STATISTICAL REPORT

ESTIMATION OF AR PARAMETERS IN TIME SERIES WITH SUDDENLY CHANGING STRUCTURE

BY

GYRID JOHNSEN AND DAG TJØSTHEIM

REPORT NO 19 JULY 1988

Department of Mathematics

UNIVERSITY OF BERGEN

Bergen, Norway

Department of Mathematics
University of Bergen
Allégt. 55
5007 Bergen
NORWAY

ISSN-0333-1865

ESTIMATION OF AR PARAMETERS IN TIME SERIES WITH SUDDENLY CHANGING STRUCTURE

BY

GYRID JOHNSEN AND DAG TJØSTHEIM

REPORT NO 19 JULY 1988

Abstract

We study autoregressive (AR) time series with suddenly changing structure. There are two interrelated estimation problems associated with this model: The estimation of shift points and the estimation of AR parameters. In this paper we study the properties of the AR parameter estimates. We prove consistency and asymptotic normality when the shift points are known. When the shift points are unknown, the parameter estimates will in general be biased, and we find an approximate expression for the bias in a simple situation. The results are checked by simulations.

Keywords

Asymptotic normality, Autoregressive, Consistency, Doubly stochastic, Shift point, Sudden change.

1 Introduction

In this paper we will study autoregressive (AR) time series with suddenly changing parameters. Such changes may be due to external events, e.g. political or stock market events influencing an economic time series (Tyssedal and Tjøstheim 1988) or a geophysical time series obtained from layered geological formations (Karlsen and Tjøstheim 1988). Various aspects of such phenomena have been considered by Maddala (1986), Millnert (1982), Picard (1985), Sclove (1983), Telknys (1986) and Wichern et al (1976), and more background material has been given in Tyssedal and Tjøstheim (1988). Quite often it is reasonable to assume that the external events themselves are regulated by a random mechanism leading to a doubly stochastic time series model (Tjøstheim 1986a). Our models are also related to the threshold models of Tong (1983), but they lack the feedback mechanism from past observations inherent in those models.

Although our results can be generalised to an AR(p) model, for simplicity we will only treat AR(1) models $\{X_t\}$ given by

$$X_t = \theta_t X_{t-1} + e_t \quad , t \geq 1 \ , \ X_0 = x_0 \tag{1.1}$$

Here $\{e_t \ , \ t \geq 1\}$ is a sequence of independent identically distributed (iid) random variables independent of X_0, and $\{\theta_t \ , \ t \geq 1\}$ is a deterministic sequence or a stochastic process. In either case each θ_t is only allowed to take k values $a_1, ..., a_k$, corresponding to k possible states for the correlation structure of $\{X_t\}$.

There are two interrelated estimation problems associated with this model: The estimation of shift points, where θ_t jumps from one value to another, and the estimation of the AR parameters $\{a_1, ..., a_k\}$. The nature of these problems are quite different. In general the shift points are stochastic variables, and, in the absence of external information, they cannot be estimated asymtotically with an arbitrary preselected level of accuracy. The estimation of the AR parameters, however, can be phrased in more traditional terms.

Some theoretical aspects of shift point estimation has been discussed in Telknys (1986), but as far as we know the properties of the AR estimates in this situation are unknown, and the main concern of this paper is to try to establish such properties. In section 2 the estimation problem will be

discussed under the assumption that the shift points are known, which is not very realistic unless they are generated by a known external mechanism. In section 3 we adopt the somewhat more realistic attitude that the shift points are unknown, but that we know a probability distribution for them. This distribution could have been furnished subjectively by an expert, or it could be the result of an estimation procedure for the shift points. The distribution of shift point estimates is not easy to derive, but the interested reader may consult Telknys (1986), where a few of the papers are dealing with this problem.

As far as applications are concerned, it is important to have good estimates of $\{a_1, ..., a_k\}$ to be able to identify and distinguish between various correlation or frequency structures. For seismic or other geophysical time series, say, one may try to link such a characterization directly to geological properties by labelling geological layers. (cf Karlsen and Tjøstheim 1988)

2 The shifts are known

We let $\{X_t\}$ be the process generated by (1.1), and we denote by δ_{it} , $i = 1, ..., k$, the ith state indicator process given by

$$\delta_{it} = \begin{cases} 1 & \text{for } \theta_t = a_i \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

We assume that the indicator processes are known. This amounts to knowing the shift points and which states are involved at each shift.

With the above notation (1.1) can be written

$$X_t - \sum_{i=1}^{k} a_i \delta_{it} X_{t-1} = e_t \tag{2.2}$$

We assume in addition that $E(e_t) = 0$ and $E(e_t^2) = \sigma^2 < \infty$. Our task is to find estimates of the AR parameters $\{a_1, ..., a_k\}$ and to evaluate their properties. To this end let $\mathcal{F}_t$ be the σ-algebra generated by $\{X_s \ , \ s \leq t\}$. Then $\widetilde{X}_{t|t-1} \triangleq E(X_t \mid \mathcal{F}_{t-1}) = \sum_i a_i \delta_{it} X_{t-1}$, and given observations $X_0, ..., X_n$, the least squares estimates of $\hat{a}_i$, i=1,...,k , obtained by minimizing $\sum_t (X_t - \widetilde{X}_{t|t-1})^2$ are given by

$$\hat{a}_i = (\sum_{t=1}^{n} \delta_{it} X_t X_{t-1}) / (\sum_{t=1}^{n} \delta_{it} X_{t-1}^2) \tag{2.3}$$

for i=1,...,k. (We assume that $\sum_{t=1}^{n} \delta_{it} X_{t-1}^2 > 0$, i.e. all states are visited. Otherwise we can just omit the corresponding a_i in (2.3)).

Theorem 2.1 *Let $\{X_t\}$ and $\{\hat{a}_i \ , \ i = 1, ..., k\}$ be as defined above. If*

$$|a_i| < 1 \ and \ \lim_{n \to \infty} inf \ (n^{-1} \sum_{t=1}^{n} \delta_{it}) \ > \ 0 \ ,$$

then $\hat{a}_i - a_i \stackrel{a.s.}{\to} 0$ as $n \to \infty$ for $i = 1, ..., k$.

Proof : In general the process $\{X_t\}$ is nonstationary, and we cannot rely on the ergodic theorem. We can prove strong consistency using the general theorem of Tjøstheim (1986b), but in this case we may as well proceed directly using martingale theory.

Using (2.2) and (2.3) and the fact that $\delta_{it}\delta_{jt} = 0 \ for \ i \neq j \ and \ \delta_{it}^2 = \delta_{it}$, we have that

$$\hat{a}_i - a_i = (n^{-1} \sum_{t=1}^{n} \delta_{it} e_t X_{t-1}) / (n^{-1} \sum_{t=1}^{n} \delta_{it} X_{t-1}^2) \tag{2.4}$$

From the iid property of $\{e_t\}$ we have $E(\delta_{it} e_t X_{t-1} \mid \mathcal{F}_{t-1}) \stackrel{a.s.}{=} 0$, and the process $\sum_t \delta_{it} e_t X_{t-1}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t\}$. Similarly conditionally on $\{\delta_{it}\}$ being known, $E\{(\delta_{it} e_t X_{t-1})^2\} = \delta_{it} \sigma^2 E(X_t^2)$. It follows from the defining equation (1.1) that

$$X_t = (\prod_{j=0}^{t} \theta_{t-j}) X_0 + e_t + \sum_{s=1}^{t-1} (\prod_{j=0}^{s-1} \theta_{t-j}) e_{t-s} \ , \tag{2.5}$$

and using the iid property of $\{e_t\}$ and the independence of X_0 we have

$$E(X_t^2) \leq \sigma^2 \{1 + \sum_{s=1}^{t-1} (\max_i |a_i|)^s\} + (\max_i |a_i|)^t E(X_0^2). \tag{2.6}$$

Since $\max_i |a_i| < 1$, there exists an $M > 0$ such that $E\{(\delta_{it} e_t X_{t-1})^2\} = \delta_{ti} \sigma^2 E(X_t^2) \leq M$ and it follows from the martingale convergence theorem (Stout 1974, Th. 3.3.8) that

$$n^{-1}\sum_{t=1}^{n}\delta_{it}e_tX_{t-1} \xrightarrow{a.s.} 0 \tag{2.7}$$

as $n \to \infty$ for i=1,...,k.

It remains to prove that

$$\lim_{n\to\infty} inf\ (n^{-1}\sum_{t=1}^{n}\delta_{it}X_{t-1}^2) \overset{a.s}{>} 0 \tag{2.8}$$

as $n \to \infty$ in (2.4). From the strong law of large numbers

$$n^{-1}\sum_{t=1}^{n}e_t^2 - \sigma^2 \xrightarrow{a.s.} 0 \tag{2.9}$$

Inserting from (2.2) we have

$$e_t^2 = X_t^2 - 2\sum_{i=1}^{k}a_i\delta_{it}X_tX_{t-1} + \sum_{i=1}^{k}a_i^2\delta_{it}X_{t-1}^2 \tag{2.10}$$

and

$$\sum_{i=1}^{k}a_i\delta_{it}X_tX_{t-1} - \sum_{i=1}^{k}a_i^2\delta_{it}X_{t-1}^2 = e_t\sum_{i=1}^{k}a_i\delta_{it}X_{t-1} \triangleq u_t \tag{2.11}$$

It is easily checked that $\{u_t\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t\}$ and since $E(u_t^2) = \sigma^2\sum_i a_i^2\delta_{it}E(X_{t-1}^2) \leq K$ for some $K > 0$, it follows from the strong law for martingales that $n^{-1}\sum u_t \xrightarrow{a.s.} 0$. Inserted in (2.9) and (2.10) this yields

$$n^{-1}\sum_{t=1}^{n}X_t^2 - n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{k}a_i^2\delta_{it}X_{t-1}^2 - \sigma^2 \xrightarrow{a.s.} 0 \tag{2.12}$$

Since $E(X_t^2) \leq M$, implies $n^{-1}X_n^2\delta_{i,n+1} \xrightarrow{a.s.} 0 \ \ as \ \ n \to \infty$, and since $\sum_i \delta_{it} = 1$ for all t, an alternative way of writing (2.12) is

$$\sum_{i=1}^{k}(1-a_i^2)n^{-1}\sum_{t=1}^{n}\delta_{it}X_{t-1}^2 - \sigma^2 \xrightarrow{a.s.} 0 \tag{2.13}$$

Taking expectations in (2.10) and (2.11) and using the same reasoning as above we have

$$\sum_{i=1}^{k}(1-a_i^2)n^{-1}\sum_{t=1}^{n}\delta_{it}E(X_{t-1}^2)-\sigma^2\to 0 \tag{2.14}$$

Combining the two we obtain

$$\sum_{i=1}^{k}(1-a_i^2)\left[n^{-1}\sum_{t=1}^{n}\delta_{it}X_{t1}^2-n^{-1}\sum_{t=1}^{n}\delta_{it}E(X_{t-1}^2)\right]$$
$$\triangleq \sum_{i=1}^{k}(1-a_i^2)Y_{in}\overset{a.s.}{\to}0 \tag{2.15}$$

Since $\max_i|a_i|<1$ and since the variables Y_{in} , $i=1,...,k$, are linearly independent , it follows that

$$n^{-1}\sum_{t=1}^{n}\delta_{it}X_{t-1}^2-n^{-1}\sum_{t=1}^{n}\delta_{it}E(X_{t-1}^2)\overset{a.s.}{\to}0 \tag{2.16}$$

for i=1,...,k. From (2.5) it is easily proved that $E(X_{t-1}^2)\geq\sigma^2$ for $t\geq 2$ and (2.8) follows from the persistency assumption on δ_{it} in the theorem and from (2.16).

Next we turn to the asymptotic distribution.

Theorem 2.2 *Let $\{X_t\}$ and $\{a_i$, $i=1,...,k\}$ be as in Theorem 2.1 and let a be the column vector defined by $a=[a_1,...,a_k]^T$. Moreover, let I_k be the identity matrix of dimension k and $diag(\cdot)$ the diagonal matrix. If the asumptions of Theorem 2.1 hold and in addition $E(e_t^4)<\infty$, then as $n\to\infty$*

$$diag(\frac{1}{\sigma}[\sum_{t=1}^{n}\delta_{it}E(X_{t-1}^2)]^{1/2})(\hat{a}-a)\overset{d}{\to}\mathcal{N}(0,I_k) \tag{2.17}$$

where $\hat{a}=[\hat{a}_1,...,\hat{a}_k]^T$ with $\hat{a}_i$ given by (2.3).

<u>*Proof*</u> : Using (2.16), it is sufficient to prove that

$$C_n\triangleq diag\left(\frac{\sum_{t=1}^{n}\delta_{it}X_{t-1}^2}{\sigma\{\sum\delta_{it}E(X_{t-1}^2)\}^{1/2}}\right)(\hat{a}-a)\overset{d}{\to}\mathcal{N}(0,I_k) \tag{2.18}$$

From (2.4) we have that the ith component C_{ni} of C_n is given by

$$C_{ni}=(\sum_{t=1}^{n}\delta_{it}e_tX_{t-1})/[\sigma\{\sum_{t=1}^{n}\delta_{it}E(X_{t-1}^2)\}^{1/2}] \tag{2.19}$$

We use a Cramer-Wold argument. For k arbitrary real numbers $\alpha_1, ..., \alpha_k$ it is then sufficient to prove

$$\sum_{i=1}^{k} \alpha_i C_{ni} \xrightarrow{d} \mathcal{N}(0, \sum_{i=1}^{k} \alpha_i^2) \tag{2.20}$$

For this purpose we introduce $F_{ni} \triangleq \sigma\{\sum_{t=1}^{n} \delta_{it} E(X_{t-1}^2)\}^{1/2}$ and

$$D_{nt} \triangleq \sum_{i=1}^{k} \alpha_i F_{ni}^{-1} \delta_{it} e_t X_{t-1}$$

With our assumptions on $\{\delta_{it}\}$ and a_i it is not difficult to show that $\{D_{nt}\}$ are martingale increments for a zero-mean square integrable martingale array. It is then sufficient to verify the following conditions (cf. Hall and Heyde 1980, Th. 3.2, where the nesting and integrability conditions of that theorem are trivially fulfilled).

(i) $\max_{1 \le t \le n} |D_{nt}| \xrightarrow{p} 0$

(ii) $\sum_{t=1}^{n} D_{nt}^2 \xrightarrow{p} \sum_{i=1}^{k} \alpha_i^2$

(iii) $E(\max_{1 \le t \le n} D_{nt}^2)$ *is bounded in* n

Using the technique of Hall and Heyde (1980, p.53), (i) is fulfilled if the Lindeberg condition

$$\sum_{t=1}^{n} E\{D_{nt}^2 1(|D_{nt}| > \epsilon)\} \to 0 \tag{2.21}$$

holds for all $\epsilon > 0$. Here $1(\cdot)$ is the indicator function. Since $\delta_{it}\delta_{jt} = 0$ for $i \neq j$ and $\delta_{it}^2 = \delta_{it}$, the left hand side of (2.21) equals

$$\sum_{i=1}^{k} \alpha_i^2 F_{ni}^{-2} \sum_{t=1}^{n} \delta_{it} E\left\{ e_t^2 X_{t-1}^2 1(\sum_{i=1}^{k} \alpha_i^2 F_{ni}^{-2} \delta_{it} e_t^2 X_{t-1}^2 > \epsilon^2) \right\}. \tag{2.22}$$

Since $\max_i |a_i| < 1$, we have

$$E\{e_t^2 X_{t-1}^2\} \le \sigma^2 E\{X_{t-1}^2\} \le \sigma^4 \{1 - \max_i (a_i^2)\}^{-1} \tag{2.23}$$

Further, using the persistency condition on δ_{it} , as n gets large

$$F_{ni}^2 = \sigma^2 \sum_{t=1}^{n} \delta_{it} E(X_{t-1}^2) \geq n\sigma^2 m \tag{2.24}$$

for some $m > 0$. It follows that for a given δ there is an n_0 such that for $n > n_0$ and all t

$$\delta_{it} E\{e_t^2 X_{t-1}^2 1(\sum_{i=1}^{k} \alpha_i^2 F_{ni}^{-2} \delta_{it} e_t^2 X_{t-1}^2 > \epsilon^2)\} < \delta \tag{2.25}$$

The relationship (2.21) now follows from (2.22)-(2.25) using standard arguments and (i) is proved.

Since $E(e_t^4) < \infty$ and $|a_i| < 1$, i=1,...,k , the expansion (2.5) implies that there exists a $K > 0$ such that $E\{(\delta_{it} e_t^2 X_{t-1}^2)^2\} = \delta_{it} E(e_t^4) E(X_{t-1}^4) \leq K$. It is easy to check that

$$D_{nt}^2 = \sum_{i=1}^{k} \alpha_i^2 F_{ni}^{-2} \delta_{it} e_t^2 X_{t-1}^2 \tag{2.26}$$

Hence (2.24), the independence of e_t from $\mathcal{F}_{t-1}$ and the definition of δ_{it} implies

$$\begin{aligned} &E(\sum_{t=1}^{n} D_{nt}^2 - \sum_{i=1}^{k} \alpha_i^2 F_{ni}^{-2} \sum_{t=1}^{n} \delta_{it} \sigma^2 X_{t-1}^2)^2 \\ &= \sum_{i=1}^{k} \alpha_i^4 F_{ni}^{-4} E(\sum_{t=1}^{n} \delta_{it} (e_t^2 - \sigma^2) X_{t-1}^2)^2 \\ &\qquad \leq n^{-1} \sigma^{-4} m^{-2} K \sum_{i=1}^{k} \alpha_i^4 \ , \end{aligned} \tag{2.27}$$

which tends to zero as n tends to infinity. On the other hand, reasoning exactly as in the proof of Theorem 2.1, corresponding to (2.16), we have

$$\sum_{i=1}^{k} \alpha_i^2 F_{ni}^{-2} \left[\sum_{t=1}^{n} \delta_{it} \sigma^2 X_{t-1}^2 - \sum_{t=1}^{n} \delta_{it} \sigma^2 E(X_{t-1}^2) \right] \xrightarrow{a.s.} 0 \tag{2.28}$$

But using (2.24) and (2.26) we have

$$\sum_{t=1}^{n} E(D_{nt}^2) = \sum_{i=1}^{k} \alpha_i^2 F_{ni}^{-2} \sum_{t=1}^{n} \delta_{it} \sigma^2 E(X_{t-1}^2) = \sum_{i=1}^{k} \alpha_i^2 \ , \tag{2.29}$$

and from (2.29), (2.28) and (2.27) combined with Chebyshev's inequality it follows that (ii) is fulfilled.

Finally,

$$E(\max_{1\le t\le n} D_{nt}^2) \le E(\sum_{t=1}^{n} D_{nt}^2) = \sum_{i=1}^{k} \alpha_i^2 \ , \tag{2.30}$$

from wich (iii) follows, and the theorem is proved.

It should be noted that we have asymptotic independence of the estimates $\hat{a}_i$, i=1,...,k , in the sense that the asymptotic covariance matrix is diagonal. For k=1, (2.17) reduces to the familiar result $\sigma^{-1}\{nE(X_t^2)\}^{1/2}(\hat{a}_1 - a_1) \xrightarrow{d} \mathcal{N}(0,1)$ valid in the ordinary AR(1) case.

Due to Theorems 2.1 and 2.2 and (2.16) it is clear that $Var(\hat{a}_i)$ can be estimated by

$$\widehat{Var(\hat{a}_i)} = \hat{\sigma}^2/(\sum_{t=1}^{n} \delta_{it} X_{t-1}^2) \tag{2.31}$$

where $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}(X_t - \sum_i \hat{a}_i \delta_{it} X_{t-1})^2$. Finally it should be noted that the conditions of Theorems 2.1 and 2.2 can be relaxed. For example it is not necessary to require that the e_t's are identically distributed. If σ^2 is replaced by $n^{-1}\sum_{t=1}^{n} E(e_t^2)$, it is not difficult to check that the general martingale arguments presented in the proof of Theorem 2.1 hold if the e_t's are independent with $E(e_t) = 0$ and $m \le E(e_t^2) \le M$ for two positiv constants m and M. This means that we may have e.g. step changes in the residual variance as well, thus generalizing the result of Tyssedal and Tjøstheim (1982) where only the variance and not the AR coefficients where allowed to change. If $E(e_t^4)$ is bounded one obtains asymptotic normality as well.

3 The shifts are unknown

We will limit ourselves to a model where $\{\theta_t\}$ of (1.1) is a stationary ergodic process independent of $\{e_t\}$ and taking only two values $\{a_1,\ a_2\}$. In addition $\{\theta_t\}$, $\{e_t\}$ and $\{X_t\}$ of (1.1) will be assumed given on $-\infty < t < \infty$. The time points where $\{\theta_t\}$ changes from one value to another will be denoted by $...T_{-1} < T_1 < T_2 < ...$, where $T_1 \ge 1$. Contrary to the situation in

Section 2 we do not know the change points, but we assume that estimates $\{\hat{T}_i\}$ are available, and that these estimates are related to the true change points by a time- invariant symmetric probability distribution of range M, depending only on the sequence $\{T_i\}$, such that

$$Pr(\hat{T}_i = T_i + k) = Pr(\hat{T}_i = T_i - k) = p_k \ , \quad 0 \leq k \leq M \ , \tag{3.1}$$

with $p_0 + 2\sum_1^M p_k = 1$.

If $|T_{i+1} - T_i| \leq 2M$, it is possible to have $\hat{T}_{i+1} \leq \hat{T}_i$. We assume that such a pair of division points (T_i, T_{i+1}) will not be detected with probability $Pr(\hat{T}_{i+1} \leq \hat{T}_i) = Pr(|T_{i+1} - T_i| \leq M)Pr(\hat{T}_{i+1} < \hat{T}_i \ \mid \ |T_{i+1} - T_i| \leq M)$. This results in a revised stationary sequence of estimates $\{\hat{T}_i'\}$, where some of the original T_i's may be missing. In the following we will omit the prime in our notation.

The sequence $\{\hat{T}_i\}$ leads to an estimated state indicator process given by

$$\hat{\delta}_{it} = \begin{cases} 1 & \text{for } \hat{\theta}_t = a_i \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

for i=1,2. In general the distribution $\{p_k\}$ in (3.1) will be unknown. In practice it will have to be assigned subjectively from a priori belief or by using the properties (possibly evaluated by simulation) of the shift point estimation method as a guide.

Since we do not know $\{\delta_{it}\}$, the least squares estimates (2.3) cannot be used, but by simple analogy we introduce

$$\hat{a}_i = (\sum_{t=1}^{n} \hat{\delta}_{it} X_t X_{t-1}) / (\sum_{t=1}^{n} \hat{\delta}_{it} X_{t-1}^2) \ . \tag{3.3}$$

Inserting from the defining equation (2.2) we have, letting i=1,2 and j=2,1

$$X_t = \{a_i(1 - \delta_{jt}) + a_j \delta_{jt}\} X_{t-1} + e_t \tag{3.4}$$

and thus

$$\begin{aligned} \hat{a}_i - a_i &= \frac{(a_j - a_i) \sum_{t=1}^n \delta_{jt} \hat{\delta}_{it} X_{t-1}^2}{\sum_{t=1}^n \hat{\delta}_{it} X_{t-1}^2} + \frac{\sum_{t=1}^n \hat{\delta}_{it} e_t X_{t-1}}{\sum_{t=1}^n \hat{\delta}_{it} X_{t-1}^2} \\ &= B_{ni} + S_{ni} \end{aligned} \tag{3.5}$$

where B_{ni} is a bias term due to the estimation error in $\widehat{\delta}_{it}$, and S_{ni} is a standard error term due to the error sequence $\{e_t\}$.

With our assumption on $\{\theta_t\}$ and $\{e_t\}$, the process $\{X_t, \widehat{\delta}_{it}\}$ is ergodic if $|a_i| < 1$, i=1,2 , with $\{X_t\}$ being represented by

$$X_t = e_t + \sum_{s=1}^{\infty}(\prod_{j=0}^{s-1} \theta_{t-j})e_{t-s} \tag{3.6}$$

We can now prove asymptotic normality for a bias adjusted version of $\widehat{a}_i$, i=1,2.

Theorem 3.1 *Let $\{X_t\}$, $\{\widehat{\delta}_{it}\}$, $\widehat{a}_i$ and S_{ni} i=1,2 be as defined in (3.4), (3.2), (3.3) and (3.5) respectively. Moreover, let $a = [a_1, a_2]^T$ and $B_n = [B_{n1}, B_{n2}]^T$. If $|a_i| < 1$ and $E(\delta_{it}X_{t-1}^2) > 0$, then*

$$diag\left(\sigma^{-1}\{E(\widehat{\delta}_{it}X_{t-1}^2)\}^{1/2}\sqrt{n}(\hat{a} - a - B_n)\right) \xrightarrow{d} \mathcal{N}(0, I_2).$$

<u>Proof:</u> Due to the ergodic theorem $n^{-1}\sum_{t=1}^{n} \widehat{\delta}_{it}X_{t-1}^2 \xrightarrow{a.s.} E(\widehat{\delta}_{it}X_{t-1}^2)$. This expectation is positiv since $E(\widehat{\delta}_{it}X_{t-1}^2) \geq p_0 E(\delta_{it}X_{t-1}^2) > 0$. Using (3.5) and a Cramer-Wold argument again it is then sufficient to prove

$$\sum_{i=1}^{2} \alpha_i \sigma^{-1}\{E(\widehat{\delta}_{it}X_{t-1}^2)\}^{-1/2} n^{-1/2} \sum_{t=1}^{n} \widehat{\delta}_{it}e_t X_{t-1} \xrightarrow{d} \mathcal{N}(0, \sum_{i=1}^{2} \alpha_i^2). \tag{3.7}$$

However, with the given assumption on $\{\widehat{\delta}_{it}\}$, we have $E(\widehat{\delta}_{it}e_t X_{t-1} \mid \mathcal{F}_{t-1}) \overset{a.s.}{=} 0$ and Billingsley's (1961 a) stationary martingale central limit theorem yields the conclusion.

Compared to Theorem 2.2 it should be noted that the assumption $E(e_t^4) < \infty$ is no longer needed. On the other hand we have assumed that $\{\widehat{\delta}_{it}\}$ is only depending on $\{T_i\}$ and not on $\{e_t\}$. Since

$$E(\delta_{it}e_t X_{t-1} \mid \mathcal{F}_{t-1}) = E\left[\delta_{it}X_{t-1}E(e_t \mid \mathcal{F}_{t-1}^x \vee \mathcal{F}_t^\theta) \mid \mathcal{F}_{t-1}^x\right] \overset{a.s.}{=} 0, \tag{3.8}$$

it is seen from the above proof that the independence of $\{\widehat{\delta}_{it}\}$ on $\{e_t\}$ may be replaced by the weaker assumtion

$$E\{(\widehat{\delta}_{it} - \delta_{it})e_t \mid \mathcal{F}_{t-1}\} \overset{a.s.}{=} 0 \tag{3.9}$$

It seems plausible, due to symmetry, that this assumption will hold for a number of shift point estimation methods.

Using the ergodic theorem we have under the conditions of Theorem 3.1

$$\hat{a}_i - a_i - B_{ni} \xrightarrow{a.s.} 0 \tag{3.10}$$

as n tends to infinity. However, our estimates will in general be biased, i.e. $E(B_{ni})$ does not tend to zero. To analyse this situation we assume in addition to ergodicity that $\{\theta_t\}$ is a Markov chain with transition matrix $Q = \{q_{ij}\}$, where $q_{ij} = P(\theta_{t+1} = a_j \mid \theta_t = a_i)$. The expected lenght of each visit in state i is $(1 - q_{ii})^{-1}$. Further, due to symmetry of the distribution (3.1), the expected number of terms of mismatch pr shiftpoint contributing to the term $\sum \delta_{it}\hat{\delta}_{it}X_{t-1}^2$ in the expression for B_{ni} in (3.5); i.e. E{ number of terms pr shiftpoint for which the true state is j , whereas the estimated state is i }, is given by $\sum_k kp_k$ Since $\{X_t\}$ is stationary it follows from (3.5) that in the long run the bias term for $\hat{a}_i$ will be approximated by

$$(a_j - a_i) \sum_{k=1}^{M} kp_k(1 - q_{ii}) \tag{3.11}$$

It is seen that if $a_i > a_j$, the bias term for $\hat{a}_i$ is negativ, while the opposite is true if $a_i < a_j$. This means that the estimated values will be closer than the true ones, but the bias decreases as the two values approach each other, so there is still a good possibility of distinguishing between them.

4 Simulation

Our results are asymptotic and it is reasonable to check them by simulations. This was done by generating a two-state ergodic Markov chain with a symmetric transition matrix $Q = \{q_{ij}\}$ for the parameter process $\{\theta_t\}$. This was done both in the case where $\{\delta_{it}\}$ is known, and where it is unknown with an estimation error governed by the distribution in (3.1) with M varying between 1 and 3. In each case the series $\{e_t\}$ was taken to be normally distributed random variables with zero mean and variance one. The process $\{e_t\}$ was generated by the random number generator NAG, and for each model 500 replicas of $\{X_t\}$ were generated with sample size varying from 250 to 4000.

Typical normal plots illustrating how the distribution approches normality is shown in Fig.1 for the case of $a_1 = 0.9$ and $a_2 = -0.3$ in (2.2), $p_0 = 0.5$ and $p_1 = 0.25$, in (3.1), and with transition matrix given by $q_{11} = q_{22} = 0.95$.

Bias and standard errors of the estimates are given in Tables 1-3. In Table 1, $\{\delta_{it}\}$ is known; i.e. $p_0 = 1$ in (3.1), in Table 2 $p_0 = 0.5$ and $p_1 = 0.25$,and in Table 3 various probability distributions with a fixed sample size of 1500 have been tried. Some of the examples of Table 3 are clearly very unreal, but they have been included to illustrate the formula in extreme situations with large skewness.

The correspondence between the observed simulated biases and those obtained using formula (3.11) is seen to be quite good in the case of small bias. There are larger disagreements for cases of large skewness. This is really not surprising since then there is a high percentage for which $\widehat{T_{i+1}} \leq \widehat{T_i}$ (in fact over 10% for the last example of Table 3). Since such shift points are removed in our estimation procedure, there will be a sizable portion of the estimated segments for each state which in fact consists of a mixture of the two states, and this will create an additional bias where the estimated coefficients are drawn against each other in the situation considered by us.

In Tyssedal and Tjøstheim (1988) is given an asymptotic expression for the standard error of $\hat{a}_i$, i=1,...,k , in the stationary Markov chain case where $\{\delta_{it}\}$ is known:

$$var(\hat{a}_i) \sim (n\pi_i)^{-1}\left[(1-a_i^2) + E\{(\delta_{i,t-1} - \delta_{it})X_{t-1}^2\}/E(\delta_{it}X_{t-1}^2)\right] \quad (4.1)$$

where π_i is the stationary probability of state i. In the two-state case $\pi_i = (1-q_{jj})(2-\sum_i q_{ii})^{-1}$, and $\pi_1 = \pi_2 = \frac{1}{2}$ if $q_{11} = q_{22}$. It is also shown in Tyssedal and Tjøstheim (1988) that for a slowly varying chain $\{\theta_t\}$, i.e. $q_{ii} - 1$ small, approximately

$$var(\hat{a}_i) \approx (n\pi_i)^{-1}(1-a_i^2), \quad (4.2)$$

and this is the formula used to obtain theoretical standard errors in Table 1, where q_{ii} is 0.9 or 0.95. It is seen that there is a very good correspondence for the low values of a_i, wheras the theoretical formula tends to underestimate the error for the large values of a_i.

Strictly speaking the above derivations from Tyssedal and Tjøstheim (1988) only refers to the case of a known $\{\delta_{it}\}$, but using Tables 2 and 3 it was found that it continues to hold to quite a good approximation in the case of $\{\delta_{it}\}$ unknown.

Finally it should be noted that a real data example concerning stock market data was given in Tyssedal and Tjøstheim (1988), whereas applications to oil well measurements were considered in Karlsen and Tjøstheim (1988). In those publications explicit methods for detecting shift points are given as well.

Acknowledgement

This research has been supported by a grant from VISTA- Statoil.

References

[1] Billingsley, P. (1961a): The Lindeberg-Levy theorem for martingales, Proc. Amer. Math. Soc. 12, 788-792.

[2] Billingsley, P. (1961b): Statistical Inference for Markov Processes, University of Chicago Press, Chicago.

[3] Hall, P. and Heyde, C.C. (1980): Martingale Limit Theory and Its Applications, Academic Press, New York.

[4] Karlsen, H. and Tjøstheim, D. (1988): Segmentation of data traces with applications to dipmeter oil well measurements, Statistical Report no. 18 , Department of Mathematics, University of Bergen.

[5] Maddala, G.S. (1986): Disequilibrium, self - selection and switching models, In Handbook of Econometrics, Vol. 3, edited by Z.Griliches and M.D. Intriligator, North Holland Amsterdam, 601-669

[6] Millnert, M (1982): Identification and control of systems subject to abrupt changes, Doctoral thesis, Department of Electrical Engineering, Linkøping University, Sweden.

[7] Picard, D. (1985): Testing and estimating change points in time series, Adv. Appl. Prob. 17, 841-867.

[8] Sclove, S.L. (1983): Time-series segmentation. A model and a method. Information Sciences, 29, 7-25.

[9] Stout, W.F. (1974): Almost Sure Convergence, Academic Press, New York.

[10] Telknys, L. (1986): Detection of changes in Random Processes, Optimization Software, Inc. Publications Division, New York.

[11] Tjøstheim, D. (1986a): Some doubly stochastic time series models, J. Time Series Analysis, 7, 51-72.

[12] Tjøstheim, D. (1986b): Estimation in nonlinear time series models, Stoch. Proc. Appl., 21, 251- 273.

[13] Tong, H. (1983): Threshold Models in Nonlinear Time Series Analysis, Lecture Notes in Statistics 21, Springer, New York.

[14] Tyssedal, J. and Tjøstheim, D. (1982): Autoregresive processes with a time dependent variance, J. Time Series Analysis 3, 209-217.

[15] Tyssedal, J. and Tjøstheim, D. (1988): An autoregressive model with suddenly changing parameters and an application to stock market prices, J. Roy. Stat. Soc. Ser. C: Applied Statistics, to appear.

[16] Wichern, D.W., Millnert, R.B. and Hsu, D.-A. (1976): Changes of variance in first order autoregressive time series models -with an application, J. Roy. Stat. Soc. Ser. C: Applied Statistics, 25, 248-256.

Figure captions

Figure 1: Normal plots of estimated values for the parameter $a_1 = 0.9$ with sample size a):250, b):500, c):1500. The state indicator process $\{\delta_{it}\}$ is unknown with $p_0 = 0.5$ and $p_1 = 0.25$ in (3.1). The normal plots are made by the program BMDP-P5D (Dixon W.J. et.al. (1983): BMDP Statistical Software). The estimated parameters $\hat{a}_1$ are plotted along the horizontal axis and the normal scores Y along the vertical axis.

Figur 1

Values from normal distribution would lie on the line indicated by the symbol / .

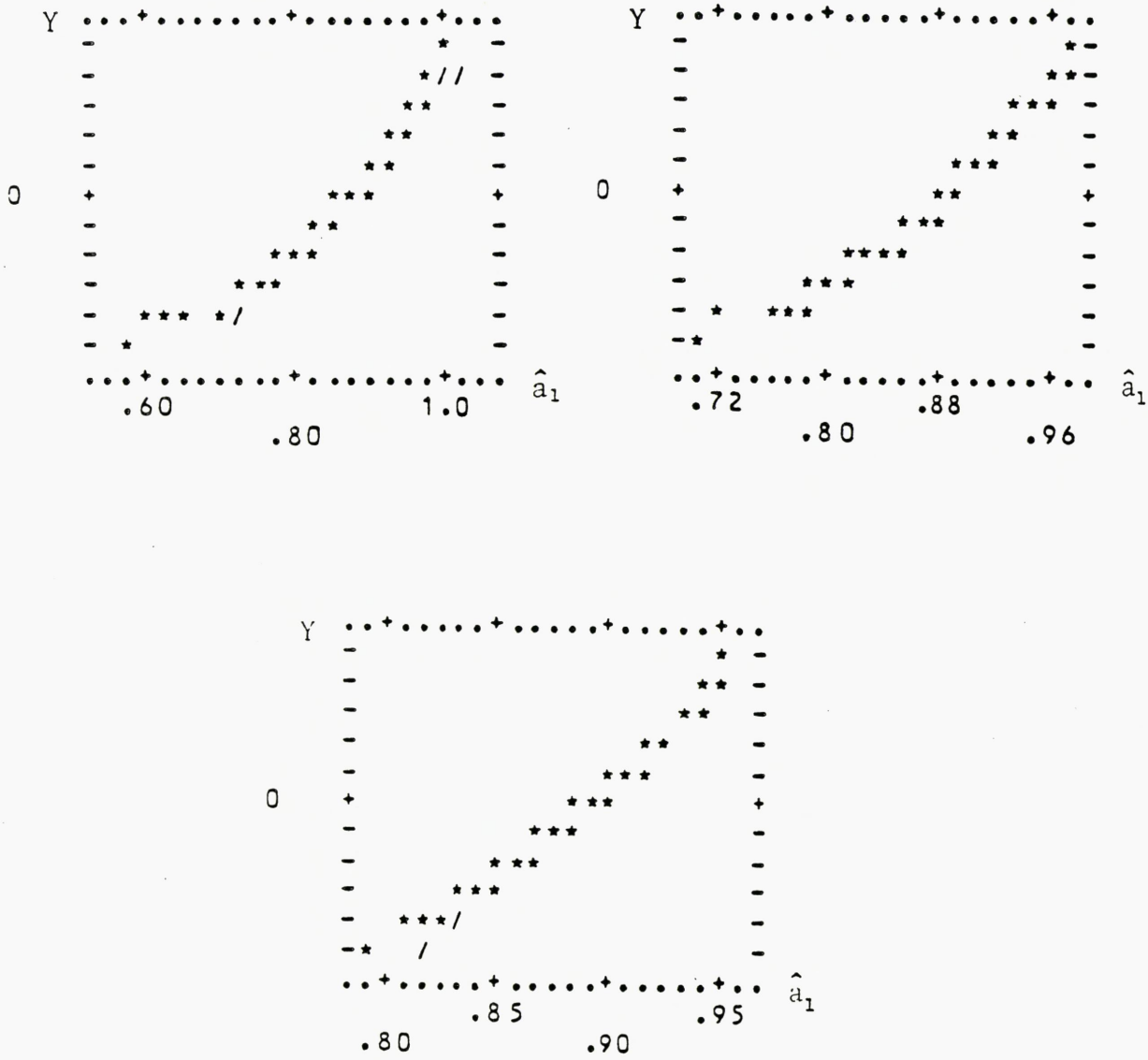

Table captions

Table 1: Simulated and theoretical values for the parameter estimates when $\{\delta_{it}\}$ is known.

Table 2: Simulated and theoretical values for the parameter estimates when $\{\delta_{it}\}$ is unknown with $p_0 = 0.5$ and $p_1 = 0.25$ in (3.1).

Table 3: Simulated and theoretical values for the parameter estimates when $\{\delta_{it}\}$ is unknown with various probability distributions in (3.1).

Table 1

Parameters	Sample size	q_{ii}	Mean value Theor.	Mean value Simulated	Standard error Theor.	Standard error Simulated
0.9	250	0.95	0.9	0.879	0.039	0.056
-0.3			-0.3	-0.298	0.085	0.094
0.9	500	0.95	0.9	0.892	0.028	0.036
-0.3			-0.3	-0.296	0.060	0.065
0.9	1500	0.95	0.9	0.897	0.016	0.022
-0.3			-0.3	-0.299	0.035	0.036
0.9	1000	0.90	0.9	0.894	0.019	0.031
-0.3			-0.3	-0.297	0.043	0.047
0.9	1500	0.90	0.9	0.897	0.016	0.025
-0.3			-0.3	-0.300	0.035	0.036
0.9	4000	0.90	0.9	0.898	0.009	0.016
-0.3			-0.3	-0.300	0.021	0.022
0.8	1500	0.95	0.8	0.798	0.022	0.025
0.4			0.4	0.399	0.033	0.036
0.8	1500	0.90	0.8	0.799	0.022	0.027
0.4			0.4	0.397	0.033	0.038

Table 2

Parameters	Sample size	q_{ii}	Mean value Theor.	Mean value Simulated	Standard error Simulated
0.9	250	0.95	0.885	0.867	0.060
-0.3			-0.285	-0.283	0.096
0.9	500	0.95	0.885	0.882	0.038
-0.3			-0.285	-0.280	0.067
0.9	1500	0.95	0.885	0.887	0.025
-0.3			-0.285	-0.281	0.036
0.9	1000	0.90	0.87	0.870	0.036
-0.3			-0.27	-0.260	0.051
0.9	1500	0.90	0.87	0.873	0.027
-0.3			-0.27	-0.265	0.036
0.9	4000	0.90	0.87	0.875	0.018
-0.3			-0.27	-0.266	0.022
0.8	1500	0.95	0.795	0.797	0.025
0.4			0.405	0.402	0.036
0.8	1500	0.90	0.79	0.793	0.027
0.4			0.41	0.406	0.036

Table 3

Parameters	Sample size	q_{ii}	Mean value Theor.	Mean value Simulated	Standard error Simulated
		$p_0 = 0.02 \;\; p_1 = 0.49$			
0.9	1500	0.90	0.84	0.852	0.029
-0.3			-0.24	-0.237	0.038
0.9	1500	0.90	0.81	0.801	0.038
-0.9			-0.81	-0.803	0.038
		$p_0 = 0.4 \;\; p_1 = 0.2 \;\; p_2 = 0.1$			
0.9	1500	0.95	0.876	0.881	0.024
-0.3			-0.276	-0.248	0.044
0.9	1500	0.95	0.864	0.849	0.033
-0.9			-0.864	-0.852	0.031
		$p_0 = 0 \;\; p_1 = 0.2 \;\; p_2 = 0.3$			
0.9	1500	0.95	0.828	0.799	0.041
-0.9			-0.828	-0.795	0.046
		$p_0 = 0 \;\; p_1 = 0 \;\; p_2 = 0.1 \;\; p_3 = 0.4$			
0.9	1500	0.95	0.774	0.701	0.064
-0.9			-0.774	-0.701	0.066