Z. Wahrscheinlichkeitstheorie verw. Gebiete
59, 491–504 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

General M-Estimates for Contaminated pth-Order Autoregressive Processes: Consistency and Asymptotic Normality

Robustness in Autoregressive Processes

Oscar H. Bustos

IMPA, Estrada Dona Castorina 110, 22460 Rio de Janeiro, Brazil

Summary. This work is concerned with simultaneous estimation of coefficients and a scale parameter of a pth-order autoregressive process (X_t). The observations are $Y_t = V_t Z_t + (1-V_t) X_t$ where (Z_t) is a contaminating process and (V_t) represents the proportion of contamination. If (X_t) or (Z_t) have heavy tails both least squares estimates and ordinary M-estimates are seriously affected. Under general conditions we prove consistency and asymptotic normality of a general class of M-estimates which contains some M-estimates studied by Denby and Martin [6].

1. Introduction

Let (X_t) be a pth-order autoregressive process $AR(p)$ with coefficient vector θ in R^p and suppose that we actually observe a "perturbed" version of (X_t), say (Y_t). This paper is concerned with the estimation of θ and of the scale of innovations when (X_t) and/or (Y_t) have heavy tailed distributions. We show the sensitivity or lack of robustness of classical estimates under certain deviations from the habitual assumptions. New estimates are studied in order to obtain more stability or robustness. We exhibit the performance of some such estimates under both, the usual assumptions and under heavy-tailed alternatives. To simplify the exposition in the beginning we assume $p=1$.

Perfectly Observed Autoregressions

Let (X_t) be an $AR(1)$; that is, a sequence of random variables which satisfies

$$X_{t+1} = \theta X_t + U_{t+1} \qquad t = 0, \ \pm 1, \ \pm 2, \dots \tag{1.1}$$

0044-3719/82/0059/0491/$02.80

where θ is an unknown parameter to be estimated and (U_t) (the "innovations") is a sequence of independent, identically distributed (i.i.d.) random variables, with common distribution F.

The L.S.E. (least squares estimates), based on a sample of size n of (X_t) is defined as θ^*_{LS} which satisfies

$$\sum_{t=0}^{n-1} (X_{t+1} - \theta^*_{LS} X_t) X_t = 0. \tag{1.2}$$

An estimate of a scale parameter of innovations is given by

$$\sigma^{*2}_{LS} = (1/n) \sum_{t=0}^{n-1} (X_{t+1} - \theta^{*2}_{LS} X_t)^2. \tag{1.3}$$

It is well-known (see Anderson [1]) that: 1) If $F = N(0, \sigma^2_U)$ then the conditional maximum likelihood estimate of (θ, σ^2_U) conditioned on $X_0 = c_0$ (with c_0 known constant) coincides with $(\theta^*_{LS}, \sigma^{*2}_{LS})$. 2) If $F = N(0, \sigma^2_U)$, then θ^*_{LS} and the conditional maximum likelihood estimate of θ are asymptotically equivalent. 3) Under the hypothesis: $EU_t = 0$, $0 < EU_t^2 = \sigma^2_U < \infty$, $|\theta| < 1$; $(\theta^*_{LS}, \sigma^{*2}_{LS})$ converge in probability to (θ, σ^2_U). 4) $\sqrt{n}(\theta^*_{LS} - \theta)$ converge in law to $N(0, 1 - \theta^2)$.

We note (Martin [13]) that the property 4) shows a kind of robustness of θ^*_{LS} since its asymptotic variance depends only on θ.

The asymptotic behavior of θ^*_{LS} under the assumptions: (U_t) are i.i.d. and $EU_t^2 = \infty$ is analysed in Yohai and Maronna (16). The main result of that work is: if (U_t) are i.i.d. with a symmetric distribution such that $E \log^+ |U_t| < \infty$, then $\sqrt{n}(\theta^*_{LS} - \theta)$ is bounded for probability. Also, Kanter and Steiger [10], and Hannan and Kanter [7] have analysed the asymptotic behavior of θ^*_{LS} under assumptions less restrictives than $EU_t^2 < \infty$. Indeed, in Kanter and Steiger [10] is proved the consistency of least squares estimates in autoregressive and finite moving average processes when F is in the domain of attraction of a stable law with characteristic exponent α. For a such F and for the autoregressive process Hannan and Kanter [7] prove $n^{1/\delta}(\theta^*_{LS} - \theta) \to 0$ a.s. for any $\delta > \alpha$.

Efficiences Issues. For finite variance innovations we can easily obtain asymptotic efficiencies.

Assume (1.1) where (U_t) is a sequence of i.i.d. with common density $g(u)$ such that $EU_t^2 = \sigma^2_U < \infty$. Let $g(u, \xi)$ be defined by $g(u, \xi) = g(u - \xi)$, for each ξ in R; let $i(g)$ be the Fisher information for ξ (i.e. $i(g) = E\{((\delta \log g(U, \xi)/\delta \xi)|_{\xi=0})^2\}$); finally, let I_θ be the asymptotic Fisher information per observation for θ (i.e. $I_\theta = E\{(\delta \log f(X_0, X_1, \theta)/\delta\theta)^2\}$ where f is the joint density of X_0, X_1). A straightforward calculation shows $I_\theta = \sigma^2_U i(g)/(1 - \theta^2)$ (for $p \geqq 2$ see Martin [13]). Taking 4) into account, the asymptotic efficiency of θ^*_{LS} under g is $\text{Eff}(\theta^*_{LS}; g) = (1/I_\theta)/$(asymptotic variance of θ^*_{LS}) $= 1/(\sigma^2_U i(g))$, which may be arbitrarily small when we allow U to have heavy tails in arbitrarily small neighborhoods of the Gaussian distribution, while $i(g)$ remains approximately stable. On the other hand, under sufficient regularity conditions, if V_{CR} is the Rao-Cramer bound ($V_{CR} = 1/I_\theta = (1 - \theta^2)/(\sigma^2_U i(g))$) then V_{CR} is attained asymptotically by the maximum likelihood estimate.

The latter fact suggests (Martin [13]) that one good alternative to θ^*_{LS} is to use M-estimates under (1.1). That is, to use the estimate θ^*_{ME} defined as a solution of

$$\sum_{t=0}^{n-1} \psi((X_{t+1}-\theta^*_{ME}X_t)/\sigma^*)\, X_t=0 \tag{1.4}$$

where $\psi: R\to R$ is a convenient function and σ^* is an estimate of a scale parameter of U_t computed either before than (1.4) has been solved or computed simultaneously with (1.4) from

$$\sum_{t=0}^{n-1} \chi((X_{t+1}-\theta^*_{ME}X_t)^2/\sigma^{*2})=0 \tag{1.5}$$

where χ: $[0, +\infty)\to R$ is a conveniently chosen function.

Under appropriate regularity conditions, consistency and asymptotic normality of θ^*_{ME} is obtained under (1.1) with: $EU_t=0$ and $EU_t^2=\sigma_U^2<\infty$. Its asymptotic variance is (Denby and Martin [6], Bustos [5]): $AV(\theta^*_{ME})=(E\psi^2(U)/(E\psi'(U))^2)((1-\theta^2)/\sigma_U^2)$. Then, the asymptotic efficiency of θ^*_{ME} under g (the density of U_t) is $\mathrm{Eff}(\theta^*_{ME};\, g)=(1/I_\theta)/(AV(\theta^*_{ME}))=(E\psi^2(U)/(E\psi'(U))^2)^{-1}/i(g)$, which is the asymptotic efficiency of the M-estimate defined by ψ in the location model when g is the density of the error. Those results suggest that using θ^*_{ME} seems better than using θ^*_{LS} if we want to control the effects of innovations' outliers.

Imperfectly Observed Autoregressions

Now, in an observation of time series turn up that a certain portion of X's are substituted or shifted by other values which correspond to an extraneous process (Z_t) that is grafted onto the process (X_t), for example, gross measurement errors or sporadic perturbations of the model. Thus, two realistic ways of taking these deviations into account are: "substitutive" outliers as defined below and "additive" outliers (Denby and Martin [6]). The following is therefore suggested as a more adequate model than (1.1) covering such cases:

$$\begin{aligned} X_{t+1}&=\theta X_t+U_{t+1} \\ Y_{t+1}&=V_{t+1}Z_{t+1}+(1-V_{t+1})X_{t+1} \end{aligned} \qquad t=0,\ \pm 1,\ \pm 2, \ldots \tag{1.6}$$

where $Y_1, Y_2, \ldots$ are the observations and (Z_t) is the contaminating process. If $V_t\equiv 1/2$ for all t, and Z_t has an appropriate distribution then the additive outliers are obtained (A.O. model). If $V_t\equiv 0$ for all t, we obtain contaminations only in innovations (I.O. model). The "substitutive" outliers (S.O. model) are obtained with (V_t) i.i.d., $V_t=0$ with probability $(1-\varepsilon)$ and $V_t=1$ with probability ε. This corresponds to Tukey [15] contaminated normal model. We note that another kind of outlier type in time series occurs which is not covered by (1.6); namely, the "pure" process (i.e. (X_t)) is substituted by a contaminating process during a short time interval (Huber [9]); more precisely: there exist $0<\alpha<\beta<1$ such that $Y_t=Z_t$ if $\alpha\leqq t/n\leqq\beta$, $Y_t=X_t$ otherwise, where $Y_1, \ldots, Y_n$ are the observations and (Z_t) is the contaminating process. This model has been studied in Brillinger [4]. The A.O., and I.O. model are studied in Denby and Martin [6] and Martin [13]. They show, in particular, the poor performance of θ^*_{ME} under the A.O. model.

The following example shows the behavior of θ^*_{LS} and θ^*_{ME} under the S.O. model. Assume (1.6) with the following specifications: (V_t), (Z_t) and (U_t) are jointly independent, each one is a sequence of i.i.d. random variables such that: $V_t=0$ with probability $1-\varepsilon$, $V_t=1$ with probability ε; Z_t is normal with mean 0 and variance τ^2 and U_t is normal with mean 0 and variance 1. Now, we obtain θ^*_{LS} and θ^*_{ME} respectively from (1.2) and (1.4) with Y_t instead of X_t. We note that $\theta^*_{\mathrm{LS}} \to \theta_{\mathrm{LS}}$ where $E\{(Y_1-\theta_{\mathrm{LS}} Y_0) Y_0\}=0$, and $\sqrt{n}(\theta^*_{\mathrm{LS}}-\theta_{\mathrm{LS}}) \to N(0, VA(\theta^*_{\mathrm{LS}}))$ for some $VA(\theta^*_{\mathrm{LS}})>0$. Moreover, $\theta_{\mathrm{LS}}=(1-\varepsilon)^2\theta/(1-\varepsilon+\varepsilon\tau^2(1-\theta^2))$ which shows that $\theta_{\mathrm{LS}}-\theta$ (the asymptotic bias) can be considerable in a small neighborhood of (X_t). On the other hand, we deduce that there exists θ_{ME} such that $\theta^*_{\mathrm{ME}} \to \theta_{\mathrm{ME}}$ and $\sqrt{n}(\theta^*_{\mathrm{ME}}-\theta_{\mathrm{ME}}) \to N(0, VA(\theta^*_{\mathrm{ME}}))$.

In a Monte Carlo study with samples of size 30 and 500 replications, we obtain the following table

Table 1

	$\theta=0.5$, $\varepsilon=0.1$, $\tau^2=12$	$\theta=0.9$, $\varepsilon=0.1$, $\tau^2=47.4$
$\theta^{\wedge}_{\mathrm{LS}}$	0.24	0.43
$V(\theta_{\mathrm{LS}})^{\wedge}$	1.53	2.67
$\theta^{\wedge}_{\mathrm{ME}}$	0.26	0.52
$V(\theta_{\mathrm{ME}})^{\wedge}$	1.36	2.55
$\theta^{\wedge}_{\Phi}$	0.33	0.69
$V(\theta_{\Phi})^{\wedge}$	1.26	1.30

where if T is an estimate, then $T^{\wedge}=\sum_{i=1}^{500} T_i/500$ and $V(T)^{\wedge}=\left(\sum_{i=1}^{500}(T_i-T^{\wedge})^2/499\right)\sqrt{30}$ (T_i is the value of the T estimate obtained in the i-the replication). We define the θ_{Φ} estimate later. The ψ of (1.4) and the χ of (1.5) were Hubers' favorite ψ and χ (1964) defined by $\psi(x)=\min(|x|, k)\operatorname{sign} x$, $\chi(x^2)=\psi^2(x)-\beta$ with $\beta=E\{\psi^2(U)\}$ (note that U is $N(0,1)$) and $k=1.5$. We note the similarity of the behavior for both θ^*_{LS} and θ^*_{ME} in this case, in particular, the bias of θ^*_{ME} (defined as $\theta^{\wedge}_{\mathrm{ME}}-\theta$) is almost as large as the bias of θ^*_{LS} (which is defined as $\theta^{\wedge}_{\mathrm{LS}}-\theta$).

To deal with this problem we naturally wish to control the outliers both in the innovations and in each observation. Building on what Maronna and Yohai [11] have already done for regression, and what Martin and other authors have done for the I.O. and the A.O. models, we studied the behavior of a very general class of estimates for θ and a scale parameter of U_t, which for convenience we will call Φ-estimates. These estimates are solutions θ^*_{Φ}, σ^*_{Φ} of the equations

$$\begin{aligned}&\sum_{t=0}^{n-1} \Phi(Y_t,(Y_{t+1}-\theta_\Phi^* Y_t)/\sigma_\Phi^*)\, Y_t=0\\&\sum_{t=0}^{n-1} \chi((Y_{t+1}-\theta_\Phi^* Y_t)^2/\sigma_\Phi^{*2})=0\end{aligned} \tag{1.7}$$

where Φ: $R\times R\to R$ and χ: $[0,+\infty)\to R$ are conveniently chosen.

Note that if we define Φ and χ appropriately, then we obtain several important examples, i.e.: $(\theta_{LS}^*, \sigma_{LS}^*)$ if $\Phi(y,u)=u$ and $\chi(u^2)=u^2-1$; $(\theta_{ME}^*, \sigma^*)$ if $\Phi(y,u)=\psi(u)$, with ψ as in (1.4) and χ is as in (1.5).

We performed the above mentioned Monte Carlo study with $\Phi(y,u)=\psi(u)\,w(y)$ where ψ was as before and $w(y)=\min(1,k/|y|)$. The results are given in the last two rows of the Table 1, they indicate that the performance of θ_Φ^* will be better than both θ_{LS}^* and θ_{ME}^* under the S.O. model.

For regression Maronna, Yohai and Bustos [12] demonstrated the good performance of another Φ-estimate, defined by $\Phi(y,u)=\psi(|y|u|)/|y|$ where ψ is as before.

The origin of the above two estimates will be pointed out later, in the general $AR(p)$ setting.

For these reasons, we considered an important issue to study consistency and asymptotic normality of Φ-estimate in the general situation exposed in Sect. 2.

1.2. Notation. We denote the s-dimensional euclidean space by R^s. Let M be a $m\times n$-matrix with entry M_{ij} in the i-th row and the j-th column ($i=1,\ldots,m$; $j=1,\ldots,n$); we also write M as $[M_{ij}]$, its transpose as M' and its norm as $|M|$. If A is a set, then A^c is the complement of A. If $\mathbf{v}=(v_0,\ldots,v_p)'$ is in R^{p+1} let $\mathbf{v}_{p-1}^{\sim}=(v_0,\ldots,v_{p-1})'$. Let F: $R^n\to R^m$ and F_i: $R^n\to R$ for $i=1,\ldots,m$ such that $F(\mathbf{x})'=(F_1(\mathbf{x}),\ldots,F_m(\mathbf{x}))$ for all $\mathbf{x}$, then D_jF_i is the j-th partial derivative of F_i, for $j=1,\ldots,n$; and $DF=[D_jF_i]$ the total derivative. If P is a probability then E_P (or simply E when this causes no confusion) is the respective expectation functional. If $S_1,\ldots,S_n,\ldots$ is a sequence of m-dimensional random vectors, P is a probability and $\mathbf{v}$ is in R^m, then $S_n\to\mathbf{v}(P)$, $n\to\infty$ means that (S_n) converge in probability to $\mathbf{v}$ under P; if (S_n) tends in law to m-variate normal with mean $\mathbf{0}$ and covariance matrix M, then we set $S_n\to N(\mathbf{0},M)(D)$, $n\to\infty$. Finally, we set the symbol $=$ for definition.

1.3. Contents of the Paper. In Sect. 2 we give the full definition of the model and Φ-estimates, and give the main results. In Sect. 3 we give the technical details of proofs of the results in Sect. 2.

The results of this paper are generalized in Bustos [5], but we preferred to present some what less general results here for the sake of clarity.

2. Definitions, Assumptions and Results

2.1. The Model

Let $(\Omega,\mathscr{F},P)$ be a probability space; S: $\Omega\to\Omega$ is a one-to-one map, measurable with measurable inverse (S^{-1}), P-preserving (that is: $P(A)=P(S^{-1}(A))$ for all A

in $\mathscr{F}$), and P-ergodic (see Breiman [3]). For each $t=1,2,\ldots, S^t=S\cdot S^{t-1}$ where S^0 is the identity map and $S^{-t}=(S^{-1})^t$.

We assume that:

M1) $(d_r)_r=0, \pm 1, \pm 2, \ldots$ is a sequence of real numbers such that $\sum_{k=1}^{\infty}\left(\sum_{|r|>k}^{\infty}|d_r|\right)^{1/2}<\infty$.

M2) U, V, and Z are real random variables defined in Ω such that $E|U|$, $E|V|$, and $E|Z|$ are finite; for each $t=0, \pm 1, \pm 2, \ldots$ let $U_t=U\cdot S^t$, $V_t=V\cdot S^t$, and $Z_t=Z\cdot S^t$; also, we suppose that the processes (U_t), (V_t), and (Z_t) are jointly independent and each one is ϕ-mixing with $\sum_{n=1}^{\infty}\phi(n)^{1/2}<\infty$ (see Billingsley [2]).

M3) X: $\Omega\to R$ is a random variable defined by $X=\sum_{r=-\infty}^{+\infty} d_r U_{-r}$; $X_t=X\cdot S^t$ for all t (therefore $X_t=\sum_{r=-\infty}^{+\infty} d_r U_{t-r}$ for all t) (we understand that the convergence is in $L^1(\Omega,P)$).

M4) Y: $\Omega\to R$ is a random variable defined by $Y=VZ+(1-V)X$; $Y_t=Y\cdot S^t$ for all t (therefore $Y_t=V_tZ_t+(1-V_t)X_t$ for all t). In practice, (Y_t) will represent the observed process.

M5) $P(aY_p+\mathbf{Y}_{p-1}^{\sim\prime}\boldsymbol{\theta}=0)=0$ for all a in R, $\boldsymbol{\theta}$ in R^p such that $|a|+|\boldsymbol{\theta}|>0$.

The following results are well-known (see Anderson [1]). Let (X_t^*) be an $AR(p)$, that is

$$X_{t+p}^*=\mathbf{X}_{t+p-1}^{*\sim\prime}\boldsymbol{\theta}+U_{t+p}^* \qquad t=0, \pm 1, \pm 2, \ldots \tag{2.1}$$

where $\boldsymbol{\theta}'=(\theta_0,\ldots,\theta_{p-1})$ is in R^p and (U_t^*) is a sequence of i.i.d. random variable having finite variance. If all the roots of the polynomial equation $x^p-\sum_{j=0}^{p-1}\theta_j x^j=0$ are less than 1 in absolute value, then $X_t^*=\sum_{r=0}^{\infty} d_r U_{t-r}^*$ for all t (convergence in $L^2(\Omega,P)$) where $(d_r)_r=0,1,2,\ldots$ satisfies M1). From this remark we derive that an $AR(p)$ is a particular case of (X_t) defined in M3). Also, Moving Average and Mixed Process (ARMA) are particular cases of (X_t) (see Billingsley [2], p. 191). Finally, note that (U_t) is more general than (U_t^*), since a sequence of i.i.d. random variables is ϕ-mixing with $\phi(n)\equiv 0$.

We note that I.O., A.O., and S.O. models are obtained if V_t is defined appropriately, as was indicated previously.

2.2. *The Φ-estimates*

Let Φ: $R^p\times R\to R$, χ: $[0,+\infty)\to R$ such that:

E1) For each $\mathbf{y}$ in R^p, $u\to\Phi(\mathbf{y},u)$ is odd, uniformly continuous and $\Phi(\mathbf{y},u)\geqq 0$ for $u\geqq 0$.

E2) $(y,u)\to\Phi(y,u)\,y$ is bounded and there exists a constant k^* such that $|\Phi(y,u)\,y-\Phi(z,u)\,z|\leqq k^*|y-z|$ for all u.

E3) For each y, $u \to \Phi(y,u)/u$ is non increasing and there exists u_0 such that $\Phi(y,u_0)/u_0 > 0$ for all y.

E4) $u \to \partial \Phi(y,u)/\partial u$ is continuous a.e. y and $(y,u) \to y \partial \Phi(y,u)/\partial u$ is bounded.

E5) $E\{\sup\{u(\partial \Phi(\mathbf{Y}_{p-1}^{\sim}, u)/\partial u) |\mathbf{Y}_{p-1}^{\sim}|: u \text{ in } R\}\} < \infty$.

E6) χ is bounded, increasing on $\{x: -a \leqq \chi(x) < b\}$ where $b = \sup \chi(x)$, $-a = \chi(0)$ $0 < a, b$; is differentiable with $x \to x D\chi(x^2)$ continuous and bounded. Also $\chi(u_0^2) > 0$.

In Maronna and Yohai [11] it is proved that under conditions weaker than E1) to E6), given a non singular $p \times p$-matrix M_0 there exist $\boldsymbol{\theta}_0$ in R^p and σ_0 in $(0, +\infty)$ such that:

$$E\{\Phi(M_0 \mathbf{Y}_{p-1}^{\sim}, r) \mathbf{Y}_{p-1}^{\sim}\} = \mathbf{0}, \tag{2.2}$$

$$E\{\chi(r^2)\} = 0 \tag{2.3}$$

where $r = (Y_p - \mathbf{Y}_{p-1}^{\sim\prime} \boldsymbol{\theta}_0)/\sigma_0$. In practice, M_0 represents a well-chosen scatter matrix of $\mathbf{Y}_{p-1}^{\sim}$.

Finally, let (V_n^*) be a sequence of non-singular $p \times p$-matrices such that $V_n^* \to M_0$ a.e., $n \to \infty$.

The Φ-estimates of $(\boldsymbol{\theta}_0, \sigma_0)$ form a sequence $(\boldsymbol{\theta}_n^*, \sigma_n^*)$ with values in $R^p \times (0, +\infty)$ such that:

$$(1/n) \sum_{t=0}^{n-1} \Phi(J(t,n), r(t,n)) \mathbf{Y}_{t+p-1}^{\sim} \to \mathbf{0} \text{ a.e.,} \quad n \to \infty, \tag{2.4}$$

$$(1/n) \sum_{t=0}^{n-1} \chi(r(t,n)^2) \to 0 \text{ a.e.,} \quad n \to \infty \tag{2.5}$$

where $r(t,n) = (Y_{t+p} - \mathbf{Y}_{t+p-1}^{\sim\prime} \boldsymbol{\theta}_n^*)/\sigma_n^*$, $J(t,n) = V_n^* \mathbf{Y}_{t+p-1}^{\sim}$.

Two examples of such Φ-estimates are: (we can see the reasons for their names and how they work in regression in Maronna, Yohai and Bustos [12]).

Mallows Estimates. They are defined by $\Phi(y,u) = \psi(u) w(y)$ where w is a positive function. In order to obtain E1) to E5) it is sufficient to assume: $\psi(u) \geqq 0$ for $u \geqq 0$; there exists u_0 such that $\psi(u_0) > 0$; ψ is odd, bounded, uniformly continuous, continuously differentiable, and $u D\psi(u) \leqq \psi(u)$ for all u; finally, $g(y) = y w(y)$ is bounded and Lipschitz. We can modify these hypothesis for obtaining asymptotic normality (see Remark at the end of this Section).

Hampel-Krasker Estimates. They are defined by: $\Phi(y,u) = \psi(|y| u)/|y|$ if $|y| > c$, $\Phi(y,u) = \psi(cu)/c$ if $|y| \leqq c$, where $c > 0$ is a constant to be chosen conveniently. To obtain the conditions E1) to E5) it is sufficient to require: ψ is odd, bounded, non decreasing in $[0, +\infty)$ with $\psi(u_0) > 0$, uniformly continuous, continuously differentiable and $u D\psi(u) \leqq \psi(u)$ for all u; finally, we require that $(y,u) \to |D\psi(|y|u)| |y|$ is bounded (this condition would not be necessary if $E|Y^2| < \infty$).

2.3. *Consistency*

Theorem 2.1. *If* $(\boldsymbol{\theta}_0, \sigma_0)$ *which satisfies* (2.2) *and* (2.3) *is unique, then* $(\boldsymbol{\theta}_n^*, \sigma_0^*) \to (\boldsymbol{\theta}_0, \sigma_0)$ *a.e.*, $n \to \infty$.

2.4. *Asymptotic Normality*

From now on we assume $V_n^* = M_0 = I$ where I is the identity $p \times p$-matrix. This restriction would not be necessary if $E|Y|^2 < \infty$. Also, we drop the condition E3).

Theorem 2.2. *Suppose* $\Lambda = D\lambda(\boldsymbol{\theta}_0, \sigma_0)$ *is non singular, where* λ: $R^p x(0, +\infty) \to R^{p+1}$ *is defined by*

$$\lambda(\boldsymbol{\theta}, \sigma)' = (E\{\Phi(\mathbf{Y}_{p-1}^{\sim}, r^*)\}', E(\chi(r^{*2}))) \tag{2.6}$$

with $r^* = (Y_p - \mathbf{Y}_{p-1}^{\sim\prime} \boldsymbol{\theta})/\sigma$. *If in addition* (2.4) *and* (2.5) *hold with* $\sqrt{n}$ *instead of* n *and*

$$(\boldsymbol{\theta}_n^*, \sigma_n^*) \to (\boldsymbol{\theta}_0, \sigma_0)(P), \qquad n \to \infty \tag{2.7}$$

Then: a) *For each* $i = 0, \ldots, p-1$, $j = 0, \ldots, p-1$ *the following series are all absolutely convergent:*

$$S_{ij} = \sum_{t=1}^{\infty} E\{\Phi(\mathbf{Y}_{p-1}^{\sim}, r)\, \Phi(\mathbf{Y}_{t+p-1}^{\sim}, r(t))\, Y_i Y_{j+t}\},$$

$$T_i = \sum_{t=1}^{\infty} E\{\Phi(\mathbf{Y}_{p-1}^{\sim}, r)\, \chi(r(t)^2)\, Y_i\},$$

$$S_i = \sum_{t=1}^{\infty} E\{\Phi(\mathbf{Y}_{t+p-1}^{\sim}, r(t))\, \chi(r^2)\, Y_{i+t}\},$$

$$\gamma = \sum_{t=1}^{\infty} E\{\chi(r^2)\, \chi(r(t)^2)\}$$

where r *is as in* (2.2) *and* (2.3), *and* $r(t) = (Y_{t+p} - \mathbf{Y}_{t+p-1}^{\sim\prime} \boldsymbol{\theta}_0)/\sigma_0$ *for all* $t = 0, 1, 2, \ldots$

b) $\sqrt{n}((\boldsymbol{\theta}^{*\prime}, \sigma_n^*)' - (\boldsymbol{\theta}_0', \sigma_0)') \to N(\mathbf{0}, \Lambda^{-1} \Gamma \Lambda'^{-1})(D)$, $n \to \infty$.

where

$$\Gamma = \begin{bmatrix} G & \mathbf{V} \\ \mathbf{V}' & N \end{bmatrix}$$

with

$$G = [E\{\Phi(\mathbf{Y}_{p-1}^{\sim}, r)^2\, Y_i Y_j\} + S_{ij} + S_{ji}],$$

$$\mathbf{V} = [E\{\Phi(\mathbf{Y}_{p-1}^{\sim}, r)\, \chi(r^2)\, Y_i\} + S_i + T_i] \quad \textit{and}$$

$$N = E\{\chi^2(r^2)\} + 2\gamma.$$

Remark. In the particular case of the Mallows estimates we can obtain asymptotic normality of $(\boldsymbol{\theta}_n^*, \sigma_n^*)$ without the restriction imposed on V_n^* and M_0 at the beginning of this subsection. More precisely: let χ: $[0, +\infty) \to R$ be as in E6) not necessarily increasing; let ψ: $R \to R$ and w: $R^p \to R$ such that: w is a positive function, ψ is bounded, differentiable with continuous and bounded derivative, $g(y) = yw(y)$ is bounded with coordinates $g_1, \ldots, g_p$ such that $y \to D_j g_i(y)$ is bounded and Lipschitz for all i, j; there exist $(\boldsymbol{\theta}_0, \sigma_0)$ in $R^p \times (0, +\infty)$ and a non singular matrix M_0 such that $E\{\psi(r) g(J)\} = \mathbf{0}$, $E\{\psi(r) Y_{i-1} Y_{k-1} D_j w(J)\} = 0$, $E\{\chi(r^2)\} = 0$ for all $1 \leqq i, j, k \leqq p$, where r is as before and $J = M_0 \mathbf{Y}_{p-1}^{\sim}$. Finally, let $(\boldsymbol{\theta}_n^*, \sigma_n^*, V_n^*)$ be a sequence of estimates of $(\boldsymbol{\theta}_0, \sigma_0, M_0)$ such that $(\sqrt{n}(V_n^* - M_0))$ is bounded in probability and V_n^* is a non singular $p \times p$-matrix.

Theorem 2.3. *Theorem 2.2 holds with $\Phi(y, u) = \psi(u) w(y)$ and $M_0 \mathbf{Y}_{p-1}^{\sim}$ instead of $\mathbf{Y}_{p-1}^{\sim}$.*

We obtain the proof by reducing this case to Theorem 2.2 by first proving that $(1/\sqrt{n}) \sum_{t=0}^{n-1} \psi(r(t,n))\, g(M_0 \mathbf{Y}_{t+p-1}^{\sim}) \to O(P)$, $n \to \infty$.

3. Proofs

For the sake of simplicity let Ψ the $(p+1)$-dimensional vector defined by

$$\Psi(\mathbf{y}, (\boldsymbol{\theta}, \sigma)) = (\Phi(\mathbf{y}_{p-1}^{\sim}, (\mathbf{y}_p - y_{p-1}^{\sim\prime} \boldsymbol{\theta})/\sigma)\, \mathbf{y}_{p-1}^{\sim}, \chi((y_0 - \mathbf{y}_{p-1}^{\sim\prime} \boldsymbol{\theta})^2/\sigma^2))$$

for all $\mathbf{y} = (y_0, \ldots, y_{p-1}, y_p)'$ in R^{p+1}, $\boldsymbol{\theta}$ in R^p and $\sigma > 0$. Also, let $T_n' = (\boldsymbol{\theta}_n^{*\prime}, \sigma_n^*)$ for all n.

Let $\mathscr{W} = \mathscr{W}(P)$ the family of all the sets W of the form

$$W = \{Q \text{ in } Z(R^{p+1}) \colon E_Q |f_i| < \infty, |E_Q f_i - E_P f_i| < \varepsilon;\ i = 1, \ldots, k\} \tag{3.1}$$

where: $Z(R^{p+1})$ is the set of all the probabilities over R^{p+1}, k is a positive integer, $\varepsilon > 0$, $f_1, \ldots, f_k$ are P-integrable functions and P denotes, by an abuse of notation, the distribution of $\mathbf{Y}_0 = (Y_0, \ldots, Y_p)'$ over R^{p+1}.

3.1. Proof of Theorem 2.1

Let $J(t) = M_0 \mathbf{Y}_{t+p-1}^{\sim}$ for all $t = 0, 1, 2, \ldots$. From E2) we have

$$\left| (1/n) \sum_{t=0}^{n-1} (\Phi(J(t,n), r(t,n))\, J(t,n) - \Phi(J(t), r(t,n))\, J(t)) \right| \leqq k^* |V_n^* - M_0| (1/n) \sum_{t=0}^{n-1} |\mathbf{Y}_{t+p-1}^{\sim}|$$

from where $(1/n) \sum_{t=0}^{n-1} \Phi(J(t), r(t,n))\, \mathbf{Y}_{t+p-1}^{\sim} \to \mathbf{0}$ a.e., $n \to \infty$. This allows the reduction of proof of Theorem 2.1 to the following:

Under conditions E1) to E3), and E6), the Theorem 2.1 holds with $V_n^* \equiv I$ for all n and $M_0 = I$ (where I is the identity $p \times p$-matrix). (3.2)

In the remainder of this subsection 3.1 we drop the hypothesis M1), M2) and M3), and we replace M4) by: Y: $\Omega \to R$ is a random variable; $Y_t = Y \cdot S^t$ for all t. Finally, we replace E2) by: $\Phi(y, u)\, y$ is bounded.

It may be easily checked that Huber's [9] conditions (B-1) and (B-2) are fulfilled. From the hypothesis of Theorem 2.1 there exists a unique $(\boldsymbol{\theta}_0, \sigma_0)$ in $R^p \times (0, +\infty)$ such that $\lambda(\boldsymbol{\theta}_0, \sigma_0) = \mathbf{0}$. From the proof in Sect. 6 of Maronna and Yohai [11] there exist: W in $\mathscr{W}$ and C a compact set in $R^p \times (0, +\infty)$ such that:

$$\inf_{Q \in W} \inf_{(\boldsymbol{\theta}, \sigma) \notin C} \max_{i=1,2} |E_Q\{\Psi(\mathbf{Y}, (\boldsymbol{\theta}, \sigma))' z_i(\boldsymbol{\theta}, \sigma)\}| > 0$$

where $z_1(\boldsymbol{\theta}, \sigma) = (\boldsymbol{\theta}, 0)/|\boldsymbol{\theta}|$ if $|\boldsymbol{\theta}| \neq 0$, $z_1(\mathbf{0}, \sigma) = (\mathbf{v}, 0)$ with $\mathbf{v} \neq \mathbf{0}$ and $z_2(\boldsymbol{\theta}, \sigma) = (\mathbf{0}, 1)$. From this fact and the ergodicity there exists $L > 0$ such that $P(\Omega^*) = 1$ where

$$\Omega^* = \liminf_n \left(\inf_{(\boldsymbol{\theta}, \sigma) \notin C} \max_i \left| (1/n) \sum_{t=0}^{n-1} \Psi(\mathbf{Y}_t, (\boldsymbol{\theta}, \sigma))' z_i(\boldsymbol{\theta}, \sigma) \right| > L \right).$$

For each $n = 1, 2, \ldots$ we define $C_n = \{\omega$ in Ω: $T_n(\omega)$ in $C\}$. Noting that $\Omega^* = (\Omega^* \cap \lim_n \inf C_n) \cup (\Omega^* \cap (\lim_n \inf C_n)^c)$ and by using (2.4) and (2.5), we see that $P(\lim_n \inf C_n) = 1$. We achieve the proof of (3.2) by using ergodicity instead of strong law of large numbers as in Huber [9].

Remark. If f_i's are restricted to be bounded and continuous, then the W's of (3.1) generate the weak topology on $Z(R^{p+1})$.

3.2. Proof of Theorem 2.2

First we state some intermediate results in the lemmas below.

Let $\Xi = R^p \times (0, +\infty)$ and $\tau_0' = (\boldsymbol{\theta}_0', \sigma_0)$. We denote a general element of Ξ by τ. Finally, $\Psi_1, \ldots, \Psi_{p+1}$ are the coordinates of Ψ.

Lemma 3.1. a) *For each* $1 \leq i, j, \leq p+1$ *integers, the following series is absolutely convergent*

$$S_{ij}^* = \sum_{t=0}^{\infty} E\{\Psi_i(\mathbf{Y}_0, \tau_0)\, \Psi_j(\mathbf{Y}_t, \tau_0)\}.$$

b) $(1/\sqrt{n}) \sum_{t=0}^{n-1} \Psi(\mathbf{Y}_t, \tau_0) \to N(\mathbf{0}, \Gamma)(D)$, $n \to \infty$, *where* $\Gamma = [\Gamma_{ij}]$ *with*

$$\Gamma_{ij} = E\{\Psi_i(\mathbf{Y}_0, \tau_0)\, \Psi_j(\mathbf{Y}_0, \tau_0)\} + S_{ij}^* + S_{ji}^*.$$

Proof. Let $(\xi_t)_{t=0, \pm 1, \pm 2, \ldots}$ the process taking values in R^3 defined by:

$$\xi_t = (\xi_t^{(1)}, \xi_t^{(2)}, \xi_t^{(3)}) = (V_t, Z_t, U_t) \qquad t = 0, \pm 1, \ldots \tag{3.3}$$

where V_t, Z_t, and U_t are as in Sect. 2. For each $i=1,2,\ldots,p+1$ and $t=0, \pm 1, \pm 2, \ldots$ we define:

$$\eta_t^{(i)} = \Psi_i(\mathbf{Y}_t, \tau_0) = \Psi_i(Y_t, \ldots, Y_{t+p}, \tau_0)$$

from the hypothesis M2) to M4) and (3.3) we have:

$$\eta_t^{(i)} = \Psi_i(\xi_t^{(1)}\xi_t^{(2)} + (1-\xi_t^{(1)}) \sum_{r=-\infty}^{+\infty} d_r \xi_{t-r}^{(3)}, \ldots, \xi_{t+p}^{(1)}\xi_{t+p}^{(2)} + (1-\xi_{t+p}^{(1)}) \sum_{r=-\infty}^{+\infty} d_r \xi_{t+p-r}^{(3)}, \tau_0)$$

then each $\eta_t^{(i)}$ is a function of the entire process $(\xi_t)_{t=0,\pm 1,\pm 2,\ldots}$. The random variables $\eta_t^{(i)}$ have approximations

$$\eta_{kt}^{(i)} = \Psi_i(\xi_t^{(1)}\xi_t^{(2)} + (1-\xi_t^{(1)}) \sum_{r=-k+p}^{k-p} d_r \xi_{t-r}^{(3)}, \ldots, \xi_{t+p}^{(1)}\xi_{t+p}^{(2)} + (1-\xi_{t+p}^{(1)}) \sum_{r=-k+p}^{k-p} d_r \xi_{t+p-r}^{(3)}, \tau_0)$$

for all $t=0, \pm 1, \pm 2, \ldots$, and $k=p, p+1, \ldots$; then each $\eta_{kt}^{(i)}$ is a function depending only on finitely many of the ξ_t. By using the definition of Ψ, (2.2) and (2.3) we have:

$$E\eta_0^{(i)} = 0 \qquad \text{for all } i=1,2,\ldots,p+1 \tag{3.4}$$

from E2) and E6) we can deduce:

$$E\eta_0^{(i)2} < \infty \qquad \text{for all } i=1,2,\ldots,p+1. \tag{3.5}$$

For each $k=p, p+1, \ldots$ and $i=1,\ldots,p+1$ if:

$$v_k(i) = E|\eta_0^{(i)} - \eta_{k0}^{(i)}|^2$$

then, by using E2), E4) and E6) we can show that there exists a constant c such that:

$$v_k^{1/2}(i) \leqq c\left(\sum_{|r|>k} |d_r|\right)^{1/2}$$

then, M1) implies

$$\sum_{k=p}^{\infty} v_k^{1/2}(i) < \infty \tag{3.6}$$

for all $i=1,2,\ldots,p+1$.

Now, we notice that M2) implies $(\xi_t)_{t=0,\pm 1,\pm 2,\ldots}$ is a φ-mixing process with $\sum_{n=1}^{\infty} \varphi(n)^{1/2} < \infty$. From this, (3.4), (3.5), and (3.6) we can show the Lemma 3.1 by applying the multivariate version of Theorem 21.1 of Billingsley [2] to the process $\eta_t = (\eta_t^{(1)}, \ldots, \eta_t^{(p+1)}) = \Psi(\mathbf{Y}_t, \tau_0)$.

Lemma 3.2. *Let C be a compact set of Ξ; let g: $R^{p+1} \times C \to R$ a function such that: for each $\mathbf{y}$ in R^{p+1}, $\tau \to g(\mathbf{y}, \tau)$ is continuous; for each τ in C, $y \to g(\mathbf{y}, \tau)$ is measurable and there exists a neighborhood U_τ of τ such that*

$$E\{\sup\{|g(\mathbf{Y}_0, \tau^*) - g(\mathbf{Y}_0, \tau)|: \tau^* \text{ in } C \cap U_\tau\}\} < \infty.$$

Then

$$\sup\left\{(1/n) \sum_{t=0}^{n-1} g(\mathbf{Y}_t, \tau) - E\{g(\mathbf{Y}, \tau)\}: \tau \text{ in } C\right\} \to 0 \text{ a.e.}, \quad n \to \infty.$$

Proof. Is an immediate consequence of Theorem 6.2 of Rao [14].

Remark. The hypothesis over Λ of Theorem 2.2 implies that there exist $a > 0$ and $b_0 > 0$ such that: if $|\tau - \tau_0| \leqq b_0$ then $|\lambda(\tau)| \geqq a|\tau - \tau_0|$.

Lemma 3.3. *Let*

$$C = \{\tau \text{ in } \Xi: |\tau - \tau_0| \leqq b_0\}; \qquad Z_n(\tau) = \left|\sum_{t=0}^{n-1} Z^*(t)\right| \Big/ (\sqrt{n} + n|\lambda(\tau)|)$$

where $Z^*(t) = \Psi(\mathbf{Y}_t, \tau) - \lambda(\tau) - \Psi(\mathbf{Y}_t, \tau_0)$, for all $n = 1, 2, \ldots$ and τ in C. Then $\sup_{\tau \in C} Z_n(\tau) \to 0$ a.e., $n \to \infty$.

Proof. Let $n \geqq 1$ be an arbitrary integer. From the Remark above and the Mean Value Theorem applied to $\tau \to \sum_{t=0}^{n-1} (\Psi(\mathbf{Y}, \tau) - \lambda(\tau))$ there exists $L < +\infty$ depending only on p such that

$$\sup_{\tau \in C} Z_n(\tau) \leqq L \max_{ij} \sup_{\tau \in C} |A_n(i,j,\tau)| / ((\sqrt{n}\, b_0)^{-1} + a)$$
$$+ L \max_{ij} |B_n(i,j)| / ((\sqrt{n}\, b_0)^{-1} + a) \tag{3.7}$$

where

$$A_n(i,j,\tau) = (1/n) \sum_{t=0}^{n-1} (D_i \Psi_j(\mathbf{Y}_t, \tau) - D_i \Psi_j(\mathbf{Y}_t, \tau_0))$$
$$- E\{D_i \Psi_j(\mathbf{Y}_0, \tau) - D_i \Psi_j(\mathbf{Y}_0, \tau_0)\}.$$
$$B_n(i,j) = (1/n) \sum_{t=0}^{n-1} D_i \Psi_j(\mathbf{Y}_t, \tau_0) - E\{D_i \Psi_j(\mathbf{Y}_0, \tau)\}.$$

The proof follows by using ergodicity in the second summand of (3.7) and by using Lemma 3.2 with $g(\mathbf{y}, \tau) = D_i \Psi_j(\mathbf{y}, \tau) - D_i \Psi_j(\mathbf{y}, \tau_0)$ for each $i = 1, \ldots, p+1$ and $j = 1, \ldots, p+1$ in the first summand.

Corollary 3.4. $\left|\sum_{t=0}^{n-1} (\Psi(\mathbf{Y}_t, \tau_0) + \lambda(T_n))\right| \Big/ (\sqrt{n} + n|\lambda(T_n)|) \to O(P)$, $n \to \infty$.

Proof. It follows from (2.4), (2.5), (2.7) and Lemma 3.3 taking into account that: if $|\tau_0 - \tau_n| \leqq b_0$ then

$$\left|\sum_{t=0}^{n-1} (\Psi(\mathbf{Y}_t, \tau_0) + \lambda(T_n))\right| \Big/ (\sqrt{n} + n|\lambda(T_n)|)$$
$$\leqq \sup_{\tau \in C} Z_n(\tau) + \left|\sum_{t=0}^{n-1} \Psi(\mathbf{Y}_t, T_n)\right| \Big/ \sqrt{n}.$$

Lemma 3.5. $(1/\sqrt{n}) \sum_{t=0}^{n-1} \Psi(\mathbf{Y}_t, \tau_0) + \sqrt{n}\,\lambda(T_n) \to \mathbf{O}(P),\ n \to \infty.$

Proof. For each $n = 1, 2, \ldots$ let $G_n = \sum_{t=0}^{n-1} (\Psi(\mathbf{Y}_t, \tau_0) + \lambda(T_n))$. Let $\varepsilon > 0$ be an arbitrary real number. From Lemma 3.1 there exists a constant $M < \infty$ such that $P(A_n) < \varepsilon$ for all n, where $A_n = \left\{(1/\sqrt{n}) \left|\sum_{t=0}^{n-1} \Psi(\mathbf{Y}_t, \tau_0)\right| > M\right\}$; let B_n be the set defined by $B_n = \{|G_n|/(\sqrt{n} + n|\lambda(T_n)|) > 1/2\}$. A straightforward calculation shows that:

$$|G_n|/\sqrt{n} \leqq 2(1 + M)\, |G_n|/(\sqrt{n} + n\,|\lambda(T_n)|)$$

on $A_n^c \cap B_n^c$, since

$$\sqrt{n}\,|\lambda(T_n)| - |G_n - n\lambda(T_n)|/\sqrt{n} \leqq |G_n|/\sqrt{n} \leqq (1 + \sqrt{n}\,|\lambda(T_n)|)/2$$

on B_n^c. From Corollary 3.4 we see that $\limsup_n P(|G_n|/\sqrt{n} > d) \leqq \varepsilon$ for all $d > 0$.

To complete the proof of Theorem 3.2 note that:

$$\sqrt{n}\,\Lambda(T_n - \tau_0) = -\sqrt{n}(\lambda(T_n) - \lambda(\tau_0) - \Lambda(T_n - \tau_0)) + \sqrt{n}(\lambda(T_n) - \lambda(\tau_0)).$$

and

$$\sqrt{n}(\lambda(T_n) - \lambda(\tau_0)) = -(1/\sqrt{n}) \sum_{t=0}^{n-1} \Psi(\mathbf{Y}_t, \tau_0)$$
$$+ (1/\sqrt{n}) \sum_{t=0}^{n-1} (\Psi(\mathbf{Y}_t, \tau_0) + \lambda(T_n)).$$

Now, it suffices to show that $\sqrt{n}(\lambda(T_n) - \lambda(\tau_0) - \Lambda(T_n - \tau_0)) \to \mathbf{O}(P),\ n \to \infty$. This is easily proved since $(\sqrt{n}\,|T_n - \tau_0|)$ is bounded in probability (see the last Remark and take into account that $(\sqrt{n}(\lambda(T_n) - \lambda(\tau_0)))$ is bounded in probability.

Acknowledgements. The author of this work is greatly indebted to his friends and professors Ricardo A. Maronna and Victor J. Yohai for their guidance and suggestions. Also he acknowledges the useful suggestions given by the referees, in particular the reference to the paper of Hannan and Kanter [7].

References

1. Anderson, T.W.: The statistical analysis of time series. New York: Wiley 1971
2. Billingsley, P.: Convergence of probability measures. New York: Wiley 1968
3. Breiman, L.: Probability. Reading: Addison-Wesley 1968
4. Brillinger, D.R.: A power spectral estimate which is insensitive to transients. Technometrics, **15**, 559–562 (1973)
5. Bustos, O.H.: Asymptotic behavior of general M-estimates for contaminated autoregressive process. Notas e Comunicações N° 5, Universidade Federal de Pernambuco, Recife, Brasil 1979
6. Denby, L., Martin, R.D.: Robust estimation of the first order autoregressive parameter. J. Amer. Statist. Ass. **74**, 140–143 (1979)
7. Hannan, E.J., Kanter, M.: Autoregressive processes with infinite variance. J. Appl. Probab. **14**, 411–415 (1977)
8. Huber, P.J.: The behavior of maximum likelihood estimates under non-standard conditions. Proc. Fifth Berkeley Sympos. Math. Statist. Probab. **1**, 221–233, University of California Press 1967
9. Huber, P.J.: Robust statistics: a review, Ann. Math. Statist. **43**, 1041–1067 (1972)
10. Kanter, M., Steiger, W.L.: Regression and autoregression with infinite variance. Adv. Appl. Probab. **6**, 768–783 (1974)
11. Maronna, R.A., Yohai, V.J.: Consistency of general M-estimates for regression and scale with random carriers. Submitted to publication to Z. Wahrscheinlichkeitstheorie verw. Geb.
12. Maronna, R.A., Yohai, V.J., Bustos, O.H.: Bias- and efficiency-robustness of general m-estimates for regression with random carriers. Smoothing Techniques for Curve Estimation (Proc. Heidelberg, 1979) eds. Gasser, Th. and Rosenblatt, M. Lectures Notes in Mathematics N° **757.** Berlin-Heidelberg-New York: Springer 1979
13. Martin, R.D.: Robust estimation for time series. Robustness in Statistics; eds. R.L. Launer and G. Wilkinson. New York: Academic Press 1979
14. Rao, R.R.: Relation between weak and uniform convergence of measures with applications. Ann. Math. Statist., **33**, 659–680 (1962)
15. Tukey, J.W.: A survey of sampling from contaminated distributions. Contributions to Probability and Statistics. Stanford; ed. I. Olkin: Stanford University Press 1960
16. Yohai, V.J., Maronna, R.A.: Asymptotic behavior of least squares estimates for autoregressive processes with infinite variances. Ann. Statist., **5**, 554–560 (1977)

Received December 5, 1979; in revised form October 15, 1981