2}-\alpha_{02} \,\vdots\, \hat{\beta}_N-\beta_0\right) \xrightarrow{\mathcal{D}} N\left[0,\ \Psi_0(\varphi)^{-1}\Psi_0\left(f_{xx}\varphi^2\right)\Psi_0(\varphi)^{-1}\right]$$

where

$$\Psi(\varphi) = (2\pi)^{-1}\int_{-\pi}^{\pi} \psi(\lambda)^*\psi(\lambda)\left|1+\alpha_1 e^{i\eta\lambda}-\alpha_2 i\lambda\right|^{-4}\varphi(\lambda)f_{zz}(\lambda)d\lambda\ ,$$

with

$$\psi(\lambda) = \left[-\alpha\beta i\lambda e^{i\eta\lambda} \,\vdots\, -\beta e^{i\eta\lambda} \,\vdots\, \beta i\lambda \,\vdots\, \beta\left\{1+\alpha_1 e^{i\eta\lambda}-\alpha_2 i\lambda\right\}\right]\ .$$

Using these results the covariance matrix was estimated in the usual way.

Table 7.1								
Economic Data								
φ	$\hat{\eta}_N$	s.e. $\hat{\eta}_N$	$\hat{\alpha}_{1N}$	s.e. $\hat{\alpha}_{1N}$	$\hat{\alpha}_{2N}$	s.e. $\hat{\alpha}_{2N}$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
1	1.922	0.322	-0.348	0.135	0.202	0.093	0.501	0.010
f	1.977	0.273	-0.092	0.048	0.023	0.043	0.847	0.027

Table 7.2								
Oceanographic Data								
φ	$\hat{\eta}_N$	s.e. $\hat{\eta}_N$	$\hat{\alpha}_{1N}$	s.e. $\hat{\alpha}_{1N}$	$\hat{\alpha}_{2N}$	s.e. $\hat{\alpha}_{2N}$	$\hat{\beta}_N$	s.e. $\hat{\beta}_N$
1	16.750	6.619	-0.032	0.075	1.456	0.277	0.931	0.071
f	15.484	1.857	-0.162	0.074	2.234	0.332	0.845	0.070

There is good agreement between Table 7.1 and Tables 5.1 and 6.1 but the large estimated difference $\hat{\eta}_N$ in Table 7.2 requires comment. If, prompted by Table 5.2, only small values of η are scanned an analytic maximum of (5.1) is reached, but the estimates there have grossly inflated standard errors because a model that includes both a derivative and a very small difference will be nearly unidentified. The estimates recorded are the optimal ones over the whole region (5.2).

In each of §§4-7 the simplest possible model has been chosen to show our methods in action, no serious consideration having been given to the most plausible *a priori* models for the data. Each additional parameter added involves, loosely speaking, the multiplication of computing time by a factor so for more complicated systems computational efficiency is of prime importance. However if the values of the lags are assumed (but not necessarily integral) the methods of §6.3 are easily modified to produce explicit estimates for the systems of §5 and §7.

BIBLIOGRAPHY

AITCHISON, J. and S.D. SILVEY (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.*, 29, 813-828.

ANDERSON, T.W. (1951). The asymptotic distribution of certain characteristic roots and vectors. *Proc. Second Berkeley Symp. Math. Statist. Prob.*, 103-130. Berkeley: Univ. of California Press.

——— (1958). *An Introduction to Multivariate Statistical Analysis.* New York: Wiley.

ANSOFF, H.I. and J.A. KRUMHANSL (1948). A general stability criterion for linear oscillating systems with constant time lag. *Quart. Appl. Math.*, 6, 337-341.

BARTLETT, M.S. (1947 a). Multivariate analysis. *J. Roy. Statist. Soc. Ser. B*, 9, 176-197.

——— (1947 b). The general canonical correlation distribution. *Ann. Math. Statist.*, 18, 1-17.

BELLMAN, R. and K.L. COOKE (1963). *Differential-Difference Equations.* New York: Academic Press.

BERGSTROM, A.R. (1966). Nonrecursive models as discrete approximations to systems of stochastic differential equations. *Econometrica*, 34, 173-182.

——— (1967). *The Construction and Use of Economic Models.* London: English Universities Press.

BERNOULLI, J. (1728). Meditationes. Dechordibus vibrantibus .... *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 3, 13-28.

BERNSTEIN, S. (1926-7). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.*, 97, 1-59.

BOCHNER, S. (1931). Allgemeine lineare differenzengleichungen mit asymptotisch konstanten koeffizienten. *Math. Zeitschrift*, 33, 426-450.

BRILLINGER, D.R. (1969). The canonical analysis of stationary time series. In *Multivariate Analysis - II* (P.R. Krishnaiah, Ed.), pp. 331-350. New York: Academic Press.

BURGER, E. (1956). On the stability of certain economic systems. *Econometrica*, 24, 488-493.

CARMICHAEL, R.D. (1933). Systems of linear difference equations and expansions in series of exponential functions. *Trans. Amer. Math. Soc.*, 35, 1-28.

CEBOTAREV, N.G. and N.N. MEIMAN (1949). The Routh-Hurwitz problem for polynomials and entire functions. *Trudy Mat. Inst. Steklov*, 26.

CONSTANTINE, A.G. and A.T. JAMES (1958). On the general canonical correlation distribution. *Ann. Math. Statist.*, 29, 1146-1166.

CRAMER, H. and M.R. LEADBETTER (1967). *Stationary and Related Stochastic Processes*. New York: Wiley.

DHRYMES, P.J. (1971 a). On the strong consistency of estimators for certain distributed lag models with autocorrelated errors. *Int. Econ. Rev.*, 12, 329-343.

——— (1971 b). *Distributed Lags. Problems of Estimation and Formulation*. San Francisco: Holden-Day.

DURBIN, J. (1961). Efficient fitting of linear models for continuous stationary time series from discrete data. *Bull. Int. Statist. Inst.*, 38, 273-282.

FISHER, F.M. (1959). Generalization of the rank and order conditions for identifiability. *Econometrica,* 27, 431-447.

——— (1966). *The Identification Problem in Econometrics.* New York: McGraw-Hill.

FRISCH, R. and H. HOLME (1935). The characteristic solutions of a mixed difference and differential equation occurring in economic dynamics. *Econometrica,* 3, 225-239.

GOLDBERGER, A.S. (1972). Maximum-Likelihood estimation of regression models containing unobservable variables. Unpublished paper, *SSRI,* Univ. of Wisconsin.

HANNAN, E.J. (1963). Regression for time series. In *Time Series Analysis* (M. Rosenblatt, Ed.), pp. 17-37. New York: Wiley.

——— (1967). The estimation of a lagged regression relation. *Biometrika,* 54, 409-418.

——— (1970). *Multiple Time Series.* New York: Wiley.

——— (1971 a). Non-linear time series regression. *J. Appl. Prob.,* 8, 767-780.

——— (1971 b). The identification problem for multiple equation systems with moving average errors. *Econometrica,* 39, 751-765.

——— (1973 a). The asymptotic theory of linear time series models. Accepted for publication in *J. Appl. Prob.*

——— (1973 b). The uniform convergence of autocovariances. Submitted for publication.

——— and P.M. ROBINSON (1973). Lagged regression with unknown lags. Accepted for publication in *J. Roy. Statist. Soc. Ser. B.*

——— and R.D. TERRELL (1973). Multiple equation systems with stationary errors. Accepted for publication in *Econometrica.*

——— and P.J. THOMSON (1972). The estimation of coherence and group delay. *Biometrika,* 58, 469-482.

HAYES, N.D. (1950). Roots of the transcendental equation associated with a certain difference-differential equation. *J. London Math. Soc.,* 25, 226-232.

HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika,* 28, 321-377.

HSU, P.L. (1939). On the distribution of the roots of certain determinantal equations. *Ann. Eugen.,* 9, 250-258.

——— (1941). On the limiting distribution of roots of a determinantal equation. *J. London Math. Soc.*, 16, 183-194.

IBRAGIMOV, I.A. and Y.V. LINNIK (1971). *Independent and Stationary Sequences of Random Variables.* Groningen: Wolters-Noordhoff.

JAMES, A.T. (1964). Distributions of matrix variables and latent roots derived from normal samples. *Ann. Math. Statist.,* 35, 475-501.

JAMES, H.M., N.B. NICHOLS and R.S. PHILLIPS (1947). *Theory of Servo-mechanisms.* New York: McGraw-Hill.

JAMES, R.W. and M.H. BELZ (1936). On a mixed difference and differential equation. *Econometrica,* 4, 157-160.

JENNRICH, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.,* 40, 633-643.

JORGENSON, D.W. (1966). Rational distributed lag functions. *Econometrica,* 32, 135-149.

KALECKI, M. (1935). A macrodynamic theory of business cycles. *Econometrica,* 3, 327-344.

——— (1943). *Studies in Economic Dynamics.* London: Allen and Unwin.

KAPLAN, W. (1962). *Operational Methods for Linear Systems.* Reading, Mass.: Addison-Wesley.

KHATRI, C.G. (1968). Some results for the singular normal multivariate regression models. *Sankhyā, Series A*, 30, 267-280.

KITAGAWA, T. (1937). On the theory of linearly translatable functional equations and Cauchy's series. *Japanese J. Math.*, 13, 233-332.

LANGER, R.E. (1931). On the zeros of exponential sums and integrals. *Bull. Amer. Math. Soc.*, 37, 213-239.

LEONTIEF, W.W. (1953). *Studies in the Structure of the American Economy*. New York: Oxford University Press.

MALINVAUD, E. (1970 a). The consistency of non-linear regressions. *Ann. Math. Statist.*, 41, 956-969.

——— (1970 b). *Statistical Methods of Econometrics*. Amsterdam: North-Holland.

MARDEN, M. (1949). *The Geometry of the Zeros*. Mathematical Surveys, Number III. New York: Amer. Math. Soc.

MARTIN, W.T. (1938). Linear difference equations with arbitrary real spans. *Acta Math.*, 69, 57-98.

PHILLIPS, A.W. (1959). The estimation of parameters in systems of stochastic differential equations. *Biometrika*, 46, 67-76.

PINCHERLE, S. (1926). Sur la résolution de l'équation fonctionelle $\sum h_\nu \varphi(x+a_\nu) = f(x)$ à coefficients constants. *Acta Math.*, 48, 279-304. (First published in 1888.)

PINNEY, E. (1958). *Ordinary Difference-Differential Equations*. Berkeley: Univ. of Calif. Press.

PONTRYAGIN, L.S. (1955). On the zeros of some elementary transcendental functions. *Amer. Math. Soc. Transl. Ser.* 2, 1, 95-110. (First published in 1942.)

RACLIS, R. (1930). Solution principale de l'équation linéaire aux différences finies. *Acta Math.*, 55, 277-394.

RAO, C.R. (1965). The use and interpretation of principal component analysis in applied research. *Sankhyā Ser. A*, 26, 329-358.

RITT, J.F. (1927). A factorization theory for functions $\sum a_i \exp(\alpha_i x)$. *Trans. Amer. Math. Soc.*, 29, 584-596.

ROBINSON, P.M. (1969). *A gravity model for telephone calls.* M.Sc. Dissertation, Statistics Dept., London School of Economics and Political Science.

——— (1972). Non-linear regression for multiple time series. Accepted for publication in *J. Appl. Prob.*

——— (1973). Generalized canonical analysis for time series. Accepted for publication in *J. Multivariate Analysis.*

ROHDE, C.A. (1965). Generalized inverses of partitioned matrices. *SIAM J. Appl. Math.*, 13, 1033-1035.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.*, 42, 43-47.

ROTHENBERG, T.J. (1971). Identification in parametric models. *Econometrica*, 39, 577-591.

ROZANOV, Y.A. (1967). *Stationary Random Processes.* San Francisco: Holden-Day.

SARGAN, J.D. (1958). The instability of the Leontief dynamic model. *Econometrica*, 26, 381-392.

——— (1968). Some discrete approximations to continuous time stochastic models. Unpublished.

SIMS, C.A. (1971). Discrete approximations to continuous time distributed lag models in econometrics. *Econometrica*, 39, 545-563.

STRODT, W. (1948). Linear difference equations and exponential polynomials. *Trans. Amer. Math. Soc.*, 64, 439-466.

TINBERGEN, J. (1935). Annual Survey: Suggestions on quantitative business cycle theory. *Econometrica*, 3, 241-308.

TUSTIN, A. (1957). *The Mechanism of Economic Systems*. London: Heinemann.

WALD, A. (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. New York: Wiley.

WEGGE, L. (1965). Identifiability criteria for a system of equations as a whole. *Austral. J. Statist.*, 7, 67-77.

WILKINSON, J.H. (1965). *The Algebraic Eigenvalue Problem.* New York: Oxford Univ. Press (Clarendon).

WYMER, C.R. (1972). Econometric estimation of stochastic differential equation systems. Accepted for publication in *Econometrica.*

ZELLNER, A. (1970). Estimation of regression relationships containing unobservable independent variables. *Int. Econ. Rev.*, 11, 441-454.

ZYGMUND, A. (1959). *Trigonometric Series Vol I.* Cambridge: Cambridge University Press.