Stochastic Processes and their Applications 21 (1986) 251-273
North-Holland

ESTIMATION IN NONLINEAR TIME SERIES MODELS

Dag TJØSTHEIM
Department of Mathematics, University of Bergen, 5000 Bergen, Norway, and Department of Statistics, University of North Carolina, Chapel Hill, NC 27514, USA

Received 10 December 1984
Revised 30 August 1985

A general framework for analyzing estimates in nonlinear time series is developed. General conditions for strong consistency and asymptotic normality are derived both for conditional least squares and maximum likelihood types estimates. Ergodic strictly stationary processes are studied in the first part and certain nonstationary processes in the last part of the paper. Examples are taken from most of the usual classes of nonlinear time series models.

asymptotic normality * conditional least square * consistency * maximum likelihood * nonlinear time series

Introduction

Recently there has been a growing interest in nonlinear time series models. Some representative references are Ǎndel (1976) and Nicholls and Quinn (1982) on random coefficient autoregressive models, Granger and Andersen (1978) and Subba Rao and Gabr (1984) on bilinear models, Haggan and Ozaki (1981) on exponential autoregressive models, Tong (1983) on threshold autoregressive models, Harrison and Stevens (1976), Ledolter (1981) on dynamic state space models and Priestley (1980) on general state dependent models. A review has been given in Tjøstheim (1985a).

To be able to use nonlinear time series models in practice one must be able to fit the models to data and estimate the parameters. Computational procedures for determining parameters for various model classes are outlined in the above references. Often these are based on a minimization of a least squares or a maximum likelihood type criterion. However, very little is known about the theoretical properties of these procedures and the resulting estimates. An exception is the class of random coefficient autoregressive processes for which a fairly extensive theory of estimation exists (Nicholls and Quinn 1982). See also the special models treated by Robinson (1977) and Aase (1983). Sometimes properties like consistency and asymptotic normality appear to be taken for granted also for other model classes, but some of the simulations performed indicate that there are reasons for being cautious.

In this paper we will try to develop a more systematic approach and discuss a general framework for nonlinear time series estimation. The approach is based on Taylor expansion of a general penalty function which is subsequently specialized to a conditional least squares and a maximum likelihood type criterion. Klimko and Nelson (1978) have previously considered such Taylor expansions in the conditional least squares case in a general (non-time series) context.

Our approach yields the estimation results of Nicholls and Quinn (1982) as special cases, and, in fact, we are able to weaken their conditions in the maximum likelihood case. The results derived are also applicable to other classes of nonlinear time series. Although the conditions for consistency and asymptotic normality are not always easy to verify, they seem to give a good indication of the specific problems that arise for each class of series.

An outline of the paper is as follows: In Section 2 we present some results on consistency and asymptotic normality using a general penalty function. In Sections 3-5 we specialize to conditional least squares and to a maximum likelihood type penalty function for stationary processes and give examples. In Sections 6 and 7 we consider conditional least squares estimates for nonstationary processes. The present paper is an abridged version of Tjøstheim (1984a, b) to which we refer for more details and complete proofs.

2. Two general results on consistency and asymptotic normality

Let $\{X_t, t \in I\}$ be a discrete time stochastic process taking values in R^d and defined on a probability space $(\Omega, \mathscr{F}, P)$. The index set I is either the set Z of all integers or the set N of all positive integers. We assume that observations $(X_1, \ldots, X_n)$ are available. We will treat the asymptotic theory of two types of estimates, namely conditional least squares and maximum likelihood type estimates. Both of these are obtained by minimizing a penalty function, and since, in our setting, the theory is quite similar for the two, we will formulate our results in terms of a general real-valued penalty function $Q_n = Q_n(\beta) = Q_n(X_1, \ldots, X_n; \beta)$ depending on the observations and on a parameter vector β.

The parameter vector $\beta = [\beta_1, \ldots, \beta_r]^{\mathrm{T}}$ will be assumed to be lying in some open set B of Euclidean r-space. Its true value will be denoted by β^0. We will assume that the penalty function Q_n is almost surely twice continuously differentiable in a neighborhood S of β^0. We will denote by $|\cdot|$ the Euclidean norm, so that $|\beta| = (\beta^{\mathrm{T}}\beta)^{1/2}$. For $\delta > 0$, we define $N_\delta = \{\beta : |\beta - \beta^0| < \delta\}$. We will use a.s. as an abbreviation for almost surely, although, when no misunderstanding can arise, it will be omitted in identities involving conditional expectations.

Theorems 2.1 and 2.2 are proved using the standard technique of Taylor expansion around β^0 (cf. Klimko and Nelson, 1978, and Hall and Heyde, 1980, Ch. 6). Let $N_\delta \subset S$. Moreover, let $\partial Q_n/\partial\beta$ be the column vector defined by $\partial Q_n/\partial\beta_i$, $i = 1, \ldots, r$, and likewise let $\partial^2 Q_n/\partial\beta^2$ be the $r \times r$ matrix defined by $\partial^2 Q_n/\partial\beta_i\,\partial\beta_j$, $i, j = 1, \ldots, r$.

Then

$$Q_n(\beta) = Q_n(\beta^0) + (\beta - \beta^0)^{\mathrm{T}} \frac{\partial Q_n}{\partial \beta}(\beta^0) + \tfrac{1}{2}(\beta - \beta^0)^{\mathrm{T}} \frac{\partial^2 Q_n(\beta^0)}{\partial \beta^2}(\beta - \beta^0)$$
$$+ \tfrac{1}{2}(\beta - \beta^0)^{\mathrm{T}} \left\{ \frac{\partial^2 Q_n}{\partial \beta^2}(\beta^*) - \frac{\partial^2 Q_n}{\partial \beta^2}(\beta^0) \right\} (\beta - \beta^0) \tag{2.1}$$

is valid for $|\beta - \beta^0| < \delta$. Here $\beta^* = \beta^*(X_1, \ldots, X_n; \beta)$ is an intermediate point between β and β^0.

Theorem 2.1. *Assume that* $\{X_t\}$ *and* Q_n *are such that, as* $n \to \infty$,

A1: $$n^{-1} \frac{\partial Q_n}{\partial \beta_i}(\beta^0) \xrightarrow{\text{a.s.}} 0, \quad i = 1, \ldots, r.$$

A2: *The symmetric matrix* $\partial^2 Q_n(\beta^0)/\partial \beta^2$ *is non-negative definite and*

$$\liminf_{n \to \infty} \lambda^n_{\min}(\beta^0) \overset{\text{a.s.}}{>} 0$$

where $\lambda^n_{\min}(\beta^0)$ *is the smallest eigenvalue of* $n^{-1} \partial^2 Q_n(\beta^0)/\partial \beta^2$.

A3: $$\limsup_{n \to \infty} \sup_{\delta \downarrow 0} (n\delta)^{-1} \left| \frac{\partial^2 Q_n}{\partial \beta_i \partial \beta_j}(\beta^*) - \frac{\partial^2 Q_n}{\partial \beta_i \partial \beta_j}(\beta^0) \right| \overset{\text{a.s.}}{<} \infty \quad \text{for } i, j = 1, \ldots, r.$$

Then there exists a sequence of estimators $\hat{\beta}_n = (\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nr})^{\mathrm{T}}$ *such that* $\hat{\beta}_n \xrightarrow{\text{a.s.}} \beta^0$ *as* $n \to \infty$, *and such that for* $\varepsilon > 0$, *there is an* E *event in* $(\Omega, \mathcal{F}, P)$ *with* $P(E) > 1 - \varepsilon$ *and an* n_0 *such that on* E *and for* $n > n_0$, $\partial Q_n(\hat{\beta}_n)/\partial \beta_i = 0$, $i = 1, \ldots, r$, *and* Q_n *attains a relative minimum at* $\hat{\beta}_n$.

The proof is as in Klimko and Nelson (1978) since it is easily checked that the argument does not depend on the special conditional least squares function used there.

When it comes to asymptotic normality it is essentially sufficient to prove asymptotic normality of $\partial Q_n(\beta^0)/\partial \beta$.

Theorem 2.2. *Assume that the conditions of Theorem* 2.1 *are fulfilled and that in addition we have that, as* $n \to \infty$,

B1: $$n^{-1} \frac{\partial^2 Q_n}{\partial \beta_i \partial \beta_j}(\beta^0) \xrightarrow{\text{a.s.}} V_{ij}$$

for $i, j = 1, \ldots, r$, *where* $V = (V_{ij})$ *is a strictly positive definite matrix, and*

B2: $$n^{-1/2} \frac{\partial Q_n}{\partial \beta}(\beta^0) \xrightarrow{\text{d}} \mathcal{N}(0, W)$$

where $\mathcal{N}(0, W)$ *is used to denote a multivariate normal distribution with a zero mean vector and covariance matrix* W. *Let* $\{\hat{\beta}_n\}$ *be the estimators obtained in Theorem* 2.1. *Then*

$$n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{\text{d}} \mathcal{N}(0, V^{-1} W V^{-1}). \tag{2.1}$$

The proof is identical to the proof of Theorem 2.2 of Klimko and Nelson (1978).

3. Conditional least squares. The stationary case

In Sections 3-5, $\{X_t\}$ will be assumed to be strictly stationary and ergodic. In addition second moments af $\{X_t\}$ will always be assumed to exist, so that $\{X_t\}$ is second order stationary as well. The task of finding nonlinear models satisfying these assumptions is far from trivial (cf. Tjøstheim, 1985a, Section 5). It should be realized that a strictly stationary model is capable of producing realizations with a distinctive nonstationary outlook (cf. e.g. Nicholls and Quinn, 1982, Section 1 and Tjøstheim, 1985a, Section 5).

We denote by $\mathscr{F}_t^X$ the sub σ-field of $\mathscr{F}$ generated by $\{X_s, s \leq t\}$, and we will use the notation $\tilde{X}_{t|t-1} = \tilde{X}_{t|t-1}(\beta)$ for the conditional expectation $E_\beta(X_t \mid \mathscr{F}_{t-1}^X)$. We will often omit β for notational convenience.

In the case where $\{X_t\}$ is defined for $t \geq 1$ only (this will be referred to as the one sided case), $\tilde{X}_{t|t-1}$ will in general depend explicitly on t and therefore $\tilde{X}_{t|t-1}$ do not define a stationary process. If the index set I of $\{X_t, t \in I\}$ comprises all the integers, then $\tilde{X}_{t|t-1}$ is stationary, but in general $\tilde{X}_{t|t-1}$ will depend on X_t's not included in the set of observations $(X_1, \ldots, X_n)$. To avoid these problems we replace $\mathscr{F}_{t-1}^X$ by $\mathscr{F}_{t-1}^X(m)$, which is the σ-field generated by $\{X_s, t-m \leq s \leq t-1\}$, and let $\tilde{X}_{t|t-1} = E\{X_t \mid \mathscr{F}_{t-1}^X(m)\}$. Here m is an integer at our disposal, and we must have $t \geq m+1$ in the one sided case.

We will use the penalty function

$$Q_n(\beta) = \sum_{t=m+1}^{n} \{X_t - \tilde{X}_{t|t-1}(\beta)\}^2 \tag{3.1}$$

and the conditional least squares estimates will be obtained by minimizing this function. In the important special case where $\tilde{X}_{t|t-1}$ only depends on $\{X_s, t-p \leq s \leq t-1\}$, i.e. $\{X_t\}$ is a nonlinear autoregressive process of order p, we can take $m = p$ and we have $E(X_t \mid \mathscr{F}_{t-1}^X) = E\{X_t \mid \mathscr{F}_{t-1}^X(m)\}$, where $t \geq m+1$ in the one sided case.

The theorems in this section are essentially obtained by reformulating and extending the arguments of Klimko and Nelson (1978) to the multivariate case. Their proofs are therefore omitted and the interested reader is referred to Tjøstheim (1984a).

Theorem 3.1. *Assume that $\{X_t\}$ is a d-dimensional strictly stationary ergodic process with $E(|X_t|^2) < \infty$ and such that $\tilde{X}_{t|t-1}(\beta) = E_\beta\{X_t \mid \mathscr{F}_{t-1}^X(m)\}$ is almost surely three times continuously differentiable in an open set B containing β^0. Moreover, suppose that*

C1: $$E\left\{\left|\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}(\beta^0)\right|^2\right\} < \infty \quad \textit{and} \quad E\left\{\left|\frac{\partial^2 \tilde{X}_{t|t-1}}{\partial \beta_i \partial \beta_j}(\beta^0)\right|^2\right\} < \infty$$

for $i, j = 1, \ldots, r$.

C2: *The vectors* $\partial\tilde{X}_{t|t-1}(\beta^0)/\partial\beta_i$, $i=1,\ldots,r$, *are linearly independent in the sense that if* $a_1,\ldots,a_r$ *are arbitrary real numbers such that*

$$E\left\{\left|\sum_{i=1}^{r} a_i \frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}(\beta^0)\right|^2\right\}=0,$$

then $a_1=a_2=\cdots=a_r=0$.

C3: *For* $\beta\in B$, *there exists functions* $G_{t-1}^{ijk}(X_1,\ldots,X_{t-1})$ *and* $H_t^{ijk}(X_1,\ldots,X_t)$ *such that*

$$\left|\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}(\beta)\frac{\partial^2\tilde{X}_{t|t-1}}{\partial\beta_j\beta_k}(\beta)\right|\leq G_{t-1}^{ijk},\quad E(G_{t-1}^{ijk})<\infty,$$

$$\left|\{X_t-\tilde{X}_{t|t-1}(\beta)\}\frac{\partial^3\tilde{X}_{t|t-1}}{\partial\beta_i\,\partial\beta_j\,\partial\beta_k}(\beta)\right|\leq H_t^{ijk},\quad E(H_t^{ijk})<\infty,$$

for $i,j,k=1,\ldots,r$.

Then there exists a sequence of estimators $\{\hat{\beta}_n\}$ *minimizing* Q_n *of* (3.1) *such that the conclusion of Theorem* 2.1 *holds.*

Let $\partial\tilde{X}_{t|t-1}/\partial\beta$ be the $d\times r$ matrix having $\partial\tilde{X}_{t|t-1}/\partial\beta_i$, $i=1,\ldots,r$, as its column vectors. We denote by $U=\frac{1}{2}V$ the $r\times r$ matrix defined by

$$U=E\left\{\frac{\partial\tilde{X}_{t|t-1}^{T}}{\partial\beta}(\beta^0)\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta}(\beta^0)\right\}. \tag{3.2}$$

The proof of asymptotic normality depends on Billingsley's (1961) martingale central limit theorem. We then need to condition with respect to an increasing sequence of σ-fields in order to obtain a martingale, and since $\{\mathcal{F}_t^X(m)\}$ is not increasing, we now assume the existence of an m such that we have ($t\geq m+1$ in one sided case)

$$E(X_t\,|\,\mathcal{F}_{t-1}^X)\overset{\text{a.s.}}{=}E(X_t\,|\,\mathcal{F}_{t-1}^X(m))$$

and

$$\begin{aligned} f_{t|t-1} &\overset{\Delta}{=} E\{(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^{\mathrm{T}}\,|\,\mathcal{F}_{t-1}^X\} \\ &\overset{\text{a.s.}}{=} E\{(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^{\mathrm{T}}\,|\,\mathcal{F}_{t-1}^X(m)\} \end{aligned} \tag{3.3}$$

where we have used $f_{t|t-1}$ to denote the $d\times d$ conditional prediction error matrix of $\{X_t\}$. The relations in (3.3) hold trivially for nonlinear AR processes.

Theorem 3.2. *Assume that* (3.3) *and the conditions of Theorem* 3.1 *are fulfilled. In addition assume that*

D1:

$$R=E\left\{\frac{\partial\tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial\beta}(\beta^0)f_{t|t-1}(\beta^0)\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta}(\beta^0)\right\}<\infty.$$

Let $\{\hat{\beta}_n\}$ *be the estimators obtained in Theorem* 3.1. *Then*

$$n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{d} \mathcal{N}(0, U^{-1}RU^{-1}). \tag{3.4}$$

For a large class of time series models (including the ordinary linear AR models) the condition D1 is implied by the condition C1 of Theorem 3.1, and hence essentially no extra condition is required to ensure asymptotic normality.

Corollary 3.1. *If* $X_t - \tilde{X}_{t|t-1}(\beta^0)$ *is independent of* $\mathscr{F}^X_{t-1}$, *then* D1 *is implied by* C1.

Proof. Under the stated independence assumption we have

$$f_{t|t-1}(\beta^0) = E[\{X_t - X_{t|t-1}(\beta^0)\}\{X_t - X_{t|t-1}(\beta^0)\}^{\mathrm{T}}] \tag{3.5}$$

and the Schwarz inequality yields the conclusion. □

4. Examples

For notational convenience we will omit the superscript 0 for the true value of the parameter vector in this section. Moreover, in all of the following $\{e_t, -\infty < t < \infty\}$ will denote a sequence of independent identically distributed (iid) (possibly vector) random variables with $E(e_t) = 0$ and $E(e_t e_t^{\mathrm{T}}) = G < \infty$.

4.1. Exponential autoregressive models

These models were introduced and studied by Ozaki (1980) and Haggan and Ozaki (1981). They have considered the problem of numerical evaluation of the parameters by minimization of the sum of squares penalty function Q_n of (3.1), and they have done simulations. However, we are not aware of any results concerning the asymptotic properties of these estimates.

To make the principles involved more transparent we will work with the first order model

$$X_t - \{\psi + \pi \exp(-\gamma X^2_{t-1})\}X_{t-1} = e_t \tag{4.1}$$

defined for $t \geq 2$ with X_1 being an initial variable.

Theorem 4.1. *Let* $\{X_t\}$ *be defined by* (4.1). *Assume that* $|\psi| + |\pi| < 1$, *and that* e_t *has a density function with infinite support such that* $E(e_t^6) < \infty$. *Then there exists a unique distribution for the initial variable* X_1 *such that* $\{X_t, t \geq 1\}$ *is strictly stationary and ergodic. Moreover, there then exists a sequence of estimators* $\{(\hat{\psi}_n, \hat{\pi}_n, \hat{\gamma}_n)\}$ *minimizing (as described in the conclusion of Theorem* 2.1) *the penalty function* Q_n *of* (3.1) *and such that* $(\hat{\psi}_n, \hat{\pi}_n, \hat{\gamma}_n) \xrightarrow{\text{a.s.}} (\psi, \pi, \gamma)$, *and* $(\hat{\psi}, \hat{\pi}_n, \hat{\gamma}_n)$ *is asymptotically normal.*

Proof. Our independence assumption on $\{e_t\}$ implies that $\{X_t, t \geq 1\}$ is a Markov process, and the problem of existence of a strictly stationary and ergodic solution to the difference equation (4.1) can then be treated using Corollary 5.2 of Tweedie (1975).

Since e_t has a density with infinite support it follows that $\{X_t\}$ is ϕ-irreducible (cf. Tweedie 1975) with ϕ being Lebesgue measure. Since for an arbitrary Borel set B we have

$$P(x, B) \stackrel{\Delta}{=} P(X_t \in B \mid X_{t-1} = x) = P(e_t \in B - a(x) \cdot x) \tag{4.2}$$

where $a(x) = \psi + \pi \exp(-\gamma x^2)$, and since the function a is continuous, it follows that $\{P(x, \cdot)\}$ is strongly continuous. Moreover, it is easily seen from (4.1) that

$$\gamma_x \stackrel{\Delta}{=} E\{(|X_t| - |X_{t-1}|) \mid X_{t-1} = x\} \leq \{|a(x)| - 1\}|x| + E(|e_t|). \tag{4.3}$$

Here, $|a(x)| \leq |\psi| + |\pi| \exp(-\gamma x^2) \leq |\psi| + |\pi|$ since $\gamma \geq 0$. Let $\alpha = E(|e_t|)/(1 - |\psi| - |\pi|)$. Then if $|\psi| + |\pi| < 1$, there exists a $c > 0$ such that $\gamma_x \leq -c$ for all x with $|x| > \alpha$. Moreover, γ_x is bounded from above for all x with $|x| \leq \alpha$. It follows from Corollary 5.2 of Tweedie (1975) that there exists a unique invariant initial distribution for X_1 such that $\{X_t, t \geq 1\}$ is strictly stationary and ergodic.

Since we have a nonlinear AR(1) process, we can take $m = 1$ in Theorems 3.1 and 3.2. The conditions stated in (3.3) will then be trivially fulfilled and we have for $t \geq 2$

$$\tilde{X}_{t|t-1} = E(X_t \mid \mathscr{F}^X_{t-1}) = \{\psi + \pi \exp(-\gamma X^2_{t-1})\} X_{t-1}. \tag{4.4}$$

Furthermore, $f_{t|t-1} = E(e_t^2) = \sigma^2$ such that D1 of Theorem 3.2 follows from C1 of Theorem 3.1, and it is sufficient to verify C1-C3.

Since any moment of $\pi \exp(-\gamma X^2_{t-1}) X_{t-1}$ exists, it follows from (4.1) and the strict stationarity of $\{X_t\}$ that $E(e_t^6) < \infty$ implies $E(X_t^6) < \infty$. From (4.4) we have

$$\begin{aligned} &\frac{\partial \tilde{X}_{t|t-1}}{\partial \psi} = X_{t-1}, \qquad \frac{\partial \tilde{X}^k_{t|t-1}}{\partial \gamma^k} = (-2)^k \pi \exp(-\gamma X^2_{t-1}) X^{k+1}_{t-1}, \\ &\frac{\partial \tilde{X}_{t|t-1}}{\partial \pi} = \exp(-\gamma X^2_{t-1}) X_{t-1}, \qquad \frac{\partial \tilde{X}^{k+1}_{t|t-1}}{\partial \gamma^k \partial \pi} = (-2)^k \exp(-\gamma X^2_{t-1}) X^{k+1}_{t-1} \end{aligned} \tag{4.5}$$

for $k = 1, \ldots,$ while the other derivatives are zero. It is easily seen that $E(X_t^6) < \infty$ implies that C1 is satisfied. Since $|\psi| + |\pi| < 1$, we have that $|X_t - X_{t|t-1}| \leq |X_t| + |X_{t-1}|$ and that the above derivatives are bounded by $|X_{t-1}|$, $2^k |X_{t-1}|^{k+1}$, $|X_{t-1}|$ and $2^k |X_{t-1}|^{k+1}$, respectively. Successive applications of the Schwarz inequality and use of $E(X_t^6) < \infty$ yield C3.

Let a_1, a_2 and a_3 be three arbitrary real numbers. Then

$$E\left(\left| a_1 \frac{\partial \tilde{X}_{t|t-1}}{\partial \psi} + a_2 \frac{\partial \tilde{X}_{t|t-1}}{\partial \pi} + a_3 \frac{\partial \tilde{X}_{t|t-1}}{\partial \gamma} \right|^2 \right) = 0 \tag{4.6}$$

implies

$$X_{t-1}[a_1 + \exp(-\gamma X_{t-1}^2)\{a_2 X_{t-1} - 2a_3\pi\}] \overset{\text{a.s.}}{=} 0, \tag{4.7}$$

and since $E(X_t^2) \geq E(e_t^2) > 0$, it follows that $a_1 = a_2 = a_3 = 0$. Hence C2 holds and the proof is completed. □

The infinite support assumption on $\{e_t\}$ can be relaxed. Moreover, it is not absolutely critical that the model (4.1) is initiated with X_1 in its stationary invariant distribution. The critical fact is the *existence* of such a distribution (cf. Klimko and Nelson, 1978, Section 4).

The general pth order model can be transformed to a first order vector autoregressive model, and essentially the same technique can be used as indicated in Tjøstheim (1984a).

4.2. *Some other models*

A related class of models is the threshold autoregressive processes (Tong and Lim 1980). Unfortunately we have not been able to establish the existence of a stationary invariant initial distribution for these processes. The transition probability $P(X, \cdot)$ is not in general strongly continuous (nor is it weakly continuous), and this makes it difficult to apply Tweedie's (1975) criterion. We will treat the threshold processes in Section 7, however.

Another class of related processes is studied by Aase (1984) (see also Jones, 1978). Results similar to those of Theorem 4.1 can be obtained (cf. Tjøstheim, 1984a).

Random coefficient autoregressive (RCA) models are defined by allowing random additive perturbations of the AR coefficients of ordinary AR models. Thus a d-dimensional RCA model of order p is defined by

$$X_t - \sum_{i=1}^{p} (a_i + b_{ti}) X_{t-i} = e_t \tag{4.8}$$

for $-\infty < t < \infty$. Here, a_i, $i = 1, \ldots, p$, are deterministic $d \times d$ matrices, whereas $\{b_t(p)\} = \{[b_{t1}, \ldots, b_{tp}]\}$ defines a $d \times pd$ zero-mean matrix process with the $b_t(p)$'s being iid and independent of $\{e_t\}$. Using the methods described in Section 3 the results of Nicholls and Quinn (1982) on least squares estimation, i.e. their theorems 3.1, 3.2, 7.1 and 7.2 can easily be derived. Again we refer to Tjøstheim (1984a) for details.

The bilinear class of models has received considerable attention recently. We refer to Granger and Andersen (1978), Subba Rao and Gabr (1984) and Bhaskara Rao et al. (1983) and references therein. We are not aware of a theory of statistical inference for these models, except in rather special cases (cf. Hall and Heyde 1980, Section 6.5). Using our general framework we have only been able to treat (Tjøstheim, 1984a) some special bilinear series studied by Guegan (1983).

In general the conditions C1, C3 and D1 essentially require mean square convergence in terms of past X_t's of such quantities as $E(e_{t-i}|\mathscr{F}_{t-1}^X)$ and their derivatives for general lags i and are thus intimately connected with the invertibility problem of bilinear models. This problem seems very complicated (cf. Granger and Andersen, 1978, Chapter 8) and until more progress is made, it appears to be difficult to make substantial headway in conditional least squares estimation of bilinear series using the present framework.

5. A maximum likelihood type penalty function

In all of the following it will be assumed that the conditional prediction error matrix $f_{t|t-1}$ is nonsingular and that there exists an m such that (3.3) holds. We introduce the likelihood type penalty function

$$L_n = \sum_{t=m+1}^{n} [\ln\{\det(f_{t|t-1})\} + (X_t - \tilde{X}_{t|t-1})^{\mathrm{T}} f_{t|t-1}^{-1} (X_t - \tilde{X}_{t|t-1})] \overset{\Delta}{=} \sum_{t=m+1}^{n} \phi_t. \tag{5.1}$$

If $\{X_t\}$ is a conditional Gaussian process, then L_n coincides with the log likelihood function except for a multiplicative constant. However, in this paper we will not restrict ourselves to Gaussian processes and a likelihood interpretation, but rather view L_n as a general penalty function which, since it has (cf. Tjøstheim, 1984a) the martingale property, e.g. $E\{\partial\phi_t(\beta^0)/\partial\beta_i|\mathscr{F}_{t-1}^X\}=0$, for a general $\{X_t\}$, it can be subjected to the kind of analysis described in Sections 2 and 3.

The analysis of L_n will differ in an essential way from that based on conditional least squares only in the case where $f_{t|t-1}$ is a genuine stochastic process; i.e. when $X_t - \tilde{X}_{t|t-1}$ is not independent of $\mathscr{F}_{t-1}^X$. For the examples treated in Section 4 this is the case only for the RCA processes. More general state space models of this type will be treated in Section 7. As will be seen, using L_n it is sometimes possible to relax moment conditions on $\{X_t\}$.

5.1. Consistency

We denote by s the number of components of the parameter vector β appearing in $L_n(\beta)$. Due to the presence of $f_{t|t-1}$ in L_n, in general $s>r$ with r defined as in Theorem 3.1. The symbol $\otimes$ denotes tensor product, while vec($\cdot$) stands for vectorization (cf. Nicholls and Quinn, 1982, Chapter 1).

Theorem 5.1. *Assume that $\{X_t\}$ is a d-dimensional strictly stationary and ergodic process with $E(|X_t|^2)<\infty$, and that $\tilde{X}_{t|t-1}(\beta)$ and $f_{t|t-1}(\beta)$ are almost surely three times continuously differentiable in an open set B containing β^0. Moreover, if ϕ_t is defined by* (5.1), *assume that*

E1: $E\left(\left|\dfrac{\partial\phi_t}{\partial\beta_i}(\beta^0)\right|\right)<\infty$ and $E\left(\left|\dfrac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta^0)\right|\right)<\infty$ *for* $i,j=1,\ldots,s$.

E2: *For arbitrary real numbers* $a_1, \ldots, a_s$ *such that, for* $\beta = \beta^0$,

$$E\left(\left|f_{t|t-1}^{-1/2} \sum_{i=1}^{s} a_i \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}\right|^2\right) + E\left[\left|f_{t|t-1}^{-1/2} \otimes f_{t|t-1}^{-1/2} \sum_{i=1}^{s} a_i \frac{\partial}{\partial \beta_i}\{\operatorname{vec}(f_{t|t-1})\}\right|^2\right] = 0, \tag{5.2}$$

then we have $a_1 = a_2 = \cdots = a_s = 0$,

E3: *For* $\beta \in B$, *there exists a function* $H_t^{ijk}(X_1, \ldots, X_t)$ *such that*

$$\left|\frac{\partial^3 \phi_t}{\partial \beta_i \partial \beta_j \partial \beta_k}(\beta)\right| \leq H_t^{ijk} \quad \text{and} \quad E(H_t^{ijk}) < \infty$$

for $i, j, k = 1, \ldots, s$.

Then there exists a sequence of estimators $\{\hat{\beta}_n\}$ *minimizing* L_n *of* (5.1) *such that the conclusion of Theorem* 2.1 *holds.*

Proof. Due to stationarity and ergodicity and the first part of E1, we have $n^{-1}\partial L_n(\beta^0)/\partial \beta_i \xrightarrow{\text{a.s.}} E\{\partial \phi_t(\beta^0)/\partial \beta_i\}$ as $n \to \infty$. However, because of the martingale increment property just mentioned for $\{\partial \phi_t(\beta^0)/\partial \beta_i\}$ we have $E\{\partial \phi_t(\beta^0)/\partial \beta_i\} = E[E\{\partial \phi_t(\beta^0)/\partial \beta_i \mid \mathscr{F}_{t-1}^X\}] = 0$ and A1 of Theorem 2.1 follows. Similarly, A3 of that theorem follows from E3, the mean value theorem and the ergodic theorem.

Using the last part of E1 and the ergodic theorem we have

$$n^{-1}\frac{\partial^2 L_n}{\partial \beta_i \partial \beta_j}(\beta^0) \xrightarrow{\text{a.s.}} E\left[E\left\{\frac{\partial^2 \phi_t}{\partial \beta_i \partial \beta_j}(\beta^0) \mid \mathscr{F}_{t-1}^X\right\}\right] \stackrel{\Delta}{=} V'_{ij}. \tag{5.3}$$

It remains to show that E2 implies that the matrix $V' = (V'_{ij})$ is positive definite. It can be shown (Tjøstheim, 1984a) that

$$E\left(\frac{\partial^2 \phi_t}{\partial \beta_i \partial \beta_j} \mid \mathscr{F}_{t-1}^X\right) = \operatorname{Tr}\left(f_{t|t-1}^{-1}\frac{\partial f_{t|t-1}}{\partial \beta_i} f_{t|t-1}^{-1}\frac{\partial f_{t|t-1}}{\partial \beta_j}\right) + 2\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i} f_{t|t-1}^{-1}\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_j}. \tag{5.4}$$

However, using standard rules about tensor products and trace operations we have for $\beta = \beta^0$ and arbitrary real numbers $a_1, \ldots, a_s$

$$\sum_{i=1}^{s}\sum_{j=1}^{s} a_i a_j E\left\{E\left(\frac{\partial^2 \phi_t}{\partial \beta_i \partial \beta_j} \mid \mathscr{F}_{t-1}^X\right)\right\} = 2E\left(\left|f_{t|t-1}^{-1/2}\sum_{i=1}^{s} a_i \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}\right|^2\right) + E\left(\left|f_{t|t-1}^{-1/2} \otimes f_{t|t-1}^{-1/2}\sum_{i=1}^{s} a_i \operatorname{vec}\left(\frac{\partial f_{t|t-1}}{\partial \beta_i}\right)\right|^2\right) \geq 0. \tag{5.5}$$

Hence the matrix V defined in (5.3) is non-negative definite, and due to the positive definiteness of $f_{t|t-1}$ it now follows from (5.5) and E2 that V is in fact positive definite and the theorem is proved. □

5.2. Asymptotic normality

To ease comparison with the results of Section 3 we introduce the matrix U' defined by $U' = \frac{1}{2}V'$, where $V' = (V'_{ij})$ is given by (5.3). Also we will only treat the scalar case. The multivariate case is considered in Tjøstheim (1984a).

Using (5.4) U' is given for $\beta = \beta^0$ by

$$U'_{ij} = E\left\{\frac{1}{f_{t|t-1}^2}\left(f_{t|t-1}\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_j} + \frac{1}{2}\frac{\partial f_{t|t-1}}{\partial \beta_i}\frac{\partial f_{t|t-1}}{\partial \beta_j}\right)\right\} \tag{5.6}$$

for $i, j, \ldots, s$. Corresponding to Theorem 3.2 we have

Theorem 5.2. *Assume that the conditions of Theorem* 5.1 *are fulfilled and that, for* $\beta = \beta^0$ *and* $i, j = 1, \ldots, s$,

$$\text{F1:}\quad S_{ij} \stackrel{\Delta}{=} \tfrac{1}{4}E\left\{\frac{1}{f_{t|t-1}^4}\left(\frac{\partial f_{t|t-1}}{\partial \beta_i}\frac{\partial f_{t|t-1}}{\partial \beta_j}[E\{(X_t - X_{t|t-1})^4 \mid \mathscr{F}_{t-1}^X\} - 3f_{t|t-1}^2]\right.\right.$$
$$+ 2E\{(X_t - X_{t|t-1})^3 \mid \mathscr{F}_{t-1}^X\} f_{t|t-1}$$
$$\left.\left.\times\left(\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}\frac{\partial f_{t|t-1}}{\partial \beta_j} + \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_j}\frac{\partial f_{t|t-1}}{\partial \beta_i}\right)\right)\right\} < \infty. \tag{5.7}$$

Let $S = (S_{ij})$, *and let* $\{\hat{\beta}_n\}$ *be the estimators obtained in Theorem* 5.1. *Then we have*

$$S_{ij} = \tfrac{1}{4}E\left(\frac{\partial \phi_t}{\partial \beta_i}\frac{\partial \phi_t}{\partial \beta_j}\right) - U'_{ij}$$

and

$$n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{\text{d}} \mathcal{N}(0, (U')^{-1} + (U')^{-1}S(U')^{-1}). \tag{5.8}$$

Proof. We use the same technique as in the proof of Theorem 3.2. From the martingale central limit theorem in the strictly stationary ergodic situation and a Cramer–Wold argument, it follows that $n^{-1/2}\partial L_n(\beta^0)/\partial\beta$ has a multivariate normal distribution as its limiting distribution if the limiting covariance of this quantity exists. Using Theorem 2.2 this implies asymptotic normality of $\hat{\beta}_n$ and what remains is to evaluate the covariance matrix.

Since $\{\partial L_n(\beta^0)/\partial\beta_i, \mathscr{F}_n^X\}$ is a martingale, it is easy to verify that

$$n^{-1}E\left\{\frac{\partial L_n}{\partial \beta_i}(\beta^0)\frac{\partial L_n}{\partial \beta_j}(\beta^0)\right\} = n^{-1}\sum_{t=1}^{n} E\left\{\frac{\partial \phi_t}{\partial \beta_i}(\beta^0)\frac{\partial \phi_t}{\partial \beta_j}(\beta^0)\right\}$$
$$= E\left[E\left\{\frac{\partial \phi_t}{\partial \beta_i}(\beta^0)\frac{\partial \phi_t}{\partial \beta_j}(\beta^0) \mid \mathscr{F}_{t-1}^X\right\}\right]. \tag{5.9}$$

Using the definition (5.1) of ϕ_t it is not difficult to show that, for $\beta = \beta^0$,

$$E\left\{E\left(\frac{\partial \phi_t}{\partial \beta_i}\frac{\partial \phi_t}{\partial \beta_j} \mid \mathscr{F}_{t-1}^X\right)\right\} = 4(S_{ij} + U'_{ij}). \tag{5.10}$$

The finiteness of $E\{n^{-1/2}\partial L_n(\beta^0)/\partial\beta_j \cdot n^{-1/2}\partial L_n(\beta^0)/\partial\beta_i\}$ now follows from the assumptions E1 and F1, while the form of the covariance matrix in (5.8) follows from (2.1) and the definition of S and U'. □

In the case where $f_{t|t-1}$ does not depend on β we have $S=0$. Under the additional assumption of Corollary 3.1 we have

$$U' = E[\partial \tilde{X}^{\mathrm{T}}_{t|t-1}(\beta^0)/\partial\beta \{E(f_{t|t-1})\}^{-1} \partial \tilde{X}_{t|t-1}(\beta^0)/\partial\beta]$$

and estimation using L_n of (5.1) or Q_n of (3.1) essentially gives identical results.

5.3. An example: RCA processes

The method used by Nicholls and Quinn (1982, Chapter 4) requires compactness of the region over which the parameter vector is allowed to vary. This necessitates rather restrictive conditions (cf. conditions (ci)-(cii), p. 64 of their monograph). On the other hand the boundedness conditions on the moments are weaker than in the conditional least squares case.

Using our general theoretical framework we are able to dispense with the compactness conditions, while retaining the same weak conditions on the moments. We assume that conditions are fulfilled so that an ergodic strictly and second order stationary $\mathscr{F}^e_t \vee \mathscr{F}^b_t$-measurable solution of (4.8) exists. Such conditions are given in Quinn and Nicholls (1982, Chapter 2). Moreover, we will again omit the superscript 0 for the true value of the parameter vector. Finally, it is clear that (3.3) is satisfied with $m=p$.

In the scalar RCA case we have from (4.8) that $\tilde{X}_{t|t-1} = Y^{\mathrm{T}}_{t-1}a$ where $Y^{\mathrm{T}}(t-1) = [X_{t-1}, \ldots, X_{t-p}]$ and $a^{\mathrm{T}} = [a_1, \ldots, a_p]$ such that $\partial X_{t|t-1}/\partial a_i = X_{t-i}$. Furthermore, it is easy to show that

$$f_{t|t-1} = Y^{\mathrm{T}}_{t-1} \Lambda Y_{t-1} + \sigma^2 \tag{5.11}$$

where $\sigma^2 = E(e_t^2)$ and where Λ is the covariance matrix of the random perturbations $\{b_t(p)\}$ of (4.8).

Theorem 6.1. *Let $\{X_t\}$ be a scalar RCA process such that the above stated conditions are satisfied. Assume that $\{e_t\}$ cannot take on only two values almost surely and that Λ is positive definite. Then there exists a sequence of estimators $\{[\hat{a}_n, \hat{\Lambda}_n, \hat{\sigma}^2_n]\}$ minimizing (as described in the conclusion of Theorem 2.1) the penalty function L_n of (5.1) and such that $[\hat{a}_n, \hat{\Lambda}_n, \hat{\sigma}^2_n] \xrightarrow{\text{a.s.}} [a, \Lambda, \sigma^2]$. The estimates $[\hat{a}_n, \hat{\Lambda}_n, \hat{\sigma}^2_n]$ are joint asymptotically normal, if, in addition, we assume $E(e_t^4)<\infty$ and $E(b_{ti}^4)<\infty$, $i=1,\ldots,p$.*

Proof. We denote by $\lambda_{\min}>0$ the minimum eigenvalue of Λ. It is seen from (5.11) that

$$f_{t|t-1} \geq \lambda_{\min} Y^{\mathrm{T}}_{t-1} Y_{t-1} + \sigma^2 \geq \begin{cases} \sigma^2, \\ \lambda_{\min} Y^{\mathrm{T}}_{t-1} Y_{t-1}, \end{cases} \tag{5.12}$$

whereas

$$\left|\frac{\partial f_{t|t-1}}{\partial \Lambda_{ij}}\right| = |2X_{t-i}X_{t-j}| \leq Y_{t-1}^{\mathrm{T}} Y_{t-1} \tag{5.13}$$

for $i, j = 1, \ldots, p$, and $\partial f_{t|t-1}/\partial \sigma^2 = 1$. It follows from the assumption on $\{e_t\}$ that $\sigma^2 > 0$, and thus $f_{t|t-1}^{-1}$ is well defined and we have from (5.12) and (5.13) that

$$\left|f_{t|t-1}^{-1} \frac{\partial f_{t|t-1}}{\partial \Lambda_{ij}}\right| \leq \frac{2}{\lambda_{\min}} \tag{5.14}$$

and

$$\left|f_{t|t-1}^{-1} \frac{\partial f_{t|t-1}}{\partial \sigma^2}\right| \leq \frac{1}{\sigma^2}. \tag{5.15}$$

Only first order derivatives of $\tilde{X}_{t|t-1}$ and $f_{t|t-1}$ are non-zero for the RCA case, and it is seen by examination on a term by term basis that each of the terms involved in evaluating $E(|\partial \phi_t/\partial \beta_i|)$ and $E(|\partial^2 \phi_t/\partial \beta_i \partial \beta_j|)$ is bounded by $KE(X_t^2)$ for some constant K. It follows that $E1$ of Theorem 5.1 is fulfilled.

Similarly, we find (Tjøstheim, 1984a) that $|\partial^3 \phi_t(\beta)/\partial \beta_i \partial \beta_j \partial \beta_k| \leq M|X_t|^2$ for a constant M and where this holds for all $\beta \in B$. Thus, since we assume that $\{X_t\}$ is second order stationary, it follows that condition E3 of Theorem 5.1 is fulfilled. It remains to verify E2. But this essentially follows (Tjøstheim, 1984a) from the linear independence properties of RCA processes.

To prove asymptotic normality, according to Theorem 5.2, we have to prove finiteness of S_{ij} with S_{ij} defined as in (5.7). We only look at the term

$$E\left[\left|f_{t|t-1}^{-4} \frac{\partial f_{t|t-1}}{\partial \beta_i} \frac{\partial f_{t|t-1}}{\partial \beta_j} E\{(X_t - \tilde{X}_{t|t-1})^4 | \mathscr{F}_{t-1}^X\}\right|\right] \stackrel{\Delta}{=} E[C_{ij}] \tag{5.16}$$

of (5.7). The other terms can be treated likewise.

Using the fact that $\{e_t\}$ and $\{b_t(p)\} = \{[b_{t1}, \ldots, b_{tp}]\}$ are independent with $E(e_t) = E\{b_t(p)\} = 0$, we have

$$\begin{aligned} E\{(X_t - \tilde{X}_{t|t-1})^4 | \mathscr{F}_{t-1}^X\} &= E\{(b_t(p) Y_{t-1} + e_t)^4 | \mathscr{F}_{t-1}^X\} \\ &= \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \sum_{m=1}^p X_{t-i}X_{t-j}X_{t-k}X_{t-m} E(b_{ti}b_{tj}b_{tk}b_{tm}) \\ &\quad + 6\sigma^2 \sum_{i=1}^p \sum_{j=1}^p X_{t-i}X_{t-j} E(b_{ti}b_{tj}) + E(e_t^4). \end{aligned} \tag{5.17}$$

From (5.13) and $E(b_{ti}^4) < \infty$, $i = 1, \ldots, p$, it follows by successive applications of the Schwarz inequality that

$$\sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \sum_{m=1}^p |X_{t-i}X_{t-j}X_{t-k}X_{t-m}| |E(b_{ti}b_{tj}b_{tk}b_{tm})| \leq M_1 p^4 (Y_t^{\mathrm{T}} Y_t)^2 \tag{5.18}$$

and

$$\sum_{i=1}^{p}\sum_{j=1}^{p}|X_{t-i}X_{t-j}||E(b_{ti}b_{tj})|\leq M_2p^2Y_t^{\mathrm{T}}Y_t \tag{5.19}$$

for some positive constants M_1 and M_2. Using $E(e_t^4)<\infty$ and (5.12), (5.14) and (5.15) it is seen that C_{ij} defined in (5.16) is bounded with probability one, and thus $E(C_{ij})<\infty$. The other terms of (5.7) are shown to have a finite expectation using identical arguments, and this completes the proof. □

6. Conditional least squares. The nonstationary case

In the rest of the paper we will try to extend our general framework to some classes of nonstationary models. We will only treat certain types of nonstationarity, such as that arising from a nonexisting stationary initial distribution, or the nonstationarity arising from a nonhomogeneous generating white noise process $\{e_t\}$. Unlike the case of consistency for stationary series, it will not be possible to condition on $\mathscr{F}_{t-1}^X(m)$, which is the σ-field generated by $\{X_s,\ t-m\leq s\leq t-1\}$. This is because we will rely more on pure martingale arguments, and then we need an increasing sequence of σ-fields. Hence, from now on we will always condition with respect to $\mathscr{F}_{t-1}^X$ and assume that (3.3) is fulfilled. For autoregressive type processes of order p it will then be possible to express $\tilde{X}_{m+1|m}$ in terms of $(X_1,\ldots,X_n)$ if $\min(n,m)\geq p$.

The following two theorems correspond to Theorems 3.1 and 3.2 and we use the penalty function Q_n defined in (3.1).

Theorem 6.1. *Assume that $\{x_t\}$ is a d-dimensional stochastic process with $E\{|X_t|^2\}<\infty$ and such that $\tilde{X}_{t|t-1}(\beta)=E_\beta\{X_t\mid\mathscr{F}_{t-1}^X\}$ is almost surely twice continuously differentiable in an open set B containing β^0. Moreover, assume that there are two positive constants M_1 and M_2 such that, for $t\geq m+1$,*

$$\text{CN1:}\quad E\left\{\frac{\partial\tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial\beta_i}(\beta^0)f_{t|t-1}(\beta^0)\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}(\beta^0)\right\}\leq M_1$$

and

$$\text{CN2:}\quad E\left\{\frac{\partial^2\tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial\beta_i\,\partial\beta_j}(\beta^0)f_{t|t-1}(\beta^0)\frac{\partial^2\tilde{X}_{t|t-1}}{\partial\beta_i\,\partial\beta_j}(\beta^0)\right\}\leq M_2$$

for $i,j=1,\ldots,r$.

$$\text{CN3:}\quad \liminf_{n\to\infty}\lambda_{\min}^n(\beta^0)\overset{\text{a.s.}}{>}0$$

where $\lambda_{\min}^n(\beta^0)$ is the smallest eigenvalue of the symmetric non-negative definite matrix $A^n(\beta^0)$ with matrix elements given by

$$A_{ij}^n(\beta^0)=\frac{1}{n}\sum_{t=m+1}^{n}\frac{\partial\tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial\beta_i}(\beta^0)\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}(\beta^0). \tag{6.1}$$

CN4: *Let $N_\delta = \{\beta: |\beta - \beta^0| < \delta\}$ be contained in B. Then*

$$\limsup_{n\to\infty,\ \delta\downarrow 0} \delta^{-1} \left| A^n_{ij}(\beta) - A^n_{ij}(\beta^0) + \frac{1}{n} \sum_{t=m+1}^{n} \left[\{X_t - \tilde{X}_{t|t-1}(\beta)\}^{\mathrm{T}} \frac{\partial^2 \tilde{X}_{t|t-1}}{\partial\beta_i\,\partial\beta_j}(\beta) - \{X_t - \tilde{X}_{t|t-1}(\beta^0)\}^{\mathrm{T}} \frac{\partial^2 \tilde{X}_{t|t-1}}{\partial\beta_i\,\partial\beta_j}(\beta^0) \right] \right| \overset{\text{a.s.}}{<} \infty$$

for $i, j = 1, \ldots, r$.

Then there exists a sequence of estimators $\{\hat{\beta}_n\} = \{[\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nr}]^{\mathrm{T}}\}$ such that $\hat{\beta}_n \xrightarrow{\text{a.s.}} \beta^0$, and such that for $\varepsilon > 0$, there is an event E in $(\Omega, \mathcal{F}, P)$ with $P(E) > 1 - \varepsilon$ and an n_0 such that on E and for $n > n_0$, $\partial Q_n(\hat{\beta}_n)/\partial\beta_i = 0$, $i = 1, \ldots, r$, and Q_n attains a relative minimum at $\hat{\beta}_n$.

Proof. From the definition of $Q_n(\beta)$ in (3.1) it is easily seen that $\{\partial Q_n(\beta^0)/\partial\beta_i, \mathcal{F}^X_n\}$ is a zero-mean martingale. The increments $U_t = \partial Q_t/\partial\beta_i - \partial Q_{t-1}/\partial\beta_i$ are such that (using CN1)

$$E(|U_t(\beta^0)|^2) = 4E\left\{ \frac{\partial \tilde{X}^{\mathrm{T}}_{t|t-1}}{\partial\beta_i}(\beta^0) f_{t|t-1}(\beta^0) \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_i}(\beta^0) \right\} \leq 4M_1 \tag{6.2}$$

and it follows from a martingale strong law of large numbers (cf. Stout 1974, Theorem 3.3.8) that $n^{-1}\partial Q_n(\beta^0)/\partial\beta_i \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$, and A1 of Theorem 2.1 is fulfilled. Computing second order derivatives we have

$$\frac{\partial^2 Q_n}{\partial\beta_i\,\partial\beta_j} = 2 \sum_{t=m+1}^{n} \frac{\partial \tilde{X}^{\mathrm{T}}_{t|t-1}}{\partial\beta_i} \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_j} - 2 \sum_{t=m+1}^{n} \frac{\partial^2 \tilde{X}^{\mathrm{T}}_{t|t-1}}{\partial\beta_i\,\partial\beta_j} (X_t - \tilde{X}_{t|t-1}). \tag{6.3}$$

Here $\{\partial^2 \tilde{X}^{\mathrm{T}}_{t|t-1}(\beta^0)/\partial\beta_i\,\partial\beta_j[X_t - \tilde{X}_{t|t-1}(\beta^0)]\}$ defines a martingale difference sequence with respect to $\{\mathcal{F}^X_t\}$ and using CN2 while reasoning as above we have

$$\frac{1}{n}\frac{\partial^2 Q_n}{\partial\beta_i\,\partial\beta_j}(\beta^0) - \frac{2}{n} \sum_{t=m+1}^{n} \frac{\partial \tilde{X}^{\mathrm{T}}_{t|t-1}}{\partial\beta_i}(\beta^0) \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_j}(\beta^0) \xrightarrow{\text{a.s.}} 0 \tag{6.4}$$

as $n \to \infty$, and hence CN3 implies A2 of Theorem 2.1. Using (6.3) it is seen that CN4 is identical to A3, and the conclusion now follows from Theorem 2.1. □

The conditions CN1 and CN2 may be weakened in two directions as indicated in Tjøstheim (1984b).

When we now turn to the asymptotic distribution of $\hat{\beta}_n$, we cannot rely on Billingsley's (1961) result for ergodic strictly stationary martingale difference sequences which essentially is used to prove Theorem 3.2. However, there are more recent results from martingale central limit theory that can be applied. Typically these require a random scaling factor.

Let $\partial\tilde{X}_{t|t-1}/\partial\beta$ be the $d \times r$ matrix having $\partial\tilde{X}_{t|t-1}/\partial\beta_i$, $i = 1, \ldots, r$, as its column vectors and let R_n be the $r \times r$ symmetric non-negative definite matrix given by

$$R_n = \sum_{t=m+1}^{n} E\left(\frac{\partial \tilde{X}^{\mathrm{T}}_{t|t-1}}{\partial\beta} f_{t|t-1} \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta} \right) \tag{6.5}$$

Moreover, denote by T_n the stochastic $r \times r$ symmetric non-negative definite matrix defined by

$$T_n = \sum_{t=m+1}^{n} \frac{\partial \tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial \beta} \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta} \tag{6.6}$$

We will denote by A^{-1} the Moore–Penrose inverse of a matrix A and by $\det(A)$ the determinant of A. Then we have

Theorem 6.2. *Assume that the conditions of Theorem* 6.1 *are fulfilled and assume in addition that*

DN1: $\liminf_{n \to \infty} n^{-r} \det\{R_n(\beta^0)\} > 0$

and

DN2: $$R_n^{-1/2}(\beta^0)\left[\sum_{t=m+1}^{n} \frac{\partial \tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial \beta}(\beta^0)\{X_t - \tilde{X}_{t|t-1}(\beta^0)\}\right.$$
$$\left.\{X_t - \tilde{X}_{t|t-1}(\beta^0)\}^{\mathrm{T}} \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta}(\beta^0)\right] R_n^{-1/2}(\beta^0) \xrightarrow{\mathrm{p}} I_r$$

where I_r is the density matrix of dimension r.

Let $\{\hat{\beta}_n\}$ be the estimators obtained in Theorem 6.1. *Then*

$$R_n^{-1/2}(\beta^0) T_n(\beta^0)(\hat{\beta}_n - \beta^0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, I_r). \tag{6.7}$$

Proof. Note first of all that $R_n(\beta^0)$ is finite from CN1. Let

$$\tfrac{1}{2}\frac{\partial Q_n}{\partial \beta} \stackrel{\Delta}{=} S_n = -\sum_{t=m+1}^{n} \frac{\partial \tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial \beta}(X_t - \tilde{X}_{t|t-1}) \stackrel{\Delta}{=} \sum_{t=m+1}^{n} \zeta_t. \tag{6.8}$$

Since we are dealing with an asymptotic result, as in the proof of Theorem 2.2 of Klimko and Nelson (1978), we may assume that $S_n(\hat{\beta}_n) = 0$. Taylor expanding S_n about β^0 and subsequently normalizing with $R_n^{1/2}(\beta^0)$ we have

$$0 = R_n^{-1/2}(\beta^0) S_n(\beta^0) + R_n^{-1/2}(\beta^0) \frac{\partial S_n}{\partial \beta}(\beta_n^*)(\hat{\beta}_n - \beta^0) \tag{6.9}$$

where β_n^* is an intermediate point between $\hat{\beta}_n$ and β^0. Again, reasoning as in the proof of Theorem 2.2 of Klimko and Nelson (1978), in the limit as $n \to \infty$ we may replace β_n^* by β^0. Moreover, using DN1, the boundedness condition CN2 and the orthogonal increment property of a martingale difference sequence, it follows from Chebyshev's inequality that there exists an n_0 such that

$$F_n(\beta^0) \stackrel{\Delta}{=} R_n^{-1/2}(\beta^0) \sum_{t=m+1}^{n} \left[\frac{\partial \zeta_t}{\partial \beta}(\beta^0) - E\left\{\frac{\partial \zeta_t}{\partial \beta}(\beta^0) \,|\, \mathcal{F}_{t-1}^X\right\}\right] \tag{6.10}$$

is bounded in probability for $n \geq n_0$. Since, from Theorem 6.1, $\hat{\beta}_n \xrightarrow{\text{P}} \beta^0$, it follows that $F_n(\beta^0)(\hat{\beta}_n - \beta^0) \xrightarrow{\text{P}} 0$, and therefore, when taking distributional limits in (6.9), $R_n^{-1/2}(\beta^0)\,\partial S_n(\beta_n^*)/\partial\beta$ may be replaced by

$$R_n^{-1/2}(\beta^0) \sum_{t=m+1}^{n} E\left\{\frac{\partial \zeta_t}{\partial \beta}(\beta^0) \,|\, \mathscr{F}_{t-1}^X\right\} = R_n^{-1/2}(\beta^0)\, T_n(\beta^0), \tag{6.11}$$

and hence from (6.7) and (6.9), the theorem will be proved if we can prove that $R_n^{-1/2}(\beta^0) S_n(\beta^0) \xrightarrow{\text{d}} \mathcal{N}(0, I_r)$.

We use a Cramer–Wold argument. For an r-dimensional vector λ of real numbers it is sufficient to prove that

$$\lambda^{\mathrm{T}} R_n^{-1/2}(\beta^0) S_n(\beta^0) \xrightarrow{\text{d}} \mathcal{N}(0, \lambda^{\mathrm{T}}\lambda). \tag{6.12}$$

For this purpose we introduce

$$\xi_{tn} = -\lambda^{\mathrm{T}} R_n^{-1/2} \frac{\partial \tilde{X}_{t|t-1}^{\mathrm{T}}}{\partial \beta}(X_t - \tilde{X}_{t|t-1}) = \lambda^{\mathrm{T}} R_n^{-1/2} \zeta_t \tag{6.13}$$

Then $\lambda^{\mathrm{T}} R_n^{-1/2} S_n = \sum_{t=m+1}^{n} \xi_{tn}$, and for $\beta = \beta^0$ we have that ξ_{tn}, $m+1 \leq t \leq n$, are martingale increments for a zero-mean square integrable martingale array $J_{in} = \sum_{t=m+1}^{i} \xi_{tn}$, $m+1 \leq i \leq n$. It is then sufficient to verify the following conditions (cf. Hall and Heyde, 1980, Theorem 3.2, where the nesting and integrability conditions of that theorem are trivially fulfilled) for $\beta = \beta^0$:

(i) $\max_{m+1 \leq t \leq n} |\xi_{tn}| \xrightarrow{\text{P}} 0$,

(ii) $\sum_{t=m+1}^{n} \xi_{tn}^2 \xrightarrow{\text{P}} \lambda^{\mathrm{T}}\lambda$,

(iii) $E\left(\max_{m+1 \leq t \leq n} \xi_{tn}^2\right)$ is bounded in n.

The condition (ii) follows trivially from the definition of ξ_{tn} and the assumption DN2. Moreover,

$$\max_{m+1 \leq t \leq n} \xi_{tn}^2 \leq \sum_{t=m+1}^{n} \xi_{tn}^2 = \lambda^{\mathrm{T}} R_n^{-1/2} \sum_{t=m+1}^{n} \zeta_t \zeta_t^{\mathrm{T}} R_n^{-1/2} \lambda, \tag{6.14}$$

and using the definition of R_n in (6.5) we have that the expectation of the extreme right hand side of (6.14) is $\lambda^{\mathrm{T}}\lambda$, and (iii) follows from this.

Also, using the technique described in Hall and Heyde (1980, p. 53), for a given $\varepsilon > 0$

$$P\left(\max_{m+1 \leq t \leq n} |\xi_{tn}| > \varepsilon\right) = P\left\{\sum_{t=m+1}^{n} \xi_{tn}^2 1(|\xi_{tn}| > \varepsilon) > \varepsilon^2\right\} \tag{6.15}$$

where $1(\cdot)$ is the indicator function. But

$$\sum_{t=m+1}^{n} E\{\xi_{tn}^2 1(|\xi_{tn}| > \varepsilon)\}$$
$$= \sum_{t=m+1}^{n} \lambda^{\mathrm{T}} R_n^{-1/2} E\{\zeta_t \zeta_t^{\mathrm{T}} 1(|\lambda^{\mathrm{T}} R_n^{-1/2} \zeta_t \zeta_t^{\mathrm{T}} R_n^{-1/2} \lambda| > \varepsilon)\} R_n^{-1/2} \lambda, \tag{6.16}$$

and using the definition of ζ_t in (6.8) and the conditions CN1 and DN1 we have that for a given $\delta > 0$, there is an n_0 such that for $n > n_0$ and all t, $m+1 \leq t \leq n$,

$$E\{\zeta_t \zeta_t^{\mathrm{T}} 1(|\lambda^{\mathrm{T}} R_n^{-1/2} \zeta_t \zeta_t^{\mathrm{T}} R_n^{-1/2} \lambda| > \varepsilon)\} < \delta \tag{6.17}$$

for $\beta = \beta^0$. Again using CN1 and DN1 there exists an n_1 such that $|R_{n,ij}^{-1}(\beta^0)| \leq kn^{-1}$ for $n \geq n_1$; $i, j = 1, \ldots, r$, and for some constant k. Let $n' = \max(n_0, n_1)$. Then from (6.16) and (6.17) we have for $\beta = \beta^0$ and for $n \geq n'$

$$\sum_{t=n'}^{n} E\{\xi_{tn}^2 1(|\xi_{tn}| > \varepsilon)\} \leq K(\lambda, k)\delta, \tag{6.18}$$

where $K(\lambda, k)$ is a constant depending on λ and k but independent of n. On the other hand, using CN1, DN1 and (6.16) it follows at once that for $\beta = \beta^0$

$$\sum_{t=m+1}^{n'} E\{\xi_{tn}^2 1(|\xi_{tn}| > \varepsilon)\} \to 0 \tag{6.19}$$

as $n \to \infty$. Using Chebyshev's inequality, (6.18) and (6.19) now imply (i), and the proof is completed. □

The matrix R_n corresponds to the number of observations in the statement of Theorem 3.2. In the stationary ergodic case $n^{-1}R_n \to R$ and $n^{-1}T_n \xrightarrow{\text{a.s.}} U$ as $n \to \infty$, where U and R are given by (3.2) and condition D1 of Theorem 3.2, and it is seen that (6.7) reduces to (3.4) then. However, in the nonstationary case we do not require the convergence of $n^{-1}R_n$ and $n^{-1}T_n$, and in fact for the examples to be treated in the next section these quantities do not always converge.

7. Examples

We will illustrate our general results on several nonlinear time series classes. The technical difficulties are larger than in the stationary ergodic case, and, partly to display the essential elements involved more clearly, we will confine ourselves to discussing scalar first order AR type models. Extensions to higher order and vector models will be relatively straightforward in some of the cases. We will generally omit the superscript 0 for the true value of the parameters.

7.1. Threshold autoregressive processes

These models were originally introduced by Tong. The underlying idea is a piecewise linearized autoregressive model obtained by introduction of a local threshold dependence on the amplitude X_t.

Tong and Lim (1980) consider the numerical evaluation of maximum likelihood estimates of the parameters of the threshold model.

We will only treat the first order AR case and we will assume that there is only *one* residual process $\{e_t\}$ consisting of zero-mean iid random variables. We can then write the threshold model as

$$X_t - \sum_{j=1}^{m} a^j X_{t-1} H_j(X_{t-1}) = e_t \tag{7.1}$$

where this equation is supposed to hold for $t \geq 2$ with X_1 as an initial variable, and where $H_j(X_{t-1}) = 1(X_{t-1} \in F_j)$, $1(\cdot)$ being the indicator function and $F_1, \ldots, F_m$ disjoint regions of R^1 such that $UF_j = R^1$. There is no explicit time dependence in (7.1). The reason that we did not treat such processes in connection with our study of stationary processes in Section 4, is that we have not been able to prove the existence of an invariant stationary distribution for the initial variables in the threshold case. For a general initial variable X_1 it is clear that the process generated by (7.1) will be nonstationary.

Theorem 7.1. *Let* $\{X_t\}$ *be defined by* (7.1). *Assume that the threshold regions* F_j *are such that there exist constants* $\alpha_j > 0$ *so that for all t,* $E\{X_t^2 H_j(X_t)\} \geq \alpha_j$, $j = 1, \ldots, m$. *Moreover, assume that* $|a^j| < 1$, $j = 1, \ldots, m$, $E(X_1^4) < \infty$ *and* $E(e_t^4) < \infty$. *Then there exists a strongly consistent sequence of estimators* $\{\hat{a}_n\} = \{[\hat{a}_n^1, \ldots, \hat{a}_n^m]^{\mathrm{T}}\}$ *for* $a = [a^1, \ldots, a^m]^{\mathrm{T}}$. *These estimates are obtained by minimizing the penalty function* Q_n *of* (3.1), *and they are jointly asymptotically normal.*

Proof. The system of equations $\partial Q_n/\partial a^j = 0$, $j = 1, \ldots, m$, is linear in $a^1, \ldots, a^m$, and it is easily verified that Q_n is minimized by taking

$$\hat{a}_n^j = \frac{\sum_{t=2}^{n} X_t X_{t-1} H_j(X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})} \tag{7.2}$$

where this exists with probability one since $E\{X_t^2 H_j(X_t)\} \geq \alpha_j$.

Using (7.1) and the independence of the e_t's we have

$$\tilde{X}_{t|t-1} = \sum_{j=1}^{m} a^j X_{t-1} H_j(X_{t-1}) \quad \text{and} \quad \frac{\partial \tilde{X}_{t|t-1}}{\partial a^j} = X_{t-1} H_j(X_{t-1}), \tag{7.3}$$

while higher order derivatives are zero. Also, it is easily shown that $f_{t|t-1} = E\{(X_t - \tilde{X}_{t|t-1})^2 | \mathscr{F}_{t-1}^X\} = E(e_t^2) = \sigma^2$. Since $\partial \tilde{X}_{t|t-1}/\partial a^j$ does not depend on a^k, $k = 1, \ldots, m$, it follows that CN2 and CN4 of Theorem 6.1 are trivially fulfilled. Moreover, using $|a^j| < 1$, $j = 1, \ldots, m$, $E(X_1^2) < \infty$ and $E(e_t^2) < \infty$, it follows from (7.1) that $E(X_t^2) \leq K$ for some constant K, and that CN1 of Theorem 6.1 holds.

From the special structure of the derivatives given in (7.3) we have that the matrix A^n in (6.1) in the present case is a diagonal matrix and is given by

$$A^n = \operatorname{diag}\left\{\frac{1}{n} \sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})\right\} \tag{7.4}$$

and using the assumption $E\{X_t^2 H_j(X_t)\} \geq \alpha_j$ we have that CN3 of Theorem 6.1 will be fulfilled if we can prove that

$$\frac{1}{n}\sum_{t=1}^{n} X_t^2 H_i(X_t) - \frac{1}{n}\sum_{t=1}^{n} E\{X_t^2 H_i(X_t)\} \xrightarrow{\text{a.s.}} 0 \tag{7.5}$$

for $i = 1, \ldots, m$. This is proved (Tjøstheim, 1984b) by exploiting that $n^{-1}\sum_{t=1}^{n} e_t^2 - \sigma^2 \xrightarrow{\text{a.s.}} 0$, and by using the martingale strong law on the martingale difference sequence $\{e_t \sum_{j=1}^{m} a^j X_{t-1} H_j(X_{t-1})\}$.

Turning now to the proof of asymptotic normality, it is not difficult to verify that the matrix R_n defined in (6.5) in the present case is given by

$$R_n = \sigma^2 \operatorname{diag}\left[\sum_{t=2}^{n} E\{X_{t-1}^2 H_j(X_{t-1})\}\right], \tag{7.6}$$

and using the assumption $E\{X_{t-1}^2 H_j(X_{t-1})\} \geq \alpha_j$ for $j = 1, \ldots, m$ it follows at once that DN1 of Theorem 6.2 is fulfilled. Moreover, the matrix in DN2 is seen to be given by

$$D_n \stackrel{\Delta}{=} \frac{1}{\sigma^2} \operatorname{diag}\left[\frac{\sum_{t=2}^{n} e_t^2 X_{t-1}^2 H_j(X_{t-1})}{E\{\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})\}}\right]. \tag{7.7}$$

Since $E(e_t^4) < \infty$ and $|a^j| < 1$, $j = 1, \ldots, m$, there exists a K_2 such that $E\{X_{t-1}^4 H_j(X_{t-1})\} \leq K_2$ for all j and t, and thus, using that e_t is independent of $\mathscr{F}_{t-1}^X$, we have $E[\{e_t^2 X_{t-1}^2 H_j(X_{t-1})\}^2] \leq K_2 E(e_t^4)$. From the martingale strong law applied to the martingale difference sequence $\{e_t^2 X_{t-1}^2 H_j(X_{t-1}) - \sigma^2 X_{t-1}^2 H_j(X_{t-1})\}$ it follows that

$$\frac{1}{n}\sum_{t=2}^{n} e_t^2 X_{t-1}^2 H_j(X_{t-1}) - \frac{\sigma^2}{n}\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1}) \xrightarrow{\text{a.s.}} 0. \tag{7.8}$$

Using (7.5) and an addition-subtraction argument in (7.7) it follows that $D_n \xrightarrow{\text{a.s.}} I_m$ as $n \to \infty$, and thus, from Theorem 6.2,

$$\operatorname{diag}\left(\frac{\sum_{t=2}^{n} X_{t-1}^2 H_j(X_{t-1})}{\sigma[\sum_{t=2}^{n} E\{X_{t-1}^2 H_j(X_{t-1})\}]^{1/2}}\right)(\hat{a}_n - a) \xrightarrow{\text{d}} \mathcal{N}(0, I_m). \quad \square \tag{7.9}$$

It is not difficult to check that the above proof applies to the case where the e_t's are independent and zero-mean, and where $m \leq E(e_t^2) \leq M$ and $E(e_t^4) \leq M'$ for some positive constants m, M and M'. It should also be noted that a similar nonstationary generalization can be made for the exponential autoregressive model treated in Section 4.1.

An example where the condition $E\{X_t^2 H_j(X_t)\} \geq \alpha_j$ is satisfied, is given in Tjøstheim (1984b).

7.2. Random coefficient autoregressive processes

We assume that $\{X_t\}$ is given on $-\infty < t < \infty$ by

$$X_t - (a + b_t)X_{t-1} = e_t \tag{7.10}$$

where $\{e_t\}$ and $\{b_t\}$ are zero-mean independent processes each consisting of independent variables such that $m_1 \leq E(e_t^2) \leq M_1$ and $E(b_t^2) \leq M_2$, where m_1, M_1 and M_2 are positive constants such that $a^2 + M_2 < 1$. These conditions guarantee that there exists a $\mathscr{F}_t^b \vee \mathscr{F}_t^e$-measurable solution of (7.10) with uniformly bounded second moments. This solution can be expressed as

$$X_t = \sum_{i=0}^{\infty} a_{ti} e_{t-i} \tag{7.11}$$

with $a_{ti} = \prod_{j=0}^{i-1} (a + b_{t-j})$ and where by definition $a_{t0} = 1$.

We consider the problem of estimating the parameter a. Since $\tilde{X}_{t|t-1} = aX_{t-1}$, it is clear that there is a unique solution to $\partial Q_n / \partial a = 0$ with Q_n as in (3.1), namely $\hat{a}_n = (\sum_{t=2}^n X_t X_{t-1}) / (\sum_{t=2}^n X_{t-1}^2)$ assuming that observations $(X_1, \ldots, X_n)$ are available. The following theorem is proved in Tjøstheim (1984b).

Theorem 7.2. *Let $\{X_t\}$ be as above. If in addition $E(X_t^4) \leq K$ for some constant K, then $\hat{a}_n \to a$. Moreover, if we also have $E(e_t^8) \leq C_1$ and $E\{(a + b_t)^8\} \leq C_2 < 1$ for two constants C_1 and C_2, then $\hat{a}_n$ is asymptotically normal.*

The main ingredients in the proof are use of Theorems 6.1 and 6.2, the martingale strong law and the mixingale convergence theorem (cf. Hall and Heyde, 1980, Th. 2.21).

7.3. Doubly stochastic processes

Random coefficient autoregressive processes are special cases of what we have termed doubly stochastic time series models in Tjøstheim (1985, a, b). In the simplest first order case these are given by

$$X_t = \theta_t X_{t-1} + e_t \tag{7.12}$$

where $\{a + b_t\}$ of (7.10) now is replaced by a more general stochastic process $\{\theta_t\}$. The process $\{\theta_t\}$ is usually assumed to be independent of $\{e_t\}$ and to be generated by a separate mechanism. Thus $\{\theta_t\}$ could be a Markov chain or it could itself be an AR process. We refer to Tjøstheim (1985a, b) for a definition and properties in the general case.

For the case where $\{\theta_t\}$ is an ARMA process, there is a close connection with Kalman type dynamic state space models (cf. Harrison and Stevens, 1976, Ledolter, 1981, and Tjøstheim, 1985b). This type of processes has attracted considerable attention lately, and there exist procedures (see e.g. Ledolter, 1981) for computation of unknown parameters, but as far as we know there are no results available concerning the properties of these estimates.

We have only considered a very special case, namely the case where $\{\theta_t\}$ is a first order MA process given by

$$\theta_t = a + \varepsilon_t + b\varepsilon_{t-1}, \tag{7.13}$$

where $\{\varepsilon_t\}$ consists of zero-mean iid random variables independent of $\{e_t\}$ and with $E(\varepsilon_t^2) < \infty$. Both $\{e_t\}$ and $\{\varepsilon_t\}$ will be assumed to be defined on $-\infty < t < \infty$.

To be able to construct Kalman-like algorithms for the predictor $\tilde{X}_{t|t-1}$, the process $\{X_t\}$ must be conditional Gaussian and this requires (Tjøstheim, 1985b) that $\{e_t\}$ and $\{\varepsilon_t\}$ be Gaussian, and that there is an initial variable X_0 such that the conditional distribution of θ_0 given X_0 is Gaussian. This last requirement is achieved here by choosing $X_0 = 0$. Obviously it implies that $\{X_t\}$ is nonstationary.

Theorem 7.3. *Let $\{X_t, t \geq 1\}$ be given by* (7.12) *and* (7.13) *under the above stated assumptions. Assume that $E(X_t^4) \leq K$ for some constant K, and that the MA parameter b is less than $\frac{1}{2}$ in absolute value. Then there exists a sequence of estimators $\{\hat{a}_n\}$ such that $\hat{a}_n \xrightarrow{\text{a.s.}} a$ as $n \to \infty$, such that $\hat{a}_n$ is obtained by minimization of Q_n in* (3.1) *as described in the conclusion of Theorem* 6.1.

The proof is given in Tjøstheim (1984b). It makes use of the same techniques as for Theorem 7.2 in addition to recursive relationships for the conditional mean and the conditional covariance of ε_t given $\mathcal{F}_{t-1}^X$.

7.4. *Some other problems*

In Tjøstheim (1984b) it is shown that also autoregressive models with deterministic time varying coefficients can in certain cases be treated within our framework. Moreover, as in Section 5, it is possible to introduce a maximum likelihood type penalty function in the nonstationary case. Again it can be shown (Tjøstheim, 1984b) that consistency and asymptotic normality can then be obtained under weaker moment conditions at least for the examples treated in Sections 7.2 and 7.3.

Acknowledgement

This work was carried out while the author was visiting the Center for Stochastic Processes, Department of Statistics, University of North Carolina. It is a pleasure to acknowledge stimulating discussions with Professor M. Pourahmadi on the problem of existence of stationary solutions.

References

K.K. Aase, Recursive estimation in non-linear time series models of autoregressive type, J. Roy. Stat. Soc. Ser. B 45 (1983) 228–237.

J. Ăndel, Autoregressive series with random parameters, Math. Operationsforsch. u. Statist. 7 (1976) 735–741.

P. Billingsley, The Lindeberg–Levy theorem for martingales, Proc. Amer. Math. Soc. 12 (1961) 788–792.

M. Bhaskara Rao, T. Subba Rao and A.M. Walker, On the existence of strictly stationary solutions to bilinear equations, J. Time Series Anal. 4 (1983) 95–110.

C.W.J. Granger and A.P. Andersen, An Introduction to Bilinear Time Series Models (Vanderhoeck and Ruprecht, Gøttingen, 1978).
D. Guegan, Une condition d'ergodicité pour des modelès bilinéaires à temps discret, C.R. Acad. Sc. Paris 297 (1983) 537-540.
V. Haggan and T. Ozaki, Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, Biometrika 68 (1981) 189-196.
P. Hall and C.C. Heyde, Martingale Limit Theory and its Application (Academic Press, New York, 1980).
P.J. Harrison and C.F. Stevens, Bayesian forecasting (with discussion), J. Roy. Stat. Soc. Ser. B 38 (1976) 205-248.
D.A. Jones, Non-linear autoregressive processes, Proc. Roy. Soc. London A, 360 (1978) 71-95.
L.A. Klimko and P.I. Nelson, On conditional least squares estimation for stochastic processes, Ann. Statist. 6 (1978) 629-642.
J. Ledolter, Recursive estimation and forecasting in ARIMA models with time varying coefficients, in: D.F. Findlay, ed., Applied Time Series Analysis II (Academic Press, New York (1981) 449-471.
D.F. Nicholls and B.G. Quinn, Random Coefficient Autoregressive Models: An Introduction, Lecture Notes in Statistics 11 (Springer, New York, 1982).
T. Ozaki, Non-linear time series models for non-linear random vibrations, J. Appl. Probab. 17 (1980) 84-93.
M.B. Priestley, State-dependent models: A general approach to time series analysis, J. Time Series Anal. 1 (1980) 47-71.
P.M. Robinson, The estimation of a nonlinear moving average model, Stoch. Proc. Appl. 5 (1977) 81-90.
T. Subba Rao and M.M. Gabr, An Introduction to Bispectral Analysis and Bilinear Time Series Models, Lecture Notes in Statistics 24 (Springer, New York, 1984).
D. Tjøstheim, Estimation in nonlinear time series models I: Stationary Series, Technical Report 70, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514 (1984a).
D. Tjøstheim, Estimation in nonlinear time series models II: Some nonstationary series, Technical Report 71, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514 (1984b).
D. Tjøstheim, Recent developments in nonlinear time series modelling, The John B. Thomas Birthday Volume of Papers on Communications, Networks and Signal Processing (Springer, New York, 1985a).
D. Tjøstheim, Some doubly stochastic time series models, to appear in J. Time Series Analysis (1985b).
H. Tong, Threshold Models in Nonlinear Time Series Analysis, Lecture Notes in Statistics, 21 (Springer, New York, 1983).
H. Tong and K.S. Lim, Threshold autoregression, limit cycles and cyclical data (with discussion), J. Roy. Stat. Soc. Ser. B 42 (1980) 245-292.
R.L. Tweedie, Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space, Stoch. Proc. Appl. 3 (1975) 385-403.