The Annals of Statistics
1979, Vol. 7, No. 3, 490–506

A CENTRAL LIMIT THEOREM FOR PARAMETER ESTIMATION IN STATIONARY VECTOR TIME SERIES AND ITS APPLICATION TO MODELS FOR A SIGNAL OBSERVED WITH NOISE

BY W. DUNSMUIR

Massachusetts Institute of Technology

A general finite parameter model for stationary ergodic nondeterministic vector time series is considered. A central limit theorem for parameter estimates, obtained by maximising frequency domain approximations to the Gaussian likelihood, is established. The treatment given extends the central limit theorem of Dunsmuir and Hannan in that the innovations covariance matrix and the linear transfer function need not be separately parameterised. Models for a stationary vector signal observed with stationary vector noise are discussed in relation to the central limit theorem and the conditions imposed for this result are related to this model. Finally, the special case of a scalar autoregressive signal observed with noise is discussed. It is shown that this model may be reparameterised so that the central limit theorem of Dunsmuir and Hannan may be applied.

1. Introduction. Let $z(n)$ be a vector time series with s components generated by a parametric model of the form

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

where $\varepsilon(n)$ is a vector sequence of white noise (i.e., $E\varepsilon(m) = 0$, $E\varepsilon(m)\varepsilon(n)' = \delta_{mn}K(\theta)$ for all m and n, and δ_{mn} is Kronecker's delta). It will be assumed that $C(0;\theta) = I_s$ (the s-rowed identity) and that the $s \times s$ matrices $C(j;\theta)$ satisfy $\operatorname{tr} \sum_{j=0}^{\infty} C(j;\theta)K(\theta)C(j;\theta)^* < \infty$. ($A^*$ will mean the complex conjugate transpose of A, $\|A\|$ will be taken as the matrix norm $(\operatorname{tr}(AA^*))^{\frac{1}{2}}$, $\operatorname{tr} A$ is the trace of A, $\det A$ is the determinant.) Then $z(n)$ has zero mean and finite variance. It will also be assumed that $z(n)$ is strictly stationary and ergodic.

In (1.1) θ denotes a vector of U unknown parameters to be estimated using observations $z(1), z(2), \cdots, z(N)$. The most common example of a parametric model of the form (1.1) is the multiple autoregressive-moving average (ARMA) model

$$\sum_{j=0}^{q} B(j)z(n-j) = \sum_{j=0}^{p} A(j)\varepsilon(n-j) \tag{1.2}$$

where $B(0) = A(0) = I_s$ and $\det h(\zeta) \neq 0$ for $|\zeta| \leqslant 1$ where $h(\zeta) = \sum_{j=0}^{q} B(j)\zeta^q$. However, there are other examples—see Bloomfield [1] for instance.

The methods of estimation to be considered are all obtained by approximating the likelihood of $z_N' = (z(1)', \cdots, z(N)')$ derived on Gaussian assumptions

Received May 1977; revised May 1978.

AMS 1970 *subject classifications.* Primary 62M10; secondary 60G10.

Key words and phrases. Linear processes, ARMA models, central limit theorems, martingales, prediction theory, signal with noise, Fourier methods.

although no such distributional assumptions will be required below. Specifically, $-(2/N)\log$ (likelihood) on this assumption is

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

where $\Gamma_N(\theta) = Ez_Nz_N'$ and has as its (m, n)th block of $s \times s$ elements the matrix $\Gamma(n - m; \theta) = E(z(m)z(n)')$. As in Dunsmuir and Hannan [4], various frequency domain approximations to (1.3) may be introduced. To define these it is useful to define the periodogram at frequency $\omega \in [-\pi, \pi]$ as $I(\omega) = W(\omega)W(\omega)^*$ where

$$W(\omega) = (2\pi N)^{-\frac{1}{2}}\Sigma_{n=1}^{N}z(n)e^{in\omega}$$

is the discrete Fourier transform of the data. Also the spectral density matrix of $z(n)$ is

(1.4)

$$f(\omega; \theta) = \frac{1}{2\pi}k(e^{i\omega}; \theta)K(\theta)k(e^{i\omega}; \theta)^* \qquad \text{where} \quad k(e^{i\omega}; \theta) = \Sigma_{j=0}^{\infty}C(j; \theta)e^{ij\omega}.$$

In terms of these quantities two approximations to $\hat{L}_N(\theta)$ may be introduced as follows.

$$\tilde{L}_N(\theta) = \log\det K(\theta) + N^{-1}\Sigma_t \operatorname{tr}\left[f^{-1}(\omega_t; \theta)I(\omega_t)\right] \tag{1.5}$$

where $\omega_t = 2\pi t/N$, $-N/2 < t \leqslant [N/2]$,

$$\bar{L}_N(\theta) = \log\det K(\theta) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left[f^{-1}(\omega; \theta)I(\omega)\right]d\omega. \tag{1.6}$$

In practice $\tilde{L}_N$ will tend to be preferred to $\bar{L}_N$ since for $\tilde{L}_N$ the $W(\omega_t)$ may be computed efficiently by using the fast Fourier transform. When $\hat{L}_N(\theta)$, $\tilde{L}_N(\theta)$, $\bar{L}_N(\theta)$ are minimised with respect to θ (belonging to a suitable parameter space—see [3] and [4] for some examples) the estimates so obtained will be referred to as $\hat{\theta}_N$, $\tilde{\theta}_N$, $\bar{\theta}_N$ respectively.

In [3] and [4] the strong law of large numbers (i.e., $\hat{\theta}_N \to_{\text{a.s.}} \theta_0$, the true value) for these estimates is established under quite general conditions. In particular, this property for the estimates of the parameters in (1.2) is discussed at some length. Also in [4] the central limit theorem (CLT) for $\hat{\theta}_N$, etc., was established under the assumption that θ is partitionable into two subvectors $\theta' = (\tau', \mu')$ where τ specifies the $C(j; \tau)$ (and thus $k(e^{i\omega}; \tau)$ in (1.4)) and μ specifies $K(\mu)$ with τ and μ independently varying. Then $N^{1/2}(\hat{\tau}_N - \tau_0)$ converges in distribution to the multivariate normal under mild conditions (condition C2.3(a) and (b) of Section 2 to follow) on the distributional properties of $\varepsilon(m)$ in (1.1). The corresponding result of [4] for $N^{1/2}(\hat{\mu}_N - \mu_0)$ requires fourth moments to exist for the elements of $\varepsilon(m)$—see condition C2.3. In this article the condition that k and K be separately parameterised, in the CLT for $\bar{\theta}_N$ and $\tilde{\theta}_N$, will be removed (see Theorem 2.1 and Corollary 2.2).

The main example which will be introduced to motivate the need for the removal of this separate parameterisation condition is the model for a signal observed with noise. In this, $z(n) = y(n) + x(n)$ is observed where the signal $y(n)$ and the noise $x(n)$ are processes of the type (1.1). In Section 3 this model is discussed and the CLT of Section 2 is related to it. For the case where $x(n)$ is white noise some prediction theory for the prediction of $y(n)$ in terms of the history of $z(n)$ is given in order to motivate one of the conditions needed in the CLT (see C2.3(a)). In Section 4 the very special case of a scalar ($s = 1$) autoregressive ($p = 0, q \geq 1$) signal plus white noise is considered. A discussion of how the CLT of Section 2 may be applied to this case is given. It is then shown that this model may be reparameterised, in a fairly natural way, so that the CLT of [4] applies to the new parameter set.

2. The central limit theorem. In this section $z(n)$ will be taken to be a stationary, ergodic time series generated by a finite parameter model of the type (1.1) wherein the $\varepsilon(m)$ satisfy C2.3 below. Since it is our intention to only discuss the CLT for the estimates $\bar{\theta}_N$, $\tilde{\theta}_N$ it will be assumed that $\bar{\theta}_N \to_P \theta_0$ and $\tilde{\theta}_N \to_P \theta_0$ where θ_0 is unique. Sufficient conditions to ensure that $\bar{\theta}_N, \tilde{\theta}_N \to_{\text{a.s.}} \theta_0$ for the general model (1.1) and for the ARMA model (1.2) are given in [3] and [4]. The true parameter vector θ_0 is assumed to belong to a twice differentiable manifold $\mathfrak{M}$ of dimension U. It will be assumed throughout that $f(\omega, \theta) > 0$ for $\omega \in [-\pi, \pi]$, $\theta \in \mathfrak{M}$. It is also necessary to introduce the following conditions:

C2.1. $f(\omega, \theta)$ has elements which are twice continuously differentiable functions of $\theta \in \mathfrak{M}$. The second derivatives of these elements are continuous in $\omega \in [-\pi, \pi]$.

C2.2. $f(\omega, \theta)$ has elements belonging to Λ_α, the Lipschitz class of degree α, where $1/2 < \alpha \leq 1$. For the definition of Λ_α see Zygmund [12].

C2.3. For all $1 \leq a, b, c, d \leq s$ and $-\infty < n < \infty$,

(a) $E(\varepsilon_a(n)|\mathfrak{F}_{n-1}) = 0$ a.s.;

(b) $E(\varepsilon_a(n)\varepsilon_b(n)|\mathfrak{F}_{n-1}) = K_{ab}(\theta_0)$ a.s.;

(c) $E(\varepsilon_a(n)\varepsilon_b(n)\varepsilon_c(n)|\mathfrak{F}_{n-1}) = \beta_{abc}$ a.s.;

(d) $E(\varepsilon_a(n)\varepsilon_b(n)\varepsilon_c(n)\varepsilon_d(n)) < \infty$,

where $K_{ab}(\theta_0)$, β_{abc} are constants and $\mathfrak{F}_n$ is the sub σ-algebra generated by the elements of $\varepsilon(m)$ for $m \leq n$. We will call the fourth cumulant between $\varepsilon_a(m)$, $\varepsilon_b(m)$, $\varepsilon_c(m)$, $\varepsilon_d(m)$, κ_{abcd}. Condition C2.1 is analogous to C1 of [4], C2.2 is a stronger smoothness condition on $f(\omega; \theta)$ than C2 of [4] but this condition is satisfied for rational spectral density matrices corresponding to ARMA models, while C2.3 (which is like independence up to third moments) is stronger than C3 of [4] in that C2.3(c) and C2.3(d) are additionally imposed. C2.3 ensures that the covariance matrix in the CLT below is of reasonably neat form (see also Hannan [6]). The relevance of C2.3(a) to linear models of the type (1.1) was pointed out in Hannan and Heyde [7]. Its relevance to models for a signal plus noise will be stressed in the next section.

The CLT for $\bar{\theta}_N$ will be given first. Corollary 2.1 contains the corresponding result for $\tilde{\theta}_N$. In what follows it is convenient to let $k_{uc}(e^{i\omega})$ denote the (u, c)th element of $k(e^{i\omega}, \theta_0)$ and $f_0(\omega)$, $\partial f_0(\omega)/\partial\theta_j$ etc., denote $f(\omega; \theta)$, $\partial f(\omega; \theta)/\partial\theta_j$, etc., evaluated at $\theta = \theta_0$. The notation $\phi_{pq}^{(j)}(\omega)$ will be used for the (p, q)th element of $\partial f_o^{-1}(\omega)/\partial\theta_j$.

Theorem 2.1. *Under conditions* C2 *the vector* $N^{1/2}(\bar{\theta}_N - \theta_0)$ *has an asymptotic normal distribution with zero mean vector and covariance matrix*

$$\Omega^{-1}(2\Omega + \Pi)\Omega^{-1} \tag{2.1}$$

where

$$\Omega_{jl} = \frac{1}{2\pi}\int \operatorname{tr}\left[f_0^{-1}(\omega)\frac{\partial f_0(\omega)}{\partial\theta_j}f_0^{-1}(\omega)\frac{\partial f_0(\omega)}{\partial\theta_l}\right]d\omega$$

and

$$\Pi_{jl} = \Sigma_{a, b, c, d=1}^{s}\kappa_{abcd}\left[\frac{1}{(2\pi)^2}\int k^*\phi^{(j)}k\,d\omega\right]_{ab}\left[\frac{1}{(2\pi)^2}\int k^*\phi^{(l)}k\,d\omega\right]_{cd}.$$

Proof. Since $\bar{\theta}_N$ minimises $\bar{L}_N(\theta)$ and eventually enters an arbitrary neighborhood of θ_0 we consider (as in the proof of Theorem 5, [4])

$$0 = N^{1/2}\frac{\partial}{\partial\theta}\bar{L}_N(\bar{\theta}_N) = N^{1/2}\frac{\partial}{\partial\theta}\bar{L}_N(\theta_0) + \left[\frac{\partial^2}{\partial\theta^2}\bar{L}_N(\dot{\theta}_N)\right]N^{1/2}(\bar{\theta}_N - \theta_0)$$

where $\|\dot{\theta}_N - \theta_0\| \leqslant \|\bar{\theta}_N - \theta_0\|$ and by $(\partial/\partial\theta)\bar{L}_N(\theta_0)$, for instance, is meant the vector of derivatives evaluated at θ_0. Thus the central limit theorem for $N^{1/2}(\bar{\theta}_N - \theta_0)$ reduces to that for $N^{1/2}[(\partial^2/\partial\theta^2)\bar{L}_N(\dot{\theta}_N)]^{-1}[(\partial/\partial\theta)\bar{L}_N(\theta_0)]$. Now the (j, k)th element of $(\partial^2/\partial\theta^2)\bar{L}_N(\dot{\theta}_N)$ converges in probability to Ω_{jk}. In fact, letting $g(\omega; \theta) = (\partial^2 f^{-1}(\omega; \theta)/\partial\theta_j\partial\theta_k)$,

$$\frac{\partial^2\bar{L}_N(\dot{\theta}_N)}{\partial\theta_j\partial\theta_k} = \frac{1}{2\pi}\int \operatorname{tr}\{I(\omega)g(\omega; \dot{\theta}_N)\}\,d\omega - \frac{1}{2\pi}\int \operatorname{tr}\left\{\frac{\partial f(\omega; \dot{\theta}_N)}{\partial\theta_j}\frac{\partial f^{-1}(\omega; \dot{\theta}_N)}{\partial\theta_k} + f(\omega; \dot{\theta}_N)g(\omega; \dot{\theta}_N)\right\}d\omega. \tag{2.2}$$

Since $\dot{\theta}_N \to_p \theta_0$ and the derivatives of f and f^{-1} appearing in this expression are continuous the second integral in (2.2) converges in probability to

$$-\frac{1}{2\pi}\int \operatorname{tr}\left\{\frac{\partial f_0(\omega)}{\partial\theta_j}\phi^{(k)}(\omega) + f_0(\omega)g(\omega; \theta_0)\right\}d\omega.$$

Consider the first term in (2.2). Since $g(\omega; \theta)$ is uniformly continuous in $(\omega, \theta) \in [-\pi, \pi] \times \mathfrak{N}_0$, where $\mathfrak{N}_0$ is any closed neighborhood of θ_0, then

$$h(\theta) = \sup_{\omega\in[-\pi, \pi]}|g(\omega; \theta) - g(\omega; \theta_0)|$$

is continuous at $\theta = \theta_0$. Then $h(\dot{\theta}_N) \to_p h(\theta_0) = 0$. But

$$\left|\frac{1}{2\pi}\int \operatorname{tr}\big(I(\omega)g(\omega;\dot{\theta}_N)\big)\,d\omega - \frac{1}{2\pi}\int \operatorname{tr}(I(\omega)g(\omega;\theta_0))\,d\omega\right| \leqslant h(\dot{\theta}_N)\operatorname{tr}\left\{\frac{1}{2\pi}\int I(\omega)\,d\omega\right\} = h(\dot{\theta}_N)\operatorname{tr} G(0) \to_p 0$$

since, by ergodicity, $\operatorname{tr} G(0) \to_{\text{a.s.}} \operatorname{tr}\Gamma(0;\theta_0) < \infty$ where $G(l) = N^{-1}\sum_{m=1}^{N-l} z(m)z(m+l)'$ is the lth sample autocovariance matrix. Approximating $g(\omega;\theta_0)$ by the Cesaro sum, $g_M(\omega;\theta_0)$, to M terms of its Fourier series we may, for any $\varepsilon > 0$, choose M sufficiently large so that

$$\sup_\omega \| g(\omega;\theta_0) - g_M(\omega;\theta_0)\| < \varepsilon.$$

Now

$$\frac{1}{2\pi}\int \operatorname{tr}\{I(\omega)g(\omega;\theta_0)\}\,d\omega = \frac{1}{2\pi}\int \operatorname{tr}\{I(\omega)[g(\omega;\theta_0) - g_M(\omega;\theta_0]\}\,d\omega + \frac{1}{2\pi}\int \operatorname{tr}\{I(\omega)g_M(\omega;\theta_0)\}\,d\omega. \tag{2.3}$$

The first integral in the right-hand side of (2.3) has modulus bounded by $\varepsilon/2\pi \operatorname{tr} G(0) \to_{\text{a.s.}} \varepsilon/2\pi \operatorname{tr}\Gamma(0;\theta_0)$ so that this first term may be made arbitrarily small by choosing M large. The second integral in (2.3) is of the form

$$\operatorname{tr}\left\{\Sigma_{l=-M}^{M}\left(1 - \frac{|l|}{M}\right)g(l;\theta_0)G(l)'\right\} / (2\pi)^2.$$

But $G(l) \to_{\text{a.s.}} \Gamma(l,\theta_0)$ for each $|l| \leqslant M$ and $g(l;\theta_0)$, the lth Fourier coefficient of $g(\omega;\theta_0)$, have norms which are uniformly bounded. Hence the last displayed expression converges a.s. to

$$\operatorname{tr}\left\{\Sigma_{-M}^{M}\left(1 - \frac{|l|}{M}\right)g(l;\theta_0)\Gamma(l;\theta_0)'\right\} / (2\pi)^2.$$

In turn this expression is arbitrarily close to $\frac{1}{2\pi}\int \operatorname{tr}\{f_0(\omega)g(\omega;\theta_0)\}\,d\omega$ as required. Hence

$$\frac{\partial^2 \bar{L}_N(\dot{\theta}_N)}{\partial\theta_j\partial\theta_k} \to_p -\frac{1}{2\pi}\int \operatorname{tr}\left\{\frac{\partial f_0(\omega)}{\partial\theta_j}\frac{\partial f_0^{-1}(\omega)}{\partial\theta_k}\right\}d\omega = \Omega_{jk}.$$

Hence the CLT for $N^{1/2}(\bar{\theta}_N - \theta_0)$ reduces to that for $N^{1/2}(\partial\bar{L}_N(\theta_0)/\partial\theta)$ the jth element of which can be written as

$$N^{1/2}\frac{\partial\bar{L}_N(\theta_0)}{\partial\theta_j} = \frac{N^{1/2}}{2\pi}\int \operatorname{tr}\big[[I(\omega) - f_0(\omega)]\phi^{(j)}(\omega)\big]\,d\omega. \tag{2.4}$$

Writing $f_N(\omega)$ for the Nth order Cesaro sum of the Fourier series for $f_0(\omega)$ we have, since $f_N(\omega) = EI(\omega)$,

$$\frac{N^{1/2}}{2\pi}\int \mathrm{tr}\Big[\big[EI(\omega) - f_0(\omega)\big]\phi^{(j)}(\omega)\Big]\,d\omega = \frac{N^{1/2}}{2\pi}\int \mathrm{tr}\Big[\big[f_N(\omega) - f_0(\omega)\big]\phi^{(j)}(\omega)\Big]\,d\omega$$
$$= 0(N^{(1/2-\alpha)})$$

by Theorem 3.15 of [12]. Hence $f_0(\omega)$ may be replaced by $EI(\omega)$ in (2.4) to arrive at

$$A_N\big(\phi^{(j)}\big) = \frac{N^{1/2}}{2\pi}\int \mathrm{tr}\Big\{\big[I(\omega) - EI(\omega)\big]\phi^{(j)}(\omega)\Big\}\,d\omega, \qquad 1 \leqslant j \leqslant U. \tag{2.5}$$

To see that $\phi^{(j)}(\omega)$ may be replaced by its Cesaro sum to a finite number of terms, M, consider $A_N(\delta)$ where $\delta(\omega) = \phi^{(j)}(\omega) - \phi_M(\omega)$, $\sup_\omega\|\delta(\omega)\| \leqslant \varepsilon$ and ϕ_M is the Cesaro sum corresponding to $\phi^{(j)}$. Then

(2.6)

$$A_N(\delta) = \Sigma_{u=1}^s\Sigma_{v=1}^s\frac{N^{1/2}}{2\pi}\int\big[I_{uv}(\omega) - EI_{uv}(\omega)\big]\,\delta_{vu}(\omega)\,d\omega$$
$$= \Sigma_u\Sigma_v\big\{N^{-1/2}\Sigma\Sigma_{m,n=1}^N\big[z_u(m)z_v(n) - \gamma_{uv}(n-m)\big]\,\delta_{vu}(m-n)\big\}/(2\pi)^2,$$

where $\delta_{uv}(m-n) = \int_{-\pi}^{\pi}\delta_{uv}(\omega)e^{i(m-n)\omega}\,d\omega$. Calling the summand in braces D_{uv} we need to consider, in evaluating the variance of $A_N(\delta)$ (see Hannan [5], pages 209–211])

(2.7)

$$E\big[D_{rt}D_{uv}\big] = N^{-1}\Sigma\Sigma\Sigma\Sigma_{m,n,p,q=1}^N\{\big[\gamma_{ru}(p-m)\gamma_{tv}(q-n) + \gamma_{rv}(q-m)\gamma_{tu}(p-n)$$
$$+\Sigma\Sigma\Sigma\Sigma_{b,c,d,e=1}^s\kappa_{bcde}\Sigma_l C_{rb}(l)C_{tc}(l+n-m)C_{ud}(l+p-m)C_{ve}(l+q-m)\big]$$
$$\times\,\delta_{tr}(m-n)\,\delta_{vu}(p-q)\}$$

where $C_{rb}(l)$ denotes the (r, b)th element of $C(l, \theta_0)$ and $\gamma_{ru}(p-m) = E(z_r(m)z_u(p))$. First consider

$$|N^{-1}\Sigma_m\Sigma_n\Sigma_p\Sigma_q\gamma_{rv}(q-m)\gamma_{tu}(p-n)\,\delta_{tr}(m-n)\,\delta_{vu}(p-q)|$$
$$= |N^{-1}\int\int\big(\Sigma_m\Sigma_n\,\delta_{tr}(m-n)e^{-i(m\lambda+n\omega)}f_{rv}(\lambda)\big)$$
$$\times\big(\Sigma_p\Sigma_q\,\delta_{vu}(p-q)e^{i(q\lambda+p\omega)}f_{tu}(\omega)\big)\,d\omega\,d\lambda|.$$
$$\leqslant (2\pi B)^2\cdot B_{tr}^{1/2}B_{uv}^{1/2},$$

where

$$B_{tr} = \frac{N^{-1}}{(2\pi)^2}\int\int|\Sigma_m\Sigma_n\,\delta_{tr}(m-n)e^{-i(m\lambda+n\omega)}|^2\,d\omega\,d\lambda$$

and each typical element, $f_{ru}(\omega)$, of $f_0(\omega)$ has modulus bounded by B because $f_0(\omega)$ is continuous. Now, by simplifying the integrals it is easily shown that

$$B_{tr} = \Sigma_{l=-N+1}^{N-1}\left(1 - \frac{|l|}{N}\right)|\delta_{tr}(l)|^2 \leqslant 2\pi\int|\delta_{tr}(\omega)|^2\,d\omega \leqslant (2\pi\varepsilon)^2.$$

Hence the contribution to $ED_{rt}D_{uv}$ from the second term under the braces in (2.7) may be made arbitrarily small. Similarly for the first term. Finally consider the term arising from the fourth cumulant in (2.7)

$$\Sigma_b\Sigma_c\Sigma_d\Sigma_e\kappa_{bcde}\{N^{-1}\Sigma_m\Sigma_n\Sigma_p\Sigma_q\Sigma_l C_{rb}(l)C_{tc}(l+n-m)C_{ud}(l+p-m)$$
$$C_{ve}(l+q-m)\,\delta_{tr}(m-n)\,\delta_{vu}(p-q)\}.$$

The modulus of the factor in braces can be written

$$\Bigg|\frac{N^{-1}}{2\pi}\int\frac{1}{2\pi}\int\Sigma_m\Sigma_n\,\delta_{rt}(n-m)e^{i(n-m)\lambda}e^{-in\omega}k_{rb}(e^{i\lambda})k_{tc}(e^{i(\omega-\lambda)})\,d\lambda$$
$$\cdot\overline{\frac{1}{2\pi}\int\Sigma_p\Sigma_q\,\delta_{uv}(q-p)e^{i(q-p)\lambda}e^{-iq\omega}k_{ud}(e^{i\lambda})k_{ve}(e^{i(\omega-\lambda)})\,d\lambda}.d\Bigg|$$
$$\leqslant B\left(\frac{N^{-1}}{2\pi}\int\frac{1}{2\pi}\int|\Sigma_m\Sigma_n\,\delta_{rt}(n-m)e^{i(n-m)\lambda}e^{-in\omega}|^2\,d\lambda\,d\omega\right)^{1/2}$$
$$\cdot\left(\frac{N^{-1}}{2\pi}\int\frac{1}{2\pi}\int|\Sigma_p\Sigma_q\,\delta_{uv}(q-p)e^{i(q-p)\lambda}e^{-iq\omega}|^2\,d\lambda\,d\omega\right)^{1/2}$$

where $\int|k_{rb}(e^{i\omega})|^4\,d\omega\leqslant B<\infty$, for each r and b, since C2.2 implies that $f(\omega;\theta_0)$ is square integrable which in turn implies that $|k_{rb}(e^{i\omega})|^4$ is integrable—see [6], page 398. As before, the square of the first factor is equal to $B_{tr}\leqslant(2\pi\varepsilon)^2$. Therefore the contribution to (2.7) from the fourth cumulant term may be made arbitrarily small. Hence the CLT for $A_N(\phi^{(j)})$ in (2.5) may be established by considering the CLT for the U quantities

$$A_N(\phi_M^{(j)})=\Sigma_r\Sigma_t\Sigma_{l=-M}^{M}\left(1-\frac{|l|}{M}\right)\phi_{rt}^{(j)}(l)N^{1/2}\left\{g_{rt}(l)-\frac{N-|l|}{N}\gamma_{rt}(l)\right\}/(2\pi)^2,$$

where $g_{rt}(l)$ is the (r,t)th element of $G(l)$. Now, as in [6], $A_N(\phi_M^{(j)})$ is asymptotically equivalent to

$$\text{(2.8)}\qquad B_N(\phi_M^{(j)})=\Sigma_r\Sigma_t\Sigma_{l=-M}^{M}\left(1-\frac{|l|}{M}\right)\phi_{rt}^{(j)}(l)\tilde{\tau}_{rt}(l)/(2\pi)^2,$$

where

$$\tilde{\tau}_{rt}(l)=N^{-1/2}\Sigma_{m=1}^{N}(z_r(m)z_t(m+l)-\gamma_{rt}(l)).$$

In [6] it is proved that when C2.3 is satisfied the necessary and sufficient condition that any finite set of the $\tilde{\tau}_{rt}(l)$ by asymptotically jointly normal is that the diagonal elements of $f_0(\omega)$ be square integrable. (In [6] C2.3(d) is replaced by the stronger condition that the conditional fourth moment is a.s. constant. This is not needed in [6] nor here.) The condition C2.2 certainly ensures this so that the theorem of [6] may be applied to $B_N(\phi_M^{(j)})$ in (2.5) since M is finite. All that remains is to evaluate the asymptotic covariance between $B_N(\phi_M^{(j)})$ and $B_N(\phi_M^{(l)})$. Now,

$$E\{B_N(\phi_M^{(j)})B_N(\phi_M^{(l)})\}$$
$$=\Sigma_r\Sigma_t\Sigma_u\Sigma_v\Sigma_{n,m=-M}^{M}\left(1-\frac{|n|}{M}\right)\left(1-\frac{|m|}{M}\right)\phi_{rt}^{(j)}(n)\phi_{uv}^{(l)}(m)E[\tilde{\tau}_{rt}(n)\tilde{\tau}_{uv}(m)]/(2\pi)^4$$

which, by using the asymptotic formula ([6], equation (3)) for $E[\tilde{\tau}_{rt}(n)\tilde{\tau}_{uv}(m)]$, converges to

$$2\operatorname{tr}\left\{\frac{1}{2\pi}\int f\phi_M^{(j)}f\phi_M^{(l)}\,d\omega\right\}$$
$$+\Sigma_a\Sigma_b\Sigma_c\Sigma_d\kappa_{abcd}\left[\frac{1}{(2\pi)^2}\int k^*\phi_M^{(j)}k\,d\omega\right]_{ab}\left[\frac{1}{(2\pi)^2}\int k^*\phi_M^{(l)}k\,d\omega\right]_{cd}$$

as $N \to \infty$. By taking M large this last line is easily seen to be arbitrarily close to $2\Omega_{jl} + \Pi_{jl}$. This completes the proof. □

The result for $\bar{\theta}_N$ may be used to prove the same result for $\tilde{\theta}_N$ when C2.2 is strengthened (in the same way that the conditions of [4], Corollary 2, are strengthened).

COROLLARY 2.2. *If, in addition to the conditions of Theorem* 2.1 *it is assumed that, for each* $1 \leqslant j \leqslant U$,

C2.4. $\partial f(\omega;\theta)/\partial\theta_j$ *has elements belonging to* Λ_α, *for* $\alpha > 1/2$,
then the conclusion of Theorem 2.1 *holds for* $N^{1/2}(\tilde{\theta}_N - \theta_0)$.

PROOF. By a similar argument to that used in the proof of Theorem 2.1 it follows that $\partial^2\tilde{L}_N(\dot{\theta}_N)/\partial\theta_j\partial\theta_k \to_P \Omega_{jk}$. Then the CLT for $N^{1/2}(\tilde{\theta}_N - \theta_0)$ follows from that for the U quantities

$$N^{1/2}\frac{\partial\tilde{L}_N(\theta_0)}{\partial\theta_j} = N^{1/2}\left\{\frac{1}{2\pi}\int \operatorname{tr}\left[f_0(\omega)\phi^{(j)}(\omega)\right]d\omega - N^{-1}\Sigma_t \operatorname{tr}\left[I(\omega_t)\phi^{(j)}(\omega)\right]\right\}$$
$$(2.9)\qquad = N^{1/2}\left\{\frac{1}{2\pi}\int \operatorname{tr}\left[f_0(\omega)\phi^{(j)}(\omega)\right]d\omega - N^{-1}\Sigma_t \operatorname{tr}\left[f_N(\omega_t)\phi^{(j)}(\omega_t)\right]\right\}$$
$$-N^{1/2}\left\{N^{-1}\Sigma_t \operatorname{tr}\left[I(\omega_t) - f_N(\omega_t)\right]\phi^{(j)}(\omega_t)\right\}$$

where $f_N(\omega)$ is as before. To establish the CLT for the $N^{1/2}\partial\tilde{L}_N(\theta_0)/\partial\theta_j$ we will show that the first term in the right-hand side of (2.9) converges to zero and the second term is arbitrarily close to (2.5) in probability. Now the first term can be rewritten as

$$(2.10)\qquad N^{1/2}\left\{\frac{1}{2\pi}\int \operatorname{tr} h(\omega)\,d\omega - N^{-1}\Sigma_t \operatorname{tr} h(\omega_t)\right\}$$
$$+N^{1/2}\left\{N^{-1}\Sigma_t \operatorname{tr}\left[\left(f_0(\omega_t) - f_N(\omega_t)\right)\phi^{(j)}(\omega_t)\right]\right\}$$

where $h(\omega) = f_0(\omega)\phi^{(j)}(\omega)$. The first term in (2.10) is dominated by

$$N^{1/2}\cdot 2\pi\cdot \sup_t \sup_{|\omega-\omega_t|<\frac{2\pi}{N}}\left|\operatorname{tr}\left[h(\omega) - h(\omega_t)\right]\right| = 0(N^{1/2-\alpha})$$

where $\alpha > 1/2$ since $h(\omega) \in \Lambda_\alpha$, $\alpha > 1/2$. Hence this term in (2.10) converges to zero as $N \to \infty$. The second term in (2.10) converges to zero by a similar argument as used to replace $f_0(\omega)$ by $EI(\omega)$ in the proof of Theorem 2.1. Hence, the first term in (2.9) converges to zero. Consider the second term in (2.9). Define

$$\tilde{A}_N(\delta) = \Sigma\Sigma_{u\,v=1}^{s} N^{-1/2}\Sigma_t[I_{uv}(\omega_t) - EI_{uv}(\omega_t)]\,\delta_{vu}(\omega_t)$$

where $\delta(\omega)$ is as in the proof of Theorem 2.1. Then, letting

$$\delta_{vu}^{(N)}(l) = N^{-1}\Sigma_t\,\delta_{vu}(\omega_t)e^{il\omega_t}$$

$\tilde{A}_N(\delta)$ may be rewritten as in (2.6) with $\delta_{vu}(m-n)$ replaced by $\delta_{vu}^{(N)}(m-n)$. Making this replacement throughout the argument (given in the proof of Theorem 2.1) used to show that $E[D_{rt}D_{uv}]$ in (2.7) can be made arbitrarily small by choosing M large in $\phi_M^{(j)}(\omega)$, the corresponding conclusion will follow here provided $\Sigma_{l=-N+1}^{N-1}(1-(|l|/N)\,\delta_{tr}^{(N)}(l)|^2 \leqslant B\varepsilon^2$, $B < \infty$. But

$$\begin{aligned}\Sigma_{l=-N+1}^{N-1}\left(1-\frac{|l|}{N}\right)|\delta_{tr}^{(N)}(l)|^2 &\leqslant \Sigma_{l=-N+1}^{N-1}|N^{-1}\Sigma_t\,\delta(\omega_t)e^{il\omega_t}|^2\\ &\leqslant 2\Sigma_{l=0}^{N-1}|N^{-1}\Sigma_t\,\delta(\omega_t)e^{il\omega_t}|^2\\ &= 2N^{-1}\Sigma_t|\delta(\omega_t)|^2 \leqslant 2\varepsilon.\end{aligned}$$

Hence $\tilde{A}_N(\phi^{(j)})$ may be replaced by $\tilde{A}_N(\phi_M)$. But

$$A_N(\phi_M) - \tilde{A}_N(\phi_M) = \Sigma_{l=-M}^{M}\left(1-\frac{|l|}{M}\right)\phi(l)N^{1/2}E(l)$$

where $E(l) = N^{-1}\Sigma_{n=N-l+1}^{N}[z(n+l-N)z(n)' - \gamma(N-l)]$ for $l \geqslant 0$ and $E(-l) = E(l)'$. But the Fourier coefficients, $\phi(l)$, of $\phi^{(j)}(\omega)$ have elements which are uniformly bounded for $|l| \leqslant M$ and each element of $N^{1/2}E(l)$ is of the form $N^{-1/2}\{z_a(l)z_b(N-k) - \gamma_{ab}(N-k-l)\}$ which converge in probability to zero for $|l|, |k| \leqslant M$. Hence the CLT for the quantities $N^{1/2}\partial\tilde{L}_N(\theta_0)/\partial\theta_j$ is the same as that for the $N^{1/2}\partial\bar{L}_N(\theta_0)/\partial\theta_j$ and the proof is complete. □

Remark 1. If $\log\det K(\theta) = 1/2\pi\int \log\det 2\pi f(\omega;\theta)\,d\omega$ in the definition of $\tilde{L}_N(\theta)$ is replaced by $N^{-1}\Sigma_t \log\det 2\pi f(\omega_t;\theta)$ (as in Davies [2], for example) then, provided the strong law holds, $\tilde{\theta}_N$ minimising the modified expression satisfies the above CLT without requiring C2.4 but only C2.1-C2.3. That this is so follows from the fact that in the expression corresponding to (2.9) the first term will be null. The remaining arguments in the proof of Corollory 2.2 only require C2.1-C2.3.

Remark 2. We have not investigated the CLT for $N^{1/2}(\hat{\theta}_N - \theta_0)$ for reasons of space. It is very likely true that the result for this case holds with C2.2 strengthened along similar lines to the extra conditions needed to prove Corollary 3 of [4].

Remark 3. When $k(e^{i\omega},\theta)$ and $K(\theta)$ are differentiable functions of θ the Π_{jl} term in the covariance matrix (2.1) simplifies to

$$\Pi_{jl} = \Sigma\Sigma\Sigma\Sigma_{a,b,c,d=1}^{s}\kappa_{abcd}\left(K_0^{-1}\frac{\partial K_0}{\partial\theta_j}K_0^{-1}\right)_{ab}\left(K_0^{-1}\frac{\partial K_0}{\partial\theta_l}K_0^{-1}\right)_{cd}$$

since $(2\pi)^{-1}\int k_0\partial k_0/\partial\theta_j\,d\omega$ is null for example.

REMARK 4. It is important, from a practical point of view, to consider the circumstances under which the fourth cumulant term (i.e., Π), in the limiting covariance matrix (2.1), vanishes. If κ_{abcd} is zero for all subscript values (as would occur if the $\varepsilon(m)$ were Gaussian) then Π is null so that the asymptotic covariance is $2\Omega^{-1}$. When θ' partitions as (τ', μ')—the case discussed in [4]—the matrix (2.1) reduces to

$$\begin{bmatrix} 2\Omega^{(1)^{-1}} & 0 \\ 0 & \Omega^{(2)^{-1}}(2\Omega^{(2)} + \Pi^{(2)})\Omega^{(2)^{-1}} \end{bmatrix} \tag{2.11}$$

where, for example, $\Omega^{(1)}$ has the dimensions of τ. In [4] the CLT for $\bar{\tau}_N$, etc., was established without assuming the existence of moments higher than the second order for $\varepsilon(m)$ with asymptotic covariance $2\Omega^{(1)^{-1}}$ which does not depend upon the κ_{abcd}. In general, for Π to be the null matrix it is necessary and sufficient (given the conditions of Remark 3) that $K_0^{-1}(\partial K_0/\partial\theta_j)K_0^{-1}$ be the null matrix for all j. (Compare Whittle [11], pages 13–14).

3. Applications to models for a signal observed with noise. Consider the following model for a stationary signal observed with noise

$$z(n) = y(n) + x(n) \tag{3.1}$$

where the "signal" $y(n)$ and the "noise" $x(n)$ are incoherent (i.e., $Ey(m)x(n)' = 0$ all m, n) and are each of the type (1.1). That is

$$\begin{aligned} y(n) &= \Sigma_{j=0}^{\infty} C_y(j;\, \theta_y)\varepsilon_y(n-j) \\ x(n) &= \Sigma_{j=0}^{\infty} C_x(j,\, \theta_x)\varepsilon_x(n-j) \end{aligned} \tag{3.2}$$

where $\varepsilon_y(n)$ and $\varepsilon_x(n)$ are at least white noise with covariance matrices $K_y(\theta_y)$ and $K_x(\theta_x)$ respectively. The spectral density matrix of $z(n)$ is

$$f_z(\omega, \theta_z) = f_y(\omega, \theta_y) + f_x(\omega;\, \theta_x) \tag{3.3}$$

where $f_y = (1/2\pi)k_y K_y k_y^*$, $f_x = (1/2\pi)k_x K_x k_x^*$ and $\theta_z' = (\theta_y', \theta_x')$ or is an equivalent reparametrisation of the problem. Since $z(n)$ is also zero mean, stationary and purely nondeterministic it has Wold decomposition

$$z(n) = \Sigma_{j=0}^{\infty} C_z(j,\, \theta_z)\varepsilon_z(n-j) \tag{3.4}$$

in which $\varepsilon_z(n)$ is white noise with covariance matrix $K_z(\theta_z)$. Here we have implicitly assumed that the $C_z(j;\, \theta_z)$ and $K_z(\theta_z)$ are matrix functions of the parameter θ_z. This will be the case in many examples. The spectral density matrix of $z(n)$ in (3.3) may now also be written as

$$f_z(\omega;\, \theta_z) = \frac{1}{2\pi} k_z(e^{i\omega};\, \theta_z)K_z(\theta_z)k(e^{i\omega};\, \theta_z)^*. \tag{3.5}$$

Using the minimisation criteria $\hat{L}_N$, $\tilde{L}_N$, $\bar{L}_N$ θ_z may be estimated using observations $z(1), \cdots, z(N)$. We will be concerned with relating the CLT of Section 2 to the above signal plus noise model and will assume that, e.g., $\bar{\theta}_{z,N}$ converges a.s. to the unique limit θ_{z0}. In particular, if $z(n)$ is ergodic, if $z(n)$ can be represented in the form (3.4), and if conditions B of [4] hold for the parameter space to which θ_z belongs and for k_z and K_z as functions of ω and θ_z, then the strong consistency of $\bar{\theta}_{z,N}$, etc., holds. Moreover, it is not required that k_z and K_z be separately parameterised for this result. In signal plus noise models of the above type it is not always true that K_x and k_x may be separately parameterised because of the complicated way in which the factorisation of f_z in (3.5) is arrived at commencing from the original parameters θ_y and θ_x. Thus the CLT of [4] may not always apply to such models (see Section 4, however, for an example where the model can be reparameterised so that the CLT just cited applies). However, when the $\varepsilon_z(m)$ satisfy all of C2.3 then the CLT of Theorem 2.1 (and Corollary 2.2) may be applied. On the other hand the condition that (at least) $\varepsilon_z(m)$ be martingale differences with respect to their past may not be appropriate since it is not always the prediction of $z(n)$ on its past which is of interest but rather the prediction of $y(n)$ on the past of $z(n)$. Recall that when the prediction of $z(n)$ is of interest C2.3(a) can be given a natural interpretation in terms of linear modelling—see [7].

A corresponding discussion of the relevance of condition C2.3(a) will now be given in the context of signal plus noise models where $x(n)$ is now taken to be at least white noise with covariance K_x. The signal $y(n)$ will be taken to be a stationary nondeterministic process with one sided representation

$$y(n) = \Sigma_{j=0}^{\infty} C_y(j)\varepsilon_y(n-j), \qquad C_y(0) = I_s \tag{3.6}$$

in which the $\varepsilon_y(n)$ form a white noise sequence with covariance matrix K_y. If $x(n)$ and $y(n)$ are incoherent then (see (3.3))

$$f_z(\omega) = (2\pi)^{-1}\{k_y(e^{i\omega})K_y k_y(e^{i\omega})^* + K_x\} \tag{3.7}$$

which may also be written in the following form (see (3.5))

$$f_z(\omega) = (2\pi)^{-1}k_z(e^{i\omega})K_z k_z(e^{i\omega})^*. \tag{3.8}$$

Reference to θ_x, θ_y, θ_z has been suppressed in the above because these parameters play no part in the following discussion.

Since only $z(n)$ is observed in (3.1) then prediction of the signal $y(n)$ must be based on $z(m)$, $m \leqslant n-1$. Once the $C_z(j)$ corresponding to k_z in (3.8) are known the optimal linear filter for predicting $y(n)$ in terms of $z(m)$, $m \leqslant n-1$, may be constructed using only these $C_z(j)$ (see the proof of the theorem to follow). It is therefore relevant, if the above model for $z(n)$ (with $y(n)$ as in (3.6)) is correct, to impose the condition that the best linear predictor of $y(n)$ is the best predictor of $y(n)$ (both based on $z(m)$, $m \leqslant n-1$ and both "best" in the least squares sense). To discuss this the following notation is convenient. Let

(3.9) $\hat{y}(n)$ = the best predictor of $y(n)$ given $z(m)$, $m \leqslant n-1$;

$\dot{y}(n)$ = the best linear predictor of $y(n)$ given $z(m)$, $m \leqslant n-1$;

$\hat{z}(n)$ = the best predictor of $z(n)$ given $z(m)$, $m \leqslant n-1$;

$\dot{z}(n)$ = the best linear predictor of $z(n)$ given $z(m)$, $m \leqslant n-1$.

Let $\mathcal{F}_z(n)$ be the σ-algebra generated by the elements of $z(m)$ (equivalently by the elements of $\varepsilon_z(m)$ given in (3.4)) for $m \leqslant n$ and let $\mathcal{F}_x(n)$, $\mathcal{F}_y(n)$ be similarly defined for $x(m)$, $y(m)$, respectively. Then $\hat{y}(n) = E(y(n)|\mathcal{F}_z(n-1))$ and $\hat{z}(n) = E(z(n)|\mathcal{F}_z(n-1))$.

THEOREM 3.1. *If* $x(n), y(n)$ *and* $z(n) = y(n) + x(n)$ *are as described above and if*
C3.1. $E(x(m)|\mathcal{F}_z(m-1)) = 0$ a.s. *all* m,
then $\hat{z}(n) = \dot{z}(n)$ *if and only if* $\hat{y}(n) = \dot{y}(n)$.

PROOF. The result will be established in two steps. The first shows that $\dot{y}(n) = \dot{z}(n)$, the second that $\hat{y}(n) = \hat{z}(n)$.

(i) $\dot{y}(n) = \dot{z}(n)$. By [5], Theorem 10[1], page 173, the response function of the optimal linear filter for $y(n)$ given $z(m)$, $m \leqslant n-1$ is

$$h(e^{i\omega}) = e^{i\omega}\Big[e^{-i\omega}f_y(e^{i\omega})k_z^*(e^{i\omega})^{-1}\Big]_+ K_z^{-1}k_z(e^{i\omega})^{-1}$$

where $[g(e^{i\omega})]_+$ denotes that only nonnegative powers of $e^{i\omega}$ are to be taken in the (matrix) series expansion of the (matrix) function $g(e^{i\omega})$. Using (2.2) f_y may be replaced in $[e^{i\omega}f_y(e^{i\omega})k_z^*(e^{i\omega})^{-1}]_+$ to reach

$$h(e^{i\omega}) = e^{i\omega}\Big[e^{-i\omega}k_z(e^{i\omega}) - e^{-i\omega}K_x k_z^*(e^{i\omega})^{-1}K_z^{-1}\Big]_+ k_z(e^{i\omega})^{-1}.$$

But the second term in $[\ ]_+$ of this expression has only negative powers of $e^{i\omega}$ so that it makes no contribution to $h(e^{i\omega})$. Thus

$$\begin{aligned} h(e^{i\omega}) &= e^{i\omega}\big(\textstyle\sum_{j=1}^{\infty} C_z(j)e^{i(j-1)\omega}\big)k_z(e^{i\omega})^{-1} \\ &= I_s - k_z(e^{i\omega})^{-1}. \end{aligned}$$

On the other hand, the transfer function giving $\dot{z}(n)$ is easily obtained (see Theorem 1, page 129 and Theorem 1″, page 163 of [5]) as $h(e^{i\omega})$. Thus $\dot{z}(n) = \dot{y}(n)$.

(ii) $\hat{y}(n) = \hat{z}(n)$. Now

$$\begin{aligned} \hat{z}(n) &= E(z(n)|\mathcal{F}_z(n-1)) \\ &= E(y(n)|\mathcal{F}_z(n-1)) + E(x(n)|\mathcal{F}_z(n-1)) \\ &= \hat{y}(n) \quad \text{a.s.}, \end{aligned}$$

since, by assumption C3.1, the second term in the previous expression is null. □

Note that $\dot{y}(n) = \dot{z}(n)$ even when C3.1 does not hold. Insofar as the requirement $\dot{y}(n) = \hat{y}(n)$ is natural (for linear modelling to be appropriate) then the additional condition C3.1 ensures that $\dot{z}(n) = \hat{z}(n)$. But $\dot{z}(n) = \hat{z}(n)$ if and only if

$$E(\varepsilon_z(n)|\mathcal{F}_z(n-1)) = 0 \quad \text{a.s. all } m, \tag{3.10}$$

[7]. This means that part (a) of C2.3 is satisfied for $\varepsilon_z(m)$. It would be of interest to determine the minimal conditions under which the remainder of C2.3 hold for $\varepsilon_z(m)$ but we have not done this. Some conditions on $y(n)$ and $x(n)$ which ensure C3.1 are as follows:

(a) If the $x(n)$ are serially independent and independent of $y(n)$ then C3.1 holds.

(b) If $\xi(n)' = (\varepsilon_y(n)', x(n)')$ and $\mathcal{F}_\xi(n)$ is the σ-algebra generated by $\xi(m)$, $m \leqslant n$, then, provided $E(\xi(m)|\mathcal{F}_\xi(m-1)) = 0$ a.s. all m, the condition C3.1 holds.

Other conditions are no doubt available to ensure C3.1.

The result of Theorem 3.1 is useful not only in justifying (part of) the condition C2.3 imposed in Theorem 2.1, (in which θ could parameterise both k and K) but also in relation to the discussion of the model considered in Section 4. There it is shown that, after suitable reparameterisation, the treatment given in [4] may be applied to that model. For the CLT of [4] to apply to the parameters (τ defined by (4.6)) specifying the linear predictors discussed above, C2.3(a) and (b) are only required for the $\varepsilon_z(m)$.

If the imposition of C2.3 cannot be justified for the *general* model (3.1) then we may proceed as follows. Let $y(n)$ and $x(n)$ be of the form (3.2), let $\eta(n)' = (\varepsilon_y(n)', \varepsilon_x(n)')$ and define $\mathcal{F}_\eta(n)$ as the σ-algebra generated by $\eta(m)$, $m \leqslant n$. We will assume that $\eta(n)$ satisfies C2.3 but will now refer to the constants in this as $K_{ab}^{(\eta)}$, $\beta_{abc}^{(\eta)}$ and $\kappa_{abcd}^{(\eta)}$, respectively. However, it need not be true that the $\varepsilon_z(m)$ in the representation (3.4) satisfy C2.3 in this case. An alternative representation of $z(n)$ is

$$z(n) = \Sigma_{j=0}^{\infty} D(j;\, \theta_z)\eta(n-j) \tag{3.11}$$

where $D(j;\, \theta_z) = [C_y(j;\, \theta_y) \vdots C_x(j;\, \theta_x)]$. If now C2.1 and C2.2 are assumed for $f_z(\omega;\, \theta_z)$ and C2.3 is assumed for $\eta(n)$ then Theorem 2.1 continues to hold. To see that this is plausible consider the following. The proof of Theorem 2.1 up to equation (2.6) is not changed by this new specification since only C2.1 and C2.2 are required in the argument to this point. Now the fourth cumulant term in (2.6) needs to be modified by increasing the upper limit of summation to $2s$, replacing κ_{bcde} by $\kappa_{bcde}^{(\eta)}$ and using $c_{rb}(l)$ to denote the (r, b)th element of the $s \times 2s$ matrix $D(l)$ given above. This replacement does not effect the validity of the steps in the proof that $A_N(\delta)$ has negligible variance. Thus the proof up to (2.8) is valid here also. By examining the proof of the theorem in [6] the same replacements as described above may be made and that theorem still applies. (That is, the proof of the theorem in [6] does not require that $D(j)$ be square.) Thus the CLT for $\bar{\theta}_{z,N}$ holds for this case. The asymptotic covariance matrix (2.1) may be obtained (see, in particular, Π_{jl}) in the form stated but with the elements of $k(e^{i\omega})$ replaced by those of the $s \times 2s$ matrix $\Sigma_0^\infty D(j)e^{ij\omega}$, the range of summation increased to $2s$ and κ_{abcd} replaced by $\kappa_{abcd}^{(\eta)}$.

When k_z and K_z are separately parameterised by τ_z and μ_z but C2.3 only holds for $\eta(n)$ and not $\varepsilon_z(n)$ then the discussion just given still applies. Furthermore the asymptotic covariance matrix of Theorem 2.1 reduces to (2.11) and, as discussed in Remark 4 above, this means that the asymptotic covariance for τ_z does not depend on fourth moments (i.e., does not depend on $\kappa_{abcd}^{(\eta)}$). It may be true that in this case the CLT for $N^{1/2}(\bar{\tau}_{z,N} - \tau_{z,0})$ can be established along similar lines to that given in [4] without the assumption that moments higher than the second exist. Finally, if the $\varepsilon_x(n)$ and $\varepsilon_y(n)$ are Gaussian and independent the fourth cumulant term Π vanishes from the asymptotic covariance (2.1). Also in this Gaussian case, the $\varepsilon_z(n)$

satisfy C2.3 directly so there is no need for the modification to the proof of Theorem 2.1 mentioned above.

4. The model for an autoregressive signal observed with noise. Recently Pagano [10] has considered the model (3.1), for $s = 1$, in which $y(n)$ satisfies the autoregression

$$\Sigma_{j=0}^{q}\beta(j)y(n-j) = \varepsilon_y(n),\ \beta(0) = 1,\ E\varepsilon_y(m)\varepsilon_y(n) = \delta_{mn}\sigma_y^2, \tag{4.1}$$

where $h(\zeta) = \Sigma_{j=0}^{q}\beta(j)\zeta^j$ has all zeros outside the unit circle (i.e., for $|\zeta| > 1$) and $x(n)$ is white noise with $Ex(m)x(n) = \delta_{mn}\sigma_x^2$. In the above, $x(n)$ and $y(n)$ are taken to be at least incoherent. Pagano [10] considers the estimation of the parameters $\beta(1), \cdots, \beta(q), \sigma_y^2, \sigma_x^2$ and establishes the consistency and asymptotic efficiency of the estimators when it is assumed that $x(n)$ and $y(n)$ are independent one to another and are Gaussian. These estimators are obtained by a nonlinear least squares regression method using consistent estimators of $\beta(1), \cdots, \beta(q)$ and of the covariances, at lags $l = 0, \cdots, q + 1$, of $z(n)$. In the following we will take a different approach and base estimation on $\hat{L}_N$, $\tilde{L}_N$, $\bar{L}_N$. It will be convenient to define the vector of parameters

$$\theta' = \{\beta(1), \cdots, \beta(q), \sigma_y^2, \sigma_x^2\}. \tag{4.2}$$

The vector θ will be taken to belong to the set

$$\Theta = \{\theta \in R^{(q+2)} : h(\zeta) \neq 0, |\zeta| \leqslant 1; \beta(q) \neq 0; \sigma_x^2 > 0, \sigma_y^2 > 0\}. \tag{4.3}$$

(The assumption that $\beta(q) \neq 0$ presupposes that the true degree, q, of the autoregression is known. This is an identification requirement for the above autoregressive signal plus noise model.)

In the model just described the spectral density of $z(n)$ (given in (3.3)) becomes

$$f_z(\omega; \theta) = \frac{1}{2\pi}\left\{\frac{\sigma_y^2}{|\Sigma_0^q\beta(j)e^{ij\omega}|^2} + \sigma_x^2\right\}. \tag{4.4}$$

As is well known (see [10], for example), this may be rewritten in the form (3.5) as

$$f_z(\omega; \theta) = \frac{\sigma_z^2}{2\pi}\cdot\frac{|\Sigma_0^q\alpha(j)e^{ij\omega}|^2}{|\Sigma_0^q\beta(j)e^{ij\omega}|^2}, \tag{4.5}$$

where σ_z^2 and $\alpha(1), \cdots, \alpha(q)$ depend on θ. In (4.5) it is always possible to choose the $\alpha(j)$ so that $g(\zeta) = \Sigma_{j=0}^{q}\alpha(j)\zeta^j$ has all zeros outside the unit circle.

The procedure $\bar{L}_N(\theta)$ (which will be the only one discussed below since the discussion for $\hat{L}_N$ and $\tilde{L}_N$ is similar) may be used to obtain the estimator of θ defined in (4.2) as $\bar{\theta}_N$. Since it is our intention to discuss only the CLT for $\bar{\theta}_N$ below and not the SLLN we will assume that $\bar{\theta}_N \to_{\text{a.s.}} \theta_0$. (If, for example, θ_0 belongs to that subset of Θ for which $0 < c \leqslant \sigma_x^2, \sigma_y^2 \leqslant b < \infty$ then the conditions for the strong law of [4] will apply.) For the parameterisation of the problem chosen (i.e., θ in (4.2)) the CLT of [4] will not apply since both σ_z^2 and $k(e^{i\omega}) = (\Sigma_0^q\beta(j)e^{ij\omega})^{-1}(\Sigma_0^q\alpha(j)e^{ij\omega})$ depend on the same vector of parameters θ. (That this joint dependence on θ is not vacuous may be seen in the simplest example of the above type, i.e., when $q = 1$, as discussed on page 171 of [5].) However, the CLT of Section 2 may be applied in this case to $N^{1/2}(\bar{\theta}_N - \theta_0)$ as follows. In the first place, conditions C2.1 and C2.2 may be shown to hold on the parameter space Θ defined in (4.3)—see also the discussion below. Secondly, when the $x(n)$ are taken as serially independent and independent of $y(n)$, Theorem 3.1 applies so that part (a) of C2.3 is not an unnatural requirement here. If, further, the remainder of C2.3 is assumed then the conditions of Theorem 2.1 hold so that $N^{1/2}(\bar{\theta}_N - \theta_0)$ has the asymptotic distribution of that theorem.

However, as will now be discussed, there is a different parameterisation of the above model which has certain advantages. We propose to take as parameters

$$\tau' = \{\beta(1), \cdots, \beta(q), \alpha(q)\} \tag{4.6}$$

and σ_z^2. The appropriate space to consider for τ and σ_z^2 is $T \times R^+$, where R^+ is the positive real line and

$$T = \{\tau \in R^{q+1} : h(\zeta) \neq 0, |\zeta| \leqslant 1;\ 0 < \alpha(q)/\beta(q) < 1;\ \beta(q) \neq 0\}. \tag{4.7}$$

To relate the different parameterisations, θ and (τ, σ_z^2), and the corresponding spaces, Θ and $T \times R^+$, we proceed as follows. Equating (4.4) and (4.5), multiplying through by $|\Sigma_0^q \beta(j)e^{ij\omega}|^2$ and taking Fourier coefficients gives

$$\sigma_z^2 \Sigma_{j=0}^{q-l} \alpha(j)\alpha(j+1) = \delta_{0l} \cdot \sigma_y^2 + \sigma_x^2 \Sigma_{j=0}^{q-l} \beta(j)\beta(j+1), \qquad 0 \leqslant l \leqslant q. \tag{4.8}$$

Now, given any $\theta \in \Theta$, there exists a unique set of $\alpha(1), \cdots, \alpha(q), \sigma_z^2$ with $g(\zeta) = \Sigma_0^q \alpha(j)\zeta^j$ having all zeros outside the unit circle. (That this is a unique set follows from the fact that there is only one such factorisation of the spectral density $\sigma_y^2 + \sigma_x^2|\Sigma_0^q \beta(j)e^{ij\omega}|^2$ which corresponds to a moving average with all zeros outside the unit circle.) However, if $\alpha(q)$ and $\beta(1), \cdots, \beta(q)$ are uniquely specified then $\alpha(1), \cdots, \alpha(q-1)$ may be obtained by solving

$$\Sigma_{j=1}^{q-1} \alpha(j)\alpha(j+1) - \frac{\alpha(q)}{\beta(q)} \Sigma_{j=0}^{q-1} \beta(j)\beta(j+1) = 0, \qquad 1 \leqslant l \leqslant q-1. \tag{4.9}$$

(These are just (4.8), for $1 \leqslant l \leqslant q-1$, with $\sigma_z^2 \alpha(q)/\beta(q) = \sigma_x^2$, from the equation (4.8) for $l = q$, substituted.) The solution to (4.9) for $\alpha(1), \cdots, \alpha(q-1)$ may be shown to be unique as follows. Call each left-hand side in (4.9) $\Phi_l(\beta(1), \cdots, \beta(q), \alpha(1), \cdots, \alpha(q-1))$ for $1 \leqslant l \leqslant q-1$. Then, by the implicit function theorem, there is a unique solution $(\alpha(1), \cdots, \alpha(q-1))$ in terms of $(\beta(1), \cdots, \beta(q), \alpha(q))$ to (4.9) provided the $(q-1) \times (q-1)$ matrix with (l, k)th element $\partial\Phi_l/\partial\alpha(k)$, for $1 \leqslant k, l \leqslant q-1$, has full rank. This is just the condition that the matrix

$$\begin{bmatrix} 1 & & & 0 \\ \alpha(1) & \ddots & & \\ \vdots & \ddots & \ddots & \\ \alpha(q-2) & \cdots & \alpha(1) & 1 \end{bmatrix} + \begin{bmatrix} \alpha(2) & \cdots & \alpha(q-1) & \alpha(q) \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \\ \alpha(q-1) & \cdot^{\cdot^{\cdot}} & & 0 \\ \alpha(q) & & & \end{bmatrix} \tag{4.10}$$

be of full rank. But, by the results in Marden [8], pages 152–155, (4.10) has nonzero determinant if the zeros of $g(\zeta) = \Sigma_0^q \alpha(j)\zeta^j$ are all outside the unit circle, which is true here. Hence, the $\alpha(j)$, for $1 \leq j \leq q-1$, in (4.5) may be written as functions only of τ and we will write $\alpha(j, \tau)$, $1 \leq j \leq q-1$, to emphasize this. Furthermore, since the Φ_l defined above are (at least) twice continuously differentiable in $\alpha(1), \cdots, \alpha(q-1)$ and (4.10) is of full rank the functions $\alpha(j; \tau)$ solving (4.9) are also twice continuously differentiable functions of τ (see Matsushima [9], page 24, for example). This remark is of relevance to the CLT to be discussed shortly. For the given $\theta \in \Theta$ there exist $\sigma_z^2, \alpha(1), \cdots, \alpha(q)$ such that (4.8) holds. But, as we have just seen $\alpha(1), \cdots, \alpha(q-1)$ may be obtained in terms of $\tau' = (\beta(1), \ldots, \beta(q), \alpha(q))$. Furthermore, this τ belongs to T since by the equation (4.8) for $l = q$, since $\sigma_z^2 > \sigma_x^2 > 0$, it follows that $\alpha(q)$ have the same sign, i.e., that $\alpha(q)/\alpha(q) > 0$ and that $\alpha(q)/\beta(q) < 1$. (Note, since $|\beta(q)| < 1$ then $|\alpha(q)| < 1$ also.) If $\bar{\theta}_N \to \theta_0$ a.s. then $\bar{\theta}_N, \bar{\sigma}_{z,N}^2$ (corresponding to $\bar{\theta}_N$) will converge a.s. to $\theta_0, \sigma_{z,0}^2$ (corresponding to θ_0). Now take a small open neighborhood of $(\tau_0, \sigma_{z,0}^2) \in T \times R^+$. For fixed τ, σ_z^2 may vary in this neighborhood and as it varies there will be a signal plus noise model of the above type (i.e., a parameter $\theta \in \Theta$) corresponding to each σ_z^2. On the other hand, fix σ_z^2 and let τ vary in this neighborhood. Then there is also a $\theta \in \Theta$ corresponding to this τ. This means that for N large (at least) $\bar{L}_N$ may be minimized as a function of τ and σ_z^2 with τ specifying $k(e^{i\omega}; \tau) = (\Sigma_0^q \beta(j) e^{ij\omega})^{-1}(\Sigma_0^q \alpha(j; \tau) e^{ij\omega})$ alone and σ_z^2 may vary freely from τ. Since $k(e^{i\omega}; \tau)$ is a twice continuously differentiable function of τ (see the argument below (4.10)) condition C1 of [4] is satisfied. We have already noted that when $x(n)$ are serially independent and independent of $y(n)$ part (a) of C2.3 for $\varepsilon_z(n)$ in (using the new notation)

$$(4.11) \qquad z(n) = \Sigma_0^\infty C(j; \tau)\varepsilon_z(n-j), \qquad k(e^{i\omega}; \tau) = \Sigma_0^\infty C(j; \tau) e^{ij\omega},$$

is not unreasonable. But C2.3 (a) is the first part of [4], C3. Hence assuming also the second part of [4], C3 the CLT of [4] may be applied to yield the asymptotic normality of $N^{1/2}(\bar{\tau}_N - \tau_0)$ without, in particular, the extra moment conditions (see C2.3) of Theorem 2.1. The CLT of [4], Theorem 5, for $N^{1/2}(\bar{\mu}_N - \mu_0)$ applies also to $N^{1/2}(\bar{\sigma}_{z,N}^2 - \sigma_{z,0}^2)$.

One advantage of the above parameterisation in terms of τ and σ_z^2 is that the CLT for the vector τ may be established under more general conditions than can the CLT for θ in (4.2). Since the transfer function giving the best linear predictor of $y(n)$ based on $z(m)$, $m \leq n-1$ is completely specified by $\alpha(1), \cdots, \alpha(q), \beta(1), \cdots, \beta(q)$ (see Section 3), that is by τ, the efficient estimation of τ may be of principal interest. Also it would appear no easier to estimate θ than (τ, σ_z^2) via $\bar{L}_N$, $\tilde{L}_N$ or $\hat{L}_N$.

Acknowledgment. The author would like to thank Professor E. J. Hannan of the Australian National University who contributed greatly to this work which, in its original form, was part of the author's Ph.D. thesis.

REFERENCES

[1] BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217–226.

[2] DAVIES, R. B. (1973). Asymptotic inference in stationary Gaussian time series. *Adv. Appl. Probability* **5** 469–497.

[3] DEISTLER, M., DUNSMUIR, W. and HANNAN, E. J. (1978). Vector linear time series models corrections and extensions. *Advances Appl. Probability.* **10** 360–372.

[4] DUNSMUIR, W. and HANNAN, E. J. (1976). Vector linear time series models. *Adv. Appl. Probability* **8** 339–364.

[5] HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

[6] HANNAN, E. J. (1976). The asymptotic distribution of serial covariances. *Ann. Statist.* **4** 396–399.

[7] HANNAN, E. J. and HEYDE, C. C. (1972). On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.* **43** 2058–2066.

[8] MARDEN, M. (1949). *The Geometry of the Zeros of a Polynomial in a Complex Variable*. American Mathematical Society, New York.

[9] MATSUSHIMA, Y. (1972). *Differentiable Manifolds*. Marcel Dekker, New York.

[10] PAGANO, M. (1974). Estimation of models of auto-regressive signal plus white noise. *Ann. Statist.* **2** 99–108.

[11] WHITTLE, P. (1961). Gaussian estimation in stationary time series. *Bull. Internat. Statist. Inst.* **33** 1–26.

[12] ZYGMUND, A. (1959). *Trigonometric Series, Vol. I.* Cambridge Univ. Press.

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139