Statistica Sinica **7**(1997), 1135-1154

BOOTSTRAPPING M-ESTIMATES IN REGRESSION AND AUTOREGRESSION WITH INFINITE VARIANCE

Richard A. Davis and Wei Wu

Colorado State University and Schering-Plough Research Institute

Abstract: The limiting distribution for M-estimates in a regression or autoregression model with heavy-tailed noise is generally intractable, which precludes its use for inference purposes. Alternatively, the bootstrap can be used to approximate the sampling distribution of the M-estimate. In this paper, we show that the bootstrap procedure is asymptotically valid for a class of M-estimates provided the bootstrap resample size m_n satisfies $m_n \to \infty$ and $m_n/n \to 0$ as the original sample size n goes to infinity.

Key words and phrases: Autoregressive processes, bootstrap, M-estimation, Poisson processes, regular variation, stable laws.

1. Introduction

Recently, there has been a great deal of interest in developing estimation procedures for statistical models designed to model heavy-tailed data. Often one assumes in these models that the regressors and/or residuals have regularly varying tail probabilities. In such cases, M-estimates, with an appropriately chosen loss function, have a number of desirable properties. While the asymptotic theory for M-estimates is well understood, the limiting distributions are generally intractable. This precludes the use of the asymptotic distribution for inference purposes such as for the construction of confidence intervals. In this paper, we investigate the bootstrap for approximating the distribution of M-estimates.

The asymptotic properties of the M-estimate have been thoroughly studied in the heavy-tailed regression and autoregression setting by Davis and Wu (1997) and Davis, Knight, and Liu (1992). For the purpose of introduction we focus on the AR(p) case. Suppose $X_1, \ldots, X_n$ are observations from the AR(p) process $\{X_t\}$ satisfying the recursions $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$, where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| \leq 1$ and $\{Z_t\}$ is an i.i.d. sequence of random variables with distribution function F which we assume belongs to the domain of attraction of an α-stable law with $\alpha \in (0, 2)$. The latter condition implies that there exists a sequence of non-negative constants $a_n \to \infty$ such that

$$nP[a_n^{-1} Z_1 \in dx] \xrightarrow{v} \lambda(dx), \tag{1.1}$$

where λ is a Lévy measure and $\xrightarrow{v}$ denotes vague convergence. (One can take a_n to be the $1-n^{-1}$ quantile of the distribution of $|Z_1|$ which has the form $n^{1/\alpha}L(n)$ where $L(\cdot)$ is a slowly varying function.)

For a given loss function $\rho(x)$, the M-estimate $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ minimizes the objective function

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

with respect to $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. The special cases $\rho(x) = x^2$ and $\rho(x) = |x|$ correspond to least squares (LS) and least absolute deviation (LAD) estimators, respectively. Under certain conditions on the loss function (which excludes the LS case), it was shown in Davis et al. (1992) that $a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} \hat{\mathbf{u}}$, where $\hat{\mathbf{u}}$ is the minimizer of some stochastic process. In order to use this result to approximate the distribution of the M-estimate, one not only needs to know the scaling constants a_n but also the quantiles of the limit random variable $\hat{\mathbf{u}}$. In general, a closed form expression for $\hat{\mathbf{u}}$ is impossible to obtain. Alternatively, one could simulate from the limit stochastic process and find the sample path minimum from which quantiles of $\hat{\mathbf{u}}$ could be estimated. However, since the limit process is a function of the underlying parameter values and the distribution of the noise, both of which are unknown, carrying out this resampling scheme is impractical. This leads to the use of the bootstrap as an alternative method for approximating the distribution of the M-estimate.

The bootstrap procedure in the AR(p) context is implemented by first generating a bootstrap replicate $X_1^*, \ldots, X_m^*$ from the fitted AR(p) model

$$X_t^* = \hat{\phi}_1 X_{t-1}^* + \cdots + \hat{\phi}_p X_{t-p}^* + Z_t^*,$$

where $\{Z_t^*\}$ is an i.i.d. sequence of random variables generated from the empirical distribution of the estimated residuals, $\hat{Z}_{p+1}, \ldots, \hat{Z}_n$ (n is the length of the observed time series). The bootstrap replicate of the M-estimate is then found by minimizing

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

In Section 3, we show that $a_m(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}})$, conditional on $X_1, \ldots, X_n$, has the same limit distribution as $a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})$ provided the bootstrap sample size m_n satisfies $m_n \to \infty$ and $m_n/n \to 0$.

The remaining hurdle in applying this bootstrap paradigm is that the normalizing constants a_n are typically unknown. This can be overcome by using random normalization, replacing a_n by the maximum of $\{|X_1|, \ldots, |X_n|\}$ and a_m in the bootstrap normalization by the maximum of $\{|X_1^*|, \ldots, |X_m^*|\}$.

The above restriction on the bootstrap sample size m is typical in bootstrapping heavy-tailed phenomena (see for example Athreya (1987)). One way to see this in the present context, is to first note that the limit behavior of the M-estimate is heavily driven by assumption (1.1). In order to reproduce the same limiting result for the bootstrap replicates, it is necessary that a similar condition holds for the distribution of Z_1^*, namely that

$$m_n P[a_{m_n}^{-1} Z_1^* \in \cdot\, | X_1, \ldots, X_n] \xrightarrow{v} \lambda(\cdot).$$

However, the left-hand side, evaluated at the fixed set B, is equal to

$$m_n \left(n^{-1} \sum_{t=1}^{n} I(a_{m_n}^{-1} \hat{Z}_t \in B) \right),$$

which converges in probability to $\lambda(B)$ if and only if $m_n/n \to 0$. On the other hand, if $m_n = n$, then the above quantity converges in distribution to a Poisson distributed random variable with mean $\lambda(B)$.

The remainder of the paper is organized as follows. In Section 2, we consider bootstrapping M-estimates for a linear regression model when the independent variables are heavy-tailed and in Section 3 we consider bootstrapping in the autoregressive case. The proofs of the more technical results in Sections 2 and 3 are contained in the Appendix.

2. Linear Regression

In this section we consider bootstrapping M-estimates in a linear regression model. We start with a simple linear model and then indicate how to extend it to the multiple regression case.

Simple Linear Model. Let $(Y_i, X_i), i = 1, \ldots, n$, be observations from the simple linear model

$$Y_i = \beta X_i + Z_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where the sequences $\{Z_i\}$ and $\{X_i\}$ are independent with $\{Z_i\}_{i=1}^n \overset{iid}{\sim} G$ and $\{X_i\}_{i=1}^n \overset{iid}{\sim} F$. It is further assumed that F belongs to the domain of attraction of a stable law with index $0 < \alpha < 2$ (denoted by $F \in D(\alpha)$ or $X_i \in \mathcal{D}(\alpha)$), i.e. there exist a slowly varying function $L(x)$ at ∞, constants $0 \le p,\ q \le 1$, $p + q = 1$, and $\alpha \in (0, 2)$, such that

$$\begin{aligned} 1 - F(x) &\sim p x^{-\alpha} L(x), \\ F(-x) &\sim q x^{-\alpha} L(x), \qquad \text{as } x \to \infty. \end{aligned} \tag{2.2}$$

Then the partial sums $\sum_{i=1}^{n} X_i$, scaled by $a_n = \inf\{x : P(|X_1| \geq x) \leq n^{-1}\}$ and centered by $nE[X_1 I(|X_1| \leq a_n)]$, converge in distribution to a stable distribution.

For a given loss function $\rho(x)$, the M-estimate $\hat{\beta}$ of the regression coefficient β is defined as any minimizer of the objective function

$$g(\phi) := \sum_{i=1}^{n} \rho(Y_i - \phi X_i) = \sum_{i=1}^{n} \rho(Z_i - (\phi - \beta)X_i). \tag{2.3}$$

As in Davis and Wu (1997), it is convenient to build the normalization into the objective function. Set $u = a_n(\phi - \beta)$ and define the sequence of stochastic processes on $C(\mathbb{R})$ by

$$W_n(u) = \sum_{i=1}^{n} (\rho(Z_i - u a_n^{-1} X_i) - \rho(Z_i)).$$

With this parameterization, the minimizer of $W_n(u)$ is given by $\hat{u}_n = a_n(\hat{\beta} - \beta)$. In Davis and Wu (1997), the stochastic processes $W_n(\cdot)$ were shown to converge in distribution to a limit stochastic process $W(\cdot)$ from which it follows that $\hat{u}_n = a_n(\hat{\beta} - \beta)$ converges in distribution to $\hat{u}$, the minimizer of $W(\cdot)$. Unfortunately, computing the distribution of $\hat{u}$ via simulation or analytically is intractable for most cases. To overcome this difficulty, we use the bootstrap to approximate the sampling distribution of $\hat{u}_n$, which we now describe.

Our bootstrap procedure involves generating a bootstrap replicate of the stochastic process $W_n(\cdot)$ and showing that it also converges in distribution to $W(\cdot)$. To define the bootstrap replicate, let

$$\hat{Z}_j = Y_j - \hat{\beta} X_j, \; j = 1, \ldots, n, \tag{2.4}$$

denote the residuals from the model fit. Let $\hat{F}_n$ be the product empirical distribution function defined by $d\hat{F}_n = d\hat{F}_{Z,n} \times d\hat{F}_{X,n}$, where $\hat{F}_{Z,n}(z) = n^{-1} \sum_{j=1}^{n} I_{[\hat{Z}_j \leq z]}$ and $\hat{F}_{X,n}(x) = n^{-1} \sum_{j=1}^{n} I_{[X_j \leq x]}$. A bootstrap replicate $\{(Y_1^*, X_1^*), \ldots, (Y_m^*, X_m^*)\}$ of the data is generated from the equations

$$Y_j^* = \hat{\beta} X_j^* + Z_j^*, \; j = 1, \ldots, m, \tag{2.5}$$

where $\{(X_j^*, Z_j^*), j = 1, \ldots, m\}$ is a random sample from $\hat{F}_n$. The bootstrap replicate $\hat{u}_m^* := a_m(\hat{\beta}^* - \hat{\beta})$ of $\hat{u}_n$ is then found by minimizing

$$W_n^*(u) = \sum_{j=1}^{m} (\rho(Z_j^* - u a_m^{-1} X_j^*) - \rho(Z_j^*)). \tag{2.6}$$

Provided the resample size $m = m_n$ is a sequence of numbers converging to infinity with $m_n/n \to 0$, the bootstrap approximation is asymptotically correct in the sense that for all continuity points x of the distribution of $\hat{u}$,

$$P[\hat{u}_m^* \le x \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} P[\hat{u} \le x], \tag{2.7}$$

where $\mathbf{X}_\infty = (X_1, X_2, \ldots,)$ and $\mathbf{Y}_\infty = (Y_1, Y_2, \ldots,)$.

In order to give a precise statement of our results, it is necessary to introduce some notation and definitions. First, let $\mathcal{M}_p(C(\mathbb{R}))$ be the space of probability measures on $C(\mathbb{R})$, the space of continuous functions on $\mathbb{R}$ where convergence is defined as uniform convergence on compact sets. Let d_0 be a metric on $\mathcal{M}_p(C(\mathbb{R}))$ which metrizes the topology of weak convergence, i.e. if $\lambda_1, \lambda_2 \in \mathcal{M}_p(C(\mathbb{R}))$, d_0 can be defined as

$$d_0(\lambda_1, \lambda_2) = \sum_{k=1}^{\infty} 2^{-k} \frac{|\int_{\mathbb{R}} g_k \, d\lambda_1 - \int_{\mathbb{R}} g_k \, d\lambda_2|}{1 + |\int_{\mathbb{R}} g_k \, d\lambda_1 - \int_{\mathbb{R}} g_k \, d\lambda_2|},$$

where $\{g_k\}_{k=1}^{\infty}$ is a dense sequence of bounded and continuous functions on $C(\mathbb{R})$. Note that if L_n and L are random elements of $\mathcal{M}_p(C(\mathbb{R}))$, then $L_n \xrightarrow{p} L$ if and only if $d_0(L_n, L) \xrightarrow{p} 0$ which is equivalent to $\int_{\mathbb{R}} g_k dL_n \xrightarrow{p} \int_{\mathbb{R}} g_k dL$ for all $k = 1, 2, \ldots$

Let d_j denote the corresponding metric on the space of probability measures on $\mathbb{R}^j$ denoted by $\mathcal{M}_p(\mathbb{R}^j)$. As above, if Q_n and Q are random elements of $\mathcal{M}_p(\mathbb{R}^j)$, then $Q_n \xrightarrow{p} Q$ if and only if $d_j(Q_n, Q) \xrightarrow{p} 0$ which is equivalent to $\int_{\mathbb{R}^j} f_k dQ_n \xrightarrow{p} \int_{\mathbb{R}^j} f_k dQ$ for all $k = 1, 2, \ldots,$ where $\{f_k\}_{k=1}^{\infty}$ is a dense sequence of bounded and uniformly continuous functions on $\mathbb{R}^j$.

Theorem 2.1. *Let $\{(Y_i, X_i)\}_{i=1}^n$ be observations from model* (2.1)*, where $\{X_i\}_{i=1}^n \overset{iid}{\sim} F$ with F satisfying* (2.2)*, $\{Z_i\}_{i=1}^n \overset{iid}{\sim} G$, and the two sequences $\{X_i\}_{i=1}^n$ and $\{Z_i\}_{i=1}^n$ are independent. Let $\rho(\cdot)$ be a loss function whose score function $\psi(x) = \rho'(x)$ satisfies:*
(a) *$\psi(\cdot)$ is Lipschitz of order τ_1,*

$$|\psi(x) - \psi(y)| \le C|x - y|^{\tau_1},$$

for some constant $\tau_1 > \max(\alpha - 1, 0)$ and some positive constant C,
(b) *$E|\psi(Z_1)|^{\tau_2} < \infty$ for some $\tau_2 > \alpha$,*
(c) *$E\psi(Z_1) = 0$ if $\alpha \ge 1$.*
Then, if $m_n \to \infty$ and $m_n/n \to 0$,

$$\begin{aligned} L_n(\cdot) &:= P[W_n^* \in \cdot | \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &\xrightarrow{p} P[W \in \cdot] \\ &=: L(\cdot), \end{aligned}$$

where $W(\cdot)$ *is the limit process defined in Theorem* 2.1 *of Davis and Wu* (1997)*. Namely,*

$$W(u) = \sum_{k=1}^{\infty} [\rho(Z_k - u\delta_k \Gamma_k^{-1/\alpha}) - \rho(Z_k)],$$

where $\{Z_k\}$, $\{\delta_k\}$, $\{\Gamma_k\}$ *are independent sequences of random variables,* $\{Z_k\} \overset{iid}{\sim} G$, $\{\delta_k\}$ *are i.i.d. with* $P(\delta_k = 1) = p = 1 - P(\delta_k = -1)$*, and* $\Gamma_k = E_1 + \cdots + E_k$*, where the* E_i*'s are i.i.d. exponential r.v.'s with mean* 1*.*

Theorem 2.2. *If* $\rho(\cdot)$ *is convex and satisfies the conditions of Theorem* 2.1 *and* $W(\cdot)$ *attains a unique minimum at* $\hat{u}$ *a.s., then*

$$\begin{aligned} Q_n(\cdot) &:= P[\hat{u}^*_{m_n} \in \cdot | \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &= P[a_{m_n}(\hat{\beta}^* - \hat{\beta}) \in \cdot | \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &\xrightarrow{p} P[\hat{u} \in \cdot] \\ &=: Q(\cdot). \end{aligned}$$

Remark 1. In case $\rho(\cdot)$ is strictly convex, i.e. $\psi(\cdot)$ strictly increasing, then $W(\cdot)$ has strictly convex sample paths and hence has a unique minimum. (See Remark 2 and Section 3 of Davis et al. (1992) for further discussion on this point.)

Proof of Theorem 2.1. If suffices to show that for any subsequence $\{n_k\}$, there exists a further subsequence $\{n_{k'}\}$ such that $L_{n_{k'}} \xrightarrow{a.s.} L$ relative to the metric d_0. This is equivalent to showing that for almost all sample paths of $\mathbf{X}_\infty, \mathbf{Y}_\infty$, $W^*_{n_{k'}} \xrightarrow{d} W$ on $C(\mathbb{R})$. Now by Lemma 4 of the Appendix, we have for any $u_1, \ldots, u_j \in \mathbb{R}$, $L_n \circ \pi^{-1}_{u_1,\ldots,u_j} \xrightarrow{p} L \circ \pi^{-1}_{u_1,\ldots,u_j}$ on $\mathcal{M}_p(\mathbb{R}^j)$, where $\pi_{u_1,\ldots,u_j} : x \to (x(u_1), \ldots, x(u_j))$ is the projection mapping. Let $\{q_1, q_2, \ldots\}$ be an enumeration of the rationals. Then using a diagonal sequence argument, there exists a subsequence $\{n_{k'}\}$ and a probability one event Ω_0 such that for all outcomes in Ω_0 and any j, $L_{n_{k'}} \circ \pi^{-1}_{q_1,\ldots,q_j} \to L \circ \pi^{-1}_{q_1,\ldots,q_j}$ or, equivalently,

$$(W^*_{n_{k'}}(q_1), \ldots, W^*_{n_{k'}}(q_j)) \xrightarrow{d} (W(q_1), \ldots, W(q_j))$$

as $k' \to \infty$. Since the limit process $W(\cdot)$ is continuous, convergence on $C(\mathbb{R})$ will follow once we show that the sequence $\{W^*_{n_{k'}}\}$ is tight for almost all sample paths of $\mathbf{X}_\infty, \mathbf{Y}_\infty$. Tightness on $C(\mathbb{R})$ is equivalent to tightness on $C([-T, T])$ for every $T > 0$, so that by Theorem 8.2 in Billingsley (1968), it is enough to check that for almost all sample paths and for every $T, \epsilon, \eta > 0$, there exists a $\delta > 0$ such that

$$P[\sup_{\substack{|t-s| \le \delta \\ |s|,|t| \le T}} |W^*_{n_{k'}}(t) - W^*_{n_{k'}}(s)| > \epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] < \eta$$

for all k'. Writing n' in place of $n_{k'}$ we have for any $s, t \in [-T, T]$,

$$\begin{aligned}
&|W^*_{n'}(s) - W^*_{n'}(t)| \\
&\le |(s-t)\sum_{i=1}^{m_{n'}} a^{-1}_{m_{n'}} X^*_i \psi(Z^*_i) + (s-t)\sum_{i=1}^{m_{n'}} a^{-1}_{m_{n'}} X^*_i (\psi(\xi^{n'}_i) - \psi(Z^*_i)|, \qquad (2.8)
\end{aligned}$$

where $|\xi^{n'}_i - Z^*_i| \le (|s| \vee |t|) a^{-1}_{m_{n'}} |X^*_i| \le T\, a^{-1}_{m_{n'}} |X^*_i|$. Using the assumptions on $\psi(\cdot)$, (2.8) may be bounded above by

$$|s-t||\sum_{i=1}^{m_{n'}} a^{-1}_{m_{n'}} X^*_i \psi(Z^*_i)| + |s-t| C\, T^{\tau_1} \sum_{i=1}^{m_{n'}} (a^{-1}_{m_{n'}} |X^*_i|)^{1+\tau_1}.$$

Applying the bootstrap results for sample means in Athreya et al. (1993), $a^{-1}_{m_{n'}} \sum_{i=1}^{m_{n'}} X^*_i \psi(Z^*_i)$ and $a^{-1-\tau_1}_{m_{n'}} \sum_{i=1}^{m_{n'}} |X^*_i|^{1+\tau_1}$ are bounded in probability for almost all sample paths of $\mathbf{X}_\infty$ and $\mathbf{Y}_\infty$. It then follows that for almost all sample paths, $W^*_{n_{k'}}$ is tight and the theorem is proved.

Proof of Theorem 2.2. By Theorem 2.1, for any subsequence $\{n_k\}$, there exists a further subsequence $\{n_{k'}\}$ such that for almost all sample paths of $\mathbf{X}_\infty, \mathbf{Y}_\infty$, $W^*_{n_{k'}} \xrightarrow{d} W$ on $C(\mathbb{R})$. It follows by the argument given for Lemma 2.2 in Davis et al. (1992) that for such sample paths, $\hat{u}^*_{n_{k'}} \xrightarrow{d} \hat{u}$, and hence $Q_{n_{k'}}(\cdot) = P[\hat{u}^*_{n_{k'}} \in \cdot \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{a.s.} P[\hat{u} \in \cdot] = Q(\cdot)$, from which the theorem is immediate.

One of the difficulties in the above formulation is that the normalizing constants $\{a_n\}$ are assumed known. This can be circumvented by using a random normalization such as the maximum of the $|X_i|$. In this formulation, $W_n(\cdot)$ and W^*_n are replaced by

$$\tilde{W}_n(u) = \sum_{i=1}^{n} (\rho(Z_i - uX_i/M_n) - \rho(Z_i))$$

and

$$\tilde{W}^*_n(u) = \sum_{i=1}^{m_n} (\rho(Z^*_i - uX^*_i/M^*_{m_n}) - \rho(Z^*_i)),$$

respectively, where $M_n = \max\{|X_1|, \ldots, |X_n|\}$ and $M^*_{m_n} = \max\{|X^*_1|, \ldots, |X^*_{m_n}|\}$. As might be expected, the distribution of the normalized M-estimate $\tilde{u}_n := M_n(\hat{\beta} - \beta)$ can be approximated by the distribution of $\tilde{u}^*_{m_n} := M^*_{m_n}(\hat{\beta}^* - \hat{\beta})$.

Theorem 2.3. *If $\rho(\cdot)$ is convex and satisfies the conditions of Theorem* 2.1 *and $W(\cdot)$ attains a unique minimum at $\hat{u}$ a.s., then*

$$P[\tilde{u}^*_{m_n} \in \cdot | \mathbf{X}_\infty, \mathbf{Y}_\infty] = P[M^*_{m_n}(\hat{\beta}^* - \hat{\beta}) \in \cdot | \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} P[\tilde{u} \in \cdot],$$

where $\tilde{u} = \hat{u}\Gamma_1^{-1/\alpha}$.

Proof. Observe that $\tilde{W}_n(u) = W_n(ua_n/M_n)$ and since $a_n^{-1}M_n \xrightarrow{d} \Gamma_1^{-1/\alpha}$, we conclude that $\tilde{W}_n(\cdot) \xrightarrow{d} \tilde{W}(\cdot)$, where $\tilde{W}(u) := W(u\Gamma_1^{1/\alpha})$. It follows that $\tilde{u}_n \xrightarrow{d} \tilde{u} = \hat{u}\Gamma_1^{-1/\alpha}$. Now, using Lemma 3 and a standard point process argument, $a_{m_n}^{-1}M_{m_n}^*$ given $\mathbf{X}_\infty, \mathbf{Y}_\infty$ converges in distribution to $\Gamma_1^{-1/\alpha}$ and the convergence is joint with that of W_n^*. We deduce that $P[\tilde{W}_n^* \in \cdot|\mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} P[\tilde{W} \in \cdot]$ and the remainder of the theorem is now argued as in the proof of Theorem 2.2.

Multiple Regression. Here the model becomes

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + Z_i, \qquad i = 1, \ldots, n,$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)'$, and $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})'$ are i.i.d. random vectors satisfying a d-variate regular variation condition (see Assumptions 1 and 2 of Davis and Wu (1997)). Specifically, we assume that there exists a sequence $a_n \to \infty$ and a Lévy measure μ on $(\mathbb{R}^d, B(\mathbb{R}^d))$, such that

$$nP(a_n^{-1}\boldsymbol{X}_1 \in \cdot) \underset{n\to\infty}{\xrightarrow{\nu}} \mu(\cdot),$$

($\xrightarrow{\nu}$ is vague convergence on $\mathbb{R}^d \setminus (0, 0, \ldots, 0)$). The M-estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ then minimizes the objective function

$$\sum_{i=1}^n \rho(Y_i - \boldsymbol{X}_i'\boldsymbol{\phi}) = \sum_{i=1}^n \rho(Z_i - \boldsymbol{X}_i'(\boldsymbol{\phi} - \boldsymbol{\beta}))$$

with respect to $\boldsymbol{\phi} \in \mathbb{R}^d$. The relevant sequence of stochastic processes, obtained by setting $\mathbf{u} = a_n(\boldsymbol{\phi} - \boldsymbol{\beta})$, is

$$W_n(\mathbf{u}) = \sum_{i=1}^n [\rho(Z_i - a_n^{-1}\boldsymbol{X}_i'\mathbf{u}) - \rho(Z_i)],$$

so that the minimizer of $W_n(\mathbf{u})$ is $\hat{\mathbf{u}}_n = a_n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. In Davis and Wu (1997), it was shown that if the loss function $\rho(x)$ satisfies conditions (a)–(c) of Theorem 2.1, then $\hat{\mathbf{u}}_n \xrightarrow{d} \hat{\mathbf{u}}$ where $\hat{\mathbf{u}}$ is the minimizer of the limit stochastic process of $W_n(\cdot)$.

The bootstrap implementation for the multiple regression case follows the development just given for the simple linear model. A bootstrap replicate $\{(Y_1^*, \mathbf{X}_1^*), \ldots, (Y_m^*, \mathbf{X}_m^*)\}$ of the data is generated from the equations $Y_j^* = \mathbf{X}_j^{*\prime}\hat{\boldsymbol{\beta}} + Z_j^*$, where $\{\mathbf{X}_j^*\}_{j=1}^m$ and $\{Z_j^*\}_{j=1}^m$ are two independent samples drawn from the empirical distributions based on $\mathbf{X}_1, \ldots, \mathbf{X}_n$, and the estimated residuals $\hat{Z}_1, \ldots, \hat{Z}_n$, respectively. The bootstrap replicate $\hat{\mathbf{u}}_m^* = a_m(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})$ is then found by minimizing

$$W_n^*(\mathbf{u}) = \sum_{i=1}^m [\rho(Z_i^* - a_m^{-1}\mathbf{X}_i^{*\prime}\mathbf{u}) - \rho(Z_i^*)].$$

Using the argument given for Theorems 2.1 and 2.2, the bootstrap replicate of $\hat{\mathbf{u}}_m$ also converges in distribution to $\hat{\mathbf{u}}$ in the sense that $P[\hat{\mathbf{u}}_m^* \in \cdot \mid \mathbf{X}_1, \ldots, \mathbf{X}_n, Y_1, \ldots, Y_n] \xrightarrow{p} P[\hat{\mathbf{u}} \in \cdot]$, provided $m_n \to \infty$ and $m/n \to 0$.

3. Autoregression

In this section, we consider bootstrapping the M-estimate of an autoregressive process. Let $\{X_t\}$ be the causal AR(p) process satisfying the recursions

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \tag{3.1}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| \leq 1$ and $\{Z_t\} \overset{iid}{\sim} F$ with F satisfying (2.2) for some $\alpha \in (0, 2)$. The causality assumption implies that X_t can be represented as the linear process

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \tag{3.2}$$

where $\{\psi_j, j = 0, 1 \ldots, \}$ are the coefficients in the power series expansion of $1/\phi(z)$. 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

$$\sum_{t=p+1}^{n} \rho(X_t - \beta_1 X_{t-1} - \cdots - \beta_p X_{t-p}) = \sum_{t=p+1}^{n} \rho(Z_t - (\beta_1 - \phi_1) X_{t-1} - \cdots - (\beta_p - \phi_p) X_{t-p}).$$

Under the reparameterization $\mathbf{u} = a_n(\boldsymbol{\beta} - \boldsymbol{\phi})$, minimizing this objective function is equivalent to minimizing

$$U_n(\mathbf{u}) = \sum_{t=p+1}^{n} \Big(\rho(Z_t - u_1 a_n^{-1} X_{t-1} - \cdots - u_p a_n^{-1} X_{t-p}) - \rho(Z_t) \Big)$$

with the minimum given by $\hat{\mathbf{u}}_n = a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})$.

In order to construct a bootstrap replicate of $\hat{\boldsymbol{\phi}}$, let $\hat{Z}_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p}$, $t = p+1, \ldots, n$, denote the residuals. If $\hat{F}_n(z) = (n-p)^{-1} \sum_{i=p+1}^{n} I_{[\hat{Z}_i \leq z]}$ denotes the empirical distribution of the residuals, a bootstrap replicate $X_1^*, \ldots, X_{m_n}^*$ of the AR process is generated from the recursions $X_t^* = \hat{\phi}_1 X_{t-1}^* + \cdots + \hat{\phi}_p X_{t-p}^* + Z_t^*$, where $\{Z_t^*\} \overset{iid}{\sim} \hat{F}_n$. (The recursions can be started by setting $X_t^* = 0$ in the distant past.) A bootstrap replicate, $\hat{\mathbf{u}}_m^* = a_m(\hat{\boldsymbol{\phi}}_{m_n}^* - \hat{\boldsymbol{\phi}}_n)$ of $\hat{\mathbf{u}}$ is then found by minimizing

$$U_n^*(\mathbf{u}) = \sum_{t=p+1}^{m} \Big(\rho(Z_t^* - u_1 a_{m_n}^{-1} X_{t-1}^* - \cdots - u_p a_{m_n}^{-1} X_{t-p}^*) - \rho(Z_t^*) \Big).$$

As in Section 2 for simple linear regression, the distribution of $\hat{\mathbf{u}}_{m_n}$ given $\mathbf{X}_n = (X_1, \ldots, X_n)$ converges to the same limit distribution as $\hat{\mathbf{u}}_n$. The following theorem summarizes the limit behavior of both U_n^* and $\hat{\mathbf{u}}_{m_n}$.

Theorem 3.1. *Let $X_1, \ldots, X_n$ be observations from the AR(p) model* (3.1), *where $\{Z_t\} \overset{iid}{\sim} F$ with F satisfying* (2.2). *Let $\rho(\cdot)$ be a loss function whose score function $\psi(x) = \rho'(x)$ satisfies:*
(a) *$\psi(\cdot)$ is Lipschitz of order τ_1,*

$$|\psi(x) - \psi(y)| \le C|x - y|^{\tau_1},$$

for some constant $\tau_1 > \max(\alpha - 1, 0)$ and some positive constant C,
(b) *$E|\psi(Z_1)| < \infty$ if $\alpha < 1$,*
(c) *$E\psi(Z_1) = 0$ and Var$(\psi(Z_1)) < \infty$ if $\alpha \ge 1$.*
Then, if $m_n \to \infty$ and $m_n/n \to 0$, $P[U_{m_n}^* \in \cdot|\mathbf{X}_n] \xrightarrow{p} P[U \in \cdot]$, where $U(\cdot)$ is the limit process

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

$\{Z_{k,i}\}$, $\{\delta_k\}$, $\{\Gamma_k\}$ are independent sequences of random variables, $\{Z_{k,i}\} \overset{iid}{\sim} F$, and $\{\delta_k\}$ and $\Gamma_k = E_1 + \cdots + E_k$ are as defined in the statement of Theorem 2.1 (*see also Davis et al.* (1992)). *Moreover, if $\rho(\cdot)$ is convex and $U(\cdot)$ attains a unique minimum at $\hat{\mathbf{u}}$ a.s., then*

$$P[\hat{\mathbf{u}}_{m_n}^* \in \cdot|\mathbf{X}_n] = P[a_{m_n}(\hat{\boldsymbol{\phi}}_{m_n}^* - \hat{\boldsymbol{\phi}}_n) \in \cdot|\mathbf{X}_n] \xrightarrow{p} P[\hat{\mathbf{u}} \in \cdot].$$

The proof of this theorem is omitted since it is almost identical to that given for Theorems 2.1 and 2.2 with Lemma 9 replacing Lemma 4.

As in Section 2, the bootstrap approximation suggested by Theorem 3.1 requires that we know the sequence of normalizing constants $\{a_n\}$ or the ratios $\{a_n/a_{m_n}\}$. Instead, random normalization, such as the maximum of the process, may be used. To incorporate random normalization, define the processes

$$\tilde{U}_n(\mathbf{u}) = \sum_{t=p+1}^{n}\left(\rho(Z_t - u_1\frac{X_{t-1}}{M_n} - \cdots - u_p\frac{X_{t-p}}{M_n}) - \rho(Z_t)\right)$$

and

$$\tilde{U}_n^*(\mathbf{u}) = \sum_{t=p+1}^{m}\left(\rho(Z_t^* - u_1\frac{X_{t-1}^*}{M_{m_n}^*} - \cdots - u_p\frac{X_{t-p}^*}{M_{m_n}^*}) - \rho(Z_t^*)\right),$$

where $M_n = \max\{|X_1|, \ldots, |X_n|\}$ and $M_{m_n}^* = \max\{|X_1^*|, \ldots, |X_{m_n}^*|\}$. Observe that $\tilde{U}_n(\mathbf{u}) = U_n(\mathbf{u}a_n/M_n)$ and $\tilde{U}_n^*(\mathbf{u}) = U_n^*(\mathbf{u}a_{m_n}/M_{m_n}^*)$. Now from the point process result in Theorem 2.4 of Davis and Resnick (1985) and Lemma 8 of the Appendix, we conclude that $a_n^{-1}M_n$ and $a_{m_n}^{-1}M_{m_n}^*$ given $\mathbf{X}_n$ converge in distribution to $\psi_+\Gamma_1^{-1/\alpha}$, where $\psi_+ := \max_{j=0}^{\infty}|\psi_j|$. It follows that $\tilde{U}_n(\cdot)\xrightarrow{d}\tilde{U}(\cdot)$ and $P[\tilde{U}_n^* \in \cdot|\mathbf{X}_n]\xrightarrow{p}P[\tilde{U} \in \cdot]$, where $\tilde{U}(\mathbf{u}) = U(\mathbf{u}\Gamma_1^{1/\alpha}/\psi_+)$. The following theorem is now an immediate consequence of this result.

Theorem 3.2. *If $\rho(\cdot)$ is convex and satisfies the conditions of Theorem 3.1 and $U(\cdot)$ attains a unique minimum at $\hat{\mathbf{u}}$ a.s., then $M_n(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})\xrightarrow{d}\tilde{\mathbf{u}}$ and $P[M_{m_n}^*(\hat{\boldsymbol{\phi}}_{m_n}^* - \hat{\boldsymbol{\phi}}_n) \in \cdot|\mathbf{X}_n]\xrightarrow{p}P[\tilde{\mathbf{u}} \in \cdot]$, where $\tilde{\mathbf{u}} = \hat{\mathbf{u}}\psi_+\Gamma_1^{-1/\alpha}$.*

Unknown Location Parameter. The bootstrap implementation described above can also be used for the case when a location parameter is included in the model. The AR(p) model with location parameter is given by $X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$ and the M-estimates $\hat{\phi}_0, \hat{\phi}_1, \ldots, \hat{\phi}_p$ are found by minimizing

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

with respect to β_0 and $\boldsymbol{\beta}$. Provided ρ is convex with a Lipschitz continuous derivative $\psi(\cdot)$, $a_n(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})$ has the same limit distribution as described in Theorem 3.1 while $n^{1/2}(\hat{\phi}_0 - \phi_0)$ is asymptotically normal with mean 0 and variance $E(\psi^2(Z_1))/(E(\psi'(Z_1)))^2$ (see Davis et al. (1992)). The two quantities are also asymptotically independent.

Now let $\hat{\phi}_0^*$ and $\hat{\boldsymbol{\phi}}^*$ be the M-estimates of the parameters based on the bootstrap replicate $X_1^*, \ldots, X_m^*$. Then, if $m/n \to 0$, the quantities $m^{1/2}(\hat{\phi}_0^* - \hat{\phi}_0)$ and $a_m(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}})$ have the same limiting distribution as the original M-estimates. The proof of this result combines a Taylor series expansion argument (see Davis et al. (1992)) with the stochastic process convergence described in Theorem 3.1. The details are omitted.

Least Absolute Deviation. The loss function $\rho(x) = |x|$ corresponding to least absolute deviation estimation does not meet the technical assumptions of Theorems 3.1 and 3.2. Nevertheless, the bootstrap procedure is still valid for the LAD estimate as described in the following theorem.

Theorem 3.3. *Let $\{X_t\}$ be an AR(p) process satisfying (3.1), where the innovations are assumed to have median 0 if $\alpha \geq 1$. Assume either*
(a) $\alpha < 1$*; or*
(b) $\alpha > 1$ *and* $E|Z_1|^\tau < \infty$ *for some* $\tau < 1 - \alpha$*; or*
(c) $\alpha = 1$ *and* $E(\ln|Z_1|) > -\infty$,
and that $W(\mathbf{u})$ has a unique minimum $\hat{\mathbf{u}}$ a.s., where $W(u) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}[|Z_{i,j} - (\psi_{i-1}u_1 + \cdots + \psi_{i-p}u_p)\delta_j\Gamma_j^{-1/\alpha}| - |Z_{i,j}|]$. If $m \to \infty$ and $m/n \to 0$, then

$P[a_m^{-1}(\hat{\boldsymbol{\phi}}_m^* - \hat{\boldsymbol{\phi}}) \in \cdot \mid \mathbf{X}_n] \xrightarrow{p} P[\hat{\mathbf{u}} \in \cdot]$, *where* $\hat{\boldsymbol{\phi}}$ *is the LAD estimate based on the original data* $X_1, \ldots, X_n$ *and* $\hat{\boldsymbol{\phi}^*}_m$ *is the LAD estimator based on the bootstrap replicate* $X_1^*, \ldots, X_m^*$.

Remark 2. A similar result is also valid for LAD estimation in the multivariate regression model of Section 2.

Acknowledgement

This research was supported by NSF Grants DMS 9100392 and 9504596.

Appendix

In this section we collect some of the technical results used throughout the paper. Much of the requisite background material on point processes, as well as notation and definitions, can be found in Davis and Resnick (1985), Resnick (1987), and Davis et al. (1992). For Lemmas 1–4 below, the assumptions of Theorem 2.1 are assumed to be met.

Lemma 1. *Let* $\mu_n(dz, dx)$ *and* $\hat{\mu}_n(dz, dx)$ *be the random measures defined on rectangles of the form* $E = A \times B \subset \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ *as*

$$\mu_n(E) = \Big(n^{-1} \sum_{i=1}^n I(Z_i \in A)\Big)\Big(m_n n^{-1} \sum_{i=1}^n I(a_{m_n}^{-1} X_i \in B)\Big)$$

and

$$\hat{\mu}_n(E) = \Big(n^{-1} \sum_{i=1}^n I(\hat{Z}_i \in A)\Big)\Big(m_n n^{-1} \sum_{i=1}^n I(a_{m_n}^{-1} X_i \in B)\Big).$$

Then if $m_n/n \to 0$,

$$\mu_n(E) \xrightarrow{p} \nu(E), \tag{A.1}$$

where $\nu(dz, dx) = G(dz) \times \lambda(dx)$, *G is the distribution function of* Z_i *and* $\lambda(dx) = \alpha(px^{-\alpha-1} I(x > 0) + (1-p)(-x)^{-\alpha-1} I(x < 0))dx$. *Moreover, if* $P[Z_1 \in \partial A] = 0$, *then*

$$\hat{\mu}_n(E) - \mu_n(E) \xrightarrow{p} 0. \tag{A.2}$$

Proof. The Laplace transform of $m_n n^{-1} \sum_{i=1}^n I(a_{m_n}^{-1} X_i \in B)$ is

$$\left[1 + \frac{1}{n}\frac{n}{m_n}(e^{-t\frac{m_n}{n}} - 1) m_n P(a_{m_n}^{-1} X_1 \in B)\right]^n$$

which by (2.2) converges to $e^{-t\lambda(B)}$ as $n \to \infty$. Thus $\mu_n(E) \xrightarrow{p} EI(Z_1 \in B)\lambda(B) = \nu(E)$.

As for the second statement, we have

$$\hat{\mu}_n(E)-\mu_n(E)=\Big(n^{-1}\sum_{j=1}^{n}\Big(I(Z_j\in A)-I(\hat{Z}_j\in A)\Big)\Big)\Big(m_n n^{-1}\sum_{j=1}^{n}I(a_{m_n}^{-1}X_j\in B)\Big).$$

By the argument above, the second term in parentheses converges in probability to $\lambda(B)$. On the other hand, the modulus of the first factor has expectation bounded by

$$E|I(Z_1-(\hat{\beta}-\beta)X_1\in A)-I(Z_1\in A)|$$

which converges to 0 since the M-estimate $\hat{\beta}$ is consistent. This proves (A.2).

Lemma 2. *Let μ_n^* be the random point measure*

$$\mu_n^*(\cdot)=\sum_{j=1}^{m_n}\epsilon_{(Z_j^*,a_{m_n}^{-1}X_j^*)}(\cdot)$$

defined on $\mathbb{R}\times(\mathbb{R}\setminus\{0\})$. For any collection of bounded disjoint rectangles $E_1,\ldots,E_d\in\mathbb{R}\times(\mathbb{R}\setminus\{0\})$,

$$P[(\mu_n^*(E_1),\ldots,\mu_n^*(E_d))\in\cdot\,|\mathbf{X}_\infty,\mathbf{Y}_\infty]\xrightarrow{p}P[(\mu(E_1),\ldots,\mu(E_d))\in\cdot)],$$

where $\mu(\cdot)$ is the Poisson process $\sum_{j=1}^{\infty}\epsilon_{(Z_j,\delta_j\Gamma_j^{-1/\alpha})}(\cdot)$.

Proof. It is enough to show that for any non-negative integers $r_1,\ldots,r_d$

$$P[\mu_n^*(E_1)=r_1,\ldots,\mu_n^*(E_d)=r_d\,|\,\mathbf{X}_\infty,\mathbf{Y}_\infty]\xrightarrow{p}P[\mu(E_1)=r_1,\ldots,\mu(E_d)=r_d].$$

Using the independence of the sample $\{(Z_i^*,X_i^*),i=1,\ldots,m_n\}$, the left-hand side is equal to

$$\begin{aligned}&\tfrac{m_n!}{r_1!\cdots r_d!(m_n-r_1-\cdots-r_d)!}\Big(\frac{\hat{\mu}_n(E_1)}{m_n}\Big)^{r_1}\cdots\Big(\frac{\hat{\mu}_n(E_d)}{m_n}\Big)^{r_d}\Big(1-\frac{\hat{\mu}_n(\cup_{j=1}^d E_j)}{m_n}\Big)^{m_n-r_1-\cdots-r_d}\\ &\xrightarrow{p}\frac{1}{r_1!\cdots r_d!}\nu^{r_1}(E_1)\cdots\nu^{r_d}(E_d)\exp\{-(\nu(E_1)+\cdots+\nu(E_d))\}\\ &=P[\mu(E_1)=r_1,\ldots,\mu(E_d)=r_d],\end{aligned}$$

where the limit follows from Lemma 1.

Lemma 3. *For any continuous function g on $\mathbb{R}\times(\mathbb{R}\setminus\{0\})$ with compact support,*

$$P[\mu_n^*(g)\in\cdot\,|\,\mathbf{X}_\infty,\mathbf{Y}_\infty]\xrightarrow{p}P[\mu(g)\in\cdot],\tag{A.3}$$

where $\mu_n^*(g)=\int g\,d\mu_n^*$ and $\mu(g)=\int g\,d\mu$.

Proof. By Lemma 2, (A.3) holds for a suitably chosen class of step functions. Now if g is continuous with support contained in a compact rectangle E, then for any $\epsilon > 0$ there exist a constant K_ϵ and a step function g_ϵ with support E such that $P[\mu(E) > K_\epsilon] < \epsilon$ and $|g(z,x) - g_\epsilon(z,x)| \le \epsilon/K_\epsilon$ for all z and x. Since

$$|\mu_n^*(g) - \mu_n^*(g_\epsilon)| \le \frac{\epsilon}{K_\epsilon}\mu_n^*(E),$$

it follows that

$$\begin{aligned} P[\mu_n^*(g) \le x \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] &\le P[\mu_n^*(g_\epsilon) \le x + \epsilon\mu_n^*(E)/K_\epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &\le P[\mu_n^*(g_\epsilon) \le x + \epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] + P[\mu_n^*(E) > K_\epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &\xrightarrow{p} P[\mu(g_\epsilon) \le x + \epsilon] + P[\mu(E) > K_\epsilon] \\ &\le P[\mu(g) \le x + 2\epsilon] + 2P[\mu(E) > K_\epsilon]. \end{aligned}$$

A similar lower bound can be obtained in exactly the same fashion. Letting $\epsilon \to 0$, we find that $P[\mu_n^*(g) \le x \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} P[\mu(g) \le x]$ from which (A.3) follows using a routine weak convergence argument.

Lemma 4. *For any* $u_1, \ldots, u_d \in \mathbb{R}$,

$$P[(W_n^*(u_1), \ldots, W_n^*(u_d)) \in \cdot \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} P[(W(u_1), \ldots, W(u_d)) \in \cdot].$$

Proof. We just provide the proof for $d = 1$ since the case $d > 1$ is similar using the Cramér-Wold device. First note that

$$W_n^*(u) = \int_{\mathbb{R}\times(\mathbb{R}\setminus\{0\})} g\, d\mu_n^*,$$

where $g(z,x) = \rho(z - ux) - \rho(z)$. Set $S_K = \{(z,x) : |z| \le K_1, K_2^{-1} \le |x| \le K_2\}$ and define $g_K = gI_{S_K}(z,x)$. Using a modification of Lemma 3, we have

$$P[\mu_n^*(g_K) \in \cdot \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} P[\mu(g_K) \in \cdot], \tag{A.4}$$

so that it just remains to replace K (K_1 and K_2) by ∞ in (A.4).

For any $\epsilon > 0$,

$$\begin{aligned} &P[|\mu_n^*(g - g_K)| > 3\epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &\le P[|\mu_n^*(gI(|x| \le K_2^{-1}))| > \epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] + P[|\mu_n^*(gI(|x| > K_2))| > \epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &\quad + P[|\mu_n^*(gI(|z| > K_1, K_2^{-1} < |x| < K_2))| > \epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\ &=: I + II + III. \end{aligned} \tag{A.5}$$

We handle each of the three terms separately. We have

$$II = P[(|\sum_{i=1}^{m_n} g(Z_i^*, a_{m_n}^{-1}X_i^*)I(a_{m_n}^{-1}|X_i^*| > K_2)| > \epsilon \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty]$$

$$
\begin{aligned}
&\leq P[\cup_{i=1}^{m_n}\{a_{m_n}^{-1}|X_i^*| > K_2\} \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\
&\leq m_n P[a_{m_n}^{-1}|X_1^*| > K_2 \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\
&= m_n n^{-1} \sum_{i=1}^{n} I(a_{m_n}^{-1}|X_i| > K_2) \\
&\xrightarrow{p} \lambda[K_2, \infty) \qquad \text{(as } n \to \infty \text{ by Lemma 1)} \\
&\to 0
\end{aligned}
$$

as $K_2 \to \infty$. Next, by Lemma 1,

$$
\begin{aligned}
III &\leq P[\cup_{i=1}^{m_n}\{|Z_i^*| > K_1, K_2^{-1} < a_{m_n}^{-1}|X_i^*| \leq K_2\} \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\
&\leq m_n P[|Z_1^*| > K_1, K_2^{-1} < a_{m_n}^{-1}|X_1^*| \leq K_2 \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\
&= \hat{\mu}_n(\{|z| > K_1\} \times \{K_2^{-1} < |x| \leq K_2\}) \\
&\xrightarrow{p} P[|Z_1| > K_1]\lambda(K_2^{-1} < |x| \leq K_2) \qquad \text{(as } n \to \infty) \\
&\to 0
\end{aligned}
$$

as $K_1 \to \infty$ and then $K_2 \to \infty$. As for the first term in (A.5), we have, using the bound $|g(z,x)| \leq |ux\psi(z)| + C|ux|^{1+\tau_1}$, that

$$
\begin{aligned}
I \leq\; & P[|ua_{m_n}^{-1} \sum_{i=1}^{m_n} X_i^* \psi(Z_i^*) I(a_{m_n}^{-1}|X_i^*| \leq K_2^{-1})| > \epsilon/2 \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \\
&+ P[C|ua_{m_n}^{-(1+\tau_1)} \sum_{i=1}^{m_n} |X_i^*|^{1+\tau_1} I(a_{m_n}^{-1}|X_i^*| \leq K_2^{-1})| > \epsilon/2 \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty]. \qquad \text{(A.6)}
\end{aligned}
$$

Using Markov's inequality and Karamata's theorem, the mean of the second term in (A.6) is bounded by

$$
\begin{aligned}
&C|u|2\epsilon^{-1} m_n a_{m_n}^{-(1+\tau_1)} E\left[E\left(|X_1^*|^{1+\tau_1} I(a_{m_n}^{-1}|X_1^*| \leq K_2^{-1}) \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty\right)\right] \\
&= (\text{const}) m_n a_{m_n}^{-(1+\tau_1)} E\Big(n^{-1} \sum_{i=1}^{n} |X_i|^{1+\tau_1} I(|X_i| \leq K_2^{-1} a_{m_n})\Big) \\
&\sim (\text{const})\alpha(1+\tau_1-\alpha)^{-1} m_n a_{m_n}^{-(1+\tau_1)} (K_2^{-1}a_{m_n})^{1+\tau_1} P[|X_1| > K_2^{-1}a_{m_n}] \\
&\to (\text{const})\alpha(1+\tau_1-\alpha)^{-1} K_2^{\alpha-(1+\tau_1)} \qquad \text{(as } n \to \infty) \\
&\to 0
\end{aligned}
$$

as $K_2 \to \infty$.

In order to show that I converges to 0 after taking a limit on $n \to \infty$ and then $K_2 \to \infty$, we require the following ancillary results:
For all $\gamma > 0$

$$
m_n P[|a_{m_n}^{-1} X_1^* \psi(Z_1^*) I(a_{m_n}^{-1}|X_1^*| \leq \delta)| \geq \gamma \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} 0, \qquad \text{(A.7)}
$$

as $n \to \infty$ and $\delta \to 0$;

$$m_n E[a_{m_n}^{-1} X_1^* \psi(Z_1^*) I(a_{m_n}^{-1}|X_1^*| \le \delta) I(|X_1^* \psi(Z_1^*)| \le a_{m_n}\gamma) \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} 0 \quad \text{(A.8)}$$

as $n \to \infty$ and then $\delta \to 0$; and

$$m_n E[a_{m_n}^{-2} (X_1^* \psi(Z_1^*))^2 I(a_{m_n}^{-1}|X_1^*| \le \delta) I(a_{m_n}^{-1}|X_1^* \psi(Z_1^*)| \le \gamma) \,|\, \mathbf{X}_\infty, \mathbf{Y}_\infty] \xrightarrow{p} 0 \quad \text{(A.9)}$$

as $n \to \infty, \delta \to 0$, and $\gamma \to 0$. The proofs of these relations are omitted since they use standard arguments which rely on Markov's inequality, Karamata's theorem, and the assumption on ψ.

To finish the proof of the lemma, write for any $\gamma > 0$

$$\begin{aligned} & a_{m_n}^{-1} \sum_{i=1}^{m_n} X_i^* \psi(Z_i^*) I(a_{m_n}^{-1}|X_i^*| \le K_2^{-1}) \\ = & a_{m_n}^{-1} \sum_{i=1}^{m_n} X_i^* \psi(Z_i^*) I(a_{m_n}^{-1}|X_i^*| \le K_2^{-1}) I(a_{m_n}^{-1}|X_i^* \psi(Z_i^*)| \le \gamma) \\ & + a_{m_n}^{-1} \sum_{i=1}^{m_n} X_i^* \psi(Z_i^*) I(a_{m_n}^{-1}|X_i^*| \le K_2^{-1}) I(a_{m_n}^{-1}|X_i^* \psi(Z_i^*)| > \gamma). \end{aligned}$$

By (A.8) and (A.9) the conditional variance and conditional mean of the first term converges to 0 in probability as $n \to \infty, K_2 \to \infty$ and $\gamma \to 0$. Similarly, by (A.7), the conditional probability that the last term is positive also converges to 0 as the same 3 indices tend to their respective limits. Hence the first term in (A.6) must converge to 0 in probability as $n \to \infty, K_2 \to \infty$ as claimed.

We now turn our attention to establishing analogues of the foregoing for the case of an autoregressive process. In the remainder of the appendix, the assumptions of Section 3, namely that $\{X_t\}$ is an AR(p) process satisfying (3.1) where $\{Z_t\} \overset{iid}{\sim} F$ with F satisfying the regular variation condition (2.2), are assumed to be met. The corresponding sequence of point processes is now given by

$$\mu_n^*(\cdot) = \sum_{t=1}^{m} \epsilon_{(Z_t^*, a_m^{-1} Y_t^*)}(\cdot),$$

where

$$Y_{t-1}^* = u_1 X_{t-1}^* + \cdots + u_p X_{t-p}^*.$$

We write P_n and E_n for the probability measure and expectation functional, respectively, conditional on $\mathbf{X}_n = (X_1, \ldots, X_n)'$. The first objective is to show that

$$P_n[\mu_n^* \in \cdot] \xrightarrow{d} P[\mu \in \cdot], \quad \text{(A.10)}$$

where μ is the random measure

$$\mu(\cdot)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\epsilon_{(Z_{i,j},(u_1\psi_{i-1}+\cdots+u_p\psi_{i-p})\delta_j\Gamma_j^{-1/\alpha})},$$

the ψ_j's are defined in (3.2) ($\psi_j:=0$ if $j<0$) and the sequences $\{Z_{i,j}\}$, $\{\delta_j\}$ and $\{\Gamma_j\}$ are as specified in the statement of Theorem 2.1. The proof of (A.10) follows the sequence of steps (see Theorem 2.4 in Davis and Resnick (1985)) used for establishing the result

$$\mu_n(\cdot):=\sum_{t=1}^{n}\epsilon_{(Z_t,a_n^{-1}Y_t)}(\cdot)\xrightarrow{d}\mu(\cdot).$$

We break up the proof of (A.10) into a series of lemmas.

Lemma 5. *For any* $k\geq 1$

$$P_n[I_{m,k}\in\cdot]\xrightarrow{p}P[I_k\in\cdot],$$

where $I_{m,k}=\sum_{t=1}^{m}\epsilon_{(Z_t^*,a_m^{-1}\mathbf{Z}_{t-1}^*)}$, $\mathbf{Z}_{t-1}^*=(Z_{t-1}^*,\ldots,Z_{t-k}^*)$, $I_k=\sum_{i=1}^{k}\sum_{j=1}^{\infty}\epsilon_{(Z_{i,j},\delta_j\gamma^{-1/\alpha}\mathbf{e}_i)}$ *and* $\mathbf{e}_i$ *is the basis element of* $\mathbb{R}^k$ *with* i*th component equal to one and the rest* 0. (*The relevant state space for the point processes is* $\mathbb{R}\times(\mathbb{R}^k\setminus(0,\ldots,0))$.)

The proof of this lemma is quite similar to the combined arguments used for Proposition 2.1 and Theorem 2.2 of Davis and Resnick (1985). The technical details are omitted.

Lemma 6. *Set* $Y^*=\sum_{j=0}^{\infty}|c_jZ_j^*|$ *where the coefficients* $\{c_j\}$ *decrease to* 0 *at an exponential rate. Then*
(i) $\limsup_n mE[P_n[Y^*>a_mx]I(|\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}|<\delta)]\leq(const)x^{-\alpha}\sum_{j=0}^{\infty}|c_j|^{\alpha}$
and
(ii) $\limsup_n ma_m^{-\gamma}E[E_n[(Y^*)^{\gamma}I(Y^*\leq a_m^{\delta})]|I(|\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}|<\delta)]\leq(const)\delta^{\gamma-\alpha}$
for all $\gamma>\alpha$.

Proof. (i) Following the argument given on p.228–230 of Resnick (1987), we have

$$\begin{aligned}
&mP_n[Y^*>a_mx]\\
&=mP_n[Y^*>a_mx,\bigvee_j|c_jZ_j^*|>a_mx]+mP_n[Y^*>a_mx,\bigvee_j|c_jZ_j^*|\leq a_mx]\\
&\leq\sum_j mP_n[|c_jZ_j^*|>a_mx]+mP_n[\sum_j|c_jZ_j^*|I(|c_jZ_j^*|\leq a_mx)>a_mx]\\
&\leq\sum_j\Big(mn^{-1}\sum_{t=1}^{n}I(|c_j\hat{Z}_t|>a_mx)\Big)+\sum_j\Big(mx^{-1}a_m^{-1}n^{-1}\sum_{t=1}^{n}|c_j\hat{Z}_t|I(|c_j\hat{Z}_t|>a_mx)\Big)\\
&=:A+B.
\end{aligned}$$

Also, $\hat{Z}_t = Z_t + (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\mathbf{X}_{t-1}$ which, on the set $|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta$ for δ small, is bounded by $|Z_t| + \delta|U_{t-1}|$ where $U_{t-1} = X_{t-1} + \cdots + X_{t-p} =: \sum_{j=0}^{\infty} d_j Z_{t-1-j}$. Using this bound, and the fact that $|Z_t| + \delta|U_{t-1}|$ has regularly varying tail probabilities, we have

$$
\begin{aligned}
E[AI(|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta)] &\leq \sum_j mP[|c_j|(|Z_1| + \delta|U_0|) > a_m x] \\
&\to \sum_j (|c_j|^\alpha (1 + \delta^\alpha d_+) x^{-\alpha}) \\
&\leq (\text{const}) x^{-\alpha} \sum_j |c_j|^\alpha,
\end{aligned}
$$

where $d_+ := \sum_j |d_j|^\alpha$. As for the second term, assume that $0 < \alpha < 1$. Then, by Karamata's theorem,

$$
\begin{aligned}
E[BI(|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta)] &\leq \sum_j (mx^{-1} a_m^{-1} |c_j| E(|Z_1| + \delta|U_0|) I(|c_j|(|Z_1| + \delta|U_0|) > a_m x)] \\
&\to (\text{const}) x^{-\alpha} \sum_j |c_j|^\alpha (1 + \delta^\alpha d_+).
\end{aligned}
$$

The case $\alpha \geq 1$ is handled using the method described in Resnick (1987).

(ii) We have

$$
\begin{aligned}
m a_m^{-\gamma} E_n[(Y^*)^\gamma I(Y^* \leq a_m^\delta)] &\leq m a_m^\gamma \int_0^{(a_m \delta)^\gamma} P_n[(Y^*)^\gamma > x] dx. \\
&= m \int_0^\delta P_n[Y^* > a_m x] dx.
\end{aligned}
$$

After taking expectations on the set $|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta$ for δ small, it can be shown using the first part of the lemma that the resulting limit (as $m \to \infty$) is

$$
(\text{const}) \sum_j |c_j|^\alpha \int_0^\delta x^{\gamma - 1 - \alpha} dx = (\text{const}) \delta^{\gamma - \alpha}.
$$

This completes the proof of the lemma.

Lemma 7. *For any* $\epsilon > 0$ *and* $\gamma > 0$

$$
\lim_{k \to \infty} \limsup_{m \to \infty} P[P_n[a_m^{-1} \bigvee_{t=1}^m |\sum_{j=0}^k \hat{\psi}_j Z_{t-j}^* - Y_{t-1}^*| > \gamma] > \epsilon] = 0.
$$

Proof. Since $\hat{\boldsymbol{\phi}}$ is weakly consistent, it suffices to consider the outside probability on the set $|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta$ for some small δ. If δ is sufficiently small, then on this set, the modulus of the coefficients $\hat{\psi}_j$ can be bounded by positive constants c_j which are exponentially decreasing. It follows that the inner-conditional probability, on the set $|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta$, is bounded by

$$mP_n[\sum_{j>k} |\hat{\psi}_j||Z_j^*| > a_m\gamma]I(|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta) \le mP_n[\sum_{j>k} |c_j||Z_j^*| > a_m\gamma]I(|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}| < \delta)$$

and the latter has the desired limit by Lemma 6(i). This proves the result.

Lemma 8. $P_n[\mu_n^* \in \cdot] \xrightarrow{d} P[\mu \in \cdot]$.

Proof. The proof of this result is omitted since it is essentially identical to the proof of Theorem 2.4 given in Davis and Resnick (1985) with Lemma 2.3 being replaced by Lemma 7.

Lemma 9. *For any* $u_1, \ldots, u_d \in \mathbb{R}$,

$$P_n[(U_n^*(u_1), \ldots, U_n^*(u_d)) \in \cdot] \xrightarrow{p} P[(U(u_1), \ldots, U(u_d)) \in \cdot].$$

Proof. The proof of this result follows the argument given for the proof of Theorem 2.1 in Davis et al. (1992). Specifically, it suffices to establish the analogues of (2.7)–(2.10) in Davis et al. (1992) which in turn were immediate consequences of their Proposition A2. In the current setting, it suffices to show that for all $\epsilon, \eta > 0$

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\Big[P_n[\Big|a_m^{-1}\sum_{t=1}^m Y_{t-1}^* I(|Y_{t-1}^*| > a_m\delta)\psi(Z_t^*)I(|Z_t^*| > K)\Big| > \eta] > \epsilon\Big] = 0,$$

$$a_m^{-(1+\tau_1)} \sum_{t=1}^m |Y_{t-1}^*|^{1+\tau_1} I(|Y_{t-1}^*| \le a_m\delta) \xrightarrow{p} 0,$$

as $m \to \infty$ and then $\delta \to 0$, and

$$a_m^{-(1+\tau_1)} \sum_{t=1}^m |Y_{t-1}^*|^{1+\tau_1} I(|Y_{t-1}^*| > a_m\delta)I(|Z_t^*| > K) \xrightarrow{p} 0,$$

as $m \to \infty$ and $K \to \infty$. The adaptation of Proposition A2 to the present framework is straightforward with Lemma 6 playing the key role. The tedious details are omitted.

References

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15**, 724-731.

Athreya, K. B., Lahiri, S. and Wu, W. (1993). Inference for heavy tailed distributions. Preprint.

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

Davis, R. A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Proc. Appl.* **40**, 145-180.

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

Davis, R. A. and Wu, W. (1997). M-estimation for linear regression and autoregression with infinite variance. *Probab. Math. Statist.* **17**, 117-136.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

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

Department of Statistics, Colorado State University, Fort Collins, CO 80523, U.S.A.

E-mail: rdavis@stat.colostate.edu

Schering-Plough Research Institute, 2015 Galloping Hill Road, k-15-2-2445, Kenilworth, NJ 07033-0539, U.S.A.

E-mail: Wei.Wu@spcorp.com

(Received February 1995; accepted March 1997)