Stochastic Processes and their Applications 40 (1992) 145-180
North-Holland

M-estimation for autoregressions with infinite variance

Richard A. Davis*
Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA

Keith Knight**
Statistics Department, University of Toronto, Toronto, Ont., Canada M5S 1A1

Jian Liu
Statistics Department, University of British Columbia, Vancouver, BC, Canada V6T 1W5

Received 26 June 1990
Revised 6 December 1990

We study the problem of estimating autoregressive parameters when the observations are from an AR process with innovations in the domain of attraction of a stable law. We show that non-degenerate limit laws exist for M-estimates if the loss function is sufficiently smooth; these results remain valid if location and scale are also estimated. For least absolute deviation (LAD) estimates, similar results hold under conditions on the innovations distribution near 0. We also discuss, under moment conditions on the innovations, consistency properties for M-estimators corresponding to the class of loss functions, $\rho(x)=|x|^\gamma$ for some $\gamma>0$.

AMS 1980 Subject Classifications: 62M09, 60G10, 62M10.

AR processes * M-estimation * least squares estimation * least absolute deviation * stable distribution * domain of attraction * point processes

1. Introduction

Let $\{X_t\}$ be the causal autoregressive AR(p) process satisfying the recursions

$$X_t=\phi_1 X_{t-1}+\cdots+\phi_p X_{t-p}+Z_t$$

where $\{Z_t\}$ is a sequence of i.i.d. random variables. Based on the data $X_1,\ldots,X_n$, the M-estimate, $\hat{\boldsymbol{\phi}}$, of $\boldsymbol{\phi}=(\phi_1,\ldots,\phi_p)'$ minimizes the objective function

$$U_n(\boldsymbol{\beta})=\sum_{t=p+1}^{n}\rho(X_t-\beta_1 X_{t-1}-\cdots-\beta_p X_{t-p})$$

with respect to $\boldsymbol{\beta}$, where $\rho(\cdot)$ is some loss function. The special cases $\rho(x)=x^2$ and $\rho(x)=|x|$ correspond to least squares (LS) and least absolute deviation (LAD)

* Research supported by NSF Grant No. MCS 8802559.
** Research partially supported by ONR Contract N00014-84-C-0169 and the Natural Sciences and Engineering Research Council of Canada.

0304-4149/92/$05.00

estimators, respectively. In this paper, we are concerned primarily with the asymptotic behavior of M-estimators when the distribution of Z_1 is in the domain of attraction of a stable distribution with index $\alpha < 2$ (written as $Z_1 \in D(\alpha)$). In other words,

$$P[|Z_1| > x] = x^{-\alpha} L(x) \tag{1.1}$$

where $L(x)$ is a slowly varying function at ∞ and

$$\lim_{x \to \infty} \frac{P[Z_1 > x]}{P[|Z_1| > x]} = p, \quad 0 \leq p \leq 1, \tag{1.2}$$

Although the AR process under assumptions (1.1) and (1.2) has an infinite variance and an infinite mean if $\alpha < 1$, the LS and LAD estimators perform surprisingly well. This phenomenon and the motivation for studying such estimators can be explained heuristically by making an analogy to linear regression. Consider, for example, the AR(1) process

$$X_t = \phi X_{t-1} + Z_t.$$

A realization of such a process of length 50 with Cauchy innovations and $\phi = 0.7$ is displayed in Figure 1. The LS and LAD estimates of ϕ are obtained by fitting a

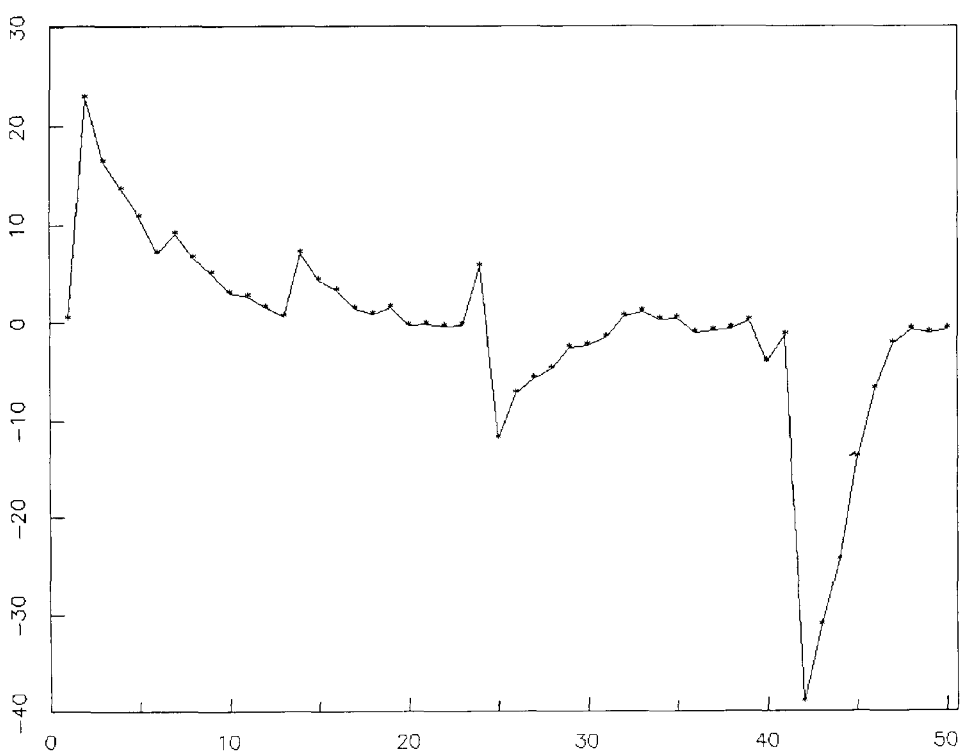

Fig. 1. 50 observations from AR(1) process $X_t = 0.7X_{t-1} + Z_t$ where Z_t is Cauchy.

straight line through the origin to the scatterplot of X_t vs X_{t-1} $(t=1,\ldots,n)$ (see Figure 2); the estimate will be the slope of the line fitted by the LS or LAD criterion. The lag 1 scatter plot has two distinct but related characteristics. First, large positive or negative values of Z_t produce points which will appear to be *outliers* (these points are labeled O in Figure 2). Second, these same Z_t produce a sequence of *leverage* points; that is, for $s>t$ we will have

$$X_s \approx \phi X_{s-1}$$

in the sense that Z_s/X_{s-1} will likely be small (see Figure 3). The beneficial effect of these leverage points compensates for the negative effect of the outliers and allows both the LAD and LS estimators to converge at a faster rate than in the finite variance setting. However, as in linear regression, the LAD estimator gives less weight to the outliers while giving essentially the same weight to the leverage points and, hence, it is reasonable to expect that the LAD estimator is more efficient. Our analysis shows that this is indeed true for a large class of heavy tailed innovation distributions.

Consistency and rates of convergence of the LS estimator $\hat{\boldsymbol{\phi}}_{\mathrm{LS}}$ for the AR(p) model have been studied by a number of people under the assumptions (1.1) and (1.2). Kanter and Steiger (1974) established weak consistency of $\hat{\boldsymbol{\phi}}_{\mathrm{LS}}$ which was

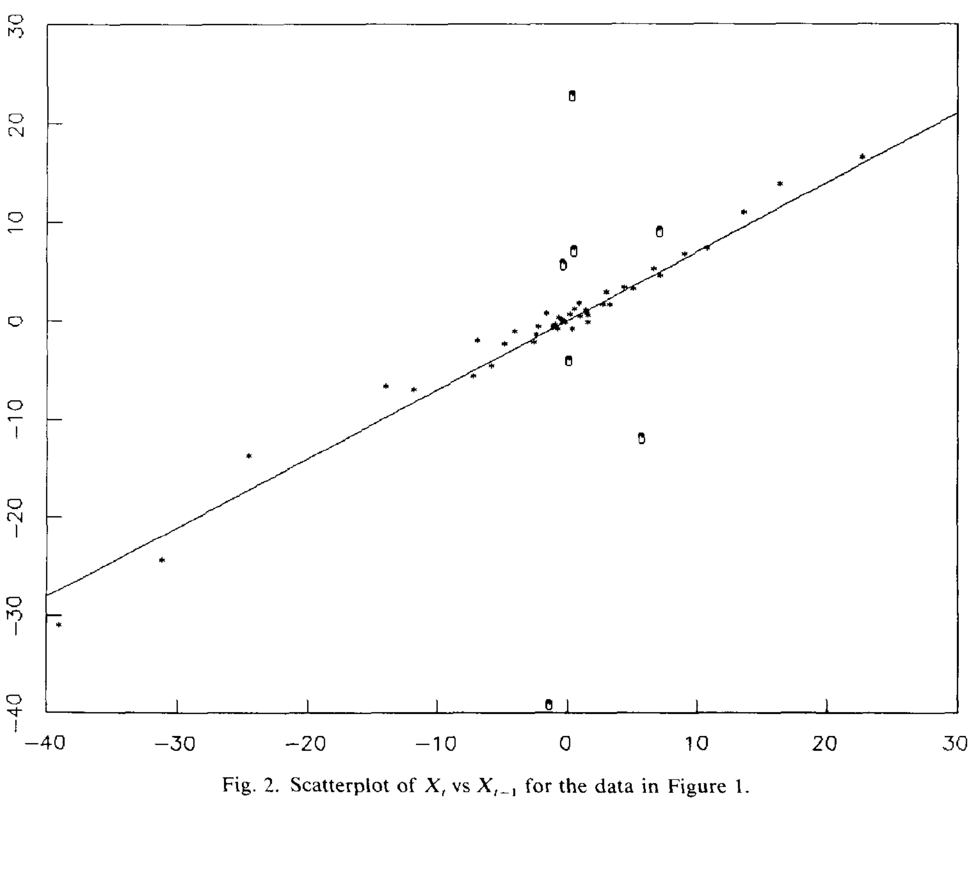

Fig. 2. Scatterplot of X_t vs X_{t-1} for the data in Figure 1.

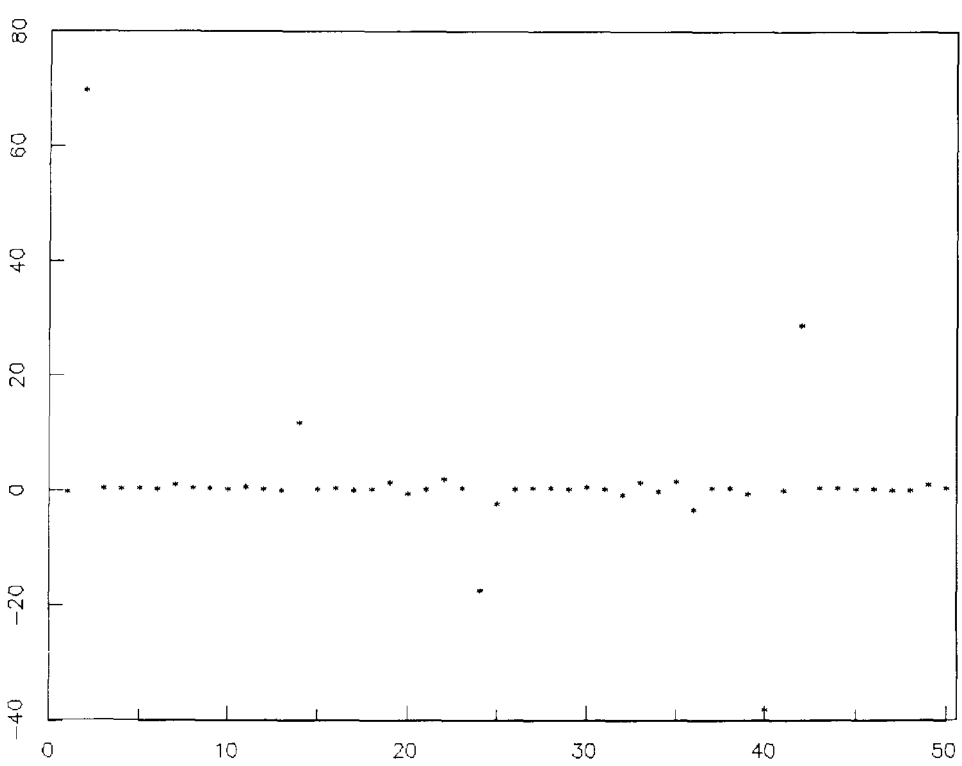

Fig. 3. Plot of Z_t/X_{t-1}.

subsequently strengthened by Hannan and Kanter (1977). They showed that if $0<\alpha<2$, and $\delta>\alpha$, then

$$n^{1/\delta}(\hat{\boldsymbol{\phi}}_{\mathrm{LS}}-\boldsymbol{\phi})\to 0 \quad \text{a.s.} \tag{1.3}$$

A similar rate was derived by Knight (1987) when an unknown location parameter is included in the model. More recently, Davis and Resnick (1985b, 1986) showed that there exists a slowly varying function $L_0(n)$ such that

$$n^{1/\alpha}L_0(n)(\hat{\boldsymbol{\phi}}_{\mathrm{LS}}-\boldsymbol{\phi})\xrightarrow{\mathrm{d}} Y \tag{1.4}$$

where Y is the ratio of two stable random variables. This immediately furnishes a convergence in probability version of (1.3).

Gross and Steiger (1979) established the strong consistency of the LAD estimator $\hat{\boldsymbol{\phi}}_{\mathrm{LAD}}$ under the assumptions that Z_1 has a unique median at zero and $E|Z_1|<\infty$. An and Chen (1982) were able to give a rate of convergence of $\hat{\boldsymbol{\phi}}_{\mathrm{LAD}}$ provided that either Z_1 has a unique median at zero and $Z_1\in D(\alpha)$, $1<\alpha<2$, or Z_1 has a Cauchy distribution centered at zero. In particular they showed that for $\delta>\alpha$,

$$n^{1/\delta}(\hat{\boldsymbol{\phi}}_{\mathrm{LAD}}-\boldsymbol{\phi})\xrightarrow{\mathrm{p}} 0. \tag{1.5}$$

Although An and Chen believed that a similar rate was valid in the $\alpha \in (0,1)$ case, they were unable to push their methods through.

In Section 2, we establish the weak convergence of the M-estimator, $\hat{\boldsymbol{\phi}}$, for the case when ρ is convex with a Lipschitz continuous derivative. Specifically, with a_n defined by

$$a_n = \inf\{x\colon P[|Z_1| > x] \leq n^{-1}\}$$

we show that

$$a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{\mathrm{d}} \boldsymbol{\xi} \tag{1.6}$$

where $\boldsymbol{\xi}$ is the minimum of a stochastic process. For Pareto-like tails, we may take $a_n = n^{1/\alpha}$ but in general $a_n = n^{1/\alpha} L_1(x)$ for some slowly varying function L_1. In Section 3, we incorporate location and scale parameters into the problem.

Unfortunately, the LAD estimators are disqualified from the theorems of Section 2 since $\rho(x) = |x|$ does not have a Lipschitz continuous derivative. Nevertheless, (1.6) remains valid for LAD estimators if $\alpha < 1$ which, in particular, proves the conjecture of An and Chen, or if $\alpha \geq 1$ and $1/Z_1$ satisfies a suitable moment condition which involves the behavior of the distribution of Z_1 near 0. Without the moment assumption, we show

$$a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) = \mathrm{O}_{\mathrm{p}}(1)$$

and in some instances this rate may be improved to $\mathrm{o}_{\mathrm{p}}(1)$. The LAD related results are contained in Section 4.

In Section 5, we focus exclusively on the case $\rho(x) = |x|^\gamma$ for some $\gamma > 0$. If $E|Z_1|^\gamma < \infty$ and $m(x) = E|Z_1 - x|^\gamma$ has a unique minimum at $x = 0$, then the resulting estimate $\hat{\boldsymbol{\phi}}$ is strongly consistent. On the other hand if $E|Z_1|^\gamma = \infty$ and Z_1 satisfies conditions (1.1) and (1.2), then $\hat{\boldsymbol{\phi}} \xrightarrow{\mathrm{P}} \boldsymbol{\phi}$.

If the class of loss functions is restricted to be of the form, $\rho(x) = |x|^\gamma$, then a natural question is that for a given $\alpha \in (0,2)$, what is an optimal choice for γ? In several simulation studies for AR(p) process with stable noise and $\alpha \in [1,2)$ (Gross and Steiger, 1979; Knight, 1986; Liu, 1987), the LAD estimator appears to be vastly superior to the LS estimator. This phenomenon has also been observed for other distributions of the noise variables by Bloomfield and Steiger (1983). The superiority of the LAD estimator over the LS estimator can partly be explained by comparing the rates in (1.4) and (1.6). If the distribution of the noise is stable or Pareto-like (see Brockwell and Davis (1987), Section 12.5) then $L_0(n)$ in (1.4) is $(\ln n)^{-1/\alpha}$ while $a_n = (\mathrm{const})n^{1/\alpha}$. Thus, since $n^{1/\alpha} L_0(n)/a_n \to 0$,

$$\|\hat{\boldsymbol{\phi}}_{\mathrm{LAD}} - \boldsymbol{\phi}\| / \|\hat{\boldsymbol{\phi}}_{\mathrm{LS}} - \boldsymbol{\phi}\| \xrightarrow{\mathrm{P}} 0$$

which *proves* the conjecture stated on p. 105 of Bloomfield and Steiger (1983). Note that this argument does not work if $E|Z_1|^\alpha < \infty$, since in this case $n^{1/\alpha} L(n) \sim a_n$ (see Davis and Resnick, 1985b). In any event, for stable noise, the simulation studies of Knight (1986) and Liu (1987) strongly suggest that for $\alpha \in [1,2)$, $\gamma = 1$ is optimal while for $\alpha \in (0, 1)$, $\gamma = \alpha$ is optimal. Certainly, there is scope for further research on this issue.

The proofs of the main results of this paper rely heavily on point process techniques for moving averages as can be found in Davis and Resnick (1985a). A discussion of these techniques and other required technical results are relegated to the appendix.

2. Limit theory for M-estimates

Let $\{X_t\}$ be the causal AR(p) process satisfying the difference equations

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t, \quad t = 0, \pm 1, \ldots, \tag{2.1}$$

where $\{Z_t\}$ is an i.i.d. sequence of r.v.'s whose common distribution belongs to the domain of attraction of a stable law with index $\alpha \in (0, 2)$, which we denote by $Z_0 \in D(\alpha)$ or $\{Z_t\} \in D(\alpha)$, and $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for all complex z with $|z| \leq 1$. The conditions

$$P(|Z_0| > x) = x^{-\alpha} L(x)$$

and

$$\lim_{x \to \infty} \frac{P(Z_1 > x)}{P(|Z_0| > x)} = p$$

where $L(x)$ is slowly varying at ∞, $\alpha > 0$, and $0 \leq p \leq 1$, are necessary and sufficient for $Z_0 \in D(\alpha)$.

The AR process (2.1) can be represented as a linear process,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

where $\{\psi_j, j = 0, 1, \ldots\}$ are the coefficients of z^j in the power series expansion of $1/\phi(z)$. It can be shown that (see Cline, 1983)

$$\lim_{x \to \infty} \frac{P(|X_t| > x)}{P(|Z_t| > x)} = \sum_{j=0}^{\infty} |\psi_j|^{\alpha}.$$

(This result applies to more general linear processes.) Thus the tails of the distribution of $\{X_t\}$ behave the same as those of $\{Z_t\}$.

The M-estimate, $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)'$, of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ minimizes the objective function

$$\sum_{t=p+1}^{n} \rho(X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}). \tag{2.2}$$

with respect to $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. The traditional approach to determining the asymptotic behavior of $\hat{\boldsymbol{\phi}}$ involves the partial derivatives of the objective function; $\hat{\phi}_1, \ldots, \hat{\phi}_p$ satisfy

$$\sum_{t=p+1}^{n} X_{t-j} \psi(X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p}) = 0 \quad \text{for } j = 1, \ldots, p,$$

where $\psi(\cdot)$ is the derivative of $\rho(\cdot)$. Making a Taylor series expansion of this equation around the true parameter vector $\boldsymbol{\phi}$, one gets, for example, in the AR(1) case,

$$0=\sum_{t=2}^{n} X_{t-1}\psi(Z_t)-(\hat{\phi}_1-\phi_1)\sum_{t=2}^{n} X_{t-1}^2\psi'(Z_t)+R_n.$$

In the finite variance case, the standard approach is to divide both sides of this equation by $\sqrt{n}$ and solve for $\sqrt{n}(\hat{\phi}_1-\phi_1)$; letting $n\to\infty$, one gets a limiting normal r.v. since $n^{-1/2}R_n \xrightarrow{\mathrm{P}} 0$ typically. The logically analogous approach in the infinite variance case would be to divide both sides by a_n and solve for $a_n(\hat{\phi}_1-\phi_1)$; the problem with this approach is that $a_n^{-1}R_n$ no longer goes to 0. For example, for any integer $k\geq 2$ and any function $h(\cdot)$ with $E(h^2(Z_1))<\infty$, it can be shown by (a) and (b) of Propositon A.2 in the Appendix that

$$a_n^{-k}\sum_{t=1}^{n} X_{t-1}^k h(Z_t)=O_p(1)$$

and, in fact, converges in distribution. Thus, an explicit representation of the limiting r.v. is not possible, in general. The approach that we will employ is similar in spirit to the approach used in obtaining the minimum Cramér–von Mises distance estimate of location; see Shorack and Wellner (1986, pp. 254–257) for details.

Note that the parameter estimates $\hat{\boldsymbol{\phi}}$ minimizing the objective function (2.2) also minimize the modified objective function

$$\sum_{t=p+1}^{n}[\rho(X_t-\beta_1X_{t-1}-\cdots-\beta_pX_{t-p})-\rho(Z_t)]$$

which can be rewritten as

$$\sum_{t=p+1}^{n}[\rho(Z_t-a_n(\beta_1-\phi_1)a_n^{-1}X_{t-1}-\cdots-a_n(\beta_p-\phi_p)a_n^{-1}X_{t-p})-\rho(Z_t)], \quad (2.3)$$

where

$$a_n=\inf\{x: P[|Z_1|>x]\leq n^{-1}\}. \quad (2.4)$$

We will consider the following sequence of stochastic processes on $\mathbb{R}^p$:

$$W_n(\boldsymbol{u})=\sum_{t=p+1}^{n}[\rho(Z_t-u_1a_n^{-1}X_{t-1}-\cdots-u_pa_n^{-1}X_{t-p})-\rho(Z_t)]$$

which, under certain conditions, will converge in distribution on $C(\mathbb{R}^p)$, the space of continuous functions mapping $\mathbb{R}^p$ to $\mathbb{R}$, to a non-trivial process $W(\cdot)$ which has a unique minimum with probability 1. (The fact that the sum is started at $t=1$ rather than $t=p+1$ is not important since

$$\sum_{t=1}^{p}\rho(Z_t-a_n^{-1}X_{t-1}u_1-\cdots-a_n^{-1}X_{t-p}u_p)\xrightarrow{\mathrm{P}}\sum_{t=1}^{p}\rho(Z_t)$$

by the continuity of $\rho(\cdot)$.) Since the vector ξ_n which minimizes $W_n(\cdot)$ is simply $a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})$, it is reasonable to expect that

$$a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{\mathrm{d}} \text{to some random vector } \boldsymbol{\xi}.$$

This will be a direct consequence of the following theorem.

Theorem 2.1. *Suppose that* $\{X_t\}$ *is the* AR *process* (2.1) *with innovations* $\{Z_t\} \in D(\alpha)$, $0 < \alpha < 2$, *and*:

(a) ψ *satisfies a Lipschitz condition of order* β; $|\psi(x) - \psi(y)| \leq K|x-y|^\beta$ *where* $\beta > \max(\alpha - 1, 0)$ *and* K *is a constant.*

(b) $E(|\psi(Z_1)|) < \infty$ *if* $\alpha < 1$.

(c) $E(\psi(Z_1)) = 0$ *and* $\mathrm{Var}(\psi(Z_1)) < \infty$ *if* $\alpha \geq 1$.

Then on $C(\mathbb{R}^p)$,

$$W_n(\cdot) \xrightarrow{\mathrm{d}} W(\cdot)$$

where

$$W(\boldsymbol{u}) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} [\rho(Z_{k,i} - (\psi_{i-1}u_1 + \cdots + \psi_{i-p}u_p)\delta_k \Gamma_k^{-1/\alpha}) - \rho(Z_{k,i})] \tag{2.5}$$

and the sequences $\{Z_{k,i}\}$, $\{\delta_k\}$, *and* $\{\Gamma_k\}$ *are as specified in Proposition* A.1 *of the Appendix. (The infinite sum defining* $W(\cdot)$ *is interpreted as a limit of partial sums provided this limit exits.)*

Note that loss function $\rho(x) = |x|^\gamma$ satisfies the assumptions of the theorem provided $\gamma > 1$ (condition (a)), $\gamma < 1 + \alpha$ if $\alpha < 1$ (condition (b)), and $\gamma < \frac{1}{2}\alpha + 1$ if $\alpha > 1$ (condition (c)). Of course the latter assumes that $E\psi(Z_1) = 0$.

Proof. First of all, we will show that the finite dimensional distributions converge weakly. For simplicity, we will deal only with the univariate distributions; the multivariate case follows by applying the Cramér–Wold device.

Let $Y_{nt}(\boldsymbol{u}) = u_1 a_n^{-1} X_{t-1} + \cdots + u_p a_n^{-1} X_{t-p}$. By (A.8), it follows that

$$W_n(\boldsymbol{u}; \delta, M) = \sum_{t=1}^{n} [\rho(Z_t - Y_{nt}(\boldsymbol{u})) - \rho(Z_t)] I(|Z_t| \leq M, |Y_{nt}| > \delta)$$

converges in distribution to

$$\begin{aligned} &W(\boldsymbol{u}; \delta, M) \\ &= \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} ([\rho(Z_{k,i} - (\psi_{i-1}u_1 + \cdots + \psi_{i-p}u_p)\delta_k \Gamma_k^{-1/\alpha}) - \rho(Z_{k,i})] \\ &\qquad \times I(|Z_{k,i}| \leq M, |\psi_{i-1}u_1 + \cdots + \psi_{i-p}u_p| \Gamma_k^{-1/\alpha} > \delta)) \end{aligned}$$

where $I(A)$ is the indicator function of the set A. It suffices to show that (A.10) and (A.11) hold with the function $f(x, y) = \rho(x-y) - \rho(x)$. A Taylor series expansion around Z_t for each term of $W_n(\cdot)$ yields

$$\begin{aligned}\sum_{t=1}^{n} [\rho(Z_t - Y_{nt}(\boldsymbol{u})) - \rho(Z_t)] &= -\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u})\psi(\xi_t^{(n)}) \\ &= -\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u})\psi(Z_t) + \sum_{t=1}^{n} Y_{nt}(\boldsymbol{u})(\psi(Z_t) - \psi(\xi_t^{(n)})) \qquad (2.6)\end{aligned}$$

where $|\xi_t^{(n)} - Z_t| \leq |Y_{nt}(\boldsymbol{u})|$. Now with $V_t = \psi(Z_t)I(|Z_t| > M)$, it follows from Proposition A.2(b) that

$$\lim_{M\to\infty} \limsup_{n\to\infty} P\left[\left|\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u})I(|Y_{nt}(\boldsymbol{u})| > \delta)\psi(Z_t)I(|Z_t| > M)\right| > \eta\right] = 0. \qquad (2.7)$$

Moreover, applying Proposition A.2(a) if $\alpha < 1$ and (c) if $\alpha \geq 1$ with $V_t = \psi(Z_t)$ we deduce

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\left[\left|\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u})\psi(Z_t)I(|Y_{nt}(\boldsymbol{u})| \leq \delta)\right| > \eta\right] = 0. \qquad (2.8)$$

Also using the Lipschitz continuity of $\psi(\cdot)$ ($|\psi(x) - \psi(y)| \leq K|x-y|^\beta$), we get

$$\begin{aligned}&\left|\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u})(\psi(\xi_t^{(n)}) - \psi(Z_t))I(|Y_{nt}(\boldsymbol{u})| \leq \delta)\right| \\ &\leq K \sum_{t=1}^{n} |Y_{nt}(\boldsymbol{u})|^{1+\beta} I(|Y_{nt}(\boldsymbol{u})| \leq \delta) \\ &\xrightarrow{\text{P}} 0 \quad \text{as } n\to\infty \text{ and then } \delta\to 0 \qquad (2.9)\end{aligned}$$

by Proposition A.2(a) and, similarly,

$$\begin{aligned}&\left|\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u})(\psi(\xi_t^{(n)}) - \psi(Z_t))I(|Y_{nt}(\boldsymbol{u})| > \delta)I(|Z_t| > M)\right| \\ &\leq K \sum_{t=1}^{n} |Y_{nt}(\boldsymbol{u})|^{1+\beta} I(|Y_{nt}(\boldsymbol{u})| > \delta)I(|Z_t| > M) \\ &\xrightarrow{\text{P}} 0 \quad \text{as } n\to\infty,\ M\to\infty, \qquad (2.10)\end{aligned}$$

by Proposition A.2(b). Combining (2.6)-(2.10) proves (A.10). Finally (A.11) is proved in a similar fashion.

Finally, we show that the distributions of $\{W_n(\cdot)\}$ are tight. First, note that

$$\begin{aligned}|W_n(\boldsymbol{u}) - W_n(\boldsymbol{v})| &= \left|\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u}-\boldsymbol{v})\psi(\xi_t^{(n)})\right| \\ &\leq \left|\sum_{t=1}^{n} Y_{nt}(\boldsymbol{u}-\boldsymbol{v})\psi(Z_t)\right| + K\sum_{t=1}^{n} |Y_{nt}(\boldsymbol{u}-\boldsymbol{v})|^{1+\beta}\end{aligned}$$

since $|\xi_t^{(n)} - Z_t| \leq |Y_{nt}(\boldsymbol{u}-\boldsymbol{v})|$ and hence $|\psi(\xi_t^{(n)}) - \psi(Z_t)| \leq K|Y_{nt}(\boldsymbol{u}-\boldsymbol{v})|^\beta$. Now by Proposition A.2(a) and (b) and a slight modification of the proof of Theorem 4.2 in Davis and Resnick (1984), we have

$$a_n^{-1} \sum_{t=1}^{n} X_{t-r}\psi(Z_t) = O_p(1)$$

and

$$a_n^{-(1+\beta)} \sum_{t=1}^{n} |X_{t-r}|^{1+\beta} = O_p(1),$$

from which it follows that

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\left[\sup_{\|\boldsymbol{u}-\boldsymbol{v}\|\leq\delta} |W_n(\boldsymbol{u}) - W_n(\boldsymbol{v})| > \eta\right] = 0$$

as required. Thus $W_n(\cdot) \xrightarrow{d} W(\cdot)$ on $C(\mathbb{R}^p)$. □

Lemma 2.2. *Let $\{V_n(\cdot)\}$ and $V(\cdot)$ be stochastic processes on $\mathbb{R}^p$ and suppose that*

$$V_n(\cdot) \xrightarrow{d} V(\cdot) \quad \text{on } C(\mathbb{R}^p).$$

Let $\boldsymbol{\xi}_n$ minimize $V_n(\cdot)$ and $\boldsymbol{\xi}$ minimize $V(\cdot)$. If $V_n(\cdot)$ is convex for each n and $\boldsymbol{\xi}$ is unique with probability 1 *then*

$$\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\xi} \quad \text{on } \mathbb{R}^p.$$

Proof. By Skorokhod's representation theorem, there exists a probability space with processes $\{V_n^*(\cdot)\}$ and $V^*(\cdot)$ having the same finite dimensional distributions as $\{V_n(\cdot)\}$ and $V(\cdot)$ such that for any given compact set K,

$$\sup_{\boldsymbol{u}\in K} |V_n^*(\boldsymbol{u}) - V^*(\boldsymbol{u})| \xrightarrow{\text{a.s.}} 0.$$

Let $\boldsymbol{\xi}_n^*$ and $\boldsymbol{\xi}^*$ minimize V_n^* and V^* respectively; if we show for these special processes that $\boldsymbol{\xi}_n^* \xrightarrow{\text{a.s.}} \boldsymbol{\xi}^*$ then, in general, we will have $\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\xi}$. Henceforth, we will argue for each ω in the special probability space for which

$$\sup_{\boldsymbol{u}\in K} |V_n^*(\boldsymbol{u}) - V^*(\boldsymbol{u})| \to 0.$$

For $\gamma > 0$, let

$$B_\gamma = \{\boldsymbol{u}: \|\boldsymbol{u} - \boldsymbol{\xi}^*\| = \gamma\}.$$

and suppose $\|\boldsymbol{\xi}_n^* - \boldsymbol{\xi}^*\| > \gamma$ for infinitely many n. Since

$$V_n^*(\boldsymbol{u}) \to V^*(\boldsymbol{u}) \quad \text{uniformly on } B_\gamma$$

and

$$V_n^*(\boldsymbol{\xi}^*) \to V^*(\boldsymbol{\xi}^*)$$

we have for infinitely many n and all $\boldsymbol{u} \in B_\gamma$,

$$V_n^*(\boldsymbol{u}) > V_n^*(\boldsymbol{\xi}^*) \geq V_n^*(\boldsymbol{\xi}_n^*).$$

But this contradicts the convexity of $V_n^*(\cdot)$ by choosing $\boldsymbol{u} \in B_\gamma$ such that the points $\boldsymbol{u}$, $\boldsymbol{\xi}^*$, $\boldsymbol{\xi}_n^*$ are collinear. □

Remark 1. Lemma 2.2 can also be generalized somewhat. If the convexity assumption on the $V_n(\cdot)$'s is removed then there exists a sequence of local minima $\{\boldsymbol{\xi}_n\}$ converging in distribution to $\boldsymbol{\xi}$, the unique minimum of $V(\cdot)$. It can also be shown that, for convex processes, weak convergence of the finite dimensional distributions implies convergence in distribution of the processes; this follows from the fact that pointwise convergence of convex functions implies uniform convergence on compact sets (see Theorem 10.8 of Rockafellar, 1970).

Theorem 2.3. *Under the conditions of Theorem* 2.1, *if ρ is convex and $W(\cdot)$ attains a unique minimum $\boldsymbol{\xi}$, a.s., then the AR parameter estimate $\hat{\boldsymbol{\phi}}$ defined by minimizing* (2.2) *satisfies*

$$a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{\mathrm{d}} \boldsymbol{\xi}.$$

Proof. The proof follows from Theorem 2.1 and Lemma 2.2. □

Remark 2. In case $\rho(\cdot)$ is strictly convex, i.e. $\psi(\cdot)$ strictly increasing, then $W(\cdot)$ will also be strictly convex and hence has a unique minimum. One can get by with weaker conditions on ρ which ensure uniqueness of the minimum of the limit process. For example, assume that with positive probability $\rho(Z_1 + \cdot)$ is strictly convex in a neighborhood of 0. Then as before, $W(\cdot)$ will be strictly convex a.s. with a unique minimum.

Theorem 2.3 shows that non-degenerate limit laws exist for certain M-estimates; unfortunately, there seems to be no easy method of calculating the limiting distribution. Moreover, this limiting distribution appears to be highly dependent on the distribution of the innovations.

3. Unknown location and scale

In general, the estimates $\hat{\beta}$ defined by minimizing (2.2) are not scale invariant as would be desirable. This lack of invariance can be remedied by modifying (2.2) to

$$\sum_{t=p+1}^{n} \rho\left(\frac{X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}}{\hat{s}}\right)$$

where $\hat{s}$ is some (scale equivariant) estimate of the innovations scale. It follows that if $n^\gamma(\hat{s} - s) = \mathrm{O}_\mathrm{p}(1)$ for some $\gamma > 0$ then $\hat{\boldsymbol{\phi}}$ will have the same asymptotic properties described in Theorem 2.3.

A variation of the basic model includes an unknown location parameter,

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t.$$

We now obtain estimates $\hat{\phi}_0, \hat{\phi}_1, \ldots, \hat{\phi}_p$ by minimizing

$$\sum_{t=p+1}^{n} \rho(X_t - \beta_0 - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}).$$

The asymptotic properties of these estimates can be studied by considering the process

$$W_n^*(\boldsymbol{u}) = \sum_{t=1}^{n} [\rho(Z_t - n^{-1/2}u_0 - a_n^{-1}X_{t-1}u_1 - \cdots - a_n^{-1}X_{t-p}u_p) - \rho(Z_t)]$$

where now $u_0 = n^{1/2}(\beta_0 - \phi_0)$. Assuming that $\rho(\cdot)$ has Lipschitz continuous derivative $\psi(\cdot)$, it is possible to show that

$$\begin{aligned} W_n^* &= \sum_{t=1}^{n} [\rho(Z_t - a_n^{-1}X_{t-1}u_1 - \cdots - a_n^{-1}X_{t-p}u_p) - \rho(Z_t)] \\ &\quad + \sum_{t=1}^{n} [\rho(Z_t - n^{-1/2}u_0) - \rho(Z_t)] + o_p(1) \\ &= W_n(u_1, \ldots, u_p) + Z_n(u_0) + o_p(1) \end{aligned}$$

uniformly over $\boldsymbol{u}$ in compact subsets of $\mathbb{R}^{p+1}$. If now $\psi(\cdot)$ has Lipschitz continuous derivative $\psi'(\cdot)$ then

$$\begin{aligned} Z_n(u_0) &= -u_0 n^{-1/2} \sum_{t=1}^{n} \psi(Z_t) + \frac{u_0^2}{2n} \sum_{t=1}^{n} \psi'(Z_t + n^{-1/2}u_0^*) \\ &= -u_0 n^{-1/2} \sum_{t=1}^{n} \psi(Z_t) + \frac{u_0^2}{2n} \sum_{t=1}^{n} \psi'(Z_t) + o_p(1) \\ &\xrightarrow{d} Z(u_0) \end{aligned}$$

where $Z(u_0)$ is a normal r.v. with mean $\frac{1}{2}u_0^2 E(\psi'(Z_1))$ and variance $u_0^2 E(\psi^2(Z_1))$; in addition, it follows easily that $Z_n(\cdot) \xrightarrow{d} Z(\cdot)$. As before, $W_n(\cdot) \xrightarrow{d} W(\cdot)$ where $W(\cdot)$ is independent of $Z(\cdot)$. Thus $a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})$ has the same limit as in Theorem 2.3 while $n^{1/2}(\hat{\phi}_0 - \phi_0)$ converges in distribution to the minimum of $Z(\cdot)$ which is normal with mean 0 and variance $E(\psi^2(Z_1))/(E(\psi'(Z_1)))^2$. Since $W(\cdot)$ and $Z(\cdot)$ are independent, the limiting distributions of the location and AR parameter estimates are also independent.

4. LAD estimates

The LAD estimate is exlcuded from Theorem 2.1 because the function $\rho(x) = |x|$ is not differentiable at 0; consequently, one would expect the asymptotic behavior of the LAD estimate to depend heavily on the behavior of the innovations distribution at or near the origin. For example, in the finite variance case, asymptotic normality of LAD estimates depends on the innovations $\{Z_t\}$ having a density $f(\cdot)$ (with respect to Lebesgue measure) which is continuous at 0.

The LAD estimates minimize the modified objective function

$$\sum_{t=p+1}^{n} [|X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}| - |Z_t|]$$

and so by analogy to the general case, we consider the following sequence of processes:

$$W_n(\boldsymbol{u}) = \sum_{t=p+1}^{n} [|Z_t - a_n^{-1} X_{t-1} u_1 - \cdots - a_n^{-1} X_{t-p} u_p| - |Z_t|].$$

where a_n is given by (2.4). The obvious limit for $W_n(\cdot)$ is

$$W(\boldsymbol{u}) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} [|Z_{k,i} - (\psi_{i-1} u_1 + \cdots + \psi_{i-p} u_p) \delta_k \Gamma_k^{-1/\alpha}| - |Z_{k,i}|]$$

if indeed this limit is well-defined. For $\alpha < 1$, $W(\cdot)$ is well-defined (see Proposition A.3) and the convergence in distribution of $W_n(\cdot)$ to $W(\cdot)$ is trivial to prove. However, when $\alpha \geqslant 1$, we need to make some further assumptions about the distribution of $\{Z_{k,i}\}$ (or $\{Z_t\}$) in order for $W(\cdot)$ to exist and for $W_n(\cdot) \xrightarrow{d} W(\cdot)$.

Theorem 4.1. *Let $\{X_t\}$ be an* AR(p) *process with innovations $\{Z_t\} \in D(\alpha)$ having median* 0 *if $\alpha \geqslant 1$.* If either

(a) $\alpha < 1$, *or*

(b) $\alpha > 1$ *and* $E(|Z_1|^\beta) < \infty$ *for some* $\beta < 1 - \alpha$, *or*

(c) $\alpha = 1$ *and* $E(\ln(|Z_1|)) > -\infty$,

then

$$W_n(\cdot) \xrightarrow{d} W(\cdot).$$

Moreover, if $W(\cdot)$ has a unique minimum a.s., then

$$a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} \boldsymbol{\xi}$$

where $\boldsymbol{\xi}$ is the minimum of $W(\cdot)$ and $\hat{\boldsymbol{\phi}}$ is the LAD estimate of $\boldsymbol{\phi}$.

The moment conditions for $\alpha \geqslant 1$ are met if Z_1 has a density f which is bounded in a neighborhood of 0. A discussion on the uniqueness of the minimum of $W(\cdot)$ follows the proof.

Proof. For $\alpha < 1$, the proof is relatively straighforward using Proposition A.3. For $\alpha > 1$, we follow the arguments given for Theorem 2.1 with $f(x, y) = |x - y| - |x|$ and

(2.6) replaced by

$$\sum_{t=p+1}^{n} (|Z_t - Y_{nt}| - |Z_t|)$$

$$= \sum_{t=p+1}^{n} (Y_{nt}(I(Z_t<0) - I(Z_t>0)))$$

$$+2 \sum_{t=p+1}^{n} ((Y_{nt} - Z_t)(I(Y_{nt}>Z_t>0) - I(Y_{nt}<Z_t<0)))$$

where Y_{nt} is as defined in the proof of Theorem 2.1. It is easy to show that the first term on the right converges in distribution. As for the second term, we have by (A.8) (with $c_i = \psi_{i-1}u_1 + \cdots + \psi_{i-p}u_p$) and the proof of Proposition A.3 that

$$2 \sum_{t=p+1}^{n} ((Y_{nt} - Z_t)I(Y_{nt}>Z_t>0) - I(Y_{nt}<Z_t<0))I(|Y_{nt}|>\delta)$$

$$\xrightarrow{\text{d}} \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} (c_i\delta_k\Gamma_k^{-1/\alpha} - Z_{k,i})$$

$$\times (I(c_i\delta_k\Gamma_k^{-1/\alpha} > Z_{k,i}>0) - I(c_i\delta_k\Gamma_k^{-1/\alpha} < Z_{k,i}<0))$$

$$\times I(|c_i\delta_k|\Gamma_k^{-1/\alpha} > \delta) \quad (\text{as } n\to\infty)$$

$$\xrightarrow{\text{a.s.}} \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} (c_i\delta_k\Gamma_k^{-1/\alpha} - Z_{k,i})$$

$$\times (I(c_i\delta_k\Gamma_k^{-1/\alpha} > Z_{k,i}>0) - I(c_i\delta_k\Gamma_k^{-1/\alpha} < Z_{k,i}<0)) \quad (\text{as } \delta\to 0).$$

It therefore suffices to prove

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left[\sum_{t=p+1}^{n} (Y_{nt}(\boldsymbol{u}) - Z_t)I(\delta > Y_{nt}(\boldsymbol{u}) > Z_t > 0) > \gamma\right] = 0 \tag{4.1}$$

and

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left[\sum_{t=p+1}^{n} (Y_{nt}(\boldsymbol{u}) - Z_t)I(-\delta < Y_{nt}(\boldsymbol{u}) < Z_t < 0) < -\gamma\right] = 0. \tag{4.2}$$

If $G(\cdot)$ is the distribution function of $u_1X_{t-1} + \cdots + u_pX_{t-p}$, then

$$E\left[\sum_{t=p+1}^{n} (Y_{nt}(\boldsymbol{u}) - Z_t)I(\delta > Y_{nt}(\boldsymbol{u}) > Z_t > 0)\right]$$

$$= (n-p)\int_0^{\delta}\int_{a_nx}^{a_n\delta}\left(\frac{y}{a_n} - x\right)G(\mathrm{d}y)F(\mathrm{d}x) \le na_n^{-1}\int_0^{\delta} H(a_nx)F(\mathrm{d}x)$$

where $H(x) = \int_x^{\infty} yG(\mathrm{d}y)$. By Karmata's theorem, $H(x) \sim (\alpha/(1-\alpha))x(1-G(x))$, and since $n(1-G(a_nx)) \to Cx^{-\alpha}$, C a constant, we have $na_n^{-1}H(a_n) \to$ const. Using Potter's theorem (see Bingham, Goldie and Teugels, 1987), we have, with β as specified in the statement of the theorem,

$$\frac{H(zx)}{H(z)} = \frac{H(zx)}{H(zx(1/x))} \le (\text{const}) \min(x^{\beta}, 1)$$

for $zx > 1$ and all z greater than some large z_0. Since $z^\beta H(z) \to \infty$ as $z \to \infty$ we have for all $zx < 1$ and $z > z_0$,

$$\frac{H(zx)}{H(z)} \leqslant \frac{H(0)}{H(z)} \leqslant \frac{(zx)^\beta H(0)}{H(z)} \leqslant \frac{x^\beta H(0)}{z^\beta H(z)} \leqslant (\text{const}) x^\beta.$$

Thus for all n large,

$$\begin{aligned} n a_n^{-1} \int_0^\delta H(a_n x) F(\mathrm{d}x) &\sim (\text{const}) \int_0^\delta \frac{H(a_n x)}{H(a_n)} F(\mathrm{d}x) \\ &\leqslant (\text{const}) \int_0^\delta x^\beta F(\mathrm{d}x) \to 0 \quad \text{as } \delta \to 0 \end{aligned}$$

which proves (4.1). We argue (4.2) in the same fashion. □

Remark. The condition that for all $\varepsilon > 0$ there exists a constant $C > 0$ such that

$$P[x < Z_1 < y] \geqslant \begin{cases} C(y-x)^{1/\alpha}, & \text{if } \alpha < 1, \\ C(y-x), & \text{if } \alpha \geqslant 1, \end{cases}$$

whenever $-\varepsilon < x < y < \varepsilon$, is sufficient for uniqueness of the minimum of $W(\cdot)$. We show this for the case $p = 1$ and $\alpha \geqslant 1$, the general case being a straightforward extension. Let $A_{k,i}(u, v)$ denote the random interval with endpoints at $u\psi_{i-1}\delta_k\Gamma_k^{-1/\alpha}$ and $v\psi_{i-1}\delta_k\Gamma_k^{-1/\alpha}$. We then have

$$\begin{aligned} &P\left[\bigcap_{i\geqslant 1}\bigcap_{k\geqslant 1}\{Z_{k,i} \notin A_{k,i}(u,v)\}\right] \\ &\quad = E\left[\prod_{i=1}^{\infty}\prod_{k=1}^{\infty}(1 - P[Z_1 \in A_{k,i}(u,v) \mid \delta_j\Gamma_j^{-1/\alpha}, j \geqslant 1])\right] \end{aligned} \tag{4.3}$$

and since for k large,

$$\begin{aligned} P[Z_1 \in A_{k,i}(u,v) \mid \delta_j\Gamma_j^{-1/\alpha}, j \geqslant 1] &\geqslant C|u-v||\psi_{i-1}|\Gamma_k^{-1/\alpha} \\ &\sim C|u-v||\psi_{i-1}|k^{-1/\alpha} \end{aligned}$$

and $\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}|\psi_{i-1}|k^{-1/\alpha} = \infty$, it follows that the probability in (4.3) is 0. Let B denote the event

$$B = \bigcup_{v\in\mathbb{Q}}\bigcup_{u\in\mathbb{Q}}\bigcap_{i\geqslant 1}\bigcap_{k\geqslant 1}\{Z_{k,i} \notin A_{k,i}(u,v)\}, \quad \mathbb{Q} = \text{rationals},$$

which, by the preceding calculation, has probability 0. Now, if on the set B^c, $W(\cdot)$ has distinct minimizers, u^* and v^*, then by the convexity of $W(\cdot)$ there exist rationals u and v such that $W(u) = W(v)$. Since $W(\frac{1}{2}(u+v)) = \frac{1}{2}W(u) + \frac{1}{2}W(v)$, we must have for all k and i,

$$\begin{aligned} &|\tfrac{1}{2}(Z_{k,i} - u\psi_{i-1}\delta_k\Gamma_k^{-1/\alpha}) + \tfrac{1}{2}(Z_{k,i} - v\psi_{i-1}\delta_k\Gamma_k^{-1/\alpha})| \\ &\quad = \tfrac{1}{2}|Z_{k,i} - u\psi_{i-1}\delta_k\Gamma_k^{-1/\alpha}| + \tfrac{1}{2}|Z_{k,i} - v\psi_{i-1}\delta_k\Gamma_k^{-1/\alpha}| \end{aligned}$$

which, in turn, implies

$$Z_{k,i} \notin A_{k,i}(u, v).$$

This contradicts the definition of B^c and, therefore, $W(\cdot)$ must have a unique minimum a.s.

Note that if the probability in (4.3) is positive for some $u < v$, then $W(\cdot)$ will be constant on the interval (u, v) and hence will not have a unique minimum.

The case when condition (b) of Theorem 4.1 is violated can introduce some interesting pathologies. Let x^+ and x^- denote the positive and negative parts of x so that $x = x^+ - x^-$. Suppose that $c_i > 0$ for all i, $\delta_k = 1$ with probability 1 and $E[(Z_1^+)^{1-\alpha}] < \infty$ while $E[(Z_1^-)^{1-\alpha}] = \infty$. It now follows from the proof of Proposition A.3 that V is finite a.s.; however, $E(|Z_1|^{1-\alpha}) = \infty$. Note that if $c_i < 0$ for all i then the partial sums defining W will not converge. To elaborate on this example, let $\{X_t\}$ be an AR(1) process with innovations $\{Z_t\} \in D(\alpha)$ for $\alpha > 1$ and suppose that

$$\lim_{x\to\infty} \frac{P(Z_t > x)}{P(|Z_t| > x)} = 1,$$

$$E[(Z_t^+)^{1-\alpha}] < \infty \quad \text{and} \quad E[(Z_t^-)^{1-\alpha}] = \infty.$$

Suppose that $\phi_1 > 0$. Then $\psi_i = \phi_1^i > 0$ and

$$W(u) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} [|Z_{k,i} - u\psi_{i-1}\delta_k \Gamma_k^{-1/\alpha}| - |Z_{k,i}|].$$

From above $W(u) = \infty$ for $u < 0$ and $W(u) < \infty$ for $u \geq 0$. Hence $P(a_n(\hat{\phi}_1 - \phi_1) < 0) \to 0$ as $n \to \infty$.

The moment conditions in Theorem 4.1 for $\alpha \geq 1$ can be dispensed with at the cost of a slightly weaker conclusion. This is the content of the following theorem.

Theorem 4.2. *Let* $\{X_t\}$ *be an* AR(p) *process with innovations* $\{Z_t\} \in D(\alpha)$ *with* $\alpha \in [1, 2)$. *If* $P[Z_1 \geq 0] = P[Z_1 \leq 0]$, *then*

$$a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) = O_p(1).$$

where $\hat{\boldsymbol{\phi}}$ *is the LAD estimate of* $\boldsymbol{\phi}$.

Proof. Let M be fixed and set $r = a_n^{-1}M/\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\|$ and $\boldsymbol{v} = (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})/\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\|$. Then by the convexity of W_n, we have on the set $\{a_n\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\| > M\}$,

$$\begin{aligned} W_n(M\boldsymbol{v}) &\leq (1-r)W_n(\boldsymbol{0}) + rW_n(a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})) \quad (M\boldsymbol{v} = ra_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})) \\ &\leq W_n(\boldsymbol{0}) = 0 \end{aligned}$$

since $W_n(a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})) \leq W_n(0)$ by definition of $\hat{\boldsymbol{\phi}}$. Thus

$$P[a_n\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\| > M] \leq P[W_n(\boldsymbol{v}) \leq 0] \leq P\left[\inf_{\|\boldsymbol{u}\|=1} W_n(\boldsymbol{u}) \leq 0\right].$$

Since

$$|a-b|-|a|=b(I(a\leqslant 0)-I(a\geqslant 0))+2|a-b|(I(0\leqslant a\leqslant b)+I(b\leqslant a\leqslant 0))$$

for $a\neq 0$ it follows that

$$W_n(\boldsymbol{u})\geqslant T_n(\boldsymbol{u};\,\delta,M)$$

where

$$\begin{aligned}T_n(\boldsymbol{u};\,\delta,M)=M&\sum_{t=p+1}^{n} a_n^{-1}Y_t(\boldsymbol{u})(I(Z_t\leqslant 0))-(I(Z_t\geqslant 0))\\&+2\sum_{t=p+1}^{n}|Z_t-Ma_n^{-1}Y_t(\boldsymbol{u})|\\&\qquad\times(I(Ma_n^{-1}Y_t(\boldsymbol{u})\leqslant Z_t\leqslant -\delta)+I(\delta\leqslant Z_t\leqslant Ma_n^{-1}Y_t))\end{aligned}$$

for some $\delta>0$ and $Y_t(\boldsymbol{u})=u_1X_{t-1}+\cdots+u_pX_{t-p}$. If δ and $-\delta$ are continuity points of the distribution of Z_1, it follows from Proposition A.4 that in $C(\{\boldsymbol{u}\colon \|\boldsymbol{u}\|=1\})$

$$T_n(\cdot\,;\,\delta,M)\xrightarrow{\mathrm{d}}T(\cdot\,;\,\delta,M)$$

where $T(\boldsymbol{u};\,\delta,M)$ is defined in Proposition A.4. Moreover, applying the proposition once again we obtain

$$\begin{aligned}\limsup_{n\to\infty}P[a_n\|\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}\|>M]&\leqslant\limsup_{n\to\infty}P\left[\inf_{\|\boldsymbol{u}\|=1}T_n(\boldsymbol{u};\,\delta,M)\leqslant 0\right]\\&\leqslant P\left[\inf_{\|\boldsymbol{u}\|=1}T(\boldsymbol{u};\,\delta,M)\leqslant 0\right]\\&\to 0\end{aligned}$$

as $M\to\infty$ and $\delta\to 0$ which proves the theorem. □

From the arguments given for Theorem 4.1 and Proposition A.3, it would appear that under certain conditions, $W_n(\boldsymbol{u})\xrightarrow{\mathrm{p}}\infty$ for all $\boldsymbol{u}\neq 0$ which implies that $a_n(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi})=o_p(1)$. In fact, more extreme results than this are possible. For example, consider the process

$$X_t=\phi_0+\phi_1X_{t-1}+\cdots+\phi_pX_{t-p}+Z_t$$

where the innovations distribution has median 0 and positive mass at 0. Suppose we estimate $\phi_0,\phi_1,\ldots,\phi_p$ by minimizing

$$\sum_{t=p+1}^{n}|X_t-\beta_0-\beta_1X_{t-1}-\cdots-\beta_pX_{t-p}|. \tag{4.4}$$

The following theorem shows that the estimate, $\hat{\boldsymbol{\phi}}$, converges at a faster rate than usual if $E(|Z_1|)<\infty$.

Theorem 4.3. *Let $\{X_t\}$ be an* AR(p) *process with innovations $\{Z_t\}$ having $E(|Z_1|)<\infty$, median* 0 *and positive mass at* 0. *If $\hat{\boldsymbol{\phi}}$ minimizes* (4.4) *then*

$$n(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi})=o_p(1).$$

Proof. Define

$$W_n^*(\boldsymbol{u})=\sum_{t=p+1}^{n}[|Z_t-n^{-1}(u_0+u_1X_{t-1}+\cdots+u_pX_{t-p})|-|Z_t|]$$

and note that

$$|x-y|-|x|=y(I(x<0)-I(x>0))$$
$$+|y|I(x=0)+2(y-x)(I(y>x>0)-I(y<x<0)).$$

Letting $Y_t(\boldsymbol{u})=u_0+u_1X_{t-1}+\cdots+u_pX_{t-p}$, we have (by the ergodic theorem),

$$\frac{1}{n}\sum_{t=p+1}^{n}Y_t(\boldsymbol{u})(I(Z_t<0)-I(Z_t>0))\xrightarrow{\text{a.s.}}0$$

and

$$\frac{1}{n}\sum_{t=p+1}^{n}|Y_t(\boldsymbol{u})|I(Z_t=0)\xrightarrow{\text{a.s.}}P(Z_t=0)E(|u_0+u_1X_p+\cdots+u_pX_1|)$$

uniformly over $\boldsymbol{u}$ in compact subsets of $\mathbb{R}^{p+1}$. Finally,

$$\begin{aligned}nE((n^{-1}Y_t(\boldsymbol{u})-Z_t)I(n^{-1}Y_t(\boldsymbol{u})>Z_t>0))&\leq E(Y_t(\boldsymbol{u})I(n^{-1}Y_t(\boldsymbol{u})>Z_t>0))\\&=\int_0^\infty\int_{nx}^\infty yG(\mathrm{d}y)F(\mathrm{d}x)\\&=\int_0^\infty E(Y_t(\boldsymbol{u})I(Y_t(\boldsymbol{u})>nx))F(\mathrm{d}x)\\&\to 0\quad\text{as } n\to\infty\end{aligned}$$

by applying the dominated convergence theorem. The same statement holds for the other term and so

$$W_n^*(\boldsymbol{u})\xrightarrow{\text{p}}P(Z_t=0)E(|u_0+u_1X_p+\cdots+u_pX_1|)$$

uniformly over compact subsets of $\mathbb{R}^{p+1}$ and this limit is minimized at $\boldsymbol{u}=0$. The conclusion now follows from Lemma 2.2. □

5. The case $\rho(x)=|x|^\gamma$

In this section, we discuss consistency properties of the M-estimator corresponding to the loss function $\rho(x)=|x|^\gamma$, $\gamma>0$. As before $\{X_t\}$ will denote a causal AR(p) process satisfying the difference equations,

$$X_t=\phi_0+\phi_1X_{t-1}+\cdots+\phi_pX_{t-p}+Z_t \tag{5.1}$$

where $\{Z_t\}$ is a sequence of i.i.d. random variables. Here, the M-estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_p)'$ is defined to be any minimum of the objective function

$$U_n(\boldsymbol{\beta}) = \sum_{t=1}^{n} |X_t - \beta_0 - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|^\gamma \tag{5.2}$$

where $\gamma > 0$ is a preassigned constant.

We shall consider consistency properties of $\hat{\boldsymbol{\phi}}$ in two settings. First we show in Theorem 5.1 below that $\hat{\boldsymbol{\phi}}$ is strongly consistent provided that $E|Z_t|^\gamma < \infty$ and that the function $m(x) = E|Z_t - x|^\gamma$ has a unique minimum. This result includes the LAD case studied by Gross and Steiger (1979) and the classical LS case. The argument used by Gross and Steiger in the LAD case relies on the fact that the function to be minimized is convex. Such an argument cannot be used in the general case $\gamma > 0$ treated below. If $E|Z_t|^\delta < \infty$ for some $\delta > 1$, then the LAD estimate is strongly consistent provided the $\text{med}(Z_t)$ is unique. In contrast, any $\hat{\boldsymbol{\phi}}$ will always be strongly consistent for $\gamma \in (1, \delta]$.

In the second case, we assume that $\phi_0 = 0$ and that Z_0 has regularly varying tail probabilities with exponent $\alpha > 2$. Under these assumptions, it is shown that $\hat{\boldsymbol{\phi}}$ is weakly consistent for all $\gamma > 0$.

The approach taken in both cases will be to show that the objective function, $U_n(\boldsymbol{\beta})$, suitably rescaled by either a sequence of constants or random variables which are independent of $\boldsymbol{\beta}$, converges almost surely or in probability to a nonrandom function whose minimum occurs at $\boldsymbol{\beta} = \boldsymbol{\phi}$. We begin with the almost sure convergence part.

Theorem 5.1. *Let $\{X_t\}$ be the* AR(p) *process given by* (5.1) *with $E|Z_1|^\gamma < \infty$ for some $\gamma > 0$. If the function $m(x) = E|Z_1 - x|^\gamma$ has a unique minimum at $x = \tilde{x}$, then*

$$\hat{\boldsymbol{\phi}} \to (\phi_0 + \tilde{x}, \phi_1, \ldots, \phi_p)' \quad a.s.$$

where $\hat{\boldsymbol{\phi}}$ minimizes (5.2).

Remark. The condition that $m(x)$ has a unique minimum is automatically satisfied if $\gamma > 1$. Of course if $\gamma = 1$, then this condition is equivalent to the existence of a unique median. For the $\gamma < 1$ case, if Z_1 has a symmetric probability density function which is strictly decreasing on $[0, \infty)$, then it is not difficult to show that $m(x)$ will have a unique minimum.

The proof of this theorem is broken up into a series of lemmas, the first of which ensures the existence of at least one minimum to (5.2).

Lemma 5.2. *Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a real function defined by*

$$\psi(\boldsymbol{x}) = \sum_{i=1}^{n} |x_i|^\gamma$$

for some $\gamma > 0$ *where* $\boldsymbol{x} = (x_1, \ldots, x_n)'$. *Set*

$$f(\boldsymbol{c}) = \psi(\boldsymbol{a} - A\boldsymbol{c})$$

where $\boldsymbol{a} \in \mathbb{R}^n$, A is an $n \times m$ *matrix and* $\boldsymbol{c} \in \mathbb{R}^m$. *Then the function* $f(\cdot)$ *has at least one (global) minimizer.*

Proof. If $\operatorname{rank}(A) = m$, then $\|A\boldsymbol{c}\| \to \infty$ as $\|\boldsymbol{c}\| \to \infty$ which implies, by the continuity of f, that the minimum of f occurs on the set $\{\boldsymbol{c}: \|\boldsymbol{c}\| \le M\}$ for some $M > 0$. If $\operatorname{rank}(A) = r < m$, then there exists an $n \times r$ matrix A_1 such that $\{A\boldsymbol{c}: \boldsymbol{c} \in \mathbb{R}^m\} = \{A_1\boldsymbol{d}: \boldsymbol{d} \in \mathbb{R}^r\}$. Applying the full rank case to the function $g(\boldsymbol{d}) := \psi(\boldsymbol{a} - A_1\boldsymbol{d})$, we conclude that g has a minimum on $\mathbb{R}^r$ and hence so does f on $\mathbb{R}^m$ since $g(\boldsymbol{d}) = f(\boldsymbol{c})$ where $A\boldsymbol{c} = A_1\boldsymbol{d}$. $\square$

Lemma 5.3. *Under the hypotheses of Theorem* 5.1, *the function*

$$U(\boldsymbol{\beta}) = E|X_{1+p} - \beta_0 - \beta_1 X_p - \cdots - \beta_p X_1|^\gamma$$

has a unique minimum at $\boldsymbol{\beta}^* = (\phi_0 + \tilde{x}, \phi_1, \ldots, \phi_p)'$.

Proof. Observe that

$$Eg(\boldsymbol{\beta}) = U(\boldsymbol{\beta})$$

where

$$\begin{aligned} g(\boldsymbol{\beta}) &= E[|X_{1+p} - \beta_0 - \beta_1 X_p - \cdots - \beta_p X_1|^\gamma | X_1, \ldots, X_p] \\ &= E[|Z_{1+p} - (\beta_0 - \phi_0) - (\beta_1 - \phi_1) X_p - \cdots - (\beta_p - \phi_p) X_1|^\gamma | X_1, \ldots, X_p]. \end{aligned}$$

Moreover, since $\tilde{x}$ is the unique minimum of $m(x) = E|Z_{1+p} - x|^\gamma$, and Z_{1+p} is independent of $X_1, \ldots, X_p$,

$$g(\boldsymbol{\beta}) \ge g(\boldsymbol{\beta})^* = U(\boldsymbol{\beta}^*) \quad \text{a.s.}$$

with equality holding if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. This $\boldsymbol{\beta}^*$ must be the unique minimum of $U(\cdot)$. $\square$

Lemma 5.4. *Under the hypotheses of Theorem* 5.1, *there is a set S with probability one such that on S,*

$$\begin{aligned} h_n(\boldsymbol{\delta}) &:= n^{-1} \sum_{t=1}^{n} |\delta_{-1} X_t - \delta_0 - \delta_1 X_{t-1} - \cdots - \delta_p X_{t-p}|^\gamma \\ &\to h(\boldsymbol{\delta}) = E|\delta_{-1} X_{p+1} - \delta_0 - \delta_1 X_p - \cdots - \delta_p X_1|^\gamma \end{aligned}$$

for every $\boldsymbol{\delta} \in \mathbb{R}^{2+p}$ *and on every compact subset of* $\mathbb{R}^{2+p}$, *the convergence is uniform.*

Proof. From the ergodic theorem, $h_n(\boldsymbol{\delta}) \to h(\boldsymbol{\delta})$ a.s. and by restricting attention to rationals, this convergence holds simultaneously for all rationals $\boldsymbol{\delta}$ in $\mathbb{R}^{2+p}$ a.s. We next show that for any compact subset K of $\mathbb{R}^{2+p}$, the sequence $\{h_n\}$ is equicontinuous on K a.s. For $\boldsymbol{\delta}, \boldsymbol{\delta}' \in K$, we have

$$|h_n(\boldsymbol{\delta}) - h_n(\boldsymbol{\delta}')| \leq \bigvee_{i=-1}^{p} |\delta_i' - \delta|^\gamma n^{-1} \sum_{t=1}^{n} (1 + |X_t|^\gamma + \cdots + |X_{t-p}|^\gamma) \quad \text{if } \gamma \leq 1$$

and

$$|h_n(\boldsymbol{\delta}) - h_n(\boldsymbol{\delta}')| \leq \max\left(\bigvee_{i=-1}^{p} |\delta_i|^{\gamma-1}, \bigvee_{i=-1}^{p} |\delta_i'|^{\gamma-1} \right) \left(\bigvee_{i=-1}^{p} |\delta_i - \delta_i'|^{\gamma-1} \right)$$
$$\times \left(\gamma n^{-1} \sum_{t=1}^{n} (1 + |X_t|^\gamma + \cdots + |X_{t-p}|^\gamma) \right) \quad \text{if } 1 < \gamma$$

where in the $\gamma > 1$ case, the inequality

$$\big||a|^\gamma - |b|^\gamma\big| \leq \gamma(|a| \vee |b|)^{\gamma-1} |a - b| \tag{5.3}$$

was used. The a.s. equicontinuity of $\{h_n\}$ now follows easily from the ergodic theorem. Moreover, by choosing an increasing sequence of compact sets $K_m \uparrow \mathbb{R}^{2+p}$, we have with probability one that $\{h_n\}$ is equicontinuous on any compact set and since $h(\cdot)$ is continuous, $h_n(\boldsymbol{\delta}) \to h(\boldsymbol{\delta})$ for all $\boldsymbol{\delta} \in \mathbb{R}^{2+p}$. The conclusion of the lemma now follows by an application of the Arzelà–Ascoli theorem. □

Proof of Theorem 5.1. Let h_n, h and S be as in the statement of Lemma 5.4. We first show that there exists an M such that for each outcome in S,

$$\|\hat{\boldsymbol{\phi}}\|^2 < M \tag{5.4}$$

for n large. For the remainder of the argument fix an outcome in S. Suppose the minimum of h on the compact set $K = \{\boldsymbol{\delta}: \|\boldsymbol{\delta}\| = 1\}$ occurs at $\boldsymbol{\delta} = \boldsymbol{\delta}^\dagger$. Then, since $\{X_t\}$ is causal, the argument given for Lemma 5.3 may be used to show that $h(\boldsymbol{\delta}^\dagger) > 0$. We therefore have with $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ and $\boldsymbol{\delta} = (1, \beta_0, \beta_1, \ldots, \beta_p)'$,

$$\begin{aligned}
n^{-1} U_n(\boldsymbol{\beta}) &= n^{-1} \sum_{t=1}^{n} |X_t - \beta_0 - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|^\gamma \\
&= h_n(\boldsymbol{\delta}) \\
&= \|\boldsymbol{\delta}\|^\gamma h_n\left(\frac{\boldsymbol{\delta}}{\|\boldsymbol{\delta}\|}\right) \\
&\geq \|\boldsymbol{\delta}\|^\gamma \left(h_n\left(\frac{\boldsymbol{\delta}}{\|\boldsymbol{\delta}\|}\right) - h\left(\frac{\boldsymbol{\delta}}{\|\boldsymbol{\delta}\|}\right) \right) + \|\boldsymbol{\delta}\|^\gamma h(\boldsymbol{\delta}^\dagger) \\
&\geq \tfrac{1}{2} \|\boldsymbol{\delta}\|^\gamma h(\boldsymbol{\delta}^\dagger)
\end{aligned}$$

for n large and uniformly in $\boldsymbol{\beta}$. Consequently, if $\|\boldsymbol{\beta}\|^2 > M := (2(E|X_1|^\gamma + 1)/h(\boldsymbol{\delta}^\dagger))^{2/\gamma} - 1$, then

$$n^{-1} U_n(\boldsymbol{\beta}) > E|X_1|^\gamma + 1.$$

But also, $n^{-1} U_n(0) \to E|X_1|^\gamma$ and hence for n large,

$$n^{-1} U_n(\boldsymbol{\beta}) > n^{-1} U_n(0)$$

whenever $\|\boldsymbol{\beta}\|^2 > M$. This implies (5.4). Now choose M so large that (5.4) holds and

$$\|(\phi_0 + \tilde{x}, \phi_1, \ldots, \phi_p)'\|^2 < M.$$

Let $U(\boldsymbol{\beta}) = h(1, \beta_0, \ldots, \beta_p)$. Then by Lemma 5.4, $n^{-1} U_n(\cdot) \to U(\cdot)$ uniformly on the compact set $\{\boldsymbol{\beta}: \|\boldsymbol{\beta}\|^2 \leq M\}$. However since U is a continuous function and has a unique minimum on this set, it follows by a standard compactness argument that for each outcome in S,

$$\hat{\boldsymbol{\phi}} \to (\phi_0 + \tilde{x}, \phi_1, \ldots, \phi_p)'$$

as desired. □

We now turn to the case that the noise $\{Z_t\}$ has regularly varying tail probabilities, i.e. the distribution of Z_1 satisfies properties (1.1) and (1.2) for some $\alpha > 0$, and $\phi_0 = 0$. In an effort to conserve notation, we continue to let $U_n^{(\gamma)}(\cdot)$ denote the objective function which now becomes

$$U_n^{(\gamma)}(\boldsymbol{\beta}) = \sum_{t=1}^{n} |X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|^\gamma. \tag{5.5}$$

If $\gamma < \alpha$, then $E|Z_0|^\gamma < \infty$ so that by appealing to Theorem 5.1, $\hat{\boldsymbol{\phi}} \to \boldsymbol{\phi}$ a.s. provided $m(x) = E|Z_1 - x|^\gamma$ has a unique minimum at $x = 0$. Consequently, we confine our attention in the remainder of this section to the case $\gamma \geq \alpha$.

Define the function

$$h^{(\gamma)}(\boldsymbol{\beta}) = \sum_{i=0}^{\infty} |\psi_j - \beta_1 \psi_{i-1} - \cdots - \beta_p \psi_{i-p}|^\gamma \tag{5.6}$$

where the ψ_j are the coefficients in the causal representation of $\{X_t\}$. These coefficients satisfy the recursion (see Brockwell and Davis, 1987, p. 91)

$$\psi_j = \sum_{i=1}^{p} \phi_i \psi_{j-i}, \quad i \geq 1,$$

with $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$.

Lemma 5.5. *Let $\{X_t\}$ be the AR(p) process (5.1) where $\{Z_t\}$ satisfies (1.1) and (1.2) for some $\alpha > 0$. If $\gamma > \alpha$, then on the space of continuous functions from $\mathbb{R}^p$ to $\mathbb{R}$,*

$$a_n^{-\gamma} U_n^{(\gamma)}(\cdot) \xrightarrow{\text{d}} U^{(\gamma)}(\cdot)$$

where

$$U^{(\gamma)}(\boldsymbol{\beta}) = h^\gamma(\boldsymbol{\beta}) \sum_{k=1}^{\infty} \Gamma_k^{-\gamma/\alpha}$$

and the sequence $\{\Gamma_k\}$ is as described in Proposition A.1. *Consequently, for all compact sets $K \subset \mathbb{R}^p$,*

$$\sup_{\boldsymbol{\beta}\in K}\left|\frac{U_n^{(\gamma)}(\boldsymbol{\beta})}{U_n^{(\gamma)}(\boldsymbol{\phi})} - h^{(\gamma)}(\boldsymbol{\beta})\right| \xrightarrow{\text{P}} 0.$$

Proof. From the point process result, Theorem 2.4 in Davis and Resnick (1985),

$$\sum_{k=1}^{n} \varepsilon_{(a_n^{-1}(X_k-\beta_1 X_{k-1}-\cdots-\beta_p X_{k-p}))} \xrightarrow{\text{d}} \sum_{i=0}^{\infty}\sum_{k=1}^{\infty} \varepsilon_{(a_n^{-1}\Gamma_n^{-1/\alpha}(\psi_i-\beta_1\psi_{i-1}-\cdots-\beta_p\psi_{i-p}))}$$

from which it easily follows, using the same ideas as in the proof of Theorem 4.1, that for each $\boldsymbol{\beta}$,

$$a_n^{-\gamma} U_n^{(\gamma)}(\boldsymbol{\beta}) \to U^{(\gamma)}(\boldsymbol{\beta})$$

in $\mathbb{R}$. This result clearly extends to finite dimensional distributions and thus it remains to check tightness of $\{a_n^{-\gamma} U_n^{(\gamma)}(\cdot)\}$ i.e. for $\boldsymbol{\beta}, \boldsymbol{\beta}' \in K$, a compact set,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left[\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}'\|\le\delta} a_n^{-\gamma}|U_n^{(\gamma)}(\boldsymbol{\beta}) - U_n^{(\gamma)}(\boldsymbol{\beta}')| > \varepsilon\right] = 0.$$

However

$$a_n^{-\gamma}|U_n^{(\gamma)}(\boldsymbol{\beta}) - U_n^{(\gamma)}(\boldsymbol{\beta}')| \le K_\gamma(\boldsymbol{\beta},\boldsymbol{\beta}')\|\boldsymbol{\beta}-\boldsymbol{\beta}'\| a_n^{-\gamma}\sum_{t=1}^{n}(|X_t|+\cdots+|X_{t-p}|)^\gamma$$

where $K_\gamma(\boldsymbol{\beta},\boldsymbol{\beta}')$ is bounded for $\boldsymbol{\beta}$, $\boldsymbol{\beta}'$ in the compact set K. Since

$$a_n^{-\gamma}\sum_{t=1}^{n}(|X_t|+\cdots+|X_{t-p}|)^\gamma = O_p(1)$$

the conclusion follows. □

If $\gamma=\alpha$ and $E|Z_1|^\alpha=\infty$, then matters would seem to become more complicated since it is no longer true that

$$a_n^{-\alpha}\sum_{t=1}^{n}|X_t-\beta_1 X_{t-1}-\cdots-\beta_p X_{t-p}|^\alpha = O_p(1).$$

However in this case the function $L(t) := E|Z_1|^\alpha 1_{[|Z_1|^\alpha\le t]}$ is slowly varying, so that by the weak law of large numbers (cf. Feller, 1971, p. 236)

$$b_n^{-1}\sum_{t=1}^{n}|Z_t|^\alpha \xrightarrow{\text{P}} 1$$

where $\{b_n\}$ satisfies $b_n^{-1}nL(b_n)\to 1$. This suggests that

$$b_n^{-1}\sum_{t=1}^{n}|X_t-\beta_1 X_{t-1}-\cdots-\beta_p X_{t-p}|^\alpha$$

may converge in probability to some function which attains its minimum at the true parameter vector.

Lemma 5.6. *Let $\{Y_t\}$ be the linear process*

$$Y_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$$

where $\{Z_t\}$ satisfies (1.1) *and* (1.2) *with $E|Z_1|^\alpha = \infty$, and $\{c_j\}$ satisfies the summability condition* (A.4). *Then*

$$b_n^{-1} \sum_{t=1}^{n} |Y_t|^\alpha \xrightarrow{\text{P}} \sum_{j=0}^{\infty} |c_j|^\alpha$$

where $\{b_n\}$ is a sequence of constants such that $b_n^{-1} nL(b_n) \to 1$.

Proof. First note that since $E|Z_1|^\alpha = \infty$, it follows by essentially Karamata's theorem and (A.7) (cf. Cline, 1983) that for all $\eta \geq \alpha$,

$$\frac{E(|Y_1|^\eta 1_{[|Y_1|^\alpha \leq t]})}{E(|Z_1|^\eta 1_{[|Z_1|^\alpha \leq t]})} \to \sum_{j=0}^{\infty} |c_j|^\alpha. \tag{5.7}$$

Also, since $b_n P[|Z_1|^\alpha > b_n]/L(b_n) \to 0$ (see Feller, 1971, p. 236), we have

$$nP[|Y_1| > b_n] \sim n \sum_{j=0}^{\infty} |c_j|^\alpha P[|Z_1|^\alpha > b_n] \to 0 \tag{5.8}$$

by (A.5) and the choice of b_n.

Now for $q > 0$ fixed, let Y_{tq} be the MA(q) process

$$Y_{tq} = \sum_{j=0}^{q} c_j Z_{t-j}$$

and set $\mu_{nq} = E|Y_{1q}|^\alpha 1_{[|Y_{1q}|^\alpha \leq b_n]}$. We next show

$$b_n^{-1} \sum_{t=1}^{n} |Y_{tq}|^\alpha \xrightarrow{\text{P}} \sum_{j=0}^{q} |c_j|^\alpha$$

and since $b_n^{-1} n\mu_{nq} \to \sum_{i=0}^{q} |c_i|^\alpha$ by (5.7), it is enough to show

$$b_n^{-1} \sum_{t=1}^{n} (|Y_{tq}|^\alpha 1_{[|Y_{tq}|^\alpha \leq b_n]} - \mu_{nq}) \xrightarrow{\text{P}} 0 \tag{5.9}$$

and

$$b_n^{-1} \sum_{t=1}^{n} |Y_{tq}|^\alpha 1_{[|Y_{tq}|^\alpha > b_n]} \xrightarrow{\text{P}} 0. \tag{5.10}$$

Because Y_{0q} and Y_{hq} are independent for $|h| > q$, the variance of the expression in (5.9) is bounded by

$$2(q+1) n b_n^{-2} E|Y_{1q}|^{2\alpha} 1_{[|Y_{1q}|^\alpha \leq b_n]}.$$

However by Karamata's theorem, this is asymptotically equivalent to $(\text{const})nP[|Y_{1q}|^\alpha > b_n]$ which converges to zero by (5.8). As for (5.10),

$$P\left[b_n^{-1}\sum_{t=1}^n |Y_{tq}|^\alpha 1_{[|Y_{tq}|^\alpha > b_n]} > \varepsilon\right] \le P\left[\bigcup_{t=1}^n [|Y_{tq}|^\alpha > b_n]\right]$$
$$\le nP[|Y_{1q}|^\alpha > b_n]$$
$$\to 0.$$

Since $\sum_{j=0}^q |c_j|^\alpha \to \sum_{j=0}^\infty |c_j|^\alpha$, the proof of the lemma will be complete once we show

$$\lim_{q\to\infty}\limsup_{n\to\infty} P\left[b_n^{-1}\sum_{t=1}^n ||Y_t|^\alpha - |Y_{tq}|^\alpha| > \varepsilon\right] = 0. \tag{5.11}$$

Observe that from (5.7) and (5.8),

$$P\left[b_n^{-1}\sum_{t=1}^n |Y_t - Y_{tq}|^\alpha > \varepsilon\right] \le 2\varepsilon^{-1} n b_n^{-1} E\left(\left|\sum_{j>q} c_j Z_{-j}\right|^\alpha 1_{[|\sum_{j>q} c_j Z_{-j}| \le b_n]}\right)$$
$$+ nP\left[\left|\sum_{j>q} c_j Z_{-j}\right|^\alpha > b_n\right]$$
$$\xrightarrow{(n\to\infty)} 2\varepsilon^{-1}\sum_{j>q}|c_j|^\alpha + 0$$
$$\xrightarrow{(q\to\infty)} 0.$$

Thus, if $\alpha \le 1$, (5.11) is immediate. Now if $\alpha > 1$, then by (5.3),

$$||Y_t|^\alpha - |Y_{tq}|^\alpha| \le \alpha(|Y_t|^{\alpha-1} + |Y_{tq}|^{\alpha-1})|Y_t - Y_{tq}|$$
$$\le (\text{const})(|Y_{tq}|^{\alpha-1}|Y_t - Y_{tq}| + |Y_t - Y_{tq}|^\alpha)$$

Since Y_{tq} and $Y_t - Y_{tq}$ are independent and $nb_n^{-1} \to 0$,

$$E\left[b_n^{-1}\sum_{t=1}^n |Y_{tq}|^{\alpha-1}|Y_t - Y_{tq}|\right] = nb_n^{-1}E|Y_{1q}|^{\alpha-1}E|Y_1 - Y_{1q}| \to 0$$

so that (5.11) now follows easily for the case $\alpha > 1$. □

Corollary 5.7. *Let $\{X_t\}$ be the* AR(p) *process* (5.1) *where $\{Z_t\}$ satisfies* (1.1) *and* (1.2) *with $E|Z_1|^\alpha = \infty$. If $\{b_n\}$ is as specified in Lemma* 5.6, *then for any compact set $K \subset \mathbb{R}^p$,*

$$\sup_{\beta\in K}\left|\frac{U_n^{(\alpha)}(\boldsymbol{\beta})}{U_n^{(\alpha)}(\boldsymbol{\phi})} - h^{(\alpha)}(\boldsymbol{\beta})\right| \xrightarrow{\text{P}} 0.$$

Proof. We have

$$X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p} = \sum_{i=0}^\infty (\psi_i - \beta_1\psi_{i-1} - \cdots - \beta_p\psi_{i-p})Z_{t-i}$$

and so the convergence in probability for each $\boldsymbol{\beta}$ follows from Lemma 5.6, and the fact that $b_n^{-1}U_n(\boldsymbol{\phi}) \xrightarrow{\text{P}} 1$. The uniformity of convergence in probability follows from a tightness condition similar to that given in the proof of Lemma 5.5. □

The following theorem summarizes the behavior of $\hat{\boldsymbol{\phi}}$ estimator for all values of γ.

Theorem 5.8. *Let $\{X_t\}$ be the AR(p) process (5.1) where $\{Z_t\}$ satisfies (1.1) and (1.2) and suppose $\hat{\boldsymbol{\phi}}$ minimizes (5.2).*

(i) *If $E|Z_1|^\gamma < \infty$ and $m(x) = E|Z_1 - x|^\gamma$ has a unique minimum at $x = 0$, then*

$$\hat{\boldsymbol{\phi}} \to \boldsymbol{\phi} \quad \text{a.s.}$$

(ii) *If $E|Z_1|^\gamma = \infty$, then*

$$\hat{\boldsymbol{\phi}} \xrightarrow{\text{P}} \boldsymbol{\phi}.$$

Proof. We only need to prove (ii) since (i) is Theorem 5.1. Write $\hat{\boldsymbol{\phi}}_n = \hat{\boldsymbol{\phi}}$. We show that for any subsequence $\{\hat{\boldsymbol{\phi}}_{n'}\}$ there exists a further subsequence $\{\hat{\boldsymbol{\phi}}_{n''}\}$ such that $\hat{\boldsymbol{\phi}}_{n''} \to \boldsymbol{\phi}$ a.s. By Lemma 5.5 or Corollary 5.7 and a diagonal subsequence argument, there exist a subsequence $\{U_{n''}^{(\gamma)}\}$ and a set S, such that $P(S) = 1$ and for any compact set $K \subset \mathbb{R}^p$,

$$\sup_{\boldsymbol{\beta} \in K} \left| \frac{U_{n''}^{(\gamma)}(\boldsymbol{\beta})}{U_{n''}^{(\gamma)}(\boldsymbol{\phi})} - h^{(\gamma)}(\boldsymbol{\beta}) \right| \to 0 \quad \text{a.s.}$$

Since

$$\inf_{\|\mathbf{c}\|=1} \sum_{i=0}^{\infty} |c_1\psi_i - c_2\psi_{i-1} - \cdots - c_{p+1}\psi_{i-p}|^\gamma > 0.$$

and $h_\gamma(\boldsymbol{\beta})$ has a unique minimum at $\boldsymbol{\beta} = \boldsymbol{\phi}$, the remainder of the proof is basically identical to the proof of Theorem 5.1 and is omitted. □

The limiting case as $\gamma \to \infty$ is of course, the L^∞ estimator where $\hat{\boldsymbol{\phi}}$ is chosen to minimize the maximum absolute residual, $|X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|$. Define $\{U_n^\infty(\cdot)\}$ to be this maximum absolute residual so that

$$a_n^{-1}U_n^{(\infty)}(\boldsymbol{\beta}) = a_n^{-1} \max_{1 \le t \le n} |X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|$$

$$= a_n^{-1} \lim_{\gamma \to \infty} (a_n^{-1} U_n^{(\gamma)}(\boldsymbol{\beta}))^{1/\gamma}.$$

Now for each β, it follows that

$$a_n^{-1}U_n^{(\infty)}(\boldsymbol{\beta}) \xrightarrow{\text{d}} \Gamma_1^{-1/\alpha} \max_{0 \le i < \infty} |\psi_i - \beta_1\psi_{i-1} - \cdots - \beta_p\psi_{i-p}|$$

and so

$$\frac{U_n^{(\infty)}(\boldsymbol{\beta})}{U_n^{(\infty)}(\boldsymbol{\phi})} \xrightarrow{\text{P}} \max_{0\le i<\infty} |\psi_i - \beta_1\psi_{i-1} - \cdots - \beta_p\psi_{i-p}| = h^{(\infty)}(\boldsymbol{\beta}).$$

However, while $h^{(\infty)}(\cdot)$ achieves a minimum at $\boldsymbol{\beta} = \boldsymbol{\phi}$, this minimum is not unique. For example, consider the AR(1) case

$$X_t = \phi X_{t-1} + Z_t.$$

Here $\psi_i - \beta\psi_{i-1} = 1$ if $i = 0$ and $\psi_i - \beta\psi_{i-1} = \phi^{i-1}(\phi - \beta)$ if $i \ge 1$; hence $h^{(\infty)}(\beta) = 1$ when $|\beta - \phi| \le 1$. Thus while $\hat{\phi}$ will be $O_p(1)$, it need not be consistent.

A similar but unrelated problem exists when we try to estimate a location parameter along with the AR parameters; that is, choose $\hat{\mu}$, $\hat{\boldsymbol{\phi}}$ to minimize

$$\sum_{t=1}^{n} |X_t - \mu - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|^\gamma.$$

Note that our results for $\gamma > \alpha$ do not depend on the location of $\{Z_t\}$; hence

$$a_n^{-\gamma} \sum_{t=1}^{n} |X_t - \mu - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|^\gamma \xrightarrow{\text{d}} h^{(\gamma)}(\boldsymbol{\beta}) \sum_{k=1}^{\infty} \Gamma_k^{-\gamma/\alpha}$$

independent of the value of μ. It follows that if, for each fixed $\boldsymbol{\beta}$, we choose $\hat{\mu}(\boldsymbol{\beta})$ to minimize

$$\sum_{t=1}^{n} |X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p} - \mu|^\gamma$$

and if $\hat{\mu}(\boldsymbol{\beta}) \xrightarrow{\text{P}} \mu(\boldsymbol{\beta})$ then we should have that $\hat{\mu} \xrightarrow{\text{P}} \mu(\boldsymbol{\phi})$ and $\hat{\boldsymbol{\phi}} \xrightarrow{\text{P}} \boldsymbol{\phi}$. However, in general $\hat{\mu}(\boldsymbol{\beta})$ will not converge in probability to any single value; thus $\hat{\mu}$ does not converge in probability. For example, when $\gamma = 2$ and $\alpha < 1$, $\hat{\mu}$ is equivalent to $\bar{X}$ which does not converge in probability. However, in this case, the AR parameter estimates (least squares estimates) are nonetheless consistent. We conjecture that if $\hat{\mu}$, $\hat{\boldsymbol{\phi}}$ minimize

$$\sum_{t=1}^{n} |X_t - \mu - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}|^\gamma,$$

then $\hat{\boldsymbol{\phi}} \xrightarrow{\text{P}} \boldsymbol{\phi}$ even though $\hat{\mu}$ need not converge in probability to any single constant.

Appendix

In this appendix, we collect the technical results required in earlier portions of the paper. Many of our results rest on point process methods for moving averages as developed in Davis and Resnick (1985). We begin with notation and definitions. For further background on point processes, see Resnick (1987, Chapter 3).

Let E be a state space, which for our purposes will be a subset of Euclidean space. Let $\mathscr{E}$ be the σ-algebra generated by the open subsets of E. For each $x \in E$, define a set function $\varepsilon_x(\cdot)$ on $\mathscr{E}$ as follows:

$$\varepsilon_x(B) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{otherwise,} \end{cases}$$

where $B \in \mathscr{E}$. A point measure m is defined to be a measure of the form

$$m(\cdot) = \sum_{i \in I} \varepsilon_{x_i}(\cdot)$$

such that m is finite on relatively compact sets of E (i.e., subsets B such that the closure $\bar{B}$ is compact). The class of such point measures is denoted by $M_{\mathrm{p}}(E)$ and $\mathscr{M}_{\mathrm{p}}(E)$ is defined to be the smallest σ-algebra which makes the evaluation maps $m \to m(B)$ measurable where $m \in M_{\mathrm{p}}(E)$ and $B \in \mathscr{E}$. A *point process* on E is then a measurable map from a probability space $(\Omega, \mathscr{F}, P)$ into $(M_{\mathrm{p}}(E), \mathscr{M}_{\mathrm{p}}(E))$.

A useful topology for $M_{\mathrm{p}}(E)$ is the vague topology which renders $M_{\mathrm{p}}(E)$ a complete separable metric space. If $\mu_n \in M_{\mathrm{p}}(E)$, $n \geq 0$, then μ_n is said to converge vaguely to μ_0 (written $\mu_n \xrightarrow{v} \mu_0$) if $\mu_m(f) \to \mu_0(f)$ for all $f \in C_K^+(E) :=$ the space of continuous functions $E \to \mathbb{R}^+$ with compact support where $\mu_n(f) = \int f \, d\mu_n$. The σ-algebra generated by the vague topology in $M_{\mathrm{p}}(E)$ coincides with $\mathscr{M}_{\mathrm{p}}(E)$ (see Kallenberg, 1983). Because $M_{\mathrm{p}}(E)$ is a complete separable metric we can speak of convergence in distribution of point processes.

Throughout this section let $\{Z_t\}$ be an i.i.d. sequence of r.v.'s with regularly varying tail probabilities, i.e. assume

$$P(|Z_1| > x) = x^{-\alpha} L(x) \tag{A.1}$$

where $L(x)$ is slowly varying at ∞, $\alpha > 0$ and

$$\lim_{x \to \infty} \frac{P(Z_1 > x)}{P(|Z_1| > x)} = p \tag{A.2}$$

where $0 \leq p \leq 1$. Define the linear process $\{Y_t\}$ by

$$Y_t = \sum_{j=1}^{\infty} c_j Z_{t-j} \tag{A.3}$$

where $\{c_j\}$ is a sequence of constants satisfying

$$\sum_{j=1}^{\infty} |c_j|^\delta < \infty \quad \text{for some } \delta \text{ such that } \delta < \min(\alpha, 1). \tag{A.4}$$

Under these conditions the infinite series (A.3) converges a.s. and

$$\lim_{x \to \infty} \frac{P(|Y_1| > x)}{P(|Z_1| > x)} = \sum_{j=1}^{\infty} |c_j|^\alpha. \tag{A.5}$$

(cf. Cline, 1985). With a_n defined by

$$a_n := \inf\{x \colon P[|Z_t| > x] \leq n^{-1}\} \tag{A.6}$$

it follows from (A.5) that

$$\lim_{n\to\infty} nP[|Y_1|>a_n x]=\sum_{j=1}^{\infty}|c_j|^{\alpha}x^{-\alpha} \quad \text{for all } x>0. \tag{A.7}$$

Proposition A.1. *Suppose* $\{Y_t\}$ *is the process given by* (A.3). *Then*

$$\zeta_n := \sum_{k=1}^{n} \varepsilon_{(Z_k, a_n^{-1}Y_k)} \xrightarrow{\text{d}} \zeta := \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} \varepsilon_{(Z_{k,i}, \delta_k c_i \Gamma_k^{-1/\alpha})}$$

in $M_{\text{p}}(\mathbb{R}\times(\bar{\mathbb{R}}\backslash\{0\}))$ *where* $\bar{\mathbb{R}}=[-\infty,\infty]$ *and*

(a) $\{Z_{k,i}\}$ *is i.i.d. with* $Z_{k,i} \stackrel{\text{d}}{=} Z_1$,

(b) $\{\delta_k\}$ *is i.i.d. with* $P[\delta_k=1]=p$ *and* $P[\delta_k=-1]=1-p$,

(c) $\Gamma_k=E_1+\cdots+E_k$ *where* $\{E_k\}$ *is an i.i.d. sequence of unit exponential r.v.'s*,

(d) $\{Z_{k,i}\}$, $\{\delta_k\}$ *and* $\{E_k\}$ *are independent.* □

The proof of this proposition is a rather straightforward extension of Theorem 2.4 in Davis and Resnick (1985a) (see also Knight, 1986) and hence is omitted. However it is worth remarking that $\sum_{k=1}^{\infty}\varepsilon_{\delta_k\Gamma_k^{-1/\alpha}}$ is a *Poisson process* on $\bar{\mathbb{R}}\backslash\{0\}$ with intensity measure

$$\nu(\text{d}x)=\alpha(px^{-\alpha-1}1_{(0,\infty)}(x)+(1-p)(-x)^{-\alpha-1}1_{(-\infty,0)}(x))\,\text{d}x.$$

As an immediate corollary to this proposition, we have for all continuous functions f on $\mathbb{R}\times(\bar{\mathbb{R}}\backslash\{0\})$ with compact support

$$\sum_{t=1}^{n} f(Z_t, Y_{nt}) \xrightarrow{\text{d}} \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} f(Z_{k,i}, c_i\delta_k\Gamma_k^{-1/\alpha}) \tag{A.8}$$

where $c_i=0$ for $i\leqslant 0$ and

$$Y_{nt}=a_n^{-1}Y_t. \tag{A.9}$$

Note that a compact subset of $\bar{\mathbb{R}}\backslash\{0\}$ is closed and bounded away from 0.

Oftentimes, one would like to extend A.8 to a larger class of functions. For example, in our applications, $f(x,y)=|x+y|^{\gamma}$ or $f(x,y)=\rho(x+y)-\rho(x)$ for some loss function $\rho(\cdot)$. If f is continuous, then A.8 will hold for the function $f(x,y)I(|x|<M)I(|y|>\delta)$ so that by Theorem 4.10 in Billingsley (1967), convergence will follow provided we can show that for all $\varepsilon>0$,

$$\lim_{\delta\to 0}\lim_{M\to\infty}\limsup_{n\to\infty} P\left(\left|\sum_{t=1}^{n} f(Z_t, Y_{nt})(1-I(|Z_t|\leqslant M)I(|Y_{nt}|>\delta))\right|>\varepsilon\right)=0 \tag{A.10}$$

and

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty} f(Z_{k,i}, c_i\delta_k\Gamma_k^{-1/\alpha})I(|Z_{k,i}|\leqslant M)I(|c_i\delta_k\Gamma_k^{-1/\alpha}|>\delta)$$

$$\xrightarrow{\text{P}} \sum_{i=1}^{\infty}\sum_{k=1}^{\infty} f(Z_{k,i}, c_i\delta_k\Gamma_k^{-1/\alpha}) \tag{A.11}$$

as $\delta \to 0$ and $M \to \infty$. The next proposition provides useful bounds which are instrumental in establishing (A.10) for the functions we have in mind.

Proposition A.2. *Suppose that* $\{Y_t\}$ *and* $\{Y_{nt}\}$ *are given by* (A.3) *and* (A.9), *respectively. Let* $\{V_t\}$ *be an i.i.d. sequence of r.v.'s with finite mean such that for every* t, V_t *and* Y_t *are independent. Then for all* $\delta > 0$ *and* $\eta > 0$,

$$\text{(a)} \quad \limsup_{n\to\infty} P\left[\sum_{t=1}^{n} |V_t||Y_{nt}|^{\gamma} I(|Y_{nt}| \leq \delta) > \eta\right]$$

$$\leq \eta^{-1} C_1 E|V_1|\delta^{\gamma-\alpha} \quad \textit{for all } \gamma > \alpha,$$

$$\text{(b)} \quad \limsup_{n\to\infty} P\left[\sum_{t=1}^{n} |V_t||Y_{nt}|^{\gamma} I(|Y_{nt}| > \delta) > \eta\right]$$

$$\leq C_2 \delta^{-\alpha} P[|V_1| > 0] \quad \textit{for all } \gamma > 0,$$

where C_1 *and* C_2 *are constants. If in addition* V_1 *has zero mean and finite variance* σ^2 *and* $1 \leq \alpha < 2$, *then*

$$\text{(c)} \quad \operatorname{Var}\left(\sum_{t=1}^{n} V_t Y_{nt} I(|Y_{nt}| \leq \delta)\right) = n a_n^{-2} E[Y_1^2 I(|Y_1| \leq a_n \delta)] E V_1^2 \to 0$$

as $n \to \infty$ *and then* $\delta \to 0$.

Proof. (a) The probability is bounded by

$$\eta^{-1} E\left[\sum_{t=1}^{n} |V_t||Y_{nt}|^{\gamma} I(|Y_{nt}| \leq \delta)\right] = \eta^{-1} E|V_1|(n a_n^{-\gamma} E|Y_1|^{\gamma} I(|Y_1| \leq a_n \delta))$$

$$\to C_1 E|V_1|\delta^{\gamma-\alpha}$$

as $n \to \infty$ by Karamata's theorem (cf. p. 283, Feller, 1971).

(b) Clearly,

$$P\left[\sum_{t=1}^{n} |V_t||Y_{nt}|^{\gamma} I(|Y_{nt}| > \delta) > \eta\right] \leq P\left[\bigcup_{t=1}^{n} (\{|Y_t| > a_n \delta\} \cap \{|V_t| > 0\})\right]$$

$$\leq n P[|Y_1| > a_n \delta] P[|V_1| > 0]$$

$$\to C_2 \delta^{-\alpha} P[|V_1| > 0]$$

by (A.7).

(c) Since the summands are uncorrelated, the result is immediate from Karamata's theorem. □

The following results are required for studying the asymptotic behavior of LAD estimates.

Proposition A.3. *Let $\{Z_{k,i}\}$ be an array of i.i.d. symmetric r.v.'s which are independent of $\{\delta_k\}$ and $\{\Gamma_k^{-1/\alpha}\}$. Define*

$$V = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} [|Z_{k,i} - c_i \delta_k \Gamma_k^{-1/\alpha}| - |Z_{k,i}|]$$

where $\sum |c_i| < \infty$. Then

(a) *for $\alpha < 1$, V is finite with probability 1,*

(b) *for $\alpha > 1$, V is finite with probability 1 if and only if, $E(|Z_{1,1}|^{1-\alpha}) < \infty$,*

(c) *for $\alpha = 1$, V is finite with probability 1 if and only if, $E(\ln(|Z_{1,1}|)) > -\infty$.*

For $\alpha \geq 1$ the moment conditions are still sufficient for the finiteness of V provided $Z_{k,i}$ has median 0. This will be clear from the proof.

Proof. The case where $\alpha < 1$ is trivial since $\Gamma_k^{-1/\alpha} = O(k^{-1/\alpha})$ and so the series defining V is absolutely convergent. For $x \neq 0$,

$$|x-y| - |x| = y(I(x<0) - I(x>0)) + 2(y-x)(I(y>x>0) - I(y<x<0)).$$

Now because $E[I(Z_{k,i}<0) - I(Z_{k,i}>0)] = 0$, the random variables, $\{c_i \delta_k \Gamma_k^{-1/\alpha}(I(Z_{k,i}<0) - I(Z_{k,i}>0)),\ i \geq 1,\ k \geq 1\}$, are uncorrelated and since $E\Gamma_k^{-2/\alpha} = \Gamma(k-2/\alpha)/\Gamma(k) \leq (\text{const})k^{-2/\alpha}$, we have

$$\sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \mathrm{Var}(c_i \delta_k \Gamma_k^{-1/\alpha}(I(Z_{k,i}<0) - I(Z_{k,i}>0))) < \infty$$

whence

$$\sum_{i=1}^{\infty} \sum_{k=1}^{\infty} (c_i \delta_k \Gamma_k^{-1/\alpha}(I(Z_{k,i}<0) - I(Z_{k,i}>0))) < \infty \quad \text{a.s.}$$

Therefore the convergence of V depends on the convergence of

$$\sum_{i=1}^{\infty} \sum_{k=1}^{\infty} (c_i \delta_k \Gamma_k^{-1/\alpha} - Z_{k,i}) I(c_i \delta_k \Gamma_k^{-1/\alpha} > Z_{k,i} > 0)$$

and

$$\sum_{i=1}^{\infty} \sum_{k=1}^{\infty} (c_i \delta_k \Gamma_k^{-1/\alpha} - Z_{k,i}) I(c_i \delta_k \Gamma_k^{-1/\alpha} < Z_{k,i} < 0).$$

From a Taylor series expansion and the law of the iterated logarithm applied to Γ_k, it follows that

$$|\Gamma_k^{-1/\alpha} - k^{-1/\alpha}| = O(k^{-(1+\alpha/2)}(\log\log k)^{1/2})$$

and, consequently,

$$\sum_{k=1}^{\infty} |\Gamma_k^{-1/\alpha} - k^{-1/\alpha}| < \infty \quad \text{a.s.}$$

so that $\Gamma_k^{-1/\alpha}$ may be replaced by $k^{-1/\alpha}$ in the definition of V. It now follows, by the symmetry of the $Z_{k,i}$'s, that both of the series above will converge if, and only if,

$$E\left(\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}(|c_i|k^{-1/\alpha}-Z_{k,i})I(|c_i|k^{-1/\alpha}>Z_{k,i}>0)\right)<\infty.$$

If $F(\cdot)$ is the distribution function of $Z_{k,i}$, then

$$\begin{aligned}E&\left(\sum_{k=1}^{\infty}(|c_i|k^{-1/\alpha}-Z_{k,i})I(|c_i|k^{-1/\alpha}>Z_{k,i}>0)\right)\\&=\sum_{k=1}^{\infty}\int_0^{|c_i|k^{-1/\alpha}}(|c_i|k^{-1/\alpha}-x)F(\mathrm{d}x)\\&\approx\int_1^{\infty}\int_0^{|c_i|k^{-1/\alpha}}(|c_i|y^{-1/\alpha}-x)F(\mathrm{d}x)\,\mathrm{d}y\\&=\alpha|c_i|^{\alpha}\int_0^{|c_i|}\int_x^{|c_i|}(u-x)u^{-\alpha-1}\,\mathrm{d}u\,F(\mathrm{d}x).\end{aligned}$$

It is easy to see that the double integral is finite if and only if,

$$\int_{0+}x^{1-\alpha}F(\mathrm{d}x)<\infty\quad\text{for }\alpha>1$$

or

$$\int_{0+}\ln(x)F(\mathrm{d}x)>-\infty\quad\text{for }\alpha=1$$

and the result is now immediate by the absolute summability of the c_i. □

Proposition A.4. *Let* $\{X_t\}$ *be an* AR(p) *process with innovations* $\{Z_t\}\in D(\alpha)$, $\alpha\in[1,2)$, *and* $P[Z_1\geqslant 0]=P[Z_1\leqslant 0]$. *Set*

$$Y_t(\boldsymbol{u})=u_1X_{t-1}+\cdots+u_pX_{t-p}=\sum_{j=1}^{\infty}c_jZ_{t-j}$$

where

$$c_j=c_j(\boldsymbol{u})=\psi_{j-1}u_1+\cdots+\psi_{j-p}u_p$$

and the ψ_j *are the coefficients in the causal representation of the* AR *process. Then for all* $M>0$ *and* $\delta>0$, *a continuity point of the distribution of* Z_1,

$$T_n(\cdot\,;\delta,M)\xrightarrow{\mathrm{d}}T(\cdot\,;\delta,M)\tag{A.12}$$

in $C(\mathbb{R}^p)$ *where*

$$\begin{aligned}T_n(\boldsymbol{u},\delta,M)=&\sum_{t=1}^{n}a_n^{-1}MY_t(\boldsymbol{u})(I(Z_t\leqslant 0)-I(Z_t\geqslant 0))\\&+2\sum_{t=1}^{n}|Z_t-a_n^{-1}MY_t(\boldsymbol{u})|(I(a_n^{-1}MY_t(\boldsymbol{u})\leqslant Z_t\leqslant-\delta)\\&\qquad+I(a_n^{-1}MY_t(\boldsymbol{u})\geqslant Z_t\geqslant\delta))\end{aligned}$$

and

$$T(\boldsymbol{u}, \delta, M) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} c_i \delta_k M \Gamma_k^{-1/\alpha}(I(Z_{k,i} \leqslant 0) - I(Z_{k,i} \geqslant 0))$$

$$+ 2 \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} |Z_{k,i} - c_i \delta_k M \Gamma_k^{-1/\alpha}| (I(c_i \delta_k M \Gamma_k^{-1/\alpha} \leqslant Z_{k,i} \leqslant -\delta)$$

$$+ I(c_i \delta_k M \Gamma_k^{-1/\alpha} \geqslant Z_{k,i} \geqslant \delta)).$$

Moreover,

$$\lim_{\delta \to 0} \lim_{M \to \infty} P\left[\inf_{\|\boldsymbol{u}\|=1} T(\boldsymbol{u}, \delta, M) \leqslant 0 \right] = 0. \tag{A.13}$$

Proof. Using (A.8) and the method of proof of Theorem A.3, it is straightforward to show convergence of the finite dimensional distributions of $T_n(\cdot\,; \delta, M)$. As for tightness, since $Y_t(\boldsymbol{u})$ is linear in $\boldsymbol{u}$ and

$$\big| |a-b| 1_{[\delta \leqslant a \leqslant b]} - |a-c| 1_{[\delta \leqslant a \leqslant c]} \big| \leqslant 2|b-c| 1_{[\delta \leqslant a \leqslant b \vee c]}$$

we have for $\|\boldsymbol{u}-\boldsymbol{v}\| \leqslant \eta$ and $\|\boldsymbol{u}\|, \|\boldsymbol{v}\| \leqslant K$, K large,

$$|T_n(\boldsymbol{u}; \delta, M) - T_n(\boldsymbol{v}; \delta, M)|$$

$$\leqslant \eta \sum_{j=1}^{p} \left| \sum_{t=1}^{n} a_n^{-1} M X_{t-j}(I(Z_t \leqslant 0) - I(Z_t \geqslant 0)) \right|$$

$$+ 8 M a_n^{-1} \sum_{t=1}^{n} |Y_t(\boldsymbol{u}-\boldsymbol{v})| I(\delta \leqslant a_n^{-1} M(|Y_t(\boldsymbol{u})| \vee |Y_t(\boldsymbol{v})|)$$

$$=: A + B.$$

By Proposition A.2(b) and (c),

$$\left| \sum_{t=1}^{n} a_n^{-1} X_{t-j}(I(Z_t \leqslant 0) - I(Z_t \geqslant 0)) \right| = O_p(1) \quad \text{for } j = 1, \ldots, p,$$

whence

$$\lim_{\eta \to 0} \limsup_{n \to \infty} P\left[\sup_{\|\boldsymbol{u}-\boldsymbol{v}\| \leqslant \eta} A > \varepsilon \right] = 0. \tag{A.14}$$

For B, since $|Y_t(\boldsymbol{u})| \leqslant Y_t^* := K \sum_{j=1}^{\infty} (|\psi_{j-1}| + \cdots + |\psi_{j-p}|)|Z_{t-j}|$, we have

$$E\left(\sup_{\|\boldsymbol{u}-\boldsymbol{v}\| \leqslant \eta} B^{\alpha/2} \right) \leqslant 8^{\alpha/2} M^{\alpha/2} \eta n a_n^{-\alpha/2} E((Y_1^*)^{\alpha/2} I(\delta < a_n^{-1} M Y_1^*))$$

and by (A.7) and Karamata's theorem,

$$\limsup_{n \to \infty} n a_n^{-\alpha/2} E((Y_1^*)^{\alpha/2} I(\delta < a_n^{-1} M Y_1^*)) < \infty.$$

Therefore,

$$\limsup_{n\to\infty} P\left[\sup_{\|\boldsymbol{u}-\boldsymbol{v}\|\leq\eta} B>\varepsilon\right]\to 0$$

as $\eta\to 0$. This combined with (A.14) proves tightness.

We now turn to the proof of (A.13). For notational convenience, write $I^{\pm}_{k,i}=I(\pm Z_{k,i}\geq 0)$, $I_{k,i}=I(Z_{k,i}\neq 0)$ and let $d_i=|\psi_{i-1}|+\cdots+|\psi_{i-p}|$ so that on $\|\boldsymbol{u}\|=1$, $|c_i(\boldsymbol{u})|\leq d_i$. We first show that given $\varepsilon>0$,

$$\lim_{N\to\infty} P\left[\sup_{\|\boldsymbol{u}\|=1}\sum_{i=1}^{\infty}|c_i|\left|\sum_{k=N}^{\infty}\delta_k\Gamma_k^{-1/\alpha}(I^-_{k,i}-I^+_{k,i})\right|>\varepsilon\right]=0 \tag{A.15}$$

and

$$\lim_{N\to\infty} P\left[\sup_{\|\boldsymbol{u}\|=1}\sum_{i=N}^{\infty}|c_i|\left|\sum_{k=1}^{\infty}\delta_k\Gamma_k^{-1/\alpha}(I^-_{k,i}-I^+_{k,i})\right|>\varepsilon\right]=0. \tag{A.16}$$

The sequence $\{\delta_k\Gamma_k^{-1/\alpha}(I^-_{k,i}-I^+_{k,i})\}$ is uncorrelated with zero mean and finite variance so that with $S=\sum_i|d_i|^{1/2}$, the probability in (A.15) is bounded by

$$\begin{aligned}
&\sum_{i=1}^{\infty} P\left[|d_i|\left|\sum_{k=N}^{\infty}\delta_k\Gamma_k^{-1/\alpha}(I^-_{k,i}-I^+_{k,i})\right|>\varepsilon|d_i|^{1/2}/S\right]\\
&\quad\leq (S/\varepsilon)^2\sum_{i=1}^{\infty}|d_i|\sum_{k=N}^{\infty}E\Gamma_k^{-2/\alpha}\\
&\quad=(S/\varepsilon)^2\sum_{i=1}^{\infty}|d_i|\sum_{k=N}^{\infty}\Gamma(k-2/\alpha)/\Gamma(k)\\
&\quad\to 0
\end{aligned}$$

as $N\to\infty$. The second term is handled in the same fashion.

Let $L=\min_{\|\boldsymbol{u}\|=1}\bigvee_{i=1}^{p}|c_i(\boldsymbol{u})|$. If $L=0$ then there exists a $\boldsymbol{u}$ such that $c_i(\boldsymbol{u})=0$, $i=1,\ldots,p$. But this implies $\|\boldsymbol{u}\|=0$, a contradiction. Choose i^* and $\boldsymbol{u}^*$, a unit vector, such that $L=|c_{i^*}(\boldsymbol{u}^*)|$. Since

$$P\left[\sum_{k=1}^{N} I_{k,i^*}>0\right]=1-P^N[Z_1=0]\to 0\quad\text{as } n\to\infty,$$

it follows from (A.15) and (A.16) that given $\varepsilon>0$ there exists N large such that

$$P\left[\left\{\sup_{\|\boldsymbol{u}\|=1}\left|\sum_{i\vee k>N} d_i\delta_k\Gamma_k^{-1/\alpha}(I^-_{k,i}-I^+_{k,i})\right|\leq\frac{L}{4}\sum_{k=1}^{N} I_{k,i^*}\Gamma_k^{-1/\alpha}\right\}\cap\left\{\sum_{k=1}^{N} I_{k,i^*}>0\right\}\right]>1-\tfrac{1}{4}\varepsilon. \tag{A.17}$$

Now choose $M_0>1$ and $\delta>0$ such that for $i,k=1,\ldots,N$,

$$P\left[\left\{|Z_{k,i}|\leq\frac{\sum_{j=1}^{N} L\Gamma_j^{-1/\alpha}M_0 I_{j,i^*}}{4N^2}\right\}\cap\left\{\sum_{j=1}^{N} I_{j,i^*}>0\right\}\right]>1-\frac{\varepsilon}{4N^2} \tag{A.18}$$

and

$$P[\{\delta<|Z_{k,i}|\}\cup\{Z_{k,i}=0\}]>1-\frac{\varepsilon}{4N^2}. \tag{A.19}$$

Let A denote the intersection of the $2N^2+1$ sets described in (A.17)-(A.19). Then clearly $P[A]>1-\varepsilon$ and from the inequalities,

$$-aI(b\geqslant 0)+2|b-a|I(\delta\leqslant b\leqslant a)\geqslant |a|-2|b| \quad \text{for } b>\delta$$

and

$$aI(b\leqslant 0)+2|b-a|I(a\leqslant b\leqslant -\delta)\geqslant |a|-2|b| \quad \text{for } b<-\delta$$

we have on A and for $M>M_0$,

$$\begin{aligned}\inf_{\|u\|=1} T(u;\delta,M) &\geqslant \sum_{k=1}^{N}\inf_{\|u\|=1}\left(\sum_{i=1}^{N}(M|c_i(u)|\Gamma_k^{-1/\alpha}-2|Z_{k,i}|)I_{k,i}\right)\\ &\quad -\frac{LM}{4}\sum_{k=1}^{N} I_{k,i^*}\Gamma_k^{-1/\alpha}\\ &\geqslant \sum_{k=1}^{N}\left(ML-\frac{M_0L}{4}-\frac{ML}{4}\right)\Gamma_k^{-1/\alpha}I_{k,i^*}\\ &=\frac{M_0L}{4}\sum_{k=1}^{N} I_{k,i^*}\Gamma_k^{-1/\alpha}>0.\end{aligned}$$

Thus $P[\inf_{\|u\|=1} T(u;\delta,M)>0]>1-\varepsilon$ for $M>M_0$ and δ small, as was to be shown. □

References

H.Z. An and Z.G. Chen, On convergence of LAD estimates in autoregression with infinite variance, J. Multivariate Anal. 12 (1982) 335-345.

P. Billingsley, Convergence of Probability Measures (Wiley, New York, 1968).

N.H. Bingham, C.M. Goldie and J.L. Teugels, Regular Variation (Cambridge Univ. Press, Cambridge, 1987).

P. Bloomfield and W. Stieger, Least Absolute Deviations: Theory, Applications and Algorithms (Birkhäuser, Boston, MA, 1983).

P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods (Springer, New York, 1987).

D. Cline, Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data, Ph.D. thesis, Dept. of Statist., Colorado State Univ. (Fort Collins, CO, 1983).

R.A. Davis and S. Resnick, Limit theory for moving averages of random variables with regularly varying tail probabilities, Ann. Probab. 13 (1985a) 179-195.

R.A. Davis and S. Resnick, More limit theory for the sample covariance and correlation functions of moving averages, Stochastic Process. Appl. 20 (1985b) 257-279.

R.A. Davis and S. Resnick, Limit theory for the sample correlation function of moving averages, Ann. Statist. 14 (1986) 533-558.

W. Feller, An Introduction to Probability Theory and its Applications, Vol. 2 (Wiley, New York, 1971, 2nd ed.).

S. Gross and W.L. Steiger, Least absolute deviation estimates in autoregression with infinite variance, J. Appl. Probab. 16 (1979) 104–116.

E.J. Hannan and M. Kanter, Autoregressive processes with infinite variance, J. Appl. Probab. 14 (1977) 411–415.

O. Kallenberg, Random Measures (Akademie, Berlin, 1983).

M. Kanter and W.L. Steiger, Regression and autoregression with infinite variance, Adv. Appl. Probab. 6 (1974) 768–783.

K. Knight, Estimation for infinite variance autoregressive processes, Ph.D. thesis, Dept. of Statist., Univ. of Washington (Seattle, WA, 1986).

K. Knight, Rate of convergence of centered estimates of autoregressive parameters for infinite variance autoregressions, J. Time Ser. Anal. 8 (1987) 51–60.

J. Liu, Infinite variance and nonlinear time series models, Ph.D. thesis, Dept. of Statist., Colorado State Univ. (Fort Collins, CO, 1987).

S. Resnick, Extreme Values, Regular Variation and Point Processes (Springer, New York, 1987).

R.T. Rockafellar, Convex Analysis (Princeton Univ. Press, Princeton, NJ, 1970).

G. Shorack and J. Wellner, Empirical Processes with Applications to Statistics (Wiley, New York, 1986).