COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 298

NON-LINEAR TIME SERIES REGRESSION

E.J. Hannan

September 14, 1970

Non-Linear Time Series Regression

E. J. Hannan
Australian National University*

1. Introduction

In Jennrich (1969) the model

$$(1) \qquad y(n) = z(n;\ \theta_o) + x(n)\ , \qquad n = 1, \ldots, N$$

is considered, where $x(n)$ is a sequence of i.i.d. $(0, \sigma^2)$ random variables and $z(n;\ \theta)$ is a continuous but non-linear function of $\theta \in \Theta$, Θ being a compact set in R^p. We shall use a second subscript when referring to a particular coordinate of θ_o so that θ_{oj} is the jth coordinate. Of course $z(n;\ \theta)$ must also satisfy other requirements, which we discuss below.

Our main purpose here is to extend these results to the case where $x(n)$ is generated by a stationary time series. Our main theorems are true under the following assumptions.

A. The sequence $x(n)$ is of the form

$$x(n) = \sum_{-\infty}^{\infty} \alpha(j)\epsilon(n-j)\ , \qquad \sum_{-\infty}^{\infty} \alpha(j)^2 < \infty$$

---

* This research was done while the author was at the Cowles Foundation for Research in Economics at Yale University and was supported by the National Science Foundation and the Ford Foundation.

where the $\alpha(j)$ are i.i.d. (0, 1). The spectrum, $f(\lambda) =$

$$(2\pi)^{-1}\left|\sum_{-\infty}^{\infty}\alpha(j)\exp i\, j\lambda\right|^2 ,$$ is a continuous function.

An alternative specification, which is in some ways more plausible, is that $x(n)$ is stationary with absolutely continuous spectrum and continuous spectral density function and satisfies a uniform mixing condition (see Rosenblatt (1956) or Rozanov (1967) p. 180. Rozanov calls a process satisfying this condition, completely regular). We shall speak of A'' when we refer to the condition of this last sentence. Rather than state two forms of our theorems (or complicated forms of them) we shall merely give some comments below each theorem relating to this alternative specification or to a weakening of A.

We have a number of examples in mind which are of the form of (1) and are specifically of a time series nature. Other examples not so closely related to time series are given in Jennrich (1969).

(i) $y(n) = \beta\sum_{0}^{\infty}\theta^j z(n-j)+x(n), \qquad |\theta| \leqq 1-\delta, \qquad \delta > 0$

(ii) $y(n) = \beta z(n-\theta)+x(n), \qquad |\theta| \leqq 1$

(iii) A linear time series regression in which some subset, θ, of the regression coefficients is constrained to lie within a compact set while the remaining coefficients, β, are not constrained.

(iv) $y(n) = \beta_1+\beta_2\cos n\theta+\beta_3\sin n\theta+x(n), \qquad |\theta| \leqq \pi$.

The examples (i), (ii), (iv) may be greatly generalised and we have considered them in these special forms only for simplicity of understanding. In these examples the parameter set has been divided into a subset β and a subset θ, the former occurring linearly. Such prior information about β as exists may be too vague, in relation to the evidence of the data, to be worth using and it may be preferable, or necessary, to leave β unconstrained since this may reduce the calculations. To avoid too complex a notation involved in a (rather trivally) more general treatment we shall continue to act as though θ comprises all parameters and lies in a compactum. However all of the theorems stated below continue to hold if a subset does not lie in a compact set provided that subset occurs linearly.

We adopt the following conditions, which are extensions of those introduced in Jennrich (1969).

B. Uniformly in θ', θ'' the following limits exist almost surely[1] and independently of the $x(n)$.

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} z(m;\theta')z(m+n;\theta''), \qquad n = 0, \pm 1, \pm 2, \ldots .$$

We shall call this limit $\gamma(n;\theta',\theta'')$ and $\gamma(n;\theta)$ when $\theta' = \theta'' = \theta$.

These conditions on $z(n)$ do not appear unreasonable in relation to examples (i), (ii) and (iii). For example in (i)[2] if the limits,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} z(m)z(m+n) = \gamma_z(n), \qquad n = 0, \pm 1, \ldots$$

[1] In future we shall omit the words almost surely when we speak of this mode of convergence.

[2] We hope to consider (ii) in more detail later. In practice only $z(n)$ and not $z(n-\theta)$ may be observable and this makes necessary further considerations.

exist, as would be the case for a stationary ergodic process with finite second moments, then

$$\lim_{N\to\infty} \frac{1}{N}\sum_{1}^{N} z(m,\theta')z(m+n,\theta'') = \lim_{N\to\infty} \sum_{0}^{\infty}\Sigma(\theta')^{j}(\theta'')^{k}\frac{1}{N}\sum_{1}^{N} z(m-j)z(m-k)$$

$$= \sum_{0}^{\infty}\Sigma(\theta')^{j}(\theta'')^{k}\gamma_{z}(j-k) ,$$

by dominated convergence, and this limit is uniform in θ' , θ'' for $|\theta'|$, $|\theta''| \leq 1-\delta$, $\delta > 0$.

However the example (iv) does not satisfy the condition B since, for example, $N^{-1}\Sigma\cos\theta' n\cos\theta'' n$ converges to zero for $\theta' \neq \theta''$ (mod 2π), to 2 for $\theta' = \theta'' \neq 0,\pi$ (mod 2π) and to unity otherwise. Nevertheless all of our theorems are true for example (iv) (and its generalisation to more than one frequency). In order to show how the conditions of the theorems may be generalised and also because of its importance we shall give a separtate proof for this example, under the slightly stronger condition $\Sigma|\alpha(j)| < \infty$. (It does not seem easy to state simple conditions which usefully generalise B, and C below, so as to include (iv)).[1]

[1]Since completing this work I have seen Walker (1969) which also treats the estimation problem of example (iv), following Whittle (1952). (Walker does not consider our general case). However the two treatments are substantially different. Walker assumes that $\epsilon(n)$ has finite fourth moment and requires a faster rate of convergence for the $\alpha(j)$. He also establishes only convergence in probability. The problem is also somewhat differently formulated. Walker considers the simultaneous estimation of β_1, β_2, β_3, θ and parameters specifying the $\alpha(j)$. Since we show that "asymptotically efficient estimation" (see below) of the β_1, β_2, β_3, θ does not depend on detailed knowledge of the structure of the process generating $x(n)$ we do not consider this part of the problem. Of course some such knowledge would be needed for a measure of the precision of the estimates but this could be obtained from a "smoothed" estimate of the spectrum of $x(n)$ at the estimate of θ_o .

In connection with the proof of asymptotic normality we also need to assume the following.

C. The function $z(n;\ \theta)$ is twice differentiable in θ. Call $z_j'(n;\ \theta)$, $z_{jk}''(n;\ \theta)$ the derivatives with respect to θ_j and to θ_j, θ_k. Then uniformly in θ and θ', θ'' the following limits exist

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} z_j'(m;\theta) z_k'(m+n;\theta), \qquad \lim_{N\to\infty} \frac{1}{N} \sum_{1}^{N} z(m,\theta') z_{jk}''(m+n,\theta'') .$$

It follows immediately that the first of these limits is $\{\partial^2 \gamma(n;\theta',\theta'')/\partial\theta_j'\partial\theta_k''\}$ evaluated at $\theta' = \theta'' = \theta$. We shall call this $\tilde{\gamma}_{jk}(n;\theta)$.

Again these conditions are not met in example (iv) but again we shall establish separately the validity of the theorem.

The generalised nature of $x(n)$ leads to the need to use a generalised form of least squares estimation, given that estimation is to be based on a quadratic function of the data. Experience suggests that a more tractable formulation will be achieved if the data is first Fourier transformed. We thus introduce

$$w_y(\omega_t) = (2\pi N)^{-\frac{1}{2}} \sum_{1}^{N} y(n) e^{in\omega_t}, \qquad w_z(\omega_t;\theta) = (2\pi N)^{-\frac{1}{2}} \sum_{1}^{N} z(n;\theta) e^{in\omega_t},$$

$$\omega_t = 2\pi t/N, \qquad t = 0, 1, \ldots, N-1 .$$

We put $I_y(\omega_t) = |w_y(\omega_t)|^2$, $I_{yz}(\omega_t;\theta) = w_y(\omega_t)\overline{w_z(\omega_t;\theta)}$, $I_z(\omega_t;\theta) = |w_z(\omega_t;\theta)|^2$. We shall omit the argument variable ω_t or the argument variable θ if this will not cause confusion.

We introduce the quadratic function

$$Q_{\Phi N}(\theta) = \frac{1}{N} \sum_t \left\{ \left[I_y(\omega_t) + I_z(\omega_t;\theta) - 2\,\mathcal{R}\,(I_{yz}(\omega_t;\theta)) \right] \Phi(\omega_t) \right\}$$

where $\Phi(\lambda)$ is a continuous even function of λ satisfying $\Phi(\lambda) \geq 0$, $\lambda \in [0, \pi]$. Of course $\Phi(\lambda) \equiv 1$ leads to

$$Q_N(\theta) = \frac{1}{2\pi N} \sum_1^N (y(n) - z(n;\theta))^2 .$$

If $f(\lambda) > 0$, $\lambda \in [0,\pi]$ then $f(\lambda)^{-1}$, if known, could be used and evidently this will be an optimal choice. However the need first to estimate θ in order to estimate $f(\lambda)$ makes it necessary to consider more general cases. We shall show that, subject to A and B, and uniformly in θ we have

$$\text{(3)} \qquad \lim_{N\to\infty} Q_{\Phi N}(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi(\lambda) f(\lambda) d\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi(\lambda) d\left\{ F(\lambda;\theta) + F(\lambda;\theta_o) - 2F(\lambda;\theta,\theta_o) \right\}$$

where $F(\lambda;\theta',\theta'')$ and $F(\lambda;\theta) = F(\lambda;\theta,\theta)$ are defined by

$$\gamma(n;\theta',\theta'') = \int_{-\pi}^{\pi} e^{in\lambda} dF(\lambda;\theta',\theta'') .$$

We call the right side of (3) $Q(\theta)$. The existence and essential uniqueness of these functions follow from Bochner's theorem. The matrix

$$\begin{bmatrix} F(\lambda;\theta') & F(\lambda;\theta',\theta'') \\ F(\lambda;\theta'',\theta') & F(\lambda,\theta'') \end{bmatrix}$$

has Hermitian non-negative increments over any interval of λ values. When C is satisfied then

$$\tilde{\gamma}_{jk}(n;\theta) = \int_{-\pi}^{\pi} e^{in\lambda} d\tilde{F}_{jk}(\lambda;\theta)$$

where

$$\tilde{F}_{jk}(\lambda;\theta) = \left\{ \partial^2 F(\lambda;\theta', \theta'')/\partial\theta'_j \partial\theta''_k \right\}_{\theta' = \theta'' = \theta} .$$

Finally we require the following.

D. $\gamma_z(0;\theta)+\gamma_z(0;\theta_o)-2\gamma_z(0,\theta,\theta_o) > 0$; $\theta \neq \theta_o$.

This condition seems rather unimportant in the sense that if it fails for some $\theta \neq \theta_o$ then one can hardly expect effectively to distinguish between θ and θ_o by means of a quadratic criterion. If we call $R_\Phi(\theta)$ the second term on the right in (3) then D evidently implies $R_\Phi(\theta) > R_\Phi(\theta_o) = 0$, $\theta \neq \theta_o$ if $\Phi(\lambda) > 0$, $\lambda \in [0,\pi]$. Of course D fails in example (iv) if $\beta_2^2+\beta_3^2 = 0$, so that our theorems give no information about this case.

2. The Strong Law of Large Numbers

We first point out that under A and B

$$c_{xz}(n;\theta) = N^{-1} \sum_{m=1}^{N} x(m)z(m+n;\theta)$$

converges, uniformly in θ , to zero. Indeed uniform convergence will follow from pointwise convergence and the equicontinuity in n of the sequence $c_{xz}(n;\theta)$. The latter follows, as in Jennrich (1969), from the evaluation

$$|c_{xz}(n;\theta')-c_{xz}(n;\theta'')|^2 \leq N^{-1}\sum_1^N x(m)^2 \left\{N^{-1}\sum_1^N z(m+n;\theta')^2 + N^{-1}\sum_1^N z(m+n;\theta'')^2 - 2N^{-1}\sum z(m+n;\theta')z(m+n;\theta'')\right\}$$

$$\leq K\left\{\epsilon + \gamma_z(0;\theta')+\gamma_z(0;\theta'')-2\gamma(0;\theta',\theta'')\right\}$$

for any $\epsilon > 0$ and N sufficiently large. To establish the pointwise convergence we prove the more general Theorem 1 below, which is of some independent interest.[1]

Theorem 1. Let $x(n)$ be weakly stationary with zero mean, absolutely continuous spectrum and uniformly bounded spectral density. If

$$\overline{\lim_{N\to\infty}}\frac{1}{N}\sum_1^N y(n)^2 = a < \infty$$

then

$$\lim_{N\to\infty}\frac{1}{N}\sum_1^N y(n)x(n) = 0 , \qquad \text{a.s.}$$

Proof. We first observe that the condition on the $y(n)$ implies that

$$\sum_1^\infty \left\{y(n)^2/n^{1+\delta}\right\} < \infty , \qquad \delta > 0 .$$

Indeed if we put $s(N) = N^{-1}\sum_1^N y(n)^2$ then

$$s(n)-s(n-1) = y(n)^2/n - n^{-1}s(n-1)$$

and the result follows from the convergence of $\Sigma\{s(n-1)/n^{1+\delta}\}$ and of

[1]The required result follows from Hannan (1970), Chapter IV, Theorem 9. However the proof of that theorem is faulty. The statement of that theorem appears to need modification though its essential content is valid as can be seen from Theorem 1 below and the comments following its proof.

$$\sum_{1}^{\infty} \frac{\{s(n)-s(n-1)\}}{n^{\delta}}$$

(the latter because $\sum_{1}^{N}(s(n)-s(n-1))$ remains bounded and $n^{-\delta}$ converges monotonically to zero).

In the second place the variance of $c(N) = N^{-1}\Sigma y(n)x(n)$ is, putting $\gamma(n) = E\{x(m)x(m+n)\}$,

$$N^{-2}\sum_{1}^{N}\Sigma y(m)y(n)\gamma(n-m) = N^{-2}\int_{-\pi}^{\pi}\left|\Sigma y(m)e^{im\lambda}\right|^{2} f(\gamma)d\lambda$$

$$\leqq KN^{-1}\left\{N^{-1}\Sigma y(n)^{2}\right\} .$$

Now consider $c(M^2)$, $M = 1, 2, \ldots$. Then by Chebyshev's inequality and the Borel-Cantelli lemma this converges to zero. Moreover

$$E\left[\left(\sup_{M^2<N\leqq(M+1)^2}\left|c(N)-(M^2/N)c(M^2)\right|\right)^{2}\right]$$

$$\leqq E\left[\left(M^{-2}\sum_{M^2+1}^{(M+1)^2}\left|x(n)y(n)\right|\right)^{2}\right]$$

$$\leqq KM^{-4}\left[(M+1)^2-M^2\right]\sum_{M^2+1}^{(M+1)^2} y(n)^2 \leqq KM^{-3}\sum_{M^2+1}^{(M+1)^2} y(n)^2 ,$$

where K may be taken independent of M .

However

$$\sum_{M=1}^{\infty} M^{-3}\left\{\sum_{M^2+1}^{(M+1)^2} y(n)^2\right\} \leqq K \sum_{n=1}^{\infty}\frac{y(n)^2}{n^{3/2}} < \infty$$

so that, again by Chebyshev's inequality and the Borel-Cantelli lemma (and the fact that $((M+1)/M \to 1)$

$$\lim_{M\to\infty}\left\{\sup_{M^2<N\leqq(M+1)^2}|c(N)-c(M^2)|\right\} = 0 .$$

Thus the theorem is established.

The theorem may be considerably extended. For example if $x(n)$ is weakly stationary, $c(N)$ has variance $O(N^{-\alpha})$, $\alpha > 0$, it continues to hold, by much the same proof. Alternatively if $x(n)$ is an in the theorem and $N^{1-\alpha}s(N) \leqq K$, $\alpha > \frac{1}{2}$, then $N^{1-\alpha}C(N)$ converges to zero, again by much the same proof.

We may now establish (3). Consider

$$N^{-1}\Sigma I_{xz}\Phi \tag{4}$$

where $I_{xz} = w_x(\omega_t)\overline{w_z(\omega_t;\theta)}$. For any $\epsilon > 0$ we may find M such that

$$\sup_{\lambda}\left|\Phi(\lambda) - \sum_{-M}^{M}\left(1-\frac{|n|}{M}\right)\delta(n)e^{in\lambda}\right| < \epsilon ,$$

where $\delta(n)$ is the n^{th} Fourier coefficient of $\Phi(\lambda)$, since that function is continuous. Thus (4) is dominated by

$$\epsilon N^{-1}\Sigma|I_{xz}| + \left|\frac{1}{2\pi}\sum_{-M}^{M}\delta(n)\left(1-\frac{|n|}{M}\right)c_{xz}(n;\theta)\right| .$$

The second term now converges to zero uniformly in θ while by Schwartz's inequality the first is dominated by $\epsilon(c_x(o)c_z(o;\theta))^{\frac{1}{2}}$. Thus (4) converges to zero uniformly in θ.

In almost the same way we may show that, uniformly in θ,

(5) $$\lim_{N\to\infty} \frac{1}{N} \Sigma\Phi w_z(\omega_t;\theta)w_z;\theta_o) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(\lambda)dF(\lambda;\theta,\theta_o)$$

while if $x(n)$ is ergodic (with finite variance)

(6) $$\lim_{N\to\infty} \frac{1}{N} \Sigma\Phi I_x(\omega_t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(\lambda)f(\lambda)d\lambda .$$

Thus the validity of (3) is established. Now we have the following result.

Theorem 2. If $x(n)$ is as in Theorem 1 and B, D are satisfied, $\Phi(\lambda) > 0$, $\lambda \in [0, \pi]$ and $\hat{\theta}_{\Phi N}$ minimises $Q_{\Phi N}(\theta)$ then $\hat{\theta}_{\Phi N}$ converges $\hat{\theta}_o$. If A is satisfied $Q_{\Phi N}(\hat{\theta}_{\Phi N})$ converges to $Q(\theta_o)$.

Proof. Consider $Q_{\Phi N}(\hat{\theta}_{\Phi N})-Q_{\Phi N}(\theta_o) \leq 0$. Let $\hat{\theta}_{\Phi m}$ be a subsequence converging to $\theta' \neq \theta_o$. From the convergence of (4) and the left hand side of (5), uniformly in θ, it follows that $Q_{\Phi m}(\hat{\theta}_{\Phi m})-Q\Phi_m(\theta_o)$ converges to $Q(\theta')-Q(\theta_o) > 0$. This shows that $\theta' = \theta_o$ and establishes the first part of the theorem. If A holds $x(n)$ is ergodic and (6) holds also and since $Q_{\Phi N}(\theta_o)$ then converges to $Q(\theta_o)$ the second part of the theorem follows also.

Of course it is only the ergodicity of $x(n)$ which is required for the second part of the theorem and this is certainly implied by the uniform mixing condition mentioned in the introduction. For the truth of the theorem it may also be observed that it is only the uniformity of convergence of $N^{-1}\Sigma z(m;\theta')z(m+n,\theta'')$ in any compact subset of $\Theta \times \Theta$, excluding the diagonal $\theta' = \theta''$, which is used so far as (5) is concerned. Thus for example (iv) the failure of the uniformity of the convergence of this quantity at $\theta' = \theta''$ is not of concern in relation to (5). However the proof given above that $c_{xz}(n;\theta)$ converges uniformly to zero does now break down since the failure of B prevents us from using the method of proof adopted above of the equicontinuity of the $c_{xz}(n;\theta)$. However, under A, and $\Sigma|\alpha(j)| < \infty$ we now establish that

$$\sup_{|\lambda|\leq\pi} \left|\frac{1}{N}\sum_1^N x(n)e^{in\lambda}\right| \tag{7}$$

converges to zero and this will establish the truth of Theorem 2 for example (iv) also.

Now

$$N^{-1}\sum_1^N x(n)e^{in\lambda} = N^{-1}\sum_{-\infty}^{\infty}\alpha(j)e^{ij\lambda}\sum_1^N \epsilon(n-j)e^{i(n-j)\lambda} .$$

Thus

$$\left\{E\sup_\lambda\left|N^{-1}\sum_1^N x(n)e^{in\lambda}\right|^2\right\}^{\frac{1}{2}} \leq N^{-1}\sum_{-\infty}^{\infty}|\alpha(j)|\left\{E\sup_\lambda\left|\sum_1^N \epsilon(n-j)e^{i(n-j)\lambda}\right|^2\right\}^{\frac{1}{2}} \tag{8}$$

$$\leq N^{-1}\sum_{-\infty}^{\infty}|\alpha(j)|\left\{N+\sum_{m=N+1}^{N-1}{}' E\left(\left|\Sigma_m \epsilon(m)\epsilon(m+n)\right|\right)\right\}^{\frac{1}{2}}$$

where the sum Σ' omits the term for $n = 0$ and the term Σ_m is over

$N-|n|$ terms (dependent on j). Since $\Sigma' E(|\Sigma_m \epsilon(m)\epsilon(m+n)|) \leq$ $\Sigma' \{E(\Sigma_m \epsilon(m)\epsilon(m+n))^2\}^{\frac{1}{2}} = O(N^{3/2})$ (uniformly in j) we see that (8) is $O(N^{-1/4})$. Now choose $\beta > 2$ and put $N(M)$ equal to the smallest integer not greater than M^β . Then as N proceeds through the sequence $N(M)$, $M = 1, 2, \ldots$ the expression (7) converges to zero. However also

$$\sup_{\lambda} \sup_{N(M)<N\leq N(M+1)} \left|N^{-1} \sum_1^N x(n)e^{in\lambda} - N^{-1} \sum_1^{N(M)} x(n)e^{in\lambda}\right|$$

$$\leq N(M)^{-1} \sum_{N(M)+1}^{N(M+1)} |x(n)|$$

whose mean square is dominated by $KN(M)^{-2}\{N(M+1)-N(M)\}^2 = O(M^{-2})$. Since also $N(M)/N(M+1) \to 1$ we see that (7) converges to zero and the following theorem is established.

<u>Theorem 2'</u>. If A is satisfied, $\Phi(\theta_o) > 0$ and $\hat{\theta}_{\Phi N}$, $\hat{\beta}_{\Phi N}$ minimise $Q_{\Phi N}(\theta,\beta)$ for example (iv) then $\hat{\theta}_{\Phi N}$, $\hat{\beta}_{\Phi N}$ converges to θ_o, β_o and $Q_{\Phi N}(\hat{\theta}_{\Phi N}, \hat{\beta}_{\Phi N})$ converges to $Q(\theta_o, \beta_o)$.

The condition A could be replaced by A' but in this case a fourth moment condition seems also to be needed. Thus if $k(n,p,q)$ is the fourth cumulant between $x(m)$, $x(m+n)$, $x(m+p)$, $x(m+q)$ we might require that

E. $$\sum_{npq=-\infty}^{\infty}\sum\sum k(n,p,q) < \infty .$$

(If A is satisfied and $\epsilon(n)$ has finite fourth moment then E is satisfied also). If $A' + E$ holds (in place of A) then Theorem 2' continues to hold also. We omit the proof.

3. The Asymptotic Normality of the Estimate

We consider first the quantities

$$d_{\Phi j} = N^{-1} \sum_{1}^{N} \Phi I_{xz}^{(j)}(\omega_t;\ \theta)$$

where the j superscript indicates differentiation with respect to θ_j. Now if A and C are satisfied and $\Phi(\lambda)$ has an absolutely convergent Fourier series the $N^{1/2} d_{\Phi j}$ are jointly asymptotically normal with zero means and covariance matrix with typical element.

$$\int_{-\pi}^{\pi} 2\pi f(\lambda)\Phi(\lambda)^2 dF_{jk}(\lambda;\theta) \quad . \tag{9}$$

Indeed

$$N^{\frac{1}{2}} d_{\Phi j} = N^{\frac{1}{2}}/2\pi \sum_{-N+1}^{N-1} \left\{ c_{xz}^{(j)}(n;\theta) \sum_{k=-\infty}^{\infty} \delta(n+2kN) \right\} \quad .$$

This follows from the orthogonality of the $\exp in\omega_t$ under summation over $t = 0, 1, \ldots, N-1$. We have

$$E\left(c_{xz}^{(j)}(n;\theta)\right) = 0 \ , \quad NE\left\{\left(c_{xz}^{(j)}(n;\theta)\right)^2\right\} \leq Kc_z^{(j)}(0;\theta)$$

where $c_z^{(j)}(0;\theta)$ is the sample mean square of $z_j'(n;\theta)$. On the other hand for any $\epsilon > 0$ there is an M so that

$$\sum_{|n|>M} \left| \sum_{k=-\infty}^{\infty} \delta(n+2kN) \right| < \epsilon \ .$$

The covariance of the $N^{\frac{1}{2}} d_{\Phi j}$ may be shown to converge to (9) precisely as in Hannan (1970) Chapter 7. Thus the asymptotic normality follows from that of the $N^{1/2} c_{xz}^{(j)}(0,\theta)$, $|n| \leq M$, and this follows from Hannan (1970) Theorem 10. We now have the following theorem.

<u>Theorem 3</u>. If A, B, C, D are satisfied for θ_o interior to Θ and $\Phi(\lambda)$ is as in Theorem 2 and has an absolutely convergent Fourier series then $N^{\frac{1}{2}}(\hat{\theta}_{\Phi N}-\theta_o)$ is asymptotically normal with zero mean and covariance matrix $A^{-1}BA'^{-1}$ where B has typical element (9), for $\theta = \theta_o$, and A has typical element

$$\int_{-\pi}^{\pi} \Phi(\lambda) d\tilde{F}_{jk}(\lambda;\theta_o) .$$

The proof of this theorem is now virtually the same as that in Jennrich (1969) Theorem 7 and will not be repeated.

If $f(\lambda)$ has an absolutely convergent Fourier series and is never null in $[-\pi, \pi]$ then $f(\lambda)^{-1}$ also has an absolutely convergent Fourier series. (This will be so if $\Sigma|\alpha(j)| < \infty$ and $f(\lambda) \neq 0$, $\lambda \in [-\pi,\pi]$.) In this case we may put $\Phi = f^{-1}$ which will minimise $A^{-1}BA'^{-1}$ (in the usual ordering of symmetric matrices). For example in example (ii) with this choice of Φ the variance in the limiting distribution will be

$$\left\{\frac{1}{2\pi} \int \lambda^2 f(\lambda)^{-1} f_z(\lambda) d\lambda \right\}^{-1} .$$

If the condition A is replaced by A′ together with E the theorem will again hold provided the $z_j'(n;\theta)$ are either generated by an ergodic process with finite variance or satisfy the condition (b)′ of Hannan (1970) p. 219. Again we shall not discuss this in detail. Instead we consider example (iv) once more. We then have the following theorem.

<u>Theorem 3′</u>. If, for example (iv), A and D hold, Φ satisfies the conditions of Theorem 3 and $\Sigma|\alpha(j)| < \infty$ then the partitioned vector

$$\begin{pmatrix} N^{\frac{1}{2}} & (\hat{\beta}_{\Phi N}-\beta_o) \\ \cdots & \cdots \\ N^{3/2} & (\hat{\theta}_{\Phi N}-\theta_o) \end{pmatrix}$$

is asymptotically normal with zero mean vector and covariance matrix

$$2\pi f(\theta_o)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & \beta_{03}/4 \\ 0 & 0 & 1/2 & -\beta_{02}/4 \\ 0 & \beta_{o3}/4 & -\beta_{o2}/4 & (\beta_{o2}^2+\beta_{o3}^2)/6 \end{bmatrix}, \quad \theta_o \neq 0, \pi .$$

For $\theta_o = 0, \pi$ when there is no β_{o3}, the factors 1/2, 1/4 are to be replaced by unity.

This is, of course, the result obtained by Walker (1969), under somewhat stronger conditions on $x(n)$ and for $\Phi \equiv 1$. Of course the limiting distribution is independent of Φ.

For simplicity we give the proof for the slightly different case where $\beta_1 = 0$, $\beta_2 = 1$, $\beta_3 = 0$, and this is known, and $\theta_o \neq 0, \pi$ since the only essential difficulty is with $\hat{\theta}_{\Phi N}$. We call this $\hat{\theta}_N$, for short, and shall use $\bar{\theta}_N$ for a random variable, such that $|\bar{\theta}_N - \theta_o| \leq |\hat{\theta}_N - \theta_o|$, for which we have the following relation, at least for N large enough,

$$N^{3/2} \sum_t \Phi I_{xz}^{(j)}(\omega_t;\theta_o) = \left\{ N^{-3} \sum_t \Phi \left[w'(\omega_t;\bar{\theta}_N)\overline{w'(\omega_t;\bar{\theta}_N)} - w''(\omega_t;\bar{\theta}_N)\overline{w(\omega_t;\bar{\theta}_N)} + w''(\omega_t;\bar{\theta}_N)\overline{w(\omega_t;\theta_o)} + w''(\omega_t;\bar{\theta}_N)\overline{w_x(\omega_t)} \right] \right\} N^{3/2}(\hat{\theta}_N - \theta_o) .$$

This is got by expanding $dQ_{\Phi N}(\theta)/d\theta$ in the first two terms of its Taylor series about $\theta = \hat{\theta}_N$. The existence of the random variable $\bar{\theta}_N$ follows as in Jennrich (1969), lemma 3 .

The proofs that the first two terms in the bracket on the right produce contributions which cancel and that the fourth produces a contribution converging to zero are not difficult and we omit them.[1] Before turning to the third term we point out that the left hand side is asymptotically normal with zero mean and variance $(12\pi)^{-1}\Phi(\theta_o)^2 f(\theta_o)$. This follows from Hannan (1970) Chapter IV Theorems 10 exactly as in the proof of Theorem 3.

[1]The convergence of the fourth term to zero is proved in almost precisely the same way as the proof given in Theorem 2'.

We now first observe that $N(\hat{\theta}_N - \theta_o)$ converges to zero. We know that $Q_{\Phi N}(\hat{\theta}_N) \to Q(\theta_o)$ so that

$$\frac{1}{N}\sum_1^N \Phi(\omega_t)\left\{\left[I_z(\omega_t;\theta_o)+I_z(\omega_t;\hat{\theta}_N)-2\,R\left[w_z(\omega_t;\hat{\theta}_N)\overline{w_z(\omega_t;\theta_o)}\right]\right]\right\}$$

converges to zero. Now it is easily seen that the sum of the first two terms converges to $\Phi(\theta_o)/2\pi$. Also

(11) $$\frac{1}{N}\sum_1^N \Phi(\omega_t) w_z(\omega_t;\hat{\theta}_N)\overline{w_z(\omega_t;\theta_o)}$$

$$= \frac{1}{2\pi}\sum_{n=N+1}^{N-1}\left\{\left[N^{-1}\Sigma_m \cos m\hat{\theta}_N \cos(m+n)\theta_o\right]\right\}\sum_{-\infty}^{\infty}\delta(n+2kN)$$

where the inner sum is over m such that m and $m+n$ lie between 1 and N, inclusive. Replacing $\cos(m+n)\theta_o$ by $\cos m\theta_o \cos n_o - \sin m\theta_o \sin n\theta_o$ and recalling that

$$\sum_n \left|\sum_{-\infty}^{\infty}\delta(n+2kN)\right| < \infty$$

while

$$N^{-1}\Sigma \cos m\theta \sin n\theta_o$$

converges to zero uniformly in θ we see that (11) may be replaced by

$$(\Phi(\theta_o)/2\pi)\ N^{-1}\sum_{m=1}^{N}\cos m\hat{\theta}_N \cos m\theta_o$$

which may be replaced in turn by

$$(\Phi(\theta_o)/2\pi)\ N^{-1} \sum_{m=1}^{N} \cos m(\hat{\theta}_N - \theta_o)$$

However the second factor is

$$\left(\frac{\sin\{(N+\frac{1}{2})(\hat{\theta}_N - \theta_o)\}}{2N\sin\{\frac{1}{2}(\hat{\theta}_N - \theta_o)\}} \right) - (2N)^{-1}$$

so that the first term must converge to unity. This can happen only if $N(\hat{\theta}_N - \theta_o)$ converges to zero.

Next consider (10). We need to show that

$$(12) \qquad N^{-3} \sum_t \Phi w''(\omega_t; \bar{\theta}_N)\overline{w(\omega_t; \theta_o)}$$

converges to a non zero limit and to evaluate that limit. However in almost the same way as in the proof that $N(\hat{\theta}_N - \theta_o)$ converges to zero we see that (12) may be replaced by

$$(\Phi(\theta_o)/2\pi)\ N^{-3} \sum_{m=1}^{N} m^2 \cos m\bar{\theta}_N \cos m\theta_o \quad .$$

We know that $N(\bar{\theta}_N - \theta_o)$ converges to zero. The second factor is easily evaluated as asymptotically equivalent to

$$\frac{1}{2}\left\{ \frac{\sin N\theta_N}{N\theta_N} + 2\ \frac{N\theta_N \cos N\theta_N - \sin N\theta_N}{N^3 \theta^3_N} \right\} ,$$

where we have put $\theta_N = (\bar{\theta}_N - \theta_o)$. Now as $N\theta_N \to 0$ this converges to 1/6 . Thus we have proved we require, namely that $N^{3/2}(\hat{\theta}_N - \theta_o)$ is asymptotically normal with zero mean and variance $12\pi f(\theta_o)$.

The theorem gives, incidentally, the almost obvious result that at least asymptotically we may as well use $\Phi(\lambda) \equiv 1$ in the case of example (iv).

4. Estimation of $f(\lambda)$

Theorem 2 shows that $\Phi(\lambda) = f(\lambda)^{-1}$ is optimal. (As before we now assume that $f(\lambda) > 0$, $\lambda \in [-\pi,\pi]$ and that it has an absolutely convergent Fourier series). We do not need so much to consider the primed theorems because of the comment at the end of section 3 and we restrict ourselves to the other case.

It follows from the proof of Theorem 2 that $Q_{\Psi N}(\hat{\theta}_{\Phi N})$ converges to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Psi(\lambda)f(\lambda)d\lambda \quad .$$

Thus taking $\Psi(\lambda) = \cos n\lambda$ we see that we may obtain estimates converging to the autocovariances of the $x(n)$ sequence. Hence if $f(\lambda)$ is prescribed as having a rational spectrum (of prescribed degree for numerator and denominator) then we may obtain, from an initial estimate $\hat{\theta}_{\Phi N}$, an estimate, $\hat{f}_N(\lambda)$, converging uniformly to $f(\lambda)$. If $Q_N(\theta)$, $\hat{Q}_N(\theta)$ are $Q_{\Phi N}(\theta)$ for Φ respectively f^{-1} and $\hat{f}_N^{-1}$ then it is almost immediate that $Q_N(\theta)-\hat{Q}_N(\theta)$ converges uniformly to zero. If $\hat{\theta}_N$ minimises $\hat{Q}_N(\theta)$ it follows that $\hat{\theta}_N$ converges to θ_o by almost the same argument as was used to prove Theorem 2 . Non parametric estimates of $f(\lambda)$ converging almost surely have not been widely discussed (see, however, Parthasarathy (1960)). It is much easier to find conditions under which we shall have, for such an estimate,

$$\lim_{N\to\infty} P \left\{ \sup_{\lambda} |\hat{f}_N(\lambda) - f(\lambda)| > \epsilon \right\} = 0 \qquad \epsilon > 0 \quad .$$

Insofar as that can be established then

$$\sup_{\theta} |Q_N(\theta) - \hat{Q}_N(\theta)|$$

will converge in probability to zero and $\hat{\theta}_N$ will converge in probability to θ_o .

So far as Theorem 3 is concerned we consider only

$$N^{-\frac{1}{2}} \sum_t f(\omega_t)^{-1} I_{xz}^{(j)}(\omega_t; \theta_o)$$

for the remainder of the proof of Theorem 3 would not be affected by the replacement of f by $\hat{f}_N$. Let us arrange the ω_t into 2M sets, S_u , of approximately $m = [N/2M]$ adjacent values centred around $2M$ frequencies λ_u , with $\lambda_o = 0$, $\lambda_M = \pi$, $\lambda_u = -\lambda_{-u}$, $u = 1, \ldots, M-1$. We say approximately since some of these sets might need to contain $m+1$ values. Then, if $\omega_t \in S_u$, $|f(\omega_t)^{-1} - f(\lambda_u)^{-1}| < Km^{-1}$. Call $\hat{f}_{zx}^{(j)}(\lambda_u; \theta_o)$ the average of the $I_{zx}^{(j)}(\omega_t; \theta_o)$ for $\omega_t \in S_u$. Then this has mean zero and standard deviation dominated by $Km^{-1/2}$. Thus

$$(12) \qquad N^{-\frac{1}{2}} \sum_t f(\omega_t)^{-1} I_{zx}^{(j)}(\omega_t; \theta_o) - (N^{\frac{1}{2}}/2M) \sum_u f(\lambda_u)^{-1} \hat{f}_{zx}^{(j)}(\lambda_u; \theta_o)$$

has standard deviation dominated by $(KN^{\frac{1}{2}}/m)$. Now assume that $M = o(N^{\frac{1}{2}})$ in which case (12) converges in probability to zero. Now replace $f(\lambda_u)^{-1}$ by $\hat{f}_N(\lambda_u)^{-1}$. By a suitable choice of M as a function of N we may ensure that this has mean square error which is $O(m^{-1})$. (See Hannan (1970) Chapter V). Then

$$(13) \qquad (N^{\frac{1}{2}}/2M) \sum_u \left\{ \left[f(\lambda_u)^{-1} - \hat{f}_N(\lambda_u)^{-1} \right] \right\} \hat{f}_{zx}^{(j)}(\lambda_u;\theta_0)$$

will converge in probability to zero. Indeed outside of a set, S, in the space of all histories of $x(n)$, dependent on N, whose probability content may be made arbitrarily small for N sufficiently large, $\hat{f}_N(\lambda_u)^{-1} \geqq a > 0$ since $\hat{f}_N(\lambda)$ converges (either almost surely or in probability) uniformly to $f(\lambda)$. Conditional upon the history being within S then the root mean squre error of (13) will be dominated by $N^{\frac{1}{2}}m^{-1}$ since $\{E[(2M)^{-1}\sum_u|\hat{f}_{zx}^{(j)}(\lambda_u)|^2]\}^{\frac{1}{2}}$ is $O(M/N)$. (See for example Hann (1970), first appendix to Chapter 7). Since $N^{\frac{1}{2}}/m$ converges to zero the replacement of $f(\lambda)$ by $\hat{f}_N(\lambda)$ does not affect the conclusion of Theorem 3.

Correspondingly, of course, these results indicate how the covariance matrix needed for the application of Theorem 3 may be estimated. The replacement of the individual $I(\omega_t;\theta_o)$ by averages of them over the sets S_u would reduce the calculations materially and our just preceding discussion shows that this is permissible. For that matter smoothed estimates of spectra might replace these averages (see Hannan (1970) Chapter v) leading to a further reduction in calculations but we do not consider that here.

REFERENCES

1. Hannan, E.J. (1970), Multiple Time Series. Wiley, New York.

2. Jennrich, R.I. (1969), Non-linear least squares estimators. Ann. Math. Statist., 40, 633-643.

3. Parthasarathy, K.R. (1960), On the estimation of the spectrum of a stationary stochastic process. Ann. Math. Statist., 31, 568-573.

4. Rosenblatt, M. (1956), A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci., U.S.A., 42, 43-47.

5. Rozanov, Yu.A., (1967), Stationary Random Processes. Holden-Day, San Francisco.

6. Walker, A.M. (1969), On the estimation of a harmonic component in a time series with stationary residuals. Proc. Int. Stat. Inst., 43, 374-376.

7. Whittle, P. (1952), The simultaneous estimation of a time series' harmonic components and covariance structure. Trab. Estad., 3, 43-57.