The Annals of Statistics
1980, Vol. 8, No. 4, 762–777

RECURSIVE ESTIMATION BASED ON ARMA MODELS

BY E. J. HANNAN

The Australian National University

A recursive estimate of the stochastic structure of a stationary time series is constructed based on the assumption that the true structure is ARMA, i.e., has a rational spectrum. The estimate is recursive in the sense that each successive estimate is obtained from the previous one by a relatively simple adjustment, that could be effected in a "real time" situation. The procedure is basically that of updating a regression when all variates involved are constructed from previous estimates of the parameter vector. The strong convergence of the estimate to the true value is established as well as a result relating to the rate of convergence.

1. Introduction. Consider initially a scalar ARMA model for an observed sequence $y(t)$, namely

$$\Sigma_0^p \alpha_j y(t-j) = \Sigma_0^q \beta_j \varepsilon(t-j),\ \alpha(0) = \beta(0) = 1. \tag{1.1}$$

The sequence, $\varepsilon(t)$, whenever it occurs, will be taken to be stationary and ergodic and to satisfy

$$\mathcal{E}\{\varepsilon(t)|\mathcal{F}_{t-1}\} = 0,\ \mathcal{E}\{\varepsilon(t)^2|\mathcal{F}_{t-1}\} = \sigma^2. \tag{1.2}$$

Here $\mathcal{F}_t$ is the σ-algebra determined by $\varepsilon(j), j \leqslant t$. Introduce the generating functions

$$g(z) = \Sigma_0^p \alpha_j z^j,\ h(z) = \Sigma_0^q \beta_j z^j,$$

and let $\mathfrak{R}$ be the region in R^{p+q} wherein $g(z) \neq 0$, $h(z) \neq 0$, $|z| \leqslant 1$ and g, h have no common zero. Then $\mathfrak{R}$ is open. (See [3] for example.) The symbol θ will be used for a typical point in $\mathfrak{R}$ and the notation g_θ, h_θ will then also be used. Also we shall put $\theta' = (\alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)$. Below cases will be considered where α_j, β_j depend on t, in which case we shall write g_t, h_t for $\Sigma\alpha_j(t)z^j$, $\Sigma\beta_j(t)z^j$. When $\theta \in \mathfrak{R}$ the $\varepsilon(t)$ are the linear innovations.

Recently attention has been paid to recursive estimation of (1.1) i.e., sequences of estimates, $\theta(n)$, constructed from the data to time n where $\theta(n)$ is obtained from $\theta(n-1)$ by a relatively simple adjustment. (See [4], [5], [7], [14], [15] for example.)

However we have in mind situations where $y(t)$ does not satisfy equation (1.1), which is why dependence upon θ has been emphasized. (This point of view is emphasized in [7], as also is the prediction error criterion shortly to be introduced.) To begin with let us point out that a recursion might be constructed commencing

Received June 1978; revised January 1979.

AMS 1970 *subject classifications.* Primary 62M10, 62N15; secondary 62L12.

Key words and phrases. ARMA models, recursive estimation, strong convergence, martingales, real time calculation.

from a rearrangement of (1.1), namely,

$$\hat{\varepsilon}(t) = -\Sigma_1^q \beta_j(t-1)\hat{\varepsilon}(t-j) + \Sigma_0^p \alpha_j(t-1)y(t-j), \tag{1.3}$$

where the vector of parameter estimates, $\theta(t-1)$, has yet to be defined. Putting

$$v(t)' = \{-y(t-1), \cdots, -y(t-p), \hat{\varepsilon}(t-1), \cdots, \hat{\varepsilon}(t-q)\}$$

the equation (1.3) may be rewritten as

$$\hat{\varepsilon}(t) = y(t) - \theta(t-1)'v(t). \tag{1.4}$$

This suggests computing $\theta(t)$ as a regression of $y(t)$ on $v(t)$ where this regression is recursively constructed using Plackett's algorithm (See [1], for example). Interpreting (1.4) as a regression, this algorithm would be written in the form

$$\theta(n) = \theta(n-1) + P(n)v(n)\hat{\varepsilon}(n), \qquad \theta(0) = 0 \tag{1.5}$$

(1.6)

$$P(n) = P(n-1) - \{1 + v(n)'P(n-1)v(n)\}^{-1}P(n-1)v(n)v(n)'P(n-1).$$

Now it is evident how the recursively defined regression may be computed since at time n all components in the right sides of (1.5), (1.6) are available. To complete the definition of the recursion initial values for $P(0)$ and $\hat{\varepsilon}(t)$, $1-q \leqslant t \leqslant 0$ and $y(t)$, $1-p \leqslant t \leqslant 0$ must be prescribed. Let us assume, for example, that the latter are observed and that the initiating $\hat{\varepsilon}(t)$ are put equal to zero. It will be convenient to assume that $P(0)$ is positive definite. As will be seen from the proof given below the effects of these initiating choices asymptotically vanish. The equations (1.5), (1.6) are equivalent to

$$P(n) = \{\Sigma_1^n v(t)v(t)' + P(0)^{-1}\}^{-1}$$

$$\theta(n) = P(n)^{-1}\Sigma_1^n y(t)v(t).$$

It will be convenient to put

$$\hat{K}(n) = \frac{1}{n}\Sigma_1^n v(t)v(t)'$$

so that

$$\hat{K}(n) = n^{-1}P(n)^{-1} - n^{-1}P(0)^{-1}.$$

The recursion, (1.5), (1.6), has been called RML_1 (or AML), these corresponding to recursive (or approximate) maximum likelihood but it is in no sense a maximum likelihood method on Gaussian assumptions. This does not invalidate the recursion since the data will be neither Gaussian nor precisely satisfy (1.1). In [12] it is shown that a, *slightly changed*, form of RML_1 does converge almost surely (under certain conditions) and results contained in [13] indicate that, if (1.1) holds, it is not asymptotically efficient. By this is meant that the asymptotic distribution of $n^{\frac{1}{2}}(\theta(n) - \theta)$ will not be that which would obtain for the maximum likelihood estimator, computed on Gaussian assumptions. The "certain conditions" just mentioned, which impose restrictions on g, h for (1.1), will be indicated in Section

3. Partly for these reasons a second recursion, called RML_2, will mainly be discussed in this paper. As indicated earlier this is derived from the "prediction error" criterion, $\Sigma\varepsilon_\theta(t)^2$. It is shown in [13], using the results of this paper, that the RML_2 recursion is asymptotically efficient. To avoid confusion with RML_1 we, for the moment, indicate the output of this new algorithm by $\theta_M(n)$. Then

$$(1.7) \qquad \theta_M(n) = \theta_M(n-1) + P_M(n)z(n)\hat{\varepsilon}_M(n), \qquad \theta_M(0) = 0$$

$$(1.8) \qquad P_M(n) = P_M(n-1) - \{1 + z(n)'P_M(n-1)z(n)\}^{-1} \times P_M(n-1)z(n)z(n)'P_M(n-1)$$

$$(1.9) \qquad \hat{\varepsilon}_M(t) = y(t) - \theta_M(t-1)'v_M(t).$$

In this last equation $v_M(t)$ is defined as was $v(t)$ but with $\hat{\varepsilon}_M(t)$ replacing $\hat{\varepsilon}(t)$. However (1.7), (1.8) differs from (1.5), (1.6) because therein $v(t)'$ has been replaced by

$$z(t)' = \{-\eta(t-1), \cdots, -\eta(t-p), \xi(t-1), \cdots, \xi(t-q)\}$$

wherein

$$(1.10) \quad \Sigma_0^p \alpha_{M,j}(t-1)\eta(t-j) = \hat{\varepsilon}_M(t), \qquad \Sigma_0^q \beta_{M,j}(t-1)\xi(t-j) = \hat{\varepsilon}_M(t).$$

Here $\alpha_{M,j}(t)$ is the estimate of α_j made at time t and similarly for $\beta_{M,j}(t)$. This recursion differs slightly from the definition of RML_2 given in [11] because of (1.10). This definition is preferred here because it is slightly more symmetrical and results in the relation

$$(1.11) \qquad \eta(t) - \xi(t) = \theta_M(t-1)'z(t),$$

as is easily checked. Again initiating values will be needed and again we take $P(0) > 0$. We also put

$$\hat{K}_M(n) = \frac{1}{n}\Sigma_1^n z(t)z(t)'.$$

Then once more RML_2 may be represented as a regression, except for the initiation of $P(n)$, but now of $y_M(t) = \hat{\varepsilon}_M(t) + \eta(t) - \xi(t)$ on $z(t)$, $t = 1, \cdots, n$. Thus

$$\theta_M(n) = \{P(0)^{-1} + \Sigma_1^n z(t)z(t)'\}^{-1}\Sigma_1^n z(t)\hat{y}_M(t).$$

To see this it is only necessary to observe that if $\hat{y}_M(t)$ is defined as $\hat{\varepsilon}_M(t) + \theta_M(t-1)'z(t)$ then (1.7), (1.8) are of the same form as (1.5), (1.6), with $\hat{y}_M(t)$, $z(t)$ replacing $y(t)$, $v(t)$. The formula for $\hat{y}_M(t)$, given earlier, now follows from (1.11).

The purpose here is to study these algorithms, and especially RML_2, by the techniques of Hannan (1976). However while the algorithm constructed above has been based on the model (1.1) it will *not* be assumed that the data is generated by (1.1), but rather, more generally, that $y(t)$ is stationary and

$$(1.12) \qquad y(t) = \Sigma_0^\infty \kappa_j \varepsilon(t-j), \; \Sigma_0^\infty |\kappa_j| < \infty.$$

Putting $k_0(z) = \Sigma \kappa_j z^j$ it is assumed that $k_0(z) \neq 0$, $z < 1$, which is needed since we require that the $\varepsilon(t)$ be the linear innovations. If (1.1) does hold then $k_0 = g^{-1}h$, of course, but (1.12) is more general in that k_0 is not required to be rational in z. The process $y(t)$ generated by (1.12) has spectral density

$$f_0(\omega) = \frac{\sigma^2}{2\pi}|k_0(e^{i\omega})|^2,$$

which is continuous because of the second part of (1.12).

Because we shall be concentrating our attention on RML_2 *in the remainder of this and the next section we shall now eliminate the M subscript.* Unfortunately it has been found impossible to analyze this algorithm without further modification because $\theta(n)$ may leave $\mathfrak{R}$. For this reason regions $\mathfrak{R}_2$, $\mathfrak{R}_1$ satisfying $\mathfrak{R}_2 \subset \mathfrak{R}_1 \subset \mathfrak{R}$ are introduced. $\mathfrak{R}_2$ will be prescribed later. $\mathfrak{R}_1$ is such that for $\theta \varepsilon \mathfrak{R}_1$, $g_\theta, h_\theta \neq 0$, $|z| \leqslant 1 + \delta$, $\delta > 0$ and any pair of zeros, one from g_θ and one from h_θ, are at least δ apart in the complex plane. Alternatively (and equivalently) the resultant of g_θ and h_θ could be bounded away from zero. Let $\tilde{\theta}(n)$ be the output of the new recursion (defined in (1.13) below). Then $\hat{\varepsilon}(t)$, $\eta(t)$, $\xi(t)$ and hence $z(t)$, $K(t)$ are computed from $\theta(n)$ where $\theta(n)$ is $\tilde{\theta}(n)$ until $\tilde{\theta}(n)$ first leaves $\mathfrak{R}_1$ when $\theta(n)$ is kept at a fixed point in $\mathfrak{R}_2$ until $\tilde{\theta}(n)$ enters $\mathfrak{R}_2$ when $\theta(n)$ is put equal to $\tilde{\theta}(n)$ once more, and so on. The formula for $\tilde{\theta}(n)$ is

(1.13) $\tilde{\theta}(n) = \tilde{\theta}(n-1) + P(n)z(n)\tilde{\varepsilon}(n)$, $\tilde{\varepsilon}(n) = \hat{y}(n) - \tilde{\theta}(n-1)'z(n)$.

The equation for $P(n)$ is as before. Thus $\tilde{\theta}(n)$ is still of the nature of a regression

$$\tilde{\theta}(n) = \left\{P(0)^{-1} + \Sigma_1^n z(t)z(t)'\right\}^{-1}\Sigma_1^n \hat{y}(t)z(t)$$

but $z(t)$, $\hat{y}(t)$ are defined in terms of $\theta(n)$ not the output $\tilde{\theta}(n)$. Rules such as this have been used in practice. An alternative, discussed in [9], is

$$\tilde{\theta}(n) = \theta(n-1) + P(n)z(n)\hat{\varepsilon}(n)$$

where all terms on the right are constructed from $\theta(n)$. This recursion cannot be summed as a regression and seems to be more difficult to analyse.

2. The almost sure convergence of RML_2. It will easily be seen how the discussion of RML_2 extends to RML_1 and this and other questions will be treated in Section 3. All convergence will be almost sure but often the symbol a.s. will be omitted, for brevity.

Let $\Theta_0 \subset \mathfrak{R}$ be the set of points that minimize

(2.1) $$\int_{-\pi}^{\pi} f_0(\omega)|k_\theta|^{-2}\,d\omega, \qquad k_\theta = g_\theta^{-1}h_\theta.$$

Here the argument variable, exp $i\omega$, has been omitted from k_θ and this will often be done. Since all integrals are over $[-\pi, \pi]$ we henceforth omit these limits. It is assumed that Θ_0 is not empty. If (1.1) holds for $\theta = \theta_0$ then Θ_0 contains only θ_0.

Introduce the vector $\phi_\theta(z)$,

(2.2) $$\phi_\theta(z)' = \left(-h_\theta^{-1}.z, \cdots, -h_\theta^{-1}.z^p, h_\theta^{-2}g_\theta.z, \cdots, h_\theta^{-2}g_\theta.z^q\right).$$

Here g_θ, h_θ are evaluated at z but this argument variable has been omitted from the notation for brevity. The place of $\phi_\theta(z)$ in the theory can be perceived from the fact that the (vector) transfer function from $y(t)$ to $z(t)$, if $\alpha_j(t)$, $\beta_j(t)$ were fixed at the values corresponding to the point θ, would be $\phi_\theta(z)$. Put

$$K(\theta) = \int f_0(\omega)\phi_\theta\phi_\theta^* \, d\omega.$$

It is evident that $K(\theta)$ corresponds to $\hat{K}(n)$ in the sense that if $\alpha_j(t)$, $\beta_j(t)$ are held fixed at the values corresponding to θ then $\hat{K}(n) \to K(\theta)$. The condition that θ minimizes (2.1) results, since $\mathfrak{R}$ is open, in

$$\int f_0(\omega)\phi_\theta k_\theta^{-1} \, d\omega = 0, \qquad \theta \in \Theta_0,$$

as is easily checked. Hence calling ϕ, k, h, evaluated at $\theta_0 \in \Theta_0$, ϕ_1, k_1, h_1, then

$$(2.3) \quad \int f_0(\omega)\phi_\theta \bar{k}_\theta^{-1} \, d\omega = \int f_0(\omega)\left\{\phi_\theta \bar{k}_\theta^{-1} - \phi_1 \bar{k}_1^{-1}\right\} d\omega = -L(\theta, \theta_0)(\theta - \theta_0)$$

where

$$L(\theta, \theta_0) = \int f_0(\omega)\phi_2\phi_2^*\left(\bar{h}_2/\bar{h}_1\right) d\omega$$

and ϕ_2, h_2 are evaluated at θ_2, intermediate between θ and θ_0. In case (1.1) holds $\int f_0\phi\bar{k}_0^{-1} \, d\omega = 0$ since $f_0\bar{k}_0^{-1}$ is analytic within the unit circle, as is also ϕ_1 and $\phi(0) = 0$. Thus then

$$(2.4) \qquad L(\theta, \theta_0) = \int f_0\phi_\theta\phi_\theta^* \bar{h}_0^{-1}\bar{h}_\theta \, d\omega.$$

Let $\mathfrak{R}(\theta_0)$ be a closed sphere whose boundary is at least 2δ from the boundary of $\mathfrak{R}$ and such that, for $\theta \in \mathfrak{R}(\theta_0)$, $I - K(\theta)^{-1}L(\theta, \theta_0)$ is less than unity in the Euclidean norm, $\|.\|$. Let $\mathfrak{R}_2$ be the union of the $\mathfrak{R}(\theta_0)$, $\theta_0 \in \Theta_0$. Then choose $\mathfrak{R}_1$ so as to satisfy the requirements of it given in Section 1, as can clearly be done. The set $\mathfrak{R}_2$ is not vacuous since $I - K(\theta_0)L(\theta_0, \theta_0) = 0$.

THEOREM 1. *Under the above conditions* $\tilde{\theta}(n) \to \Theta_0$ a.s.

The proof is long and complicated. A first step is to prove that the smallest eigenvalue of $\hat{K}(n)$ is bounded away from zero so that $\theta(n) - \theta(n-1) = o(1)$ (i.e., converges almost surely to zero) and coefficients in the sums defining $\tilde{\theta}(n)$ as a vector of regression can be lagged into an earlier σ-algebra, thus allowing the martingale convergence theorem (mgct) for square integrable martingales ([10], page 151) to be used. Then $\tilde{\theta}(n)$ is shown to converge to a deterministic formula $\hat{\theta}(n)$. The definition of $\mathfrak{R}_2$ ensures that $\hat{\theta}(n)$ returns infinitely often to $\mathfrak{R}_2$ and the use, essentially, of a Liapounoff function ensures that $\hat{\theta}(n)$ converges to Θ_0.

We state three simply established lemmas for later use.

LEMMA 1. *If* $\Sigma\alpha_j(t)x(t-j) = e(t)$ (*or* $\Sigma\beta_j(t)x(t-j) = e(t)$) *with* $x(j)$ *prescribed for* $-p \leqslant j \leqslant 0$ (*or* $-q \leqslant j \leqslant 0$) *then*

$$(2.5) \qquad x(t) = \Sigma_0^{p-1}a_j(t)x(-j) + \Sigma_1^{t-1}b_j(t)e(t-j)$$

with

$$|a_j(t)|, |b_j(t)| < c\rho^j, \qquad 0 < \rho < 1, \qquad 0 < c.$$

Here and below c is a finite positive constant, not always the same one. Since $g_t(z) = \Sigma\alpha_j(t)z^j$ has all zeros greater than $1 + \delta$ then $g_t(z) = \{1 + z_1(t)z\}\{1 + z_2(t)z\} \cdots \{1 + z_p(t)z\}$ with $|z_j(t)| \leqslant b < 1$. The result then follows by a simple induction.

By successive substitutions using (1.9), (1.10), (1.11)

$$(2.6) \quad \hat{\varepsilon}(t) = \Sigma_0^\infty c_j(t)\varepsilon(t-j),\ \eta(t) = \Sigma_0^\infty d_j(t)\varepsilon(t-j),\ \xi(t) = \Sigma_0^\infty e_j(t)\varepsilon(t-j).$$

LEMMA 2. $|c_j(t)|, |d_j(t)|, |e_j(t)| \leqslant a_j, \Sigma_0^\infty a_j < \infty$.

The proof is fairly immediate, using Lemma 1.

LEMMA 3. *For all* $a > 1$

$$\Sigma_1^n t^{-1}\hat{\varepsilon}(t)^2, \Sigma_1^n t^{-1}\eta(t)^2, \Sigma_1^n t^{-1}\xi(t)^2 < c(\log n)^a.$$

For example, using Lemma 2,

$$\begin{aligned}\Sigma_1^n t^{-1}\hat{\varepsilon}(t)^2 &= \Sigma t^{-1}\{\Sigma_0^\infty c_j(t)\varepsilon(t-j)\}^2\\ &\leqslant c(\log n)^a \Sigma_1^n\left[t^{-1}(\log t)^{-a}\{\Sigma_0^\infty a_j\varepsilon(t-j)^2\}\right],\end{aligned}$$

and the result follows since the factor in braces is stationary with finite mean.

Let $0 < \lambda_1(n) \leqslant \cdots \leqslant \lambda_{p+q}(n)$ be the eigenvalues of $n^{-1}P(n)^{-1}$ and assume $\liminf \lambda_j(n) = 0, j = 1, \cdots, d$; $\liminf \lambda_j(n) > 0, j = d+1, \cdots, p+q$. Let $Q(n)$ be the sequence of orthogonal matrices diagonalizing $P(n)^{-1}$ and choose a subsequence along which $Q(n)$ converges to an orthogonal matrix Q. Then, because $\hat{K}(n)$ has all elements bounded, by Lemma 2, there is a subsequence n_k along which the elements in the first d rows and in the first d columns of $QP(n)^{-1}Q'$, and hence of $Q\hat{K}(n)Q'$ converge to zero. On the other hand, the matrix in the last $p + q - d$ rows and columns of $n^{-1}QP(n)^{-1}Q'$ has its smallest eigenvalue with positive limit inferior. Put $\{Q\theta(n)\}' = (\theta_1(n)':\theta_2(n)')$ where the partition is after the dth element. After $\tilde{\theta}(n)$ has left $\mathfrak{R}_1$, and before it reenters $\mathfrak{R}_2$, $\theta(n)$ and hence $\theta_2(n)$ does not change. It may "jump" at points of exit and reentry. At the remaining points $\theta(n) = \tilde{\theta}(n)$ and $\theta(n) - \theta(n-1) = \tilde{\theta}(n) - \tilde{\theta}(n-1)$. Consider

$$\frac{1}{n}\{QP(n)^{-1}Q'\}Q\theta(n) = \frac{1}{n}\{QP(n)^{-1}Q'\}Q\theta(n-1) + \frac{1}{n}Qz(n)\hat{\varepsilon}(n).$$

At such remaining points we see that $\theta_2(n) - \theta_2(n-1) \to 0$ as $n, n-1$ increase through a sequence of points excluding points of jump. This is because $n^{-1}Qz(n)\hat{\varepsilon}(n) \to 0$, in virtue of Lemma 2, and the fact that $n^{-1}\varepsilon(n)^2 \to 0$. The proof then follows by dividing these "remaining" points into $d + 1$ sets according as $\lambda_1(n) \geqslant \varepsilon > 0$; $\lambda_1(n) < \varepsilon, \lambda_2(n) \geqslant \varepsilon; \cdots; \lambda_{d-1}(n) < \varepsilon, \lambda_d(n) \geqslant \varepsilon; \lambda_d(n) < \varepsilon$. As n increases through each set, $|\theta_2(n) - \theta_2(n-1)| \leqslant c\varepsilon + o(1)$ which establishes the result since ε is arbitrary.

Let e_u be the uth point of exit from $\mathfrak{R}_1$ and r_u be the subsequent point of reentry into $\mathfrak{R}_2$. Were lim inf $\lambda_1(n) > 0$ it could be shown (see below) that these points e_u, r_u are very sparse by showing that it must take an increasingly long time for $\theta(n)$ to move $\mathfrak{R}_2$ out of $\mathfrak{R}_1$ (and from entry into $\mathfrak{R}_1$ back into $\mathfrak{R}_2$). However in the present circumstance it is possible that $\theta(n)$ might move from $\mathfrak{R}_2$ out of $\mathfrak{R}_1$ (for example) along a line whereon $\theta_2(n)$ hardly changed, and very rapidly, since $\theta_1(n)$ might be changing rapidly because $\lambda_j(n) \to 0, j \leqslant d$. On the other hand when $\tilde{\theta}(n)$ is outside of $\mathfrak{R}_1$ it can, in principle, become arbitrarily large. Consider, therefore, first points e_u, r_u for which

$$\sup_{e_u \leqslant n \leqslant r_u} \|\tilde{\theta}(n)\| \leqslant A < \infty \tag{2.7}$$

and one or both of

$$\|\theta_2(e_u) - \theta_2(e_u - 1)\| > \nu, \qquad \|\theta_2(r_u - 1) - \theta_2(r_u)\| > \nu,$$

hold. *We shall not introduce a special notation for such points "of type A" but for the moment only these are considered.* To the vth such point, $\theta_2(n)$ will have changed by at most

$$c\Sigma_1^{r_v} t^{-1}\|z(t)\tilde{\varepsilon}(t)\| + c \leqslant c(\log r_v)^a \tag{2.8}$$

using the boundedness of $\tilde{\theta}(n)$, of the elements of $P(n)$ and the boundedness away from zero of the norm of the submatrix of $PK(n)P'$ in the last $p + q - d$ rows and columns. The bound on the right in (2.8) comes from Lemma 3. Thus

$$v\nu \leqslant c(\log r_v)^a. \tag{2.9}$$

Next consider points "of type B", for which the opposite of (2.7) holds i.e., $\|\tilde{\theta}(n)\|$ eventually exceeds A.

$$\|\tilde{\theta}(n) - \tilde{\theta}(n-1)\| \leqslant c\lambda_1(n)^{-1}n^{-1}\|z(n)\tilde{\varepsilon}(n)\|.$$

Since $n\lambda_1(n) \geqslant c$ then when $\|\tilde{\theta}(n-1)\| \geqslant c$

$$\log\left\{\frac{\|\tilde{\theta}(n)\|}{\|\tilde{\theta}(n-1)\|}\right\} \leqslant \left|\frac{\|\tilde{\theta}(n)\|}{\|\tilde{\theta}(n-1)\|} - 1\right| \leqslant c\|z(n)\|\,|\hat{\varepsilon}(n) + \eta(n) - \xi(n)| + cz(n)'z(n).$$

Thus up to the conclusion of the vth excursion of type B

$$v \log A \leqslant c\{r_v + o(r_v)\}.$$

This follows from Lemma 2 and the ergodic theorem. Thus

$$\limsup\{v/r_v\} \leqslant c/\log A.$$

Now taking A arbitrarily large it is evident that the proportion of points of type B in the first n points may be made as small as is desired and this is also true for points of type A for any $\nu > 0$, as follows, even more strongly, from (2.9). We shall omit these points, and N points (N fixed) before and after them. This will asymptotically make no difference to $x'\hat{K}(n_k)x$, where x' is one of the first d rows of Q, since that is converging to zero in any case and the contribution of each term to that quantity is asymptotically zero. Nevertheless we later need the fact that the points omitted can be taken to be a proportion of all points to time n_k that converges to zero. Let Σ' indicate a sum wherein these points are omitted. In

$$n_k^{-1}\Sigma_1^{n_k} x'z(t)z(t)'x \tag{2.10}$$

we shall repeatedly substitute for $\xi(t)$ by means of (1.11), expressed in terms only of the $\eta(t-j)$. Now

$$x'z(t) = \Sigma_0^s f_j(t)\eta(t-j) + r_s(t)$$

where

$$\frac{1}{n}\Sigma_1^n r_s(t)^2$$

may be made as small as is desired by taking s large. This follows from Lemmas 1 and 2. We shall show that in

$$n_k^{-1}\Sigma_1^{n_k}\{\Sigma_0^s f_j(t)\eta(t-j)\}^2 \tag{2.11}$$

any coefficient may be lagged, i.e., replaced by $f_j(t-l)$, $l \leqslant N'$, where N' is arbitrarily large but fixed, so as to make a negligible difference, asymptotically, to (2.11). Indeed, each $f_j(t)$ is a sum of products of elements of $Q\theta(t-j)$, up to a fixed lag determined by s. Recall, from (1.11), that $\xi(t) = \eta(t) - \theta'(t-1)z(t) = \eta(t) - \{Q\theta(t-1)\}'Qz(t)$. Thus the factors comprising a term in $f_j(t)$ can be decomposed into components in $\theta_1(t-j)$ and $\theta_2(t-j)$, for various j. Since s and N' are fixed, we may choose N so large, but fixed, so that no time point, $t-j$, at which a jump occurs, nor any time point before and after that, arises in this process if the sum in (2.11) is restricted to Σ'. So far as components in $\theta_2(t)$ are concerned, we have already seen that these may be lagged as desired so as to make an asymptotically negligible difference to (2.11). Thus if $a(t)$ is a term in $f_j(t)$ we have $a(t) - a(t-1) = o(1)$ and hence the effect of replacing $a(t)$ by $a(t-1)$ in $f_j(t)\eta(t-j)$ gives a contribution to (2.11) that is $o(1)$, using Schwartz's inequality and the ergodic theorem. Consider then components in $\theta_1(t)$. Say $b(t)$ is such a component and it occurs in the substitution process for the first time as a coefficient of $\zeta(t)$, which is necessarily a component in $Qz(t-j)$ for some j. Then a typical term arising at this moment will be of the form $a(t)b(t)\zeta(t)$. However we know that

$$n_k^{-1}\Sigma_1^{n_k}\zeta(t)^2 \to 0,$$

and hence it will make no asymptotic difference if we lag $b(t)$ in any way (remembering that $a(t)$, $b(t)$ are uniformly bounded). Now we may finally reinsert the omitted points of jump, and associated points, whose coefficients may certainly be lagged arbitrarily since the proportion of these points in the first n_k converges to zero. Thus the desired result concerning (2.11) and the lagging (or leading for that matter) of its coefficients is established. Now it will be shown that

$$n_k^{-1}\Sigma_1^{n_k} f_j(t)^2 \to 0 \qquad j = 1, 2, \cdots. \tag{2.12}$$

To see this observe that (2.11) is

$$n_k^{-1}\Sigma_1^{n_k}\left[f_0(t)\varepsilon(t) + \{f_0(t)\psi(t) + \Sigma_1^s f_j(t)\eta(t-j)\}\right]^2$$

where $\eta(t) = \varepsilon(t) + \psi(t)$ and $\psi(t)$ is measurable $\mathcal{F}_{t-1}$. By the mgct, using our ability to lag $f_j(t)$ "into an earlier σ-algebra", it follows that

$$\lim_{k\to\infty} n_k^{-1}\Sigma_1^{n_k} f_0(t)\varepsilon(t)\{f_0(t)\psi(t) + \Sigma_1^s f_j(t)\eta(t-j)\} = 0.$$

Hence

$$\lim_{k\to\infty} n_k^{-1}\Sigma_1^{n_k} f_0^2(t)\varepsilon(t)^2 = 0$$

because (2.10) converges to zero. However because of the second part of (1.2), putting $\varepsilon^{(1)}(t)^2 = \varepsilon(t)^2$, for $|\varepsilon(t)|^2 < c$ and $\varepsilon^{(2)}(t)^2 = \varepsilon(t)^2 - \varepsilon^{(1)}(t)^2$ then

$$n_k^{-1}\Sigma_1^{n_k} f_0(t)^2\{\varepsilon(t)^2 - \sigma^2\} = n_k^{-1}\Sigma f_0(t)^2\left[\varepsilon^{(1)}(t)^2 - \mathcal{E}\{\varepsilon^{(1)}(t)^2|\mathcal{F}_{t-1}\}\right]$$
$$+ n_k^{-1}\Sigma f_0(t)^2\left[\varepsilon^{(2)}(t)^2 - \mathcal{E}\{\varepsilon^2(t)^2|\mathcal{F}_{t-1}\}\right].$$

The first term on the right converges to zero by mgct (using the ability to lag the $f_0(t)^2$). The second may be made as small as is desired by taking c large by the ergodic theorem combined with the boundedness of the $f_0(t)$. Thus $\sigma^2 n_k^{-1}\Sigma f_0(t)^2 \to 0$ and (2.12) is established for $j = 0$. Now consider (2.11) with the term for $j = 0$ omitted and repeat the argument. This establishes (2.12) for $j = 0, 1, \cdots, s$ and since $f_j(t)$ converges geometrically to zero as j increases, uniformly in t, by Lemma 1, (2.12) holds in general. Now $f_j(t)$ is a sum of products of $\alpha_u(t-v)$, $\beta_u(t-v)$ and we know that these may be lagged or lead up to any arbitrarily large but fixed lag and that $f_j(t)$ converge geometrically to zero at a rate independent of t. Thus in (2.12) $f_j(t)$ may be evaluated as if all $\alpha_u(t-v)$, $\beta_u(t-v)$ were replaced by $\alpha_u(t)$, $\beta_u(t)$. When this is done $f_j(t)$ becomes the coefficient of z^j in the expansion of $h_t(z)\phi_t(z)'x$ (where $\phi_\theta(z)$ is defined in (2.2) and ϕ_t, h_t are ϕ_θ, h_θ for $\theta = \theta(t)$). Indeed in the process of substitution the relation now being used is $\eta(t-j) - \theta'(t)z(t-j) = \xi(t-j)$, for all j. This means that $\xi(t-j) = h_t^{-1}g_t\eta(t-j)$, where z, in h_t, g_t, is being interpreted as the lag operator. Thus $x'z(t)$ may be written as

$$-\Sigma x_j^{(1)}\eta(t-j) + \Sigma x_j^{(2)}k_t^{-1}\eta(t-j)$$

where the $x_j^{(1)}$ are the first p components in x and the $x_j^{(2)}$ are the last q. This is of the form $\{h_t x'\phi_t\}\eta(t)$ which establishes what we wish. Thus

$$\text{(2.13)}\qquad \lim_{k\to\infty}\Sigma_{j=0}^{\infty} n_k^{-1}\Sigma_1^{n_k} f_j(t)^2 = \lim_{k\to\infty} n_k^{-1}\Sigma_1^{n_k}\frac{1}{2\pi}\int x'\phi_t\phi_t^* x|h_t|^2\,d\omega.$$

However $|h_t|^2 x'\phi_t\phi_t' x$ is

$$|\psi_{p-1}(e^{i\omega})h_t(e^{i\omega}) - \chi_{q-1}(e^{i\omega})g_t(e^{i\omega})|^2|h_t|^{-2}$$

where ψ_{p-1}, χ_{q-1} are polynomials (with coefficients the elements of x) of the indicated degrees. Now, secondly, since $|h_t| \geqslant \delta > 0$, we may consider

$$\frac{1}{2\pi}\int|\psi_{p-1}(e^{i\omega})h_t(e^{i\omega}) - \chi_{q-1}(e^{i\omega})g_t(e^{i\omega})|^2\,d\omega$$

which is of the form $\|A_t x\|^2$ where det A_t is the resultant of the polynomials g_t, h_t and since this resultant is bounded away from zero the result follows. Thus (2.13) and hence (2.10) cannot converge to zero and a contradiction has been reached. Thus we have

PROPOSITION 1. $\liminf \lambda_1(n) > 0$, a.s.

Sum the recursion so that

$$\tilde{\theta}(n) = \hat{K}(n)^{-1}\frac{1}{n}\Sigma_1^n\{\hat{\varepsilon}(t) + \eta(t) - \xi(t)\}z(t) + o(1).$$

Now

$$\{\log e_v\}^a \geqslant cv, \qquad a > 1$$

the proof being as before, but simpler, since $\theta(n)$ must change appreciably when moving from $\mathfrak{R}_2$ out of $\mathfrak{R}_1$. (See (2.9)). Using Lemma 2 the contribution from points of exit and entry to a typical element in the averages forming $\tilde{\theta}(n)$ can be dominated by

$$cn^{-1}\Sigma_1'^n\{\Sigma_0^\infty a_j^2\varepsilon(t-j)^2\}$$

the sum being over points of entry and exit. Taking points of entry for example, and putting $b = a^{-1}$,

$$\Sigma' t^{-1}\varepsilon(t-j)^2 \leqslant c\Sigma_v \exp(cb^b)\varepsilon(v\nu - j)^2 < \infty$$

so that, using Kronecker's lemma, the contribution from these points to $\tilde{\theta}(n)$ converges to zero and this will be true also of the N points before and after them. Now the argument of the type that established (2.13) may be repeated. Thus express any element of $z(t)$ or $\hat{\varepsilon}(t)$ as, for example, $\hat{\varepsilon}(t) = \Sigma c_j(t)\varepsilon(t-j)$, observe that the $c_j(t)$ may be lagged up to any fixed amount, that Lemma 2 holds and then Proposition 2 below follows. In this proposition

$$(2.14)\quad \hat{\theta}(n) = \left[\frac{1}{n}\Sigma_1^n\int f_0(\omega)\phi_t\phi_t^*\,d\omega\right]^{-1}\frac{1}{n}\Sigma_1^n\int f_0(\omega)\phi_t\left\{\frac{g_t}{h_t} - \frac{g_t}{h_t^2} + \frac{1}{h_t}\right\}^* d\omega.$$

In (2.14) ϕ_t, g_t etc. are functions of exp $i\omega$ and a star has been used also for conjugation, for convenience.

PROPOSITION 2. $\lim_{n\to\infty}\{\tilde{\theta}(n) - \hat{\theta}(n)\} = 0$ a.s.

The form of $u\hat{\theta}(n)$ *is simple* in that ϕ_t; $\{h_t^{-1}g_t - h_t^{-2}g_t + h_t^{-1}\}$ are just the instantaneous z transforms that express $z(t), \hat{y}(t)$ in terms of $y(t-j), j \geqslant 0$. Thus $z(t)z(t)'$, for example, has been replaced by its expectation as if it were obtained from $y(t)$ by a filter with fixed coefficients having z-transform ϕ_t. Assume that $\tilde{\theta}(n)$ and hence $\hat{\theta}(n)$ enters $\mathfrak{R}_2$ only finitely often and ultimately remains outside of $\mathfrak{R}_2$ after last leaving $\mathfrak{R}_1$. Then

$$\hat{\theta}(n) \to \hat{\theta} = K(\theta)^{-1}\int f_0(\omega)\phi_\theta\left\{\frac{g_\theta}{h_\theta} - \frac{g_\theta}{h_\theta^2} + \frac{1}{h_\theta}\right\}^* d\omega$$

where $K(\theta)$ was defined above (2.2) and θ is the value at which $\theta(n)$ is held after the last exit of $\tilde{\theta}(n)$ from $\mathfrak{R}_1$. Then (see (2.3))

$$(2.15)\quad \hat{\theta} = \theta + K(\theta)^{-1}\int f_0(\omega)\phi_\theta\left(\frac{g_\theta}{h_\theta}\right)^* d\omega = \theta - K(\theta)^{-1}L(\theta, \theta_0)(\theta - \theta_0)$$

for any $\theta_0 \in \Theta_0$. Thus

$$(2.16)\qquad (\hat{\theta} - \theta_0) = \{I - K(\theta)^{-1}L(\theta, \theta_0)\}(\theta - \theta_0)$$

and a contradiction is reached because $\|I - K(\theta)^{-1}L(\theta, \theta_0)\| < 1$ for $\theta \in \mathfrak{R}(\theta_0)$ and θ does lie in one such set. Thus we have the following.

PROPOSITION 3. *$\tilde{\theta}(n)$ returns infinitely often to $\mathfrak{R}_2$ or eventually remains permanently in $\mathfrak{R}_1$*, a.s.

The proof may now be completed by the construction, essentially, of a Liapounoff function (that was suggested to the author by L. Ljung). Consider

$$(2.17)\qquad \int f_0(\omega)\left\{\left|\frac{\hat{g}_n}{\hat{h}_n}\right|^2 - \left|\frac{\hat{g}_{n-1}}{\hat{h}_{n-1}}\right|^2\right\} d\omega$$

where $\hat{g}_n$, $\hat{h}_n$ are constructed from $\hat{\theta}(n)$. Now when $\tilde{\theta}(n) = \theta(n)$ then $\hat{\theta}(n) - \theta(n) = o(1)$ and thus when $\tilde{\theta}(n)$, i.e., $\theta(n)$, is in $\mathfrak{R}_1$ before leaving that region,

$$(2.18)\qquad \hat{\theta}(n) = \tilde{K}(n)^{-1}\frac{1}{n}\Sigma_1^n\int f_0(\omega)\hat{\phi}_t\left\{\frac{\hat{g}_t}{\hat{h}_t} - \frac{\hat{g}_t}{\hat{h}_t^2} + \frac{1}{\hat{h}_t}\right\}^* d\omega + o(1)$$

where $\hat{\phi}_t$ is ϕ_t constructed from $\hat{\theta}_t$ and $\hat{K}(n)^{-1}$ is the first factor in (2.14) but with g_t, h_t replaced by $\hat{g}_t$, $\hat{h}_t$. Thus it will be sufficient to show that the first term on the right converges to Θ_0. *For simplicity we shall now call the first term on the right $\hat{\theta}(n)$.* Now, from (2.3), when $\theta(n)$, i.e., $\hat{\theta}(n)$, is in $\mathfrak{R}_1$ before leaving that region,

$$\hat{\theta}(n) - \hat{\theta}(n-1) = \frac{1}{n}\tilde{K}(n)^{-1}\int f_0(\omega)\hat{\phi}_n\left(\frac{\hat{g}_n}{\hat{h}_n}\right)^* d\omega$$
$$= -\frac{1}{n}\tilde{K}(n)^{-1}L(\hat{\theta}(n), \theta_0)\{\hat{\theta}(n) - \theta_0\}.$$

Also (2.17) is

$$\int f_0(\omega)\left\{\partial\theta\left|\frac{\hat{g}}{\hat{h}}\right|^2\right\}_{\hat{\theta}(n)} d\omega\{\hat{\theta}(n) - \hat{\theta}(n-1)\} + o(n^{-2})$$

where $\partial\theta$ indicates the gradient vector with respect to θ and we have used the fact that $\hat{\theta}(n) - \hat{\theta}(n-1) = 0(n^{-1})$ to show that the neglected term is of the indicated order. Thus (2.17) is

$$(2.19)\quad 2\int f_0(\omega)\hat{\phi}_n^*\left(\frac{\hat{g}_n}{\hat{h}_n}\right) d\omega\{\hat{\theta}(n) - \hat{\theta}(n-1)\} + 0(n^{-2})$$
$$= -\frac{2}{n}[\hat{\theta}(n) - \theta_0\}'L(\hat{\theta}(n), \theta_0)^*\hat{K}(n)^{-1}L(\hat{\theta}(n), \theta_0)\{\hat{\theta}(n) - \theta_0\} + 0(n^{-2}).$$

The factor multiplying n^{-1} is $\leqslant \delta(\varepsilon) < 0$ if $\hat{\theta}(n) \in \mathfrak{R}(\theta_0)$ and $\|\hat{\theta}(n) - \theta_0\| \geqslant \varepsilon$. This is because $\|I - K(\theta)^{-1}L(\theta, \theta_0)\| < 1$ and $\|K(\theta)\| \geqslant \delta_2 > 0$. Also

$$(2.20) \qquad \int f_0(\omega)\left|\frac{\hat{g}_n}{\hat{h}_n}\right|^2 d\omega \geqslant \sigma^2.$$

Now for n sufficiently large $\hat{\theta}(n)$ will have returned inside $\mathfrak{R}_2$ and will be an indefinitely long time in $\mathfrak{R}_1$ thereafter. If $\|\hat{\theta}(n) - \theta_0\| \geqslant \varepsilon$ for all $\theta_0 \in \Theta_0$ then for some such θ_0, the right side of (2.19) is not greater than $-\delta(\varepsilon)n^{-1} + 0(n^{-2})$. Since Σn^{-1} diverges a contradiction to (2.20) is reached and the theorem is proved.

THEOREM 2. *Let the conditions of Theorem* 1 *hold and* $y(t)$ *be generated by* (1.1) *and additionally let* $\mathcal{E}\{\varepsilon(n)^4\} < \infty$. *Then* $\lim_{n\to\infty} n^a(\tilde{\theta}(n) - \theta_0) = 0$, a.s., $a < \frac{1}{2}$.

In Lemma 2, now, $a_j < c\rho^j$, $0 < \rho < 1$.

There is an n_0, $P(n_0 < \infty) = 1$, so that $\tilde{\theta}(n) = \theta(n)$, $n > n_0$. Also since $\tilde{\theta}(n) - \tilde{\theta}(n-1) = n^{-1}\hat{K}(n)^{-1}z(n)\tilde{\varepsilon}(n)$ and $n^{-\frac{1}{4}}\varepsilon(n) \to 0$ then almost precisely as before $\tilde{\theta}(n) - \tilde{\theta}(n-1) = o(n^{-\frac{1}{2}})$. For $n > n_0$

$$\theta(n) = \hat{K}(n)^{-1}\frac{1}{n}\Sigma_1^n \hat{y}(t)z(t)$$

and a typical element of $\hat{K}(n)$ or of $n^{-1}\Sigma\hat{y}(t)z(t)$ is of the form

$$(2.21) \quad \frac{1}{n}\Sigma_1^n\{\Sigma_0^\infty a_j(t)\varepsilon(t-j)\Sigma_0^\infty b_j(t)\varepsilon(t-j)\};\ |a_j(t)|, |b_j(t)| < c\rho^j, 0 < \rho < 1.$$

Consider the part of (2.21) comprised by

$$\frac{1}{n}\Sigma_1^n\{\Sigma\Sigma_{j>k=0}^\infty a_j(t)b_k(t)\varepsilon(t-j)\varepsilon(t-k)\}.$$

Let us consider the error incurred by replacing $a_j(t)$, $b_k(t)$ by $a_j(t-j)$, $b_k(t-k)$. The error is the sum of three terms, of which we take one for example

$$(2.22) \qquad \frac{1}{n}\Sigma_1^n\Sigma\Sigma_{j>k=0}^\infty\{a_j(t) - a_j(t-j)\}b_k(t)\varepsilon(t-j)\varepsilon(t-k).$$

Take $n_1 \gg n_0$ and $J(n) = b\log n$, so that $J(n)\log\rho < -\log n$. Then (2.22) is dominated by

$$cn^{-\frac{1}{2}} + \left|\frac{1}{n}\Sigma_{n_1}^n\Sigma_{j>k=0}^{J(n)-1}\{a_j(t) - a_j(t-j)\}b_k(t)\varepsilon(t-j)\varepsilon(t-k)\right|$$
$$+\left|\frac{1}{n}\Sigma_{n_1}^n\Sigma_{j>k=J(n)}^\infty\{a_j(t) - a_j(t-j)\}b_k(t)\varepsilon(t-j)\varepsilon(t-k)\right|.$$

The last term is dominated by

$$cn^{-3}\Sigma_{n_1}^n\Sigma_{j>k=J(n)}^\infty\rho^{j+k-2J(n)}|\varepsilon(t-j)\varepsilon(t-k)|$$

which is evidently $0(n^{-2})$. The middle term is dominated by

$$(2.23) \qquad cn^{-1}(\log n)^2\Sigma_{n_1}^n t^{-\frac{1}{2}}\Sigma_{j>k=0}^{J(n)}\rho^k|\varepsilon(t-j)\varepsilon(t-k)|$$

since $a_j(t)$ is a sum of at most cj terms each a product of cj factors that are elements of $\theta(s)$, $s \leqslant t$, so that $|a_j(t) - a_j(t-j)|$ is dominated by $cj^2 t^{-\frac{1}{2}}$. However (2.23) is dominated by

$$c(\log n)^2 \Sigma_{j=0}^{J(n)} \frac{1}{n} \Sigma_1^n \{ t^{-\frac{1}{2}} |\varepsilon(t-j)| \Sigma_0^\infty \rho^k |\varepsilon(t-k)| \}$$
$$\leqslant c(\log n)^3 n^{-\frac{1}{2}+\delta} \Big[\Sigma_{-\infty}^{\infty} t^{-|1+\delta|} \varepsilon(t)^2 \Sigma_0^\infty t^{-(1+\delta)} \{ \Sigma_0^\infty \rho^k |\varepsilon(t-k)| \}^2 \Big]^{\frac{1}{2}}.$$

for $\delta > 0$. This is eventually $o(n^{-a})$, $a < \frac{1}{2}$. The other two error terms resulting from the replacement of $a_j(t)$, $b_k(t)$ by $a_j(t-j)$, $b_k(t-k)$ may be treated similarly. Now

$$X_n = \Sigma_1^n \Sigma\Sigma_{j<k=0}^\infty a_j(t-j) b_k(t-k) \varepsilon(t-j) \varepsilon(t-k)$$

is a martingale and the increasing process associated with X_n^2 (see [10] page 148) is

$$A_n = \Sigma_1^n \big[\Sigma\Sigma_{j<k=1}^\infty a_j(t-k) b_k(t-k) \varepsilon(t-j) \varepsilon(t-k) \big]^2$$

which is clearly $0(n)$. Thus $n^{-1} X_n$ is $o(n^{-a})$, $a < \frac{1}{2}$, by [10] (page 151). The terms in (2.21) for $j < k$ thus are $o(n^{-a})$ and the same is true for $j > k$. Hence to order n^{-a} we need consider only

$$\frac{1}{n} \Sigma_1^n \Sigma_0^\infty a_j(t) b_j(t) \varepsilon(t-j)^2 = \frac{1}{n} \Sigma_1^n \Sigma_0^\infty a_j(t) b_j(t)$$
$$+ \frac{1}{n} \Sigma_1^n \Sigma_0^\infty a_j(t) b_j(t) \{ \varepsilon(t-j)^2 - \sigma^2 \}.$$

Since $\varepsilon(t)^2 - \sigma^2$ is a sequence of square integrable martingale differences we see that the second term is $o(n^{-a})$ and hence (2.21) is

$$\frac{1}{n} \Sigma_1^n \{ \Sigma_0^\infty a_j(t) b_j(t) \} + o(n^{-a}), \qquad a < \tfrac{1}{2}. \tag{2.24}$$

Now using the same technique as was used on (2.22) we may show that in (2.24) $a_j(t)$, $b_j(t)$ may be replaced by corresponding expressions wherein the elements of $\theta(s)$, $s < t$, wherever they occur, may be replaced by the same elements of $\theta(t)$ with an error that is $\theta(n^{-a})$, $a < \frac{1}{2}$. Hence, for $n > n_0$,

$$\theta(n) = \hat{\theta}(n) + o(n^{-a}), \qquad a < \tfrac{1}{2}.$$

Now put $\hat{Q}(n) = \{\hat{\theta}(n) - \theta_0\} \tilde{K}(n) \{\hat{\theta}(n) - \theta_0\}$. Then

$$\hat{Q}(n) = \hat{Q}(n-1)\left(1 - \frac{1}{n}\right) + \frac{1}{n} \{ \hat{\theta}(n-1) - \theta_0 \}' \{ \tilde{K}(n) - 2L(\hat{\theta}(n), \theta_0) \}$$
$$+ o(n^{-1})$$

remembering that $\hat{\theta}(n) - \theta_0 \to 0$, $\hat{\theta}(n) - \hat{\theta}(n-1) = 0(n^{-1})$. We know that eventually $\tilde{K}(n) - 2L(\hat{\theta}(n), \theta_0) > 0$. Thus

$$\hat{Q}(n) = \hat{Q}(n-1)\left(1 - \frac{1}{n}\right) + o(n^{-1}), \tag{2.25}$$

and hence, putting $n^{2a}\hat{Q}(n) = r(n)$

$$r(n) = r(n-1)\left(1 - \frac{b}{n}\right) + o(1)$$

where $b = (1 - 2a) > 0$. However this implies that $r(n) \to 0$ and the result follows.

3. Discussion

1. It is not easy to analyze the nature of the convergence properties of RML_1 when (1.1) does not hold because (2.1) is then not relevant. Thus we restrict ourselves to the case where (1.1) holds. Now Θ_0 is just that true point, θ_0. The proof down to Proposition 3 is hardly altered except that now (see the explanation below Proposition 2)

$$\phi_\theta(z)' = (-z, -z^2, \cdots, -z^p, k^{-1}z, \cdots, k^{-1}z^q)$$

and

$$L(\theta, \theta_0) = \int f_0 \phi_\theta \phi_\theta^* \bar{h}_0^{-1}\, d\omega,$$

while $\mathfrak{R}_2 = \mathfrak{R}(\theta_0)$. As said before, (2.1) is no longer relevant and a Liapounoff function has to be constructed in another way. Consider $Q(n) = \{\hat{\theta}(n) - \theta_0\}'\hat{K}(n)\{\hat{\theta}(n) - \theta_0\}$, as in [5]. Since

$$\hat{\theta}(n) = \hat{\theta}(n-1) - \frac{1}{n}\hat{K}(n)^{-1}\int f_0 \hat{\phi}_n \hat{\phi}_n^* \bar{h}_0^{-1}\, d\omega\{\hat{\theta}(n-1) - \theta_0\} + 0(n)^{-2}$$

then

(3.1)

$$\hat{Q}(n) = \hat{Q}(n-1)\left(1 - \frac{1}{n}\right) - \frac{1}{n}\{\hat{\theta}(n-1) - \theta_0\}'\int f_0 \hat{\phi}_n \hat{\phi}_n^* \{2\mathfrak{R}(h_0^{-1}) - 1\}\, d\omega$$
$$\times \{\hat{\theta}(n-1) - \theta_0\} + 0(n^{-2}),$$

where now $\mathfrak{R}(h_0^{-1})$ indicates the real part of h_0^{-1}. Now

$$2(h_0^{-1}) - 1 = |h_0|^{-2}\{1 - |\Sigma_1^q \beta_{0j} e^{i,\omega}|^2\}$$

so that a sufficient condition for $\hat{Q}(n)$ to converge to zero is that $\Sigma|\beta_{0j}| < 1$. Of course the convergence of $\hat{Q}(n)$ to zero is necessary and sufficient for the convergence of $\hat{\theta}(n)$, and hence $\tilde{\theta}(n)$, to θ_0. If $q = 1$ then $|\beta_{01}| < 1$ so that convergence does then take place. (In fact it is shown then, in [5], that no regions $\mathfrak{R}_1$, $\mathfrak{R}_2$ need then be introduced.) However for $q > 1$ the matrix of the quadratic form in (3.1) is not necessarily positive definite and in fact simulations (see [11]) show that $\tilde{\theta}(n)$ need not converge to θ_0.

2. It may be shown by detailed calculation, for RML_2, that when $p = 0$, $q = 1$, and (1.1) holds, when $\theta \in \mathfrak{R}$ then so necessarily does $\hat{\theta}$. (Thus $\mathfrak{R}_1$, $\mathfrak{R}_2$ may then be much more generously defined.) However this is not so in general. Indeed let $h_0 = (1 - \rho_0 z)^{-1}$, $h = (1 - \rho z)^{-1}$, $0 < \rho, \rho_0 < 1$. Then

(3.2)
$$\int \frac{|h_0|^2}{|h|^2}\left|2 - \frac{\hat{h}}{h}\right|^2 d\omega$$

is minimized, if $\hat{h}$ is not restricted to be a polynomial, at $\hat{h} = h\{2 - h_0^{-1}h\}$. Assume $2\rho_0 - \rho > 1$. Then $\hat{h}$ has a zero inside the unit circle. We may now find two sequences of polynomials h_{0q}, h_q that have no zeros within the unit circle and that converge uniformly to h_0, h and it may be shown that there is a corresponding sequence $\hat{h}_q$, minimizing (3.2) with h_0, h replaced by h_{0q} and h_q and of degree q, that converges uniformly to $\hat{h}$. Now by a theorem of Hurwitz (see [2] page 194]) it follows that, for sufficiently large q, $\hat{h}_q$ will have a zero within the unit circle.

Nevertheless it is probable that the region $\mathfrak{R}_2$ as defined above is much too restrictive. As mentioned earlier, the definition used relates closely to that which would be used in proving the iterative convergence of a Gauss-Newton algorithm for optimizing a Gaussian likelihood. Indeed $\hat{\theta}$ is just the limit to which a first iterate would converge as $n \to \infty$, if the initial value was θ.

3. The discussion can be extended to ARMAX models, i.e., models of the form

$$\Sigma_0^p \alpha_j y(t-j) = \Sigma_1^r \gamma_j u(t-j) + \Sigma_0^q \beta_j \varepsilon(t-j),$$

Introducing $j(z) = \Sigma \gamma_j z^j$ it will now be necessary to require that any three zeros, one from each g, h, j, form a set of diameters $> \delta$ in the complex plane, or, equivalently, that the resultants of the three polynomials be bounded away from zero. Also the coefficients γ_j have to be bounded in the treatment considered by us. It has not been found possible to discuss convergence unless $u(t)$ is restricted and we require that $u(t)$ be generated by a regular stationary process with zero mean. (Mean correction of the data will not affect the results. Indeed trend correction of $y(t)$, $u(t)$ could simultaneously be carried out by a separate recursive procedure.) Thus

$$u(t) = \Sigma_0^\infty \left[\mathcal{E}\{u(t)|\mathcal{F}_{t-j}(n)\} - \mathcal{E}\{u(t)|\mathcal{F}_{t-j-1}(n)\} \right],$$

and for fixed j the summand, $\xi(t-j)$ let us say, is a sequence of square integrable martingale differences. Under these conditions results entirely analogous to those given in Section 2 may be established (see [6]).

4. Theorem 2 would hold even if (1.1) was not true provided the κ_j satisfy $\kappa_j < c\rho^j$, $0 < \rho < 1$ (for example if $y(t)$ has a rational spectrum) in the sense that the distance from $\tilde{\theta}(n)$ to Θ_0 is $o(n^{-a})$, $a < \frac{1}{2}$.

5. The standard situation in which Theorem 1 applies would be the so-called over-identified situation where Θ_0 would consist of one point. However the opposite, under-identified, case can arise, at least in simulations. In such a case, Θ_0 could consist of a continuum and $\tilde{\theta}(n)$ would not converge in a conventional sense but rather could be expected to approach the continuum but "search" along it. A case that has been brought to my attention by Dr. P. C. Young is of a vector ARMAX model where data is generated according to

$$y(n) = A^{-1}Bu(n) + C^{-1}D\varepsilon(n),$$

A, B, C, D being polynomials in the lag operator of degrees p, q, r, s. If this is regarded as an unconstrained ARMAX model with generating functions of degrees

$p + r, q + r, p + s$ respectively, then too many parameters are being estimated, we are in the under-identified case, and it can be expected that the recursion will search over the parameter space in a haphazard way as it approaches the continuum Θ_0.

Acknowledgment. I am grateful to Mr. V. Solo for picking up a slip in (3.1) and for suggesting the argument via (2.25), to him and Dr. P. C. Young for other stimulating discussions and to Professor Ljung for suggesting the Liapounoff function used in the proof of Theorem 1.

REFERENCES

[1] BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc.*, B **37** 149–192.

[2] CARATHÉODORY, C. (1954). *Theory of Functions*. Chelsea, New York.

[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] GERTLER, J. and BÁNYÁSZ, C. S. (1974). A recursive (on line) maximum likelihood identification method. *IEEE Trans. Automatic Control* AC-**19** 816–819.

[5] HANNAN, E. J. (1976). The convergence of some recursions. *Ann. Statist.* **4** 1258–1270.

[6] HANNAN, E. J. and TANAKA, K. (1978). ARMAX models and recursive calculations. In *Systems Dynamics and Control in Quantitative Economics* (H. Myoken, ed.) pages 173–198. Bunshinao, Tokyo.

[7] LJUNG, L. (1976). On consistency for prediction error identification methods. In *System Identification: Advances and Case Studies*. (Mehra, R. K. and Lainiotes, D. G., eds.) Academic Press, New York.

[8] LJUNG, L (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control.* AC-**22** 551–575.

[9] LJUNG, L. (1978). Strong convergence of a stochastic approximation algorithm. *Ann. Statist.* **6** 680–696.

[10] NEVEU, J. (1975). *Discrete Parameter Martingales*. North Holland, Amsterdam.

[11] SODERSTRÖM, T., LJUNG, J., GUSTAVSSON, I. (1974). *A Comparative Study of Recursive Methods*. Report 7427, Lund Institute of Technology, Department of Automatic Control, Lund, Sweden.

[12] SOLO, V. (1979). The convergence of AML. Unpublished manuscript.

[13] SOLO, V. (1980). The second order properties of a Time Series Recursion. To appear in *Ann. Statist.*

[14] YOUNG, P. (1974). Recursive approaches to time series analysis. *Bull. Inst. Math. App.* **10** 209–224.

[15] YOUNG, P. C. and JAKEMAN, A. (1979). Refined instrumental variable methods of recursive time series analysis. *Int. J. Control.* **29** 1–30.

DEPARTMENT OF STATISTICS, IAS,
AUSTRALIAN NATIONAL UNIVERSITY,
CANBERRA, A.C.T. 2600
AUSTRALIA