The Annals of Statistics
1981, Vol. 9, No. 2, 307–317

THE SECOND ORDER PROPERTIES OF A TIME SERIES RECURSION

By V. Solo

The Australian National University

In this work we prove the asymptotic efficiency of a certain time series recursion for the parameters of an ARMAX time series model.

0. Introduction. In recent years, in the Engineering literature, there has been an enormous interest in real time methods of parameter estimation. These methods, known as recursions, have been seen as a first stage in the design of adaptive control schemes as well as a means of tracking slowly varying parameters. In this work a particular time series recursion known as RML_2 (see Söderström et al., 1974) is investigated and shown to be asymptotically efficient in that it has the same asymptotic covariance matrix as the Gaussian maximum likelihood estimate. While this is certainly an interesting result (and the first of its kind for time series recursions) it is not the final answer since in practice it seems RML_2 must be monitored to ensure its convergence; this can be a costly affair. Thus there is interest too in less efficient recursions that can be operated without monitoring (see, e.g., Solo, 1979, 1978a).

The convergence of RML_2 has been discussed by Ljung (1977) and Hannan (1978). The present discussion is relatively straightforward in conception though the details are tedious. The idea is to introduce a fictitious term (related to the recursion) which clearly obeys a central limit theorem (CLT). The term is simply subtracted from the normed up recursion and the remainder shown to converge to zero (a.s.). The idea of subtracting off a term that obeys the desired limit law has been used also by the author to discuss stochastic approximation schemes (Solo 1978a, 1978b).

The paper is structured as follows: Section 1 is devoted to introducing RML_2. In Section 2 we review the convergence behaviour of RML_2 along the lines of Hannan (1978) and Ljung (1977). Also some propositions that will be useful in the CLT proof are established. Section 3 contains a proof of the CLT. At the end of Section 3 a brief indication of an invariance principle for RML_2 is given.

1. Preliminaries. Consider then, an ARMAX model for a scalar time series with stationary exogenous sequence or input sequence, u_n and output or measured sequence y_n. Following Ljung (1976), the model is defined by a formula for its innovations or prediction error sequence as

$$e_n(\theta) = (1 + \nu(L))^{-1}((1 + \alpha(L))y_n - \gamma(L)u_n) \tag{1.1}$$

where

$$\nu(L) = \textstyle\sum_1^{n_\nu} \nu_i L^i$$

and so on, and L is the lag or backwards operator; also $\theta = (\alpha', \nu', \gamma')'$ with $\alpha = (\alpha_1 \cdots \alpha_{n_\alpha})'$ and so on, while $\theta \in R$ a compact subset of the open set

$$S = \{\theta \mid 1 + \alpha(L), 1 + \nu(L) \quad \text{have all zeroes strictly outside the unit circle in the complex L plane}\}.$$

Received December, 1978; revised December, 1979.

AMS 1970 *subject classification.* Primary 62M10; secondary 62L12.

Key words and phrases. Recursion, time series, efficiency, central limit theorem, stochastic approximation.

It is assumed there is a true value $\theta_0 \in R \ni e_n(\theta_0)$, denoted ϵ_n, is a martingale difference sequence w.r.t. $\mathscr{F}_n$ the increasing sequence of σ-algebras, generated by ϵ_n: ϵ_n is also then called a nonlinear innovations process so that the best least squares predictor of y_n from its own past is a linear predictor. This is a minimal assumption if the linear model is to make sense: cf. Hannan and Heyde (1972). Denote $e_n(\theta_0)$ by ϵ_n and assume that

$$E(\epsilon_n^2 \mid F_{n-1}) = \sigma^2. \tag{1.2}$$

Assume also $\exists$ a small $\delta > 0$ with

$$\sup_n E(|\epsilon_n|^{4/(1-3\delta)}) < K < \infty, \tag{1.3a}$$

$$\sup_n E(|u_n|^{4/(1-3\delta)}) < K < \infty, \tag{1.3b}$$

where u_n is taken to be a stationary ergodic stochastic process.

Now recursions can be constructed by sequential minimisation of an approximate likelihood such as

$$N^{-1} \textstyle\sum_1^N e_n^2(\theta).$$

For the simplest time series model, namely a linear regression where

$$e_n(\boldsymbol{\theta}) = y_n - \theta' \mathbf{x}_n$$

for some vector of regressors $\mathbf{x}_n$ the exact solution to the minimisation problem is given by Plackett's algorithm (Plackett, 1950)

$$\begin{aligned} \boldsymbol{\theta}_n &= \boldsymbol{\theta}_{n-1} + \mathbf{P}_n \mathbf{x}_n e_n \\ \mathbf{P}_n^{-1} &= \mathbf{P}_{n-1}^{-1} + \mathbf{x}_n \mathbf{x}_n' \\ e_n &= y_n - \mathbf{x}_n' \boldsymbol{\theta}_{n-1}. \end{aligned}$$

Notice for the regression that $-de_n/d\theta = \mathbf{x}_n$. For the RML_2 recursion an approximate solution to the minimization is generated by utilising the gradient vector

$$\begin{aligned} \tilde{\varphi}_n(\theta) &= -de_n(\theta)/d\theta \\ &= -(1 + \nu(L))^{-1} \varphi_n(\theta), \end{aligned}$$

where

$$\varphi_n(\theta) = (-y_{n-1} \cdots -y_{n-n_\alpha} e_{n-1}(\theta) \cdots e_{n-n_\nu}(\theta) u_{n-1} \cdots u_{n-n_\gamma}).$$

This gradient has three components

$$(-\tilde{y}_{n-1}(\theta) \cdots -\tilde{y}_{n-n_\alpha}(\theta) \tilde{e}_{n-1}(\theta) \cdots \tilde{e}_{n-n_\nu}(\theta) \tilde{u}_{n-1}(\theta) \cdots \tilde{u}_{n-n_\gamma}(\theta))$$

that obey

$$(1 + \nu(L)) \tilde{y}_n(\theta) = y_n, \tag{1.4a}$$

$$(1 + \nu(L)) \tilde{e}_n(\theta) = e_n(\theta), \tag{1.4b}$$

$$(1 + \nu(L)) \tilde{u}_n(\theta) = u_n. \tag{1.4c}$$

Notice that

$$(1 + \alpha_0(L)) \tilde{y}_n(\theta_0) = \epsilon_n + \gamma_0(L) \tilde{u}_n(\theta_0), \tag{1.4d}$$

(where the subscript zero denotes a true value). We will need to know that stationarity and ergodicity (which we assume for ϵ_n) ensure that the limit

$$\boldsymbol{R} = \lim n^{-1} p_n^{-1}(\theta_0) = \lim n^{-1} \textstyle\sum_l^n \tilde{\varphi}_s(\theta_0) \tilde{\varphi}_s'(\theta_0) \tag{1.5}$$

exists w.p.l. It is further assumed $\mathbf{R}$ is positive definite.

The recursion will be then

$$\theta_n = \theta_{n-1} + P_n \tilde{\varphi}_n e_n, \tag{1.6}$$

$$P_n^{-1} = P_{n-1}^{-1} + \tilde{\varphi}_n \tilde{\varphi}_n', \tag{1.7}$$

$$e_n = y_n - \varphi_n' \theta_{n-1} \tag{1.8}$$

where

$$\varphi_n = (-y_{n-1} \cdots -y_{n-n_\alpha} e_{n-1} \cdots e_{n-n_\nu} u_{n-1} \cdots u_{n-n_\gamma}) \tag{1.9}$$

while $\tilde{\varphi}_n$ has components $\tilde{y}_n$, $\tilde{e}_n$, $\tilde{u}_n$ that obey

$$(1 + \nu_{n-1}(L))\tilde{y}_n = y_n, \tag{1.10a}$$

$$(1 + \nu_{n-1}(L))\tilde{e}_n = e_n, \tag{1.10b}$$

$$(1 + \nu_{n-1}(L))\tilde{u}_n = u_n \tag{1.10c}$$

(with $\nu_{n-1}(L) = \sum_1^{n_\nu} \nu_{n-1,i} L^i$ where $\nu_{n-1,i}$ are the $\boldsymbol{\nu}$-coordinates of $\boldsymbol{\theta}_n$).

With (1.4d) in mind, and for an ARMA model (u_n absent) (1.10a) is replaced in Hannan (1978) (henceforth referred to as (H)) by

$$(1 + \alpha_{n-1}(L))\tilde{y}_n = e_n. \tag{1.10a)$'$}$$

Actually (H) only explicitly derives results for an ARMA case but it is pointed out there that the same results will go through for the ARMAX case, with an equivalent of (1.4d) replacing (1.10a) once more.

2. A review of the convergence behaviour of $\mathbf{RML_2}$. The RML_2 recursion as given may not converge and must be monitored. Two schemes have been suggested, one used in (H) and the other in Ljung (1977). This latter scheme is similar to that used by Albert and Gardner (1967) and also Nevel'son and Khas'minskii (1973/1976; Chapter 7). Only the first scheme is discussed here.

The region of stability is separated into an inner region R_2 and an outer region R_1 with $R_2 \subset R_1 \subset R$. Here R_1 is such that $\forall \theta \in R_1$, $1 + \alpha(L)$, $1 + \beta(L)$ have no zeroes for $|L| \leq 1 + \mu$, $\mu > 0$. Also any pair of zeroes one from $1 + \alpha(L)$, one from $1 + \beta(L)$ are at least μ apart in the complex L plane. R_2 is described in a moment. Denote the output of the modified recursion by $\tilde{\theta}_n$.

Let $\bar{\theta}$ be a fixed value inside R_2. Define an auxiliary sequence $\bar{\theta}_n$ which is $\tilde{\theta}_n$ until $\tilde{\theta}_n$ first exits from R_1 and is $\bar{\theta}$ until it returns to R_2 when $\bar{\theta}_n = \tilde{\theta}_n$ again and so on. R_2 will be chosen to ensure the event $\{\tilde{\theta}_n$ remains outside R_1 indefinitely long$\}$ has zero probability. The recursion is now

$$\tilde{\theta}_n = \tilde{\theta}_{n-1} + P_n \tilde{\varphi}_n e_n, \tag{2.1a}$$

$$P_n^{-1} = P_{n-1}^{-1} + \tilde{\varphi}_n \tilde{\varphi}_n', \tag{2.1b}$$

$$e_n = y_n - \varphi_n' \tilde{\theta}_{n-1}. \tag{2.1c}$$

However e_n in (1.9), (1.10) is replaced by

$$\bar{e}_n = y_n - \varphi_n' \bar{\theta}_{n-1} \tag{2.1d}$$

and $\nu_{n-1}(L)$ in (1.10) is $\bar{\nu}_{n-1}(L)$ and so on. Thus to (2.1) we add

$$\varphi_n = (-y_{n-1} \cdots -y_{n-n_\alpha} \bar{e}_{n-1} \cdots \bar{e}_{n-n_\nu} u_{n-1} \cdots u_{n-n_\gamma}) \tag{2.2}$$

$$(1 + \bar{\nu}_{n-1}(L))\tilde{y}_n = y_n \tag{2.3a}$$

$$(1 + \bar{\nu}_{n-1}(L))\tilde{e}_n = \bar{e}_n \tag{2.3b}$$

$$(1 + \bar{\nu}_{n-1}(L))\tilde{u}_n = u_n \tag{2.3c}$$

with

$$\bar{\nu}_{n-1}(L) = \sum^{n_\nu} \bar{\nu}_{n-1,\iota} L^\iota \qquad \text{etc.}$$

Subtract θ_0 from (2.1a), multiply through by P_n^{-1} and sum up to obtain

$$\tilde{\theta}_n - \theta_0 = P_n \sum_1^n \bar{\varphi}_s(e_s - \tilde{\varphi}_s'(\tilde{\theta}_{s-1} - \theta_0)).$$

If we notice the Taylor series

$$e_n(\theta) = \epsilon_n + \tilde{\varphi}_n(\theta^\cdot)'(\theta - \theta_0)$$

then it is intuitively clear that we can expect ultimately to be dealing with an expression like

$$P_n \sum_1^n \tilde{\varphi}_s \epsilon_s$$

and so expect convergence and asymptotic efficiency. In any case, by repeated use of the martingale convergence theorem (in the form presented by Neveu (1975, page 151)). Hannan (1976) is able to evaluate $\tilde{\theta}_n$ as

(2.4a) $$\tilde{\theta}_n = \hat{\theta}_n + o_n(1) \qquad \text{a.s.}$$

where

(2.4b) $$\hat{\theta}_n = n^{-1} K_n^{-1} \sum_1^n a(\bar{\theta}_{s-1})$$

where

(2.4c) $$K_n = n^{-1} \sum_1^n G(\bar{\theta}_{s-1})$$

with

$$G(\theta) = E(\tilde{\varphi}_n(\theta)\tilde{\varphi}_n'(\theta))$$

in (H) this is denoted by $K(\theta)$ and

(2.5) $$a(\theta) = E(\tilde{\varphi}_n(\theta)(e_n(\theta) - \tilde{\varphi}_n'(\theta)\theta))$$

(cf. expression (2.14) of (H) where this is written in a different form; recall also that from (2.1a),

$$\tilde{\theta}_n = P_n \sum_1^n \tilde{\varphi}_s(e_s - \tilde{\varphi}_s'\tilde{\theta}_{s-1}))$$

We reorganise these expressions as follows. Write

$$a(\theta) = f(\theta) - G(\theta)\theta$$

with

$$f(\theta) = E(\tilde{\varphi}_n(\theta)e_n(\theta)).$$

Now we may write (2.4b) as

$$\hat{\theta}_n = \hat{\theta}_{n-1} + n^{-1}K_n^{-1}(a(\bar{\theta}_{n-1}) + G(\bar{\theta}_{n-1})\hat{e}_{n-1}).$$

Once it is known that the recursion converges so that $\tilde{\theta}_n = \bar{\theta}_n + o_n(1)$ then this expression may be evaluated to $o_n(1)$ a.s. as

(2.6) $$\hat{\theta}_n = \hat{\theta}_{n-1} + n^{-1}K_n^{-1}f(\bar{\theta}_{n-1}).$$

Also (2.4c) may be reorganised as

(2.7) $$K_n - K_{n-1} = n^{-1}(G(\bar{\theta}_{n-1}) - K_{n-1}).$$

Now setting $\tau_n = \sum_1^n s^{-1}$ and making the notational transformation $\hat{\theta}_n \leftrightarrow \theta(\tau)$ we are led to consider the behaviour of the ordinary differential equations

$$d\theta(\tau)/d\tau = R^{-1}(\tau)f(\theta(\tau)),$$
$$dR(\tau)/d\tau = G(\theta(\tau)) - R(\tau),$$

which are, in fact, the pair presented by Ljung (1977): see also Söderström et al. (1974).

The monitoring of the recursion enables the following conclusion to be drawn.

PROPOSITION 1 (P1). $\liminf n^{-1}P_n^{-1} > 0$ a.s.

Now we can see how the region R_2 is defined. Suppose $\bar{\theta}_n$ remains outside of R_2 indefinitely long then clearly

$$\hat{\theta}_n \to \hat{\theta} = G^{-1}(\bar{\theta})a(\bar{\theta})$$
$$= G^{-1}(\bar{\theta})f(\bar{\theta}) + \theta.$$

Recall

$$f(\theta) = E(\tilde{\varphi}_n(\theta)e_n(\theta)).$$

Straightforward manipulation shows the 'pseudo' Taylor series

$$e_n(\theta) = \epsilon_n - \varphi_n'(\theta, \theta_0)(\theta - \theta_0)$$

where

$$\varphi_n(\theta, \theta_0) = (1 + \nu_0(L))^{-1}\varphi_n(\theta).$$

Thus

$$f(\theta) = -L(\theta, \theta_0)(\theta - \theta_0) \tag{2.8}$$

where

$$L(\theta, \theta_0) = E(\tilde{\varphi}_n(\theta)\varphi_n'(\theta, \theta_0))$$

so that

$$\hat{\theta} - \theta_0 = (I - G^{-1}(\bar{\theta})L(\bar{\theta}, \theta_0))(\bar{\theta} - \theta_0)$$

(cf. (H) expression (2.16); also expression (2.4) for $L(\theta, \theta_0)$ and the present one are easily seen to be identical).

Now R_2 is defined so as to ensure

$$\|I - G^{-1}(\bar{\theta})L(\bar{\theta}, \theta_0)\| < 1 \tag{2.8a}$$

so that

$$\|\hat{\theta} - \theta_0\| \leq \|\bar{\theta} - \theta_0\|$$

which is a contradiction.

The following result is obtained.

PROPOSITION 2. (P2). $\bar{\theta}_n - \theta_0 \to 0$ a.s.

Actually in (H) a stronger result is proved. It is not required that $e_n(\theta_0)$ be a martingale difference sequence, only that y_n be a stationary process. Then θ_0 is defined as a value which minimises $E(e_n^2(\theta))$ or else satisfies $E(\tilde{\varphi}_n(\theta)e_n(\theta)) = 0$.

For the CLT/IP a stronger convergence result is needed namely (Theorem 2 of (H)) when $\sup E(\epsilon_n^4) < \infty$ then, PROPOSITION 3 (P3). $n^{1/2-\delta'}(\tilde{\theta}_n - \theta_0) \to 0$ a.s. $\forall \delta' > 0$.

In particular we will take $\delta' = \delta$ of (1.3a). The result is intuitively clear for the following reason. Consider the quadratic form $V_\tau = (\theta(\tau) - \theta_0)'R(\tau)(\theta(\tau) - \theta_0)$ (equivalent to $V_n = (\hat{\theta}_n - \theta_0)'K_n(\hat{\theta}_n - \theta_0)$) then it follows from the differential equations below (2.6) and (2.7), that

$$dV_\tau/d\tau = -V_\tau - 2(\theta(\tau) - \theta_0)'f(\theta(\tau)) + (\theta(\tau) - \theta_0)'G(\theta(\tau))(\theta(\tau) - \theta_0)).$$

Recalling (2.8) we see

$$dV_\tau/d\tau = -V_\tau - (\theta(\tau) - \theta_0)'(L_\tau' + L_\tau' - G(\theta(\tau))(\theta(\tau) - \theta_0)$$

with

$$L_\tau = L(\theta(\tau), \theta_0).$$

Now if $\theta(\tau) \to \theta_0$ a.s. then $L_\tau \to G(\theta_0)$ and we may expect the second term to be negative for τ large enough: we are left with

$$dV_\tau/d\tau \leq -V_\tau$$

or

$$V_n \leq V_{n-1}(1 - 1/n).$$

It clearly follows that $n^{1-\delta'}V_n \to 0$ for any $\delta' > 0$.

It will be naturally necessary in the ensuing discussion to dispose of the difference $\tilde{\theta}_n - \bar{\theta}_n$. It follows from the definition of $\tilde{\theta}_n$ and P2 that

PROPOSITION 4 (P4). *There exists a random variable* n_0 *with* $n_0 < \infty$ a.s. *and* $\tilde{\theta}_n \equiv \bar{\theta}_n \forall n > n_0$.

We will also need

PROPOSITION 5 (P5). $\lim n(P_n - P_n(\theta_0)) = 0$ a.s. *or* $\lim n^{-1}(\mathbf{P}_n^{-1} - \mathbf{P}_n^{-1}(\boldsymbol{\theta}_0)) = \mathbf{0}$ a.s.

Recall from (1.5), $\lim n^{-1}P_n^{-1}(\theta_0) = R$. The proof of this intuitively reasonable proposition is postponed.

The following rate of convergence result is also needed subsequently.

PROPOSITION 6. (P6). *Under condition* (1.3a),

$$n^{1/2+\delta}(\bar{\theta}_n - \bar{\theta}_{n-1}) \to 0 \quad \text{a.s.}$$

PROOF. Now

$$\tilde{\theta}_n - \tilde{\theta}_{n-1} = nP_n n^{-1}\tilde{\varphi}_n(y_n - \varphi_n'\tilde{\theta}_{n-1}).$$

In view of P4, P5 and expression (1.9) it is enough to show

$$n^{1/2+\delta}n^{-1}\tilde{\varphi}_n(y_n - \varphi_n'\tilde{\theta}_{n-1}) = n^{-1/2+\delta}\tilde{\varphi}_n\bar{e}_n = n^{-1/2+\delta}\tilde{\varphi}_n\epsilon_n + n^{-1/2+\delta}\tilde{\varphi}_n(\bar{e}_n - \epsilon_n) \to 0 \quad \text{a.s.}$$

This will follow if (setting $\tilde{\varphi}_n = \| \tilde{\varphi}_n \|$)

$$|\epsilon_n| n^{-1/4+\delta/2} \to 0 \quad \text{a.s.} \tag{2.9}$$

$$|\bar{e}_n - \epsilon_n| n^{-1/4+\delta/2} \to 0 \quad \text{a.s.} \tag{2.10}$$

$$\tilde{\varphi}_n n^{-1/4+\delta/2} \to 0 \quad \text{a.s.} \tag{2.11}$$

Of course we expect a much better result than (2.10) and one will subsequently be obtained.

Now (1.3a) ensures, via the Bienaymé-Chebyshev inequality and then the Borel-Cantelli lemma that (2.9) holds. We now show

$$\sup_n E(|\bar{e}_n - \epsilon_n|^{4/1-3\delta}) < \infty \tag{2.12a}$$

$$\sup_n E(\tilde{\varphi}_n^{4/1-3\delta}) < \infty. \tag{2.12b}$$

So that (2.10), (2.11) will follow by the same argument.

*Begin with (2.12*b*).* Take one component of $\tilde{\varphi}_n$ say $\tilde{e}_n$ with

$$(1 + \bar{\nu}_{n-1}(L))\tilde{e}_n = \bar{e}_n. \tag{2.3b}$$

Now $\bar{\theta}_n$ which is used to construct $\bar{\nu}_{n-1}(L)$ lies inside the stability region R, so we must have for some constant $\lambda < 1$, and another D

$$|\tilde{e}_n| \le \textstyle\sum_1^n \lambda^{n-s}|\bar{e}_s| + D\lambda^n$$

where $D\lambda^n$ accounts for initial conditions. Now

$$\begin{aligned} |\tilde{e}_n| &\le \textstyle\sum_1^n \lambda^{n-s}(|\bar{e}_n - \epsilon_s| + |\epsilon_s|) + D\lambda^n \\ &= \textstyle\sum_1^n \lambda^s(|\bar{e}_{n-s} - \epsilon_{n-s}| + |\epsilon_{n-s}|) + D\lambda^n \end{aligned}$$

whereupon (1.3a), (2.12a) and Hölder's inequality will ensure (2.12b) holds. Similar arguments follow for the other components of $\tilde{\varphi}_n$.

*Turn now to (2.12*a*).* From equations (2.1d) and (1.1)

$$(1 + \bar{\nu}_{n-1}(L))\bar{e}_n = (1 + \bar{\alpha}_{n-1}(L))y_n - \bar{\gamma}_{n-1}(L)u_n$$

$$(1 + \nu_0(L))\epsilon_n = (1 + \alpha_0(L))y_n - \gamma_0(L)u_n.$$

Thus

$$\begin{aligned} (1 + \nu_0(L))(\bar{e}_n - \epsilon_n) &= (\bar{\alpha}_{n-1}(L) - \alpha_0(L))y_n - (\bar{\nu}_{n-1}(L) - \nu_0(L))u_n \\ &\quad - (\bar{\gamma}_{n-1}(L) - \gamma_0(L))u_n \\ &= -\varphi_n'(\bar{\theta}_{n-1} - \theta_0). \end{aligned} \tag{2.13}$$

Now since $1 + \nu_0(L)$ is assumed stable we have for some constant $\lambda < 1$

$$|\bar{e}_n - \epsilon_n| \le \textstyle\sum_1^n \lambda^{n-s}\varphi_s \|\bar{\theta}_{s-1} - \theta_0\| + D\lambda^n \qquad \varphi_s = \|\varphi_s\|. \tag{2.14a}$$

Again we recall $\bar{\theta}_n$ lies in a bounded (stability) region so for some constant C

$$\begin{aligned} |\bar{e}_n - \epsilon_n| &\le C \textstyle\sum_1^n \lambda^{n-s}\varphi_s + D\lambda^n \\ &= C\textstyle\sum_1^n \lambda^s\varphi_{n-s} + D\lambda^n. \end{aligned} \tag{2.14b}$$

We note in passing that we can do better than this; since via P3 we have from (2.14a)

$$|\bar{e}_n - \epsilon_n| \le \textstyle\sum_1^n \lambda^{n-s}\varphi_s o_s(1)s^{-1/2+\delta} + D\lambda^n$$

so that if it can be shown that

$$\varphi_n n^{-1/4+\delta} \to 0 \quad \text{a.s.} \tag{2.15}$$

we will conclude, via the discrete l'Hospital rule (a special case of the Toeplitz lemma, see Bromwich, 1947, page 404)

$$n^{1/4}|\bar{e}_n - \epsilon_n| \to 0 \quad \text{a.s.} \tag{2.16}$$

Finally by the same argument that shows (1.3a) implies (2.12a) we see (2.15) follows from

$$\sup_n E(\varphi_n^{4/1-3\delta}) < \infty. \tag{2.12c}$$

Returning to (2.14b) it is clear, via Hölder's inequality, that (2.12a) will follow from (2.12c). *To see* (2.12c) observe that since $\tilde{\boldsymbol{\theta}}_n$ (used in the formation of $\boldsymbol{\varphi}_n$) lies in the stability region, $\boldsymbol{\varphi}_n$ is formed by stable linear filtering operations on y_n, u_n. Thus for some constant $\lambda < 1$

$$\begin{aligned}\varphi_n &\leq \textstyle\sum_1^n \lambda^{n-s}(|\epsilon_s| + |u_s|) + D\lambda^n \\ &= \textstyle\sum_1^n \lambda^{s}(|\epsilon_{n-s}| + |u_{n-s}|) + D\lambda^n\end{aligned}$$

whereupon (2.12c) follows from (1.3a), (1.3b) via Hölder's inequality.

The proof of P5 *is now given.* The norm of the difference in P5 is bounded by

$$2N^{-1} \textstyle\sum_1^N \bar{\varphi}_n^0 \delta_n + N^{-1} \sum_1^N \delta_n^2 \leq 2(N^{-1} \sum_1^N (\bar{\varphi}_n^0)^2 N^{-1} \sum_1^N \delta_n^2)^{1/2} + N^{-1} \sum_1^N \delta_n^2$$

where

$$\bar{\varphi}_n^0 = \| \bar{\varphi}_n(\theta_0) \|, \qquad \delta_n = \| \bar{\varphi}_n - \bar{\varphi}_n'(\theta_0) \|.$$

If we show $\lim \delta_n = 0$ then, by the discrete l'Hospital rule we can conclude $\lim N^{-1}\sum_1^N \delta_n^2 = 0$. Consider one component of $\bar{\varphi}_n$, say $\bar{e}_n$.

From (2.3b) and (1.4b) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ we deduce

$$(1 + \bar{\nu}_{n-1}(L))(\bar{e}_n - \bar{e}_n(\theta_0)) = \bar{e}_n - \epsilon_n + (\nu_0(L) - \bar{\nu}_{n-1}(L))\bar{e}_n(\theta_0).$$

Appealing yet again to the fact that $\bar{\boldsymbol{\theta}}_n$ lies in the stability region we have, via P3,

$$|\bar{e}_n - \bar{e}_n(\theta_0)| \leq \textstyle\sum_1^n \lambda^{n-s}(|\bar{e}_s(\theta_0)| s^{-1/2+\delta} o_s(1) + |\bar{e}_s - \epsilon_s|).$$

Thus, via the discrete l'Hospital rule $\lim |\bar{e}_n - \bar{e}_n(\theta_0)| = 0$ follows from

$$|\bar{e}_s - \epsilon_s| \to 0 \qquad \text{and} \qquad |\bar{e}_s(\theta_0)| s^{-1/2+\delta} \to 0 \quad \text{a.s.}$$

The first of these hold by (2.16) and the second from $\sup_n E|\bar{e}_n(\theta_0)|^{4/1-3\delta} < \infty$ in the same way that (2.9) follows from (1.3a).

3. A central limit theorem for $\mathbf{RML_2}$.

It was pointed out above that summing the recursion (2.1) yields

$$\bar{\theta}_n - \theta_0 = P_n \textstyle\sum_1^n \bar{\varphi}_s(\bar{e}_s + \bar{\varphi}_s'(\bar{\theta}_{s-1} - \theta_0)).$$

Also it was mentioned that the summand in this expression looks like

$$e_s(\boldsymbol{\theta}) - (de_s(\boldsymbol{\theta})/d\boldsymbol{\theta})'(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \qquad \text{for } \boldsymbol{\theta} = \boldsymbol{\theta}_{s-1}$$

and via a Taylor expansion this is almost $e_s(\boldsymbol{\theta}_0) = \epsilon_s$. So let us introduce

$$\mathbf{Q}_n = n^{1/2}\mathbf{P}_n \textstyle\sum_1^n \bar{\boldsymbol{\varphi}}_s \epsilon_s. \tag{3.1}$$

This quantity can easily be shown to obey a CLT so that a CLT for RML_2 will certainly follow if

$$n^{1/2}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \mathbf{Q}_n \to \mathbf{0} \quad \text{a.s.} \tag{3.2a}$$

THEOREM. *Let $\bar{\boldsymbol{\theta}}_n$ be the output of the monitored* RML_2 *recursion described in* (2.1), (2.2), (2.3) *and the paragraph above* (2.1), *where the region R_2 in that paragraph is defined so as to ensure* (2.8a) *holds. Then*

$$n^{1/2}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_{\mathscr{D}} N(\mathbf{0}, \mathbf{R}^{-1})$$

where $\mathbf{R}$ *is defined in* (1.5).

PROOF. Following the comments above it is enough to show (3.3a) and (3.2a).

(3.3a) $$\mathbf{Q}_n \to_{\mathscr{D}} N(\mathbf{0}, \mathbf{R}^{-1}).$$

Now in view of P5, (3.3a) may be replaced by

(3.3b) $$n^{-1/2} \textstyle\sum_1^n \boldsymbol{\lambda}' \tilde{\boldsymbol{\varphi}}_s \epsilon_s \to_{\mathscr{D}} N(0, \boldsymbol{\lambda}'\mathbf{R}\boldsymbol{\lambda})$$

where $\boldsymbol{\lambda}$ is an arbitrary fixed vector. This follows via P5 from the martingale central limit theorem of Scott (1973).

Now (3.2a) is just

$$n^{1/2}\mathbf{P}_n \textstyle\sum_1^n \tilde{\boldsymbol{\varphi}}_s(\bar{e}_s - \epsilon_s + \tilde{\boldsymbol{\varphi}}_s'(\bar{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta}_0)) \to \mathbf{0} \quad \text{a.s.}$$

Once more by P5 it is enough to show

$$n^{-1/2} \textstyle\sum_1^n \tilde{\varphi}_s(\bar{e}_s - \epsilon_s + \tilde{\varphi}_s'(\bar{\theta}_{s-1} - \theta_0)) \to 0 \quad \text{a.s.}$$

From P4 this may be replaced by

(3.2b) $$n^{-1/2} \textstyle\sum_1^n \tilde{\varphi}_s(\bar{e}_s - \epsilon_s + \tilde{\varphi}_s'(\bar{\theta}_{s-1} - \theta_0)) \to 0 \quad \text{a.s.}$$

Denote $f_s = | \bar{e}_s - \epsilon_s + \tilde{\boldsymbol{\varphi}}_s'(\bar{\boldsymbol{\theta}}_{s-1} - \boldsymbol{\theta}_0) |$ and $\tilde{\varphi}_d = \| \tilde{\varphi}_s \|$. Then (3.2b) will follow if $n^{-1/2} \sum_1^n \tilde{\varphi}_s f_s \to 0$ a.s. The idea is to show $\sum_1^\infty \tilde{\varphi}_s f_s s^{-1/2} < \infty$ a.s. so the result will follow from Kronecker's lemma. To bound f_s begin by recalling (2.13)

(2.13) $$(1 + \nu_0(L))(\bar{e}_n - \epsilon_n) = \varphi_n'(\bar{\theta}_{n-1} - \theta_0).$$

Furthermore it is shown in the Appendix that

(3.4a) $$(1 + \bar{\nu}_{n-1}(L))(\tilde{\varphi}_n'(\bar{\theta}_{n-1} - \theta_0)) = \varphi_n'(\bar{\theta}_{n-1} - \theta_0) + d_n$$

where

$$d_n = O(\| \bar{\theta}_n - \bar{\theta}_{n-1} \| \tilde{\varphi}_n).$$

It follows then that

(3.4b) $$(1 + \bar{\nu}_{n-1}(L))(\tilde{\varphi}_n'(\bar{\theta}_{n-1} - \theta_0) + \bar{e}_n - \epsilon_n) = (\bar{\nu}_{n-1}(L) - \nu_0(L)(\bar{e}_n - \epsilon_n) + d_n.$$

Now in view of P6

$$| d_n | = o_n(1) n^{-1/2-\delta} \tilde{\varphi}_n.$$

Finally it follows from argument already given in proving P6 that

$$\begin{aligned} f_n = | \tilde{\varphi}_n'(\bar{\theta}_{n-1} - \theta_0) + \bar{e}_n - \epsilon_n | &\le \textstyle\sum_1^n \lambda^{n-s} | d_s | + \sum_1^n \lambda^{n-s} | \bar{e}_s - \epsilon_s | o_s(1) s^{-1/2+\delta} \\ &= \textstyle\sum_1^n \lambda^{n-s}\, 0_n(1) n^{-1/2-\delta}(\tilde{\varphi}_n + | \bar{e}_n - \epsilon_n | n^{2\delta}). \end{aligned}$$

Thus set $\zeta_n = (\tilde{\varphi}_n + | \bar{e}_n - \epsilon_n | n^{2\delta})$ and consider

$$\begin{aligned} \textstyle\sum_1^N \tilde{\varphi}_s f_s s^{-1/2} &= \textstyle\sum_1^N \tilde{\varphi}_s s^{-1/2} \sum_1^s \lambda^{s-t} 0_t(1) t^{-1/2-\delta} \zeta_t \\ &= \textstyle\sum_1^N 0_t(1) \zeta_t t^{-1/2-\delta} \sum_t^N \tilde{\varphi}_s s^{-1/2} \lambda^{s-t} \\ &\le \textstyle\sum_1^N 0_t(1) \zeta_t(1) \zeta_t t^{-1-\delta} \sum_0^{N-t} \tilde{\varphi}_{r+t} \lambda^r \\ &\le \textstyle\sum_1^N 0_t(1) \zeta_t t^{-1-\delta} \psi_t \end{aligned}$$

where $\psi_t = \sum_0^\infty \tilde{\varphi}_{r+t} \lambda^r$.

Recalling the definition of ζ_n and that $| \bar{e}_n - \epsilon_n | n^{2\delta} \to 0$ (see (2.16)), we need only show

$$\textstyle\sum_1^\infty \tilde{\varphi}_s \psi_s s^{-1-\delta} < \infty, \ \sum_1^\infty \psi_s s^{-1-\delta} < \infty$$

these will hold if (see Lukacs, 1975, page 80)

$$\textstyle\sum_1^\infty E(\tilde{\varphi}_s \psi_s) s^{-1-\delta} < \infty, \ \sum_1^\infty E(\psi_s) s^{-1-\delta} < \infty$$

which will hold if

$$\sup_s E(\tilde{\varphi}_s \psi_s) < \infty, \ \sup_s E(\psi_s) < \infty.$$

Now $E(\tilde{\varphi}_s \psi_s) \leq \sqrt{E}(\tilde{\varphi}_s^2) \ \sqrt{E}(\psi_s^2)$ while

$$E^2(\psi_t) \leq E(\psi_t^2) \leq (1-\lambda)^{-1} \textstyle\sum_0^\infty \lambda^r E(\tilde{\varphi}_{r+t}^2)$$

so the required result follows from

$$\sup_s E(\tilde{\varphi}_s^2) < \infty$$

which holds in view of (2.12b).

REMARK. *An invariance principle for* RML_2. Denote $\tau_n = \sum_1^n s^{-1}$ and define

$$k_n(t) = \sup\{m : \tau_{n+m} - \tau_n \leq t\}.$$

Now (3.1) may be reorganised as

$$d\mathbf{Q}_n = \{(\tfrac{1}{2} + o(n^{-1}))\mathbf{I} - \mathbf{R}_n^{-1}\tilde{\boldsymbol{\varphi}}_n \tilde{\boldsymbol{\varphi}}_n'\}\mathbf{Q}_{n-1} d\tau_n + d\mathbf{W}_n$$

where

$$\mathbf{W}_n = \mathbf{R}_n^{-1} \textstyle\sum_1^n \tilde{\varphi}_s \epsilon_s / s^{1/2}$$

$\mathbf{R}_n = n\mathbf{P}_n^{-1}$ and $d\tau_n = \tau_n - \tau_{n-1}$ etc. Notice that $\mathrm{Var}(\mathbf{W}_n) \simeq \mathbf{R}_n^{-1}\tau_n$ so that $\mathbf{W}_n$ resembles a Wiener process on a time scale $\tau_n \simeq \ln n$. Also, recalling the definitions of $\mathbf{R}_n$ and $\mathbf{P}_n$ we might expect ultimately to deal with a process $\tilde{\mathbf{Q}}_n$ that obeys

$$d\tilde{\mathbf{Q}}_n \simeq -\tfrac{1}{2}\tilde{\mathbf{Q}}_{n-1} d\tau_n + d\mathbf{W}_n.$$

It is not surprising then that it can be shown (though not easily) that (see Solo, 1978a)

$$\mathbf{Q}_{n+k_n}(t) \Rightarrow \mathbf{Q}(t) \tag{3.5}$$

where for an arbitrary fixed vector $\boldsymbol{\lambda}$, $Q(t) = \boldsymbol{\lambda}'\mathbf{Q}(t)$ is the continuous Gaussian process on $D[0, 1]$ the space of all right continuous functions on $[0, 1]$ that have left-hand limits, with

$$Q(t) = e^{-1/2t}Q(o) + \int_0^t e^{-1/2(t-s)} dW(s)$$

and $Q(o) \sim N(0, \sigma^2)$ while $W(t)$ is the Wiener process on $D[0, 1]$ with variance $\sigma^2 = \boldsymbol{\lambda}'\mathbf{R}^{-1}\boldsymbol{\lambda}$. If we define $\hat{\mathbf{Q}}_n = n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ then once (3.5) is established the conclusion

$$\hat{\mathbf{Q}}_{n+k_n}(t) \Rightarrow \mathbf{Q}(t)$$

follows provided

$$\sup_{0 \leq t \leq 1} \| \mathbf{Q}_{n+k_n}(t) - \hat{\mathbf{Q}}_{n+k_n}(t) \| \to_p 0.$$

This will hold provided $\| \mathbf{Q}_n - \hat{\mathbf{Q}}_n \| \to_p 0$ which is however a consequence of the proof of the CLT (that is of condition (3.2b) of Section 3).

CONCLUSION. It has been shown that a monitored form of the RML_2 recursion obeys a CLT and is indeed asymptotically efficient. The result can be extended to give an invariance principle for RML_2. This enables the construction of a confidence interval for the trajectory of the recursion for all time from a certain time onward: see Ibragimov and Khas'minskii (1973) who give details for a conventional maximum likelihood estimator.

ACKNOWLEDGEMENTS. To Professor E. J. Hannan for some helpful discussion. This work forms part of the author's Ph.D. thesis submitted to the Australian National University and financed by an Australian Commonwealth Postgraduate Research Award.

APPENDIX

DERIVATION OF RELATION (3.4a). The argument is tedious but simple and is therefore illustrated for a MA(2) model only. Now

$$(1 + \bar{\nu}_{n-1}(L))(\tilde{\varphi}_n'(\bar{\theta}_{n-1} - \theta_0))$$
$$= (1 + \bar{\nu}_{1,n-1}L + \bar{\nu}_{2,n-1}L^2)(\tilde{e}_n(\bar{\nu}_{1,n-1} - \nu_1)) + (1 + \bar{\nu}_{1,n-1}L + \bar{\nu}_{2,n-1}L^2)(\tilde{e}_n(\bar{\nu}_{2,n-1} - \nu_2)).$$

Now the first term is

$$\begin{aligned}
&\tilde{e}_n(\bar{\nu}_{1,n-1} - \nu_1) + \bar{\nu}_{1,n-1}\tilde{e}_{n-1}(\bar{\nu}_{1,n-2} - \nu_1) + \bar{\nu}_{2,n-1}\tilde{e}_{n-2}(\bar{\nu}_{1,n-3} - \nu_1)\\
&\quad = \tilde{e}_n(\bar{\nu}_{1,n-1} - \nu_1) + \bar{\nu}_{1,n-1}\tilde{e}_{n-1}(\bar{\nu}_{1,n-1} - \nu_1) + \bar{\nu}_{2,n-1}\tilde{e}_{n-1}\tilde{e}_{n-2}(\bar{\nu}_{1,n-1} - \nu_1)\\
&\qquad + \bar{\nu}_{1,n-1}\tilde{e}_{n-1}(\bar{\nu}_{1,n-2} - \bar{\nu}_{1,n-1}) + \bar{\nu}_{2,n-1}\tilde{e}_{n-2}(\bar{\nu}_{1,n-3} - \bar{\nu}_{1,n-1})\\
&\quad = (\bar{\nu}_{1,n-1} - \nu_1)\left[(1 + \bar{\nu}_{1,n-1}L + \bar{\nu}_{2,n-1}L^2)\tilde{e}_n\right] + O(\|\bar{\theta}_n\|\|\tilde{\varphi}_n\|\|\bar{\theta}_n - \bar{\theta}_{n-1}\|)\\
&\quad = (\bar{\nu}_{1,n-1} - \nu_1)e_n + O(\|\tilde{\varphi}_n\|\|\bar{\theta}_n - \bar{\theta}_{n-1}\|)
\end{aligned}$$

in view of (2.3b) and the fact that $\bar{\theta}_n$ is surely bounded.

REFERENCES

ALBERT, A. E., AND GARDNER, L. A. (1967). *Stochastic Approximation and Nonlinear Regression.* Cambridge, Mass., M.I.T. Press.

BROMWICH, T. J. (1947). *An Introduction to the Theory of Infinite Series.* 2nd ed. MacMillan, London.

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

HANNAN, E. J. (1978). Recursive estimation based on ARMA models. Unpublished manuscript.

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

IBRAGIMOV, I. A. AND KHAS'MINSKII, R. Z. (1973). On the approximation of statistical estimators by sums of independent variables. *Soviet Math. Dokl.* **14** 883–887.

LJUNG, L. (1976). Prediction error identification methods. In *System Identification—Advances and Case Studies.* (R. K. Mehra, P. G. Lainiotis, eds.)

LJUNG, L. (1977). Analysis of recursive stochastic algorithms, *IEEE Trans. Autom. Control.* **22** 551–575.

LUKACS, E. (1975). *Stochastic Convergence.* 2nd ed. Academic.

NEVEL'SON, M. B., AND KHAS'MINSKII, R. Z. (1973/1976). *Stochastic Approximation and Recursive Estimation.* Trans. Math. Monographs, Vol. 47, Providence, Rhode Island, Amer. Math. Soc.

NEVEU, J. (1975). *Discrete Parameter Martingales,* North-Holland, New York.

PLACKETT, R. L. (1950). Some theorems in least squares. *Biometrika.* **37** 149.

SCOTT, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances Appl. Probability* **5** 119–137.

SÖDERSTRÖM, T., LJUNG, L., AND GUSTAVSSON, I. (1974). *A Comparative Study of Recursive Identification Methods.* Report 7427, Div. of Automatic Control, Lund Inst. of Technology, Lund, Sweden.

SOLO, V., (1978a). Time series recursions and stochastic approximation. Ph.D. Thesis submitted to Australian National Univ.

SOLO, V. (1978b), Stochastic approximation and the final value theorem. Unpublished manuscript.

SOLO, V., (1979), The convergence of AML. *IEEE. Trans. Autom. Control.* Dec. 1979.

MATHEMATICS RESEARCH CENTER
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706