The Annals of Statistics
1976, Vol. 4, No. 6, 1258–1270

THE CONVERGENCE OF SOME RECURSIONS

BY E. J. HANNAN

The Australian National University

In connection with a range of stationary time series models, particularly ARMAX models, recursive calculations of the parameter vector seem important. In these the estimate, $\theta(n)$, from observations to time n, is calculated as $\theta(n) = \theta(n-1) + k_n$ where k_n depends only on $\theta(n-1)$, $\theta(n-2), \cdots$ and the data to time n. The convergence of two recursions is proved for the simple model $x(n) = \varepsilon(n) + \alpha\varepsilon(n-1)$, $|\alpha| < 1$, where the $\varepsilon(n)$ are stationary ergodic martingale differences with $E\{\varepsilon(n)^2 \mid \mathscr{F}_{n-1}\} = \sigma^2$. The method of proof consists in reducing the study of the recursion to that of a recursion involving the data only through the $\theta(n)$. It seems that many of the recursions introduced for ARMAX models may be treated in this way and the nature of the extensions of the theory is discussed.

1. Introduction. The purpose of this work is to describe and analyse some recursive estimation methods. These are methods that proceed by adjusting a previous estimate as data for a new time point comes to hand. Such methods are important for a variety of reasons. In the first place they are essential for "real time" calculations, where the updating of the previous estimate must be done in a time period short even by the standards of modern computers. Even if there were no "real time" problem it might yet be necessary to update an estimate as each new data set came in and it might be unnecessarily costly fully to recompute each time. Most importantly these recursive methods are all, to some extent, adaptive in the sense that they continue to track the parameter values when they are changing through time. (See [2] for an example.) The degree of adaption can be varied (at some cost in terms of the rate of convergence to a stable value) but that will not be discussed here for we shall consider only the stable case.

To motivate the recursions that will be considered, some iterative methods are first described. These relate to the, so called, ARMA (autoregressive-moving average) models and their generalisations. These are of the form

$$(1) \qquad \sum_0^q \beta(j)y(n-j) = \sum_0^p \alpha(j)\varepsilon(n-j), \qquad E\{\varepsilon(n)\} = 0, \quad E\{\varepsilon(m)\varepsilon(n)\} = \delta_{mn}\sigma^2.$$

If $y(n)$ is to be stationary and the $\varepsilon(n)$ are to be the linear innovations then ([3], Chapter III) the generating functions, $g(z) = \sum \alpha(j)z^j$, $h(z) = \sum \beta(j)z^j$ must satisfy $g(z) \neq 0$, $|z| < 1$; $h(z) \neq 0$, $z \leqq 1$. The term "ARMAX" has been used for cases where a term involving a vector exogenous variable, here called $z(n)$,

Received September 1975; revised March 1976.

AMS 1970 *subject classifications.* Primary 62M10; Secondary 62L12.

Key words and phrases. Recursive calculation, ARMA models, martingale, stochastic approximation.

is included on the right side of (1). (By exogenous is meant the independence of $z(n)$ from the $\varepsilon(n)$ sequence.) In order to indicate the scope of the recursive methods, and not because they will be considered in detail here, we introduce the vector ARMAX model

$$\sum_0^q B(j)y(n-j) = \sum_1^m \Delta(j)z(n-j) + \sum_0^p A(j)\varepsilon(n-j)\,, \tag{2}$$
$$E\{\varepsilon(m)\varepsilon(n)\} = \delta_{mn}\Omega\,, \qquad E\{\varepsilon(n)\} = 0\,.$$

In order that the $A(j)$, $B(j)$, $\Delta(j)$ and Ω should be uniquely determined (i.e., identified), $y(n)$ should be stationary and $\varepsilon(n)$ the linear innovations, then further conditions are needed but these will not be discussed here (see [4]). However (see [5] and references therein), the model (2) is fully equivalent to the, so called, stable, "state space" system, namely one wherein an unobserved vector $x(n)$ is related to the observed $y(n)$, $z(n)$ by

$$\begin{aligned} &x(n+1) = Fx(n) + Gz(n) + u(n)\,, \qquad E\{u(m)u(n)'\} = \delta_{mn}P \\ &y(n) = Hx(n) + v(n)\,, \qquad E\{v(m)v(n)'\} = \delta_{mn}Q \\ &E\{u(m)v(n)'\} = \delta_{mn}R\,. \end{aligned} \tag{3}$$

The recursions considered herein ultimately extend to the recursive estimation of such systems as (2) or (3) though they seem to have been used, in practice, only in low dimensional forms of such systems. (See [10], [11].)

To explain the construction of the recursions we consider first the *iterative* estimation of systems such as (2) or (3) from a given block of data, $y(n)$, $z(n)$, $n = 1, \cdots, N$. This is spoken of as an "en bloc" or "off line" calculation. For examples, in relation to (1), see [1]. Typically such estimates are obtained from an approximate form of the likelihood function (constructed on Gaussian assumptions) but they may also be obtained in other ways, for example via moment estimators (see [3, Chapter VI]). In any case they lead to an iterative procedure for the estimate $\hat{\theta}_N$ of the parameter vector θ associated with the equations of estimation (e.g., likelihood equations) $\hat{\theta}_N = h_N(\hat{\theta}_N)$. The iteration is of the form

$$\hat{\theta}_N{}^{(j)} = h_N(\hat{\theta}_N{}^{(j-1)})\,. \tag{4}$$

Typically, in the cases to be studied below, this is effectively a regression procedure wherein, at the jth iteration, the vectors to be regressed upon, or the weights in the regression, are functions of $\hat{\theta}_N{}^{(j-1)}$. Of course the same set of likelihood equations (i.e., equations having the same solution) can be expressed in different forms (4) and result in different iterations. We illustrate by a simple case that will also be studied, in relation to recursions, below. Let $y(n)$ satisfy (1) for $q = 0$, $p = 1$. Put

$$\begin{aligned} &\varepsilon_a(n) = y(n) - a\varepsilon_a(n-1)\,, \qquad \varepsilon_a(0) = 0 \\ &\xi_a(n) = \varepsilon_a(n) - a\xi_a(n-1)\,, \qquad \xi_a(0) = 0\,. \\ &b_N(a) = -\textstyle\sum_1^{N-1} \xi_a(n+1)\xi_a(n)/\sum_1^{N-1} \xi_a(n)^2\,, \\ &c_N(a) = \textstyle\sum_1^{N-1} y(n+1)\xi_a(n)/\{\sum_1^{N-1} \xi_a(n)\varepsilon_a(n)\} \\ &d_N(a) = \textstyle\sum_1^{N-1} y(n+1)\varepsilon_a(n)/\{\sum_1^{N-1} \varepsilon_a(n)^2\}\,. \end{aligned} \tag{5}$$

Then the insertion of any of b_N, c_N or d_N in (4) with $\theta = \alpha$, leads to an iteration. The first two of these equations, $a = b_N(a)$, $a = c_N(a)$, are the same equation because $\xi_a(n+1) + a\xi_a(n) = y(n+1) - a\varepsilon_a(n)$. They are both forms of the likelihood equation. The third, $a = d_N(a)$, is not. However, the three have different iterative properties as can be seen from the fact that, using a prime to indicate differentiation with respect to a, and assuming $y(n)$ ergodic,

$$\lim_{N\to\infty} b_N'(\alpha) = 2\,, \qquad \lim_{N\to\infty} c_N'(\alpha) = -\alpha^2/(1-\alpha^2)\,,$$
$$\lim_{N\to\infty} d_N'(\alpha) = 0\,, \quad \text{a.s.}$$

In general, when $h_N(\theta)$ is differentiable,

$$\{\hat{\theta}_N^{(j)} - \theta\} = \{\partial h_N(\bar{\theta}_N^{(j-1)})/\partial\theta\}\{\hat{\theta}_N^{(j-1)} - \theta\}\,,$$

where the Jacobian, $\partial h_N/\partial\theta$ is evaluated at $\bar{\theta}_N^{(j-1)}$, intermediate between θ and $\hat{\theta}_N^{(j-1)}$. It is apparent that $\hat{\theta}_N^{(j)}$ can hardly be expected to iterate to $\hat{\theta}_N$ unless the Jacobian has a suitable norm at the true value θ, since for N large $\hat{\theta}_N$ will be near to θ. Thus the iteration based on $b_N(a)$ cannot be expected to converge while that based on $c_N(a)$ may converge only for $\alpha^2/(1-\alpha^2) < 1$ i.e., $|\alpha| < 1/2^{\frac{1}{2}}$. If $\theta = b_N(\theta)$ is replaced by $\theta = 2\theta - b_N(\theta)$ convergence can be expected to take place. Some of these observations have been borne out by experience with simulations (see [3, Chapter VI], for example).

Recursions are now discussed. These are also spoken of as "on line" calculations. All of those considered here arise from (2) when $y(n)$, $z(n)$ are scalar though it is not really difficult to see how to generalise them. Most of these recursions appear to be recursive equivalents of (4). In general these recursions are of the following form, using $\theta(n)$ for the estimate from data to time n,

$$\theta(n+1) = \theta(n) + P(n+1)\zeta(n+1)\hat{\varepsilon}(n+1) \tag{6.i}$$

$$P(n+1) = P(n) - \{1 + \phi(n+1)P(n)\zeta(n+1)\}^{-1} \times P(n)\zeta(n+1)\phi(n+1)'P(n) \tag{6.ii}$$

$$\hat{\varepsilon}(n+1) = y(n+1) - \phi(n+1)'\theta(n)\,. \tag{6.iii}$$

Here $\phi(n)$, $\zeta(n)$ are vectors that depend on the data to time n, partly through the $\theta(j)$, $j < n$. The vector $\phi(n)$ is to be identified with a regressor vector and $\zeta(n)$ with a vector of instrumental variables. Of course we may have $\phi(n) = \zeta(n)$. $P(n)^{-1}$ is essentially the matrix of sums of squares and cross products of the regressor vectors (or the corresponding matrix in an instrumental variables regression.) For some further background the reader may consult [10], for example. A detailed description of a wide range of such recursions will be found in [9], together with simulations and some analysis which accords fairly closely with that to be given here. Earlier examples of what might be regarded as recursions are provided by the techniques that occur in connection with stochastic approximation. Here we illustrate by some simple cases, studied further below, and shall further illustrate by slightly more elaborate examples in Section 3. It may be mentioned that there are examples where (6.iii) does not obtain but instead $y(n+1)$ has to be replaced by some vector, $\eta(n)$ say, that also depends on $\theta(j), j < n$ (i.e., the "dependent" variable in the regression is "manufactured" from past values of $\theta(j)$ as well as the regressor or "independent" variables). An example will be found in [2], [6]. Two first cases considered are examples of procedures called AML in [10] or RML in [9]. (In fact the first, which corresponds to $d_N(a)$ above, is not approximate or recursive maximum likelihood in any sense that we can perceive.) Consider

$$(4)' \qquad \begin{aligned} \hat{\varepsilon}(n) &= y(n) - \alpha(n-1)\hat{\varepsilon}(n-1)\,, \qquad \hat{\varepsilon}(0) = 0 \\ \xi(n) &= \hat{\varepsilon}(n) - \alpha(n-1)\xi(n-1)\,, \qquad \xi(0) = 0\,. \end{aligned}$$

If (6) is used with $\phi(n) = \zeta(n) = \hat{\varepsilon}(n-1)$, $\theta(n) = \alpha(n)$, with $\alpha(1) = 0$, then it is easily seen that

$$\alpha(n) = \textstyle\sum_3^n y(j)\hat{\varepsilon}(j-1)/\{\sum_2^{n-1} \hat{\varepsilon}^2(j)\}$$

which shows the connection with d_N. We shall call this RML_1. If (6) is used with $\theta(n) = \alpha(n)$, $\phi(n) = \hat{\varepsilon}(n-1)$, and $\zeta(n) = \xi(n-1)$ then again it is not hard to show that, given $\alpha(1) = 0$,

$$(7) \qquad \alpha(n) = \textstyle\sum_3^n y(j)\xi(j-1)/\{\sum_2^{n-1} \xi(j)\hat{\varepsilon}(j)\}$$

which is connected with $c_N(a)$. We speak of this recursion as RML_2, again following [9].

In the next section a rather complete and general proof of the convergence of RML_1, RML_2 is established for this simple case. There appears to be no comparable treatment of any case in the literature. In Section 3 the technique is extended to other cases and is discussed.

2. Two simple recursions. Any case of (1) for which $p = 0$ is simply reduced to a recursion by estimating the $\beta(j)$ via a regression of $y(n)$ on $y(n-j)$, $j = 1, \cdots, q$, $n = q+1, q+2, \cdots, N$. This regression is computed recursively by a standard procedure (see [9] for example). In this case the off line and on line procedures give the same result and the question of convergence of the recursion is not a new question. The simplest nonelementary case is therefore $q = 0$, $p = 1$. That is considered here, the recursions being RML_1, RML_2. It is required that the $\varepsilon(n)$ sequence be stationary and ergodic. Call $\mathscr{F}_n$ the Borel field determined by $y(n)$, $m \leqq n$ or, equivalently, $\varepsilon(m)$, $m \leqq n$. It is required that

$$(8) \qquad |\alpha| < 1\,, \qquad E\{\varepsilon(n) \mid \mathscr{F}_{n-1}\} = 0\,, \qquad E\{\varepsilon(n)^2 \mid \mathscr{F}_{n-1}\} = \sigma^2.$$

The requirement in $g(z)$, mentioned below (1), here amounts to $|\alpha| \leqq 1$ but the stricter condition in (8) now seems necessary. The ergodicity of $\varepsilon(n)$ is a costless assumption if (part of) only one realisation is available. The second condition in (8) amounts to saying that the best linear predictor for $y(n)$ is the best predictor (both best in the least squares sense) which seems natural in the context of linear models such as (1), (2). The last condition in (8) seems minimal, in connection with en-bloc calculations, in order to ensure that limiting distributions are of a reasonable kind. It could be replaced by a higher moment condition but it does not seem easy to replace it by an unequivocally weaker condition.

THEOREM 1. *Under the above conditions, for* (1) *with* $p = 1$, $q = 0$, $\alpha(n)$ *given by* RML_1 *converges* a.s. *to* α.

It will be convenient to take $\sigma^2 = 1$. Since

$$\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 = \frac{1}{n}\sum_1^n y(j)^2 + \frac{1}{n}\sum_1^n \alpha(j-1)^2\hat{\varepsilon}(j-1)^2 - \frac{2}{n}\sum_1^n y(j)\alpha(j-1)\hat{\varepsilon}(j-1) \tag{9}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_1^n y(j)^2 = 1+\alpha^2 \quad \text{a.s.}$$

then, using Schwartz inequality in the last term on the right in (9),

$$\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 \geqq \frac{1}{n}\sum_1^n \alpha(j-1)^2\hat{\varepsilon}(j-1)^2\{1-l_n\}^2 + o(1),$$

$$l_n = [(1+\alpha^2)/\{n^{-1}\sum \alpha(j-1)^2\hat{\varepsilon}(j-1)^2\}]^{\frac{1}{2}}.$$

Here and below the term that is $o(1)$ converges a.s. to zero. Thus if S is the set of integers whereon (say) $l_n \leqq \frac{1}{2}$ then the limit inferior of $\sum \hat{\varepsilon}(j)^2/\sum \alpha(j-1)^2\hat{\varepsilon}(j-1)^2$, on S, is not less than $\frac{1}{4}$. On the other hand

$$\begin{aligned}\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 &= 1 + \frac{1}{n}\sum_1^n\{\alpha\varepsilon(j-1) - \alpha(j-1)\hat{\varepsilon}(j-1)\}^2 \\ &\quad - 2\{\textstyle\sum_1^n \varepsilon(j)\alpha(j-1)\hat{\varepsilon}(j-1)/\sum_1^n \alpha(j-1)^2\hat{\varepsilon}(j-1)^2\} \\ &\quad \times \frac{1}{n}\sum \alpha(j-1)^2\hat{\varepsilon}(j-1)^2 + o(1).\end{aligned} \tag{10}$$

The bracketed factor in the third term on the right converges to zero when $\sum \alpha(j-1)^2\hat{\varepsilon}(j-1)^2$ diverges and otherwise converges ([8, page 150, corollary]. This reference will be repeatedly used and will be referred to as [8].) In the latter case evidently $\liminf \sum \varepsilon^2(j)/\sum\{\alpha(j-1)^2\hat{\varepsilon}(j-1)^2\} > 0$ while in the former the ratio under this limit sign is evidently, on the complement of S, bounded below by $\{4(1+\alpha^2)\}^{-1} + o(1)$. Hence

$$\liminf \sum_1^n \hat{\varepsilon}(j)^2/\sum_1^n \alpha(j-1)^2\hat{\varepsilon}(j-1)^2 > 0 \quad \text{a.s.}$$

Now from (10), replacing $\sum \alpha(j-1)^2\hat{\varepsilon}(j-1)^2$ by $\sum \hat{\varepsilon}(j)^2$ in both places in the third term, it follows that

$$\liminf \frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 \geqq 1 \quad \text{a.s.} \tag{11}$$

Next

$$\alpha(n+1) = \sum_1^n\{\varepsilon(j+1)\hat{\varepsilon}(j) - \alpha\varepsilon(j)\alpha(j-1)\hat{\varepsilon}(j-1) + \alpha\varepsilon(j)y(j)\}/\sum_1^n \hat{\varepsilon}(j)^2$$

and, by [8], the contribution from the first two terms in the numerator converges to zero so that

$$\lim_{n\to\infty}\left\{\alpha(n+1)-\frac{\alpha}{n^{-1}\sum_1^n \hat{\varepsilon}(j)^2}\right\}=0 \quad \text{a.s.} \tag{12}$$

Thus $\limsup|\alpha(n)| \leqq |\alpha|$ (see (11)). For $\alpha = 0$ this already proves the theorem so that henceforth we take $\alpha \neq 0$. From (9)

$$\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 \leqq 1+\alpha^2+\alpha^2\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 + 2|\alpha|(1+\alpha^2)^{\frac{1}{2}}\left\{\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2\right\} + o(1)\,,$$

which shows that the left side is uniformly bounded, a.s. Similarly

$$n^{-1}\hat{\varepsilon}(n)^2 \leqq n^{-1}\alpha^2\hat{\varepsilon}(n-1)^2 + n^{-1}y(n)^2 + 2|\alpha|\,|n^{-1}y(n)|\,|\hat{\varepsilon}(n-1)|$$

and since $y(n)^2/n \to 0$, a.s. then taking the limit superior on both sides, it is seen that $\hat{\varepsilon}(n)^2/n \to 0$ a.s.

Since, from (6),

$$\alpha(n+1)=\alpha(n)+\hat{\varepsilon}(n+1)\hat{\varepsilon}(n)/\textstyle\sum_1^n \hat{\varepsilon}(j)^2\,, \qquad \alpha(1)=0\,,$$

it follows from (11) that $\alpha(n+1)-\alpha(n)\to 0$ a.s. We shall repeatedly use this last result.

Put $\varepsilon^{(1)}(j)=\varepsilon(j)$ for $|\varepsilon(j)|<A$ and zero otherwise and $\varepsilon^{(2)}(j)=\varepsilon(j)-\varepsilon^{(1)}(j)$. Now

$$\frac{1}{n}\sum_1^n \alpha(j)\{\varepsilon(j)^2-1\} = \frac{1}{n}\sum_1^n[\varepsilon^{(1)}(j)^2-E\{\varepsilon^{(1)}(j)^2\,|\,\mathscr{F}_{j-1}\}]\alpha(j-1)$$
$$+\frac{1}{n}\sum_1^n[\varepsilon^{(2)}(j)^2-E\{\varepsilon^{(2)}(j)^2\,|\,\mathscr{F}_{j-1}\}]\alpha(j)+o(1)\,.$$

The first term on the right converges a.s. to zero by [8] while the second is dominated by

$$\alpha\left[\frac{1}{n}\sum_1^n \varepsilon^{(2)}(j)^2+\frac{1}{n}\sum_1^n E\{\varepsilon^{(2)}(j)^2\,|\,\mathscr{F}_{j-1}\}\right]$$

which converges to $2\alpha E\{\varepsilon^{(2)}(j)^2\}$, by the ergodic theorem, and this may be made as small as is desired by taking A large. Thus

$$\frac{1}{n}\sum_1^n \alpha(j)\varepsilon(j)^2=\frac{1}{n}\sum_1^n \alpha(j)+o(1)\,. \tag{13}$$

New from (8)

$$\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 = 1+\alpha^2+\frac{1}{n}\sum_1^n \alpha(j)^2\hat{\varepsilon}(j)^2-2\alpha\frac{1}{n}\sum_1^n \varepsilon(j)\alpha(j)\hat{\varepsilon}(j)+o(1)$$
$$=1+\alpha^2+\frac{1}{n}\sum_1^n \alpha(j)^2\hat{\varepsilon}(j)^2-2\alpha\frac{1}{n}\sum_1^n \varepsilon(j)\hat{\varepsilon}(j)\alpha(j-1)+o(1)$$
$$=1+\alpha^2+\frac{1}{n}\sum_n^1 \alpha(j)^2\hat{\varepsilon}(j)^2-2\alpha\frac{1}{n}\sum_1^n \varepsilon(j)y(j)\alpha(j-1)$$
$$-2\alpha\frac{1}{n}\sum_1^n \varepsilon(j)\alpha(j-1)^2\hat{\varepsilon}(j-1)+o(1)$$
$$=\frac{1}{n}\sum_1^n\{1+\alpha^2-2\alpha\alpha(j)\}+\frac{1}{n}\sum_1^n \alpha(j)^2\hat{\varepsilon}(j)^2+o(1)$$

using [8], again, and (13). Now repeating the process with $n^{-1}\sum \alpha(j)^2\hat{\varepsilon}(j)^2$ then, after k steps, we reach

$$\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 = \frac{1}{n}\sum_1^n [\{1+\alpha^2-2\alpha\alpha(j)\}\{1+\alpha(j)^2+\alpha(j)^4+\cdots+\alpha(j)^{2k-2}\}] + \frac{1}{n}\sum_1^n \alpha(j)^{2k}\hat{\varepsilon}(j)^2 + o(1).$$

Thus, because of $|\alpha(j)| \leqq \alpha < 1$ and of the boundedness of $n^{-1}\sum \hat{\varepsilon}(j)^2$,

$$\frac{1}{n}\sum_1^n \hat{\varepsilon}(j)^2 = \frac{1}{n}\sum_1^n \left\{\frac{1+\alpha^2-2\alpha\alpha(j)}{1-\alpha(j)^2}\right\} + o(1).$$

Thus

$$\alpha(n+1) = \alpha\left[\frac{1}{n}\sum_1^n \left\{\frac{1+\alpha^2-2\alpha\alpha(j)}{1-\alpha(j)^2}\right\}\right]^{-1} + o(1). \tag{14}$$

The proof may now be completed in various ways. The most informative is as follows. Call $\hat{\alpha}(n+1)$ the first term on the right in (14) and $K(n+1)$ its denominator. Then

$$\{\hat{\alpha}(n+1)-\alpha\} = (\hat{\alpha}(n)-\alpha)\left\{1-\frac{1}{n}K(n+1)^{-1}\frac{1-\alpha\hat{\alpha}(n)}{1-\hat{\alpha}(n)^2}\right\} + \{\alpha(n)-\hat{\alpha}(n)\}O(n^{-1}). \tag{15}$$

Since $K(n+1)^{-1}\{1-\alpha\hat{\alpha}(n)\}/\{1-\hat{\alpha}(n)^2\} \geqq c > 0$ and the last term is $o(n^{-1})$ (because $\{\alpha(n)-\hat{\alpha}(n)\} = o(1)$) it follows that $\{\hat{\alpha}(n)-\alpha\}$, and hence $\{\alpha(n)-\alpha\}$, converges to zero. Indeed, calling the last term in (15) $l_n/(n-1)$, then

$$\{\hat{\alpha}(n+1)-\alpha\} \leqq \sum_{j=0}^{n-1}\frac{|l_{n-j}|}{n-j}\prod_{k=0}^{j}\left(1-\frac{c}{n-k}\right),$$

with the coefficient of $|l_n|$ being unity. Since

$$\sum_1^n \frac{c}{j}\prod_{k=j}^{n}\left(1-\frac{c}{k}\right) \to 1, \qquad \prod_1^n\left(1-\frac{c}{k}\right) = O(n^{-c}).$$

The required result follows and the theorem is proved.

Consider next RML_2 (see (7)) for the same ARMA model, $p=1$, $q=0$. It seems probable that the method used in Theorem 1 will fully generalies but in this and more elaborate cases it has been found difficult to complete the proof unless $\theta(n)$ is bounded to lie in some "acceptable region." The following modified definition of $\alpha(n)$ is therefore used. Call $\tilde{\alpha}(n)$ the expression given by the right side of (7). Require

$$|\alpha| < 1-\delta, \qquad \delta > 0 \tag{16}$$

and put $\alpha(n) = \tilde{\alpha}(n)$ if $\tilde{\alpha}(n)$ lies in the acceptable region, $|\tilde{\alpha}(n)| < 1-\delta$, and otherwise put $\alpha(n)$ equal to the boundary value nearer to $\tilde{\alpha}(n)$. Of course $\alpha(n)$, not $\tilde{\alpha}(n)$, is used in (4)′. In terms of (6) it is the $\alpha(j)$, *not* $\tilde{\alpha}(j)$, that are used throughout

(6.ii), (6.iii)′ (e.g., via (4)′ and in the definition of $\zeta(n)$) but (6.i) now reads $\tilde{\theta}(n+1) = \tilde{\theta}(n) - P(n+1)\zeta(n+1)\hat{\varepsilon}(n+1)$.

Because of the cost of checking that $\tilde{\theta}(n)$ lies in the acceptable regions, in more general cases, it is not suggested that the restriction to this region will be used in practice. Indeed the following theorem shows that, in the present case, eventually this will not be necessary.

THEOREM 2. *If $y(n)$ satisfies the conditions of this section and also* (16) *then the modified* RML_2 *estimator $\alpha(n)$, just defined, converges* a.s. *to α.*

In almost the same way as for the latter part of the proof of Theorem 1 it may now be shown that

$$(17)\qquad \tilde{\alpha}(n+1) = \alpha\left[\frac{1}{n}\sum_1^n \frac{1+\alpha^2+\alpha\alpha(j)^3 - 3\alpha\alpha(j)}{\{1-\alpha(j)^2\}^2}\right]^{-1} + o(1)\,.$$

The proof of this will be omitted since it follows much the same lines as for Theorem 1 save that the calculations down to formula (11) are now rather trivial because of (16). In any case in the next section, in a more general context, it will be shown how the first term on the right side of (17) is formed. Now calling $\hat{\alpha}(n)$ the first term on that right side and $K(n+1)$ its denominator

$$(18)\qquad \{\hat{\alpha}(n+1)-\alpha\} = \{\hat{\alpha}(n)-\alpha\}\left\{1 - \frac{1}{n}K(n+1)^{-1}\frac{1-\alpha\alpha(n)}{1-\alpha(n)^2}\right\} + \{\alpha(n)-\hat{\alpha}(n)\}O(n^{-1})\,.$$

Now as before the second factor in the first term on the right is not greater than $(1 - c/n)$, $c > 0$. However, the proof is not so easily completed because it is $\tilde{\alpha}(n) - \hat{\alpha}(n)$ that is $o(1)$ and not $\{\hat{\alpha}(n) - \alpha(n)\} = \{\tilde{\alpha}(n) - \alpha(n)\} + o(1)$. However, returning to (17), it is evident that $\tilde{\alpha}(n)$ cannot stay indefinitely outside of the acceptable region for if that were so $\alpha(n)$ would stay indefinitely at one of the boundary points, say $1-\delta$. But then, from (17)

$$\tilde{\alpha}(n+1) = \alpha\left[\frac{1+\alpha^2+\alpha(1-\delta)^2 - 3\alpha(1-\delta)}{\{1-(1-\delta)^2\}^2}\right]^{-1} + o(1)\,.$$

However, it is easily seen that the first term on the right side is less than $(1-\delta)$. Thus $\tilde{\alpha}(n)$ returns indefinitely often to the acceptable region. When this is so, the last term in (18) is $o(n^{-1})$ and the proof may be concluded as for Theorem 1.

Following the same kind of argument it is easy to show that for $\alpha(n)$ corresponding to $b_N(a)$ (see below (5)) convergence will not take place. Incidentally while it seems that the iteration (4) will not converge for all α, for $h_N(a) = c_N(a)$, yet the corresponding recursion does converge for $|\alpha| < 1-\delta$, $\delta > 0$.

3. A more general discussion. To illustrate the general case consider (1) for $p = q = 1$. Thus

$$y(n) + \beta y(n-1) = \varepsilon(n) + \alpha\varepsilon(n-1)\,.$$

In addition to the requirement on α, β it is also necessary that $\alpha \neq \beta$, since otherwise the model is not distinguished from $p = q = 0$. Once more $\varepsilon(n)$ is required to satisfy (8) and to be ergodic. Now we define the acceptable region by

$$|\alpha|, |\beta| < 1 - \delta, \qquad |\alpha - \beta| > \delta, \qquad \delta > 0. \tag{19}$$

We consider RML_1 so that $\theta' = (\alpha, \beta)$, $\phi(n)' = \zeta(n)' = (\hat{\varepsilon}(n-1), -y(n-1))$ which completes the definition of the recursion via (6). Of course initial values must be chosen for $P(1)$, $\theta(1)$, the former reflecting the confidence felt in $\theta(1)$ in the same way as the inverse of the matrix of sums of squares and cross products in a regression does for the estimated vector of regression coefficients. Again, however we modify [6] by defining $\tilde{\theta}(n)$ to be the output of (6.i) but putting $\theta(n) = \tilde{\theta}(n)$, only if $\tilde{\theta}(n)$ lies in the acceptable region defined by (19) while otherwise it is kept at the boundary value nearest to the last value of $\tilde{\theta}(n)$ before exit from the region. Again $\theta(n)$ is used throughout (6.ii), (6.iii) and in the second term in (6.i).

Again it will be convenient to take $\sigma^2 = 1$. It is first proved that $n^{-1}P(n)^{-1}$, which we call $\hat{K}(n)$ for short, has its smaller eigenvalue bounded away from zero, a.s. To do this put $y(j) = u(j) + \varepsilon(j) = \hat{u}(j) + \hat{\varepsilon}(j)$, $u(j) = -\beta y(j-1) + \alpha\varepsilon(j-1)$, $\hat{u}(j) = -\beta(j-1)y(j-1) + \alpha(j-1)\hat{\varepsilon}(j-1)$. Clearly $u(j)$, $\hat{u}(j)$ are measurable $\mathscr{F}_{j-1}$. Using [8] it now follows that

$$\hat{K}(n+1) = gg' + L(n) + O(1),$$

$$L(n) = \frac{1}{n}\sum v(j)v(j)', \quad v(j)' = (\hat{u}(j) - u(j), u(j)),$$

where g is the vector composed of 1 and -1. Let μ_n be the smaller eigenvalue of $\hat{K}(n+1)$ and w_n the corresponding eigenvector, $w_n'w_n = 1$, and put $w_n = \xi_n g + \eta_n f$, $f'g = 0$, $f'f = 1$. Along a subsequence whereon $\mu_n \to 0$ then $\xi_n \to 0$, $|\eta_n| \to 1$ since gg' and $L(n)$ are semidefinite. Hence along such a subsequence $f'L(n)f \to 0$. But $2f'L(n)f = n^{-1}\sum \hat{u}(j)^2$ and

$$\frac{1}{n}\sum_1^n \hat{u}(j)^2 = \frac{1}{n}\sum\{\alpha(j-1) - \beta(j-1)\}^2 y(j-1)^2 + \frac{1}{n}\sum \alpha(j-1)^2\hat{u}(j-1)^2$$
$$- \frac{2}{n}\sum\{\alpha(j-1) - \beta(j-1)\}\alpha(j-1)y(j-1)\hat{u}(j-1).$$

Using Schwartz's inequality in the last term, and the fact that $|\alpha(j)| < 1$, it follows that $\liminf n^{-1}\sum \hat{u}(j)^2 = 0$ implies $\liminf n^{-1}\sum\{\alpha(j-1) - \beta(j-1)\}^2 y(j-1)^2 = 0$. This is a contradiction since $|\alpha(j) - \beta(j)| > \delta > 0$.

It now follows that $\tilde{\theta}(j) - \tilde{\theta}(j-1) = o(1)$, using (6.1) and the result just established. However it does *not* now follow that $\theta(j) - \theta(j-1) = o(1)$ though failure of this can occur only at points of reentry of $\tilde{\theta}(j)$ into the acceptable region. It is apparent also that, ultimately, $\theta(j) - \theta(j-1)$ can fail to be $o(1)$ only at such points of reentry which have been preceded by long excursions from the acceptable region. Thus the points whereat $\|\theta(j) - \theta(j-1)\| \geqq \eta > 0$

will eventually become sparse in the sense that the proportion of them in the first n points will, a.s., decrease to zero, for any $\eta > 0$.

It remains to derive the analogue of (18). It is easily seen that

$$\tilde{\theta}(n) = P(n)^{-1} \sum_1^n y(j)\phi(j) .$$

Consider $n^{-1} \sum y(j)\hat{\varepsilon}(j-1)$ for example. Now

$$\begin{aligned} \hat{\varepsilon}(j) &= y(j) - \{\alpha(j-1) - \beta(j-1)\}y(j-1) \\ &\quad + \cdots (-)^m \alpha(j-1)\alpha(j-1) \cdots \\ &\quad \alpha(j-m+1)\{\alpha(j-m) - \beta(j-m)\}(j-m) \\ &\quad - (-)^m \alpha(j-1) \cdots \alpha(j-m)\{\hat{\varepsilon}(j-m) - y(j-m)\} . \end{aligned} \tag{20}$$

$$n^{-1} \sum y(j)\hat{\varepsilon}(j-1) = \sum_{k=0}^{\infty} c(k) n^{-1} \sum_j \varepsilon(j-k)\hat{\varepsilon}(j-1) \tag{21}$$

where $\sum_0^\infty c(k)z^k = (1 + \alpha z)/(1 + \beta z)$. Using Schwartz's inequality, the ergodic theorem, the a.s. boundedness of $n^{-1} \sum \hat{\varepsilon}(j)^2$ and the geometric rate of convergence of the $c(k)$ to zero, it is evident that (21) may be evaluated term by term. Because $|\alpha(j)| < 1 - \delta$ it follows in the same way that $\hat{\varepsilon}(j-1)$ may be replaced by the first m terms in its expression, (20) with a.s., an arbitrarily small error. In turn an expression such as

$$\frac{1}{n} \sum_j \varepsilon(j-k)\{\alpha(j-1)\alpha(j-1) \cdots \alpha(j-m+1)\} \times \{\alpha(j-m) - \beta(j-m)\}y(j-m)$$

may be evaluated term by term replacing $y(j-m)$ by $\sum c(k)\varepsilon(j-m-k)$. This reduces us to expressions such as

$$\frac{1}{n} \sum \varepsilon(j-k)\varepsilon(j-l)\phi(j-1) , \tag{22}$$

$$\phi(j-1) = \{\alpha(j-1) \cdots \alpha(j-m+1)\}\{\alpha(j-m) - \beta(j-m)\} .$$

For $k > l$, $n^{-1} \sum \varepsilon(j-k)\varepsilon(j-l)\phi(j-k-1)$ converges a.s. to zero, by [8]. Thus for $k > l$ the expression (22) will converge to zero if

$$\lim_{n\to\infty} n^{-1} \sum |\phi(j-1) - \phi(j-k-1)| = 0 \quad \text{a.s.} \tag{23}$$

However $|\phi(j-1) - \phi(j-k-1)| \to 0$ on a set of density 1 and hence (23) holds ([12], Vol. II, page 181). If $k = l$ the proof that (22) converges to $n^{-1} \sum \phi(j-1)$ is accomplished via the type of argument used above (13).

This allows $\tilde{\theta}(n)$ to be evaluated, to $o(1)$, as follows.

$$\tilde{\theta}(n) = \hat{\theta}(n) + o(1)$$

$$\hat{\theta}(n) = K(n)^{-1} \frac{1}{n} \sum_1^n \tilde{E}\{y(j)\tilde{\phi}(j)\} , \quad K(n) = \frac{1}{n} \sum_1^n \tilde{E}\{\tilde{\phi}(j)\tilde{\phi}(j)'\} .$$

Here $\tilde{\phi}(j)$ is $\phi(j)$ but with all $\theta(k)$, $k < j$, used in its construction replaced by $\theta(j-1)$ while $\tilde{E}$ indicates that the expectation is taken with $\theta(j-1)$ treated as if it weve a constant. Of course $\tilde{E}\{y(j)\tilde{\phi}(j)\}$, $\tilde{E}\{\tilde{\phi}(j)\tilde{\phi}(j)'\}$ may be written down in spectral terms since $\phi(j)$ is merely composed of $y(j-1)$ and $\hat{\varepsilon}(j-1)$ and, in $\tilde{\phi}(j)$, $\hat{\varepsilon}(j-1)$ is replaced by the output $\tilde{\varepsilon}(j-1)$ of the filter with response (z-transform) $\{1+\beta(j)z\}/\{1+\alpha(j)z\}$. Thus

$$\hat{\theta}(n+1) = [\textstyle\sum_1^{n+1} \int_{-\pi}^{\pi} f(\omega) H_j(e^{i\omega}) H_j^*(e^{i\omega})\, d\omega]^{-1} \sum_1^{n+1} \int_{-\pi}^{\pi} f(\omega) H_j(e^{i\omega}) e^{i\omega}\, d\omega\,, \tag{24}$$

where $f(\omega) = (2\pi)^{-1}|1+\alpha \exp i\omega|^2/|1+\beta \exp i\omega|^2$ and

$$H_j(z)' = \left(\frac{1+\beta(j)z}{1+\alpha(j)z}\,,\ -1\right),$$

Now from (24),

$$\{\hat{\theta}(n+1)-\theta\} = \{\hat{\theta}(n)-\theta\} + \frac{1}{n+1} K(n+1)^{-1} \tilde{E}\{\tilde{\phi}(n+1)\tilde{\varepsilon}(n+1)\}\,. \tag{25}$$

However $\tilde{\varepsilon}(n+1) = \varepsilon(n+1) + \phi_0(n+1)'\theta - \tilde{\phi}(n+1)'\theta(n)$ where $\phi_0(n)$ is $\tilde{\phi}(n)$ computed using θ in place of $\theta(n)$. Moreover $\tilde{E}\{\tilde{\phi}(n+1)\varepsilon(n+1)\} = 0$. Thus

$$\begin{aligned}\tilde{E}\{\tilde{\phi}(n+1)\tilde{\varepsilon}(n+1)\} = &-\tilde{E}\{\tilde{\phi}(n+1)\tilde{\phi}(n+1)'\}\{\theta(n)-\theta\}\\ &- \tilde{E}[\tilde{\phi}(n+1)\{\tilde{\phi}(n+1)-\phi_0(n+1)\}']\theta\,.\end{aligned}$$

And since $\{\tilde{\phi}(n+1)-\phi_0(n+1)\}'\theta = -\alpha\{\theta(n)-\theta\}'\tilde{\phi}(n) + \alpha\{\phi_0(n)-\tilde{\phi}(n)\}'\theta$ we obtain by induction

$$\tilde{E}\{\tilde{\phi}(n+1)\tilde{\varepsilon}(n+1)\} = -\textstyle\sum_0^\infty (-\alpha)^j \tilde{E}\{\tilde{\phi}(n+1)\tilde{\phi}'(n+1-j)\}\{\theta(n)-\theta\}\,.$$

Thus

$$\begin{aligned}\{\hat{\theta}(n+1)-\theta\} = &\left[1 - \frac{1}{n} K(n+1)^{-1} \textstyle\sum_0^\infty (-\alpha)^j \tilde{E}\{\tilde{\phi}(n+1)\tilde{\phi}(n+1-j)'\}\right]\\ &\times \{\hat{\theta}(n)-\theta\} + O(n^{-1})\{\hat{\theta}(n)-\theta(n)\} + o(n^{-1})\,.\end{aligned}$$

A further simplification can be achieved by returning to (25). Put $||\hat{\theta}(n)-\theta||_n^2$ for $\{\hat{\theta}(n)-\theta\}'K(n)\{\hat{\theta}(n)-\theta\}$. This converges to zero if and only if $\{\hat{\theta}(n)-\theta\}$ converges to zero because the smallest eigenvalue of $K(n)$ is bounded away from zero, a.s. Thus from (25) we obtain

$$\begin{aligned}||\hat{\theta}(n+1)-\theta||_{n+1}^2 = ||\hat{\theta}(n)-\theta||_n^2 &- \frac{2}{n}\{\hat{\theta}(n)-\theta\}'\tilde{E}\{\tilde{\phi}(n+1)\tilde{\varepsilon}(n+1)\}\\ &+ \frac{1}{n}\{\hat{\theta}(n)-\theta\}\tilde{E}\{\tilde{\phi}(n+1)\tilde{\phi}(n+1)'\}\{\hat{\theta}(n)-\theta\}\\ &+ O(n^{-2})\,.\end{aligned}$$

i.e.,

$$\begin{aligned}||\hat{\theta}(n+1)-\theta||_{n+1}^2 = ||\hat{\theta}(n)-\theta||_n^2 &- \frac{1}{n}\{\hat{\theta}(n)-\theta\}'\\ &\times [2\textstyle\sum_0^\infty (-\alpha)^j \tilde{E}\{\tilde{\phi}(n+1)\tilde{\phi}(n+1-j)'\}\\ &- \tilde{E}\{\tilde{\phi}(n+1)\tilde{\phi}(n+1)'\}]\{\hat{\theta}(n)-\theta\}\\ &+ O(n^{-1})\{\hat{\theta}(n)-\theta(n)\} + o(n^{-1})\,.\end{aligned} \tag{26}$$

It is easily checked that the matrix of the quadratic form in the second term in (26) is

$$\int_{-\pi}^{\pi} f(\omega) H_{n+1}(e^{i\omega}) H_{n+1}^*(e^{i\omega}) \frac{1-\alpha^2}{|1+\alpha e^{i\omega}|^2} \, d\omega \,,$$

and that this is positive definite for all values of $\theta(n)$ in the acceptable region, with smaller eigenvalue bounded away from zero.

It remains only to prove that $\hat{\theta}(n)$ is infinitely often in the acceptable region. If this is not so then $\theta(n)$ stays indefinitely at the value, θ_e say, associated with last exit and, from (24), $\hat{\theta}(n) \to \hat{\theta}$

$$\hat{\theta} = [\int_{-\pi}^{\pi} f(\omega) H_e(e^{i\omega}) H_e^*(e^{i\omega}) \, d\omega]^{-1} \int_{-\pi}^{\pi} f(\omega) H_e(e^{i\omega}) e^{i\omega} \, d\omega$$

where H_e is H_j calculated at θ_e. It seems likely that this value, $\hat{\theta}$, does lie in the acceptable region but we have not been able to prove that, though it is easy to evaluate the expression for θ. Lack of convexity of the acceptable region, especially in the metrics $||\cdot||_n$, remains a problem.

The treatment given above seems rather general. The particular aspects are as follow. (1) The proof that $\hat{K}(n)$ has smallest eigenvalue bounded away from zero. In fact this proof seems to carry over to many cases (see [7]). (2) The proof that the quadratic form (see (26)) is positive definite. This will have to be particular to the case since for some cases it is known to be indefinite or negative definite. (See [9] and the example based on $b_N(a)$ in Section 1.) (3) The proof that $\hat{\theta}$ (see (27)) lies in the acceptable region. This has to be special also. (4) The case where an exogenous variable, $z(n)$, occurs has to be covered. For reasonable conditions on $z(n)$ which enable the analysis to be extended to this case see [7].

REFERENCES

[1] Anderson, T. W. (1975). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. *Ann. Statist.* **3** 1283–1304.

[2] Gertler, J. and Bányász, Cs. (1974). A recursive (on line) maximum likelihood identification method. *IEEE Trans. Automatic Control* **AC-19** 816–819.

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

[4] Hannan, E. J. (1971). The identification problem for multiple equation systems with moving average errors. *Econometrica* **39** 751–766.

[5] Hannan, E. J. (1976). The identification and parameterization of ARMAX and state space systems. *Econometrica* **44**.

[6] Hannan, E. J. (1976). ARMAX and state space systems and recursive calculations. In *Frontiers in Quantitative Economics, III* (M. Intriligator, ed.). North Holland, Amsterdam.

[7] Hannan, E. J. and Hatanaka, K. (1976). ARMAX models and recursive calculations. *Proceedings of Conference on Multivariable System Theory and Control of Dynamic Economic Systems, Nagoya.* To be published.

[8] Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

[9] Söderström, T., Ljung, L. and Gustavsson, I. (1974). A comparative study of recursive

methods. Report 7427 of the Lund Institute of Technology, Department of Auto. Control, Lund, Sweden.

[10] YOUNG, P. C. (1974). Recursive approaches to time series analysis. *Bull Inst. Math. and its Applications.* **10** 209-224.

[11] YOUNG, P. C. (1976). Some observations on instrumental variable methods of time series analysis. *Internat. J. Control* **23** 593-612.

[12] ZYGMUND, A. (1968). *Trigonometric Series.* Cambridge, Mass.

DEPARTMENT OF STATISTICS
INSTITUTE OF ADVANCED STUDIES
AUSTRALIAN NATIONAL UNIVERSITY
BOX 4, P.O. CANBERRA, A.C.T.
AUSTRALIA 2600