Bootstrapping $M$-Estimates in Regression and Autoregression with Infinite Variance.

Richard A. Davis*
Colorado State University
and
Wei Wu
University of Illinois

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.

*Research supported by NSF Grant DMS 9100392.

AMS 1980 Subject Classification: 62J05, 62M10

Key Words and Phrases. $M$-estimation, bootstrap, regular variation, stable laws, Poisson processes, autoregressive processes.
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 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 has been thoroughly studied in the heavy-tailed regression and autoregression setting by Davis and Wu (1994) 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

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

where $\phi(x) = 1 - \phi_1 x - \cdots - \phi_p x^p \neq 0$ for $|z| \leq 1$ and $\{Z_t\}$ is an iid sequence of random variables with distribution function $F$ which we assume belongs to the domain of attraction of an $\alpha$-stable law with $\alpha \in (0, 2)$. The latter condition implies that there exists a sequence of non-negative constants $a_n \rightarrow \infty$ such that

$$nP[a_n^{-1} Z_1 \in dx] \xrightarrow{\mu} \lambda(dx)$$

where $\lambda$ is a Lévy measure and $\xrightarrow{\mu}$ 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 $\hat{\phi}$ of $\phi = (\phi_1, \ldots, \phi_p)$ 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 $\beta = (\beta_1, \ldots, \beta_p)$. 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

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

where $\hat{u}$ 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 $\hat{u}$. In general, a closed form expression for $\hat{u}$ is impossible to obtain. Alternatively, one could simulate from the limit stochastic process and find the sample
path minimum from which quantiles of $\hat{u}$ 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 iid sequence of random variables generated from the empirical distribution of the estimated residuals. 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 $a_n(\hat{\phi} - \phi)$, conditional on $X_1, \ldots, X_n$, has the same limit distribution as $a_n(\hat{\phi} - \phi)$ 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 result 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 phenomenon (see for example Fukuchi (1994)). 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

$$m_n P[a_m^{-1} Z^*_1 \in \cdot |X_1, \ldots, X_n] \Rightarrow \lambda(\cdot).$$

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

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

which converges in probability to $\lambda(B)$ if and only if $m_n/n \to 0$. In particular, if $m_n = n$, then the above 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 simple linear regression model when the independent variables are heavy-tailed. Section 4 considers extensions of the results in Sections 2 and 3 to multiple regression, unknown location parameter and LAD estimation. Proofs of the more technical results in Sections 2 and 3 are contained in the Appendix.
2. Simple linear regression

Let \((Y_i, X_i), i = 1, \ldots, n\), be observations from the simple linear model

\[
y_i = \beta x_i + z_i, \quad i = 1, \ldots, n
\]

where \(\{z_i\}_{i=1}^n \sim^i i.d. G\) and \(\{x_i\}_{i=1}^n \sim^i i.d. F\). 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 \(X_i \in D(\alpha)\)), i.e. there exist a slowly varying function \(L(x)\) at \(\infty\), constants \(0 \leq p, q \leq 1\), \(p + q = 1\), and \(\alpha \in (0, 2)\), such that

\[
1 - F(x) \sim px^{-\alpha}L(x),
\]

\[
F(-x) \sim q x^{-\alpha}L(x), \quad \text{as } x \to \infty.
\]

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 \(nE[X_1I(|X_1| \leq a_n)]\), converge in distribution to a stable distribution.

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

\[
g(\phi) := \sum_{i=1}^n \rho(y_i - \phi x_i) = \sum_{i=1}^n \rho(z_i - (\phi - \beta) x_i).
\]

As in Davis and Wu (1994), 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 \(C(\mathbb{R})\)

\[
W_n(u) = \sum_{i=1}^n (\rho(z_i - ua_n^{-1} x_i) - \rho(z_i)).
\]

With this parameterization, the minimizer of \(W_n(u)\) is given by \(\hat{u}_n = a_n(\hat{\beta} - \beta)\). In Davis and Wu (1994), the stochastic processes \(W_n(\cdot)\) were shown to converge in distribution to a limit stochastic process \(W(\cdot)\) from which it followed that \(\hat{u}_n = a_n(\hat{\beta} - \beta)\) converged 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

\[
\hat{z}_j = y_j - \hat{\beta} x_j, \quad j = 1, \ldots, n,
\]

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 \(\hat{F}_{Z,n}(x) = n^{-1} \sum_{j=1}^n I[z_j \leq x]\) 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

\[
Y_j^* = \hat{\beta} X_j^* + \hat{z}_j^*, \quad j = 1, \ldots, m,
\]
where \( \{(X_j^*, Z_j^*) \mid j = 1, \ldots, m\} \) is a random sample from \( d\bar{F}_n \). The bootstrap replicate \( \hat{u}_m^\ast := a_m(\hat{\beta}^\ast - \hat{\beta}) \) of \( \hat{u}_n \) is then found by minimizing

\[
W_m^\ast(u) = \sum_{j=1}^{m} \left( \rho(Z_j^* - u a_m^{-1} X_j^*) - \rho(Z_j^*) \right).
\]

Provided the resample size \( m = m_n \) is 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 \( \hat{u}_n \),

\[
P[\hat{u}_m \leq x \mid X_\infty, Y_\infty] \xrightarrow{p} P[\hat{u} \leq x],
\]

where \( X_\infty = (X_1, X_2, \ldots) \) and \( Y_\infty = (Y_1, Y_2, \ldots) \).

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

\[
d_0(\lambda_1, \lambda_2) = \sum_{k=1}^{\infty} 2^{-k} \frac{|\int g_k \, d\lambda_1 - \int g_k \, d\lambda_2|}{1 + |\int g_k \, d\lambda_1 - \int g_k \, d\lambda_2|}
\]

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 \( \int g_k dL_n \xrightarrow{p} \int g_k dL \) 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 \( \int f_k dQ_n \xrightarrow{p} \int f_k dQ \) 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 \( \tau_1 \),

\[
|\psi(x) - \psi(y)| \leq C|x - y|^{\tau_1},
\]

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 \( \alpha \geq 1 \).

Then if \( m_n \to \infty \) and \( m_n/n \to 0 \)

\[
L_n(\cdot) := P[W_m^\ast \in \cdot \mid X_\infty, Y_\infty] \xrightarrow{p} P[W \in \cdot]
\]

\[=: L(\cdot) \]
where \( W(\cdot) \) is the limit process defined in Theorem 2.1 of Davis and Wu (1994). 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, \( \{Z_k\} \sim F, \{\delta_k\} \) are iid with \( P(\delta_k = 1) = p = 1 - P(\delta_k = -1), \) and \( \Gamma_k = E_1 + \cdots + E_k, \) where \( E_i \)'s are iid 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

\[
Q_n(\cdot) := P[\hat{u}_{m_n} \in \cdot | X_{\infty}, Y_{\infty}]
= P[a_{m_n}(\hat{\beta} - \hat{\beta}) \in \cdot | X_{\infty}, Y_{\infty}]
\]

\[
\leq P[\hat{u} \in \cdot]
=: Q(\cdot).
\]

**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 \( X_{\infty}, Y_{\infty}, W_{n_{k'}}^* \xrightarrow{d} W \) on \( C(\mathbb{R}) \). Now by Lemma 4 of the Appendix, we have for any \( u_1, \ldots, u_j \in \mathbb{R} \),

\[
L_n \circ \pi_{u_{1}, \ldots, u_j} \xrightarrow{p} L \circ \pi_{u_{1}, \ldots, u_j}
\]

on \( \mathcal{M}_p(\mathbb{R}^d) \) where \( \pi_{u_1, \ldots, u_j} : x \mapsto (x(u_1), \ldots, x(u_j)) \) 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 \( \Omega_0 \) such that for all outcomes in \( \Omega_0 \) and any \( j \),

\[
L_{n_{k'}} \circ \pi_{q_1, \ldots, q_j} \rightarrow L \circ \pi_{q_1, \ldots, q_j}
\]

or, equivalently,

\[
(W_{n_{k'}}^*(q_1), \ldots, W_{n_{k'}}^*(q_j)) \xrightarrow{d} (W(q_1), \ldots, W(q_j))
\]

as \( k' \rightarrow \infty \). 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 \( X_{\infty}, Y_{\infty} \). By Theorem 4.2 in Billingsley (1968), it is enough to check that for almost all sample paths and for every \( \epsilon, \eta > 0 \), there exists a \( \delta > 0 \) such that

\[
P[\sup_{|t-s| \leq \delta} |W_{n_{k'}}^*(t) - W_{n_{k'}}^*(s)| > \epsilon | X_{\infty}, Y_{\infty} | < \eta
\]
for all $k'$. Writing $n'$ in place of $n_{k'}$ we have for any $s, t \in [-M, M]$,

\begin{equation}
|W_{n'}^*(s) - W_{n'}^*(t)| \leq |(s - t) \sum_{i=1}^{m_{n'}} a_{m_{n'}}^{-1} X_i^* \psi(Z_i^*) + (s - t) \sum_{i=1}^{m_{n'}} a_{m_{n'}}^{-1} X_i^* (\psi(\xi_i^*) - \psi(Z_i^*))|
\end{equation}

where $|\xi_i^* - Z_i^*| \leq (|s| \vee |t|) a_{m_{n'}}^{-1} |X_i^*| \leq M a_{m_{n'}}^{-1} |X_i^*|$. Using the assumptions on $\psi(\cdot)$, (2.8) may be bounded above by

$$
|s - t||\sum_{i=1}^{m_{n'}} a_{m_{n'}}^{-1} X_i^* \psi(Z_i^*)| + |s - t|CM^{T_1} \sum_{i=1}^{m_{n'}} (a_{m_{n'}}^{-1} |X_i^*|)^{1 + \gamma_1}.
$$

Applying the bootstrap results for sample means in Athreya et al. (1993), $a_{m_{n'}}^{-1} \sum_{i=1}^{m_{n'}} X_i^* \psi(Z_i^*)$ and $a_{m_{n'}}^{-1} \sum_{i=1}^{m_{n'}} |X_i^*|^{1 + \gamma_1}$ are bounded in probability for almost all sample paths of $X_\infty$ and $Y_\infty$. 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 $X_\infty, Y_\infty$,

$$
W_{n_{k'}}^* \xrightarrow{d} W
$$

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

$$
Q_{n_{k'}}(\cdot) = P[\hat{u}_{n_{k'}}^* \in \cdot | X_\infty, Y_\infty] \xrightarrow{a.s.} P[\hat{u} \in \cdot] = Q(\cdot),
$$

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

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

and

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

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

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

$$P[\hat{u}^{*}_{n} \in \cdot|X_{\infty}, Y_{\infty}] = P[M^{*}_{n} (\hat{\beta}^{*} - \hat{\beta}) \in \cdot|X_{\infty}, Y_{\infty}]$$

$$\overset{P}{\rightarrow} P[\hat{u} \in \cdot$$

where $\hat{u} = \hat{u}\Gamma^{-1/\alpha}_{1}$.

Proof: Observe that

$$\tilde{W}_{n}(u) = W_{n}(u_{n}/M_{n})$$

and since $a^{-1}_{n}M_{n} \overset{d}{\rightarrow} \Gamma^{-1/\alpha}_{1}$, we conclude that $\tilde{W}_{n}(\cdot) \overset{d}{\rightarrow} \tilde{W}(\cdot)$ where $\tilde{W}(u) := W(u\Gamma_{1}^{1/\alpha})$. It follows that $\tilde{a}_{n} \overset{d}{\rightarrow} \tilde{a} = \hat{u}\Gamma_{1}^{-1/\alpha}$. Now, using Lemma 3 and a standard point process argument, $a^{-1}_{n}M_{n}^{*}$ given $X_{\infty}, Y_{\infty}$ converges in distribution to $\Gamma^{-1/\alpha}_{1}$ and the convergence is joint with that of $W_{n}^{*}$. We deduce that

$$P[\tilde{W}_{n}^{*} \in \cdot|X_{\infty}, Y_{\infty}] \overset{P}{\rightarrow} P[\tilde{W} \in \cdot]$$

and the remainder of the theorem is now argued as in the proof of Theorem 2.2. \qed

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

$$X_{t} = \phi_{1}X_{t-1} + \cdots + \phi_{p}X_{t-p} + Z_{t}$$

where $\phi(z) = 1 - \phi_{1}z - \cdots - \phi_{p}z^{p} \neq 0$ for $|z| \leq 1$ and \{Z_{t}\} \iid F. The causality assumption implies that $X_{t}$ can be represented as the linear process

$$X_{t} = \sum_{j=0}^{\infty} \psi_{j}Z_{t-j},$$

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, $\hat{\phi}$, of $\phi = (\phi_{1}, \ldots, \phi_{p})'$ minimizes the objective function

$$\sum_{t=p+1}^{n} \rho(X_{t} - \beta_{1}X_{t-1} - \cdots - \beta_{p}X_{t-p}) = \sum_{t=p+1}^{n} \rho(Z_{t} - (\beta_{1} - \phi_{1})X_{t-1} - \cdots - (\beta_{p} - \phi_{p})X_{t-p})$$

Assuming that $F$ satisfies condition (2.2) for some $\alpha \in (0, 2)$ and, as in Section 2, writing $u = a_{n}(\beta - \phi)$ this is equivalent to minimizing the objective function

$$U_{n}(u) = \sum_{t=p+1}^{n} (\rho(Z_{t} - u_{1}a_{n}^{-1}X_{t-1} - \cdots - u_{p}a_{n}^{-1}X_{t-p}) - \rho(Z_{t}))$$
with minimum given by $\hat{u}_n = a_n(\hat{\phi} - \phi)$.

In order to construct a bootstrap replicate of $\hat{\phi}$, 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 $F_n(x) = n^{-1} \sum_{i=p+1}^{n} I[\hat{Z}_i \leq x]$ denotes the empirical distribution of the residuals, a bootstrap replicate $X_1^*, \ldots, X_n^*$ 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 $\{Z_t^*\} \overset{iid}{\sim} F_n$. (The recursions are started by setting $X_t^* = 0$ for $t < M$ for some large negative integer $M$.) A bootstrap replicate, $\hat{u}_n^* = a_m(\hat{\phi}_m^* - \hat{\phi}_n)$ of $\hat{u}$ is then found by minimizing

$$U_n^*(u) = \sum_{t=p+1}^{n} \left( \rho(Z_t^* - u a_m^{-1} X_{t-1}^* - \cdots - u a_m^{-1} X_{t-p}^*) - \rho(Z_t^*) \right).$$

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

**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 $\tau_1$,

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

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 $\text{Var}(\psi(Z_1)) < \infty$ if $\alpha \geq 1$.

Then, if $m_n \to \infty$ and $m_n/n \to 0$,

$$P[U_n^* \in \cdot | X_n] \overset{p}{\to} P[U \in \cdot],$$

where $U(\cdot)$ is the limit process

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

(see Davis, Knight and Liu (1992)), $\{Z_{k,i}\}$, $\{\delta_k\}$, $\{\Gamma_k\}$ 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.

Moreover, if $\rho(\cdot)$ is convex and $U(\cdot)$ attains a unique minimum at $\hat{u}$ a.s., then

$$P[\hat{u}_n^* \in \cdot | X_n] = P[a_m(\hat{\phi}_m^* - \hat{\phi}_n) \in \cdot | X_n] \overset{p}{\to} P[\hat{u} \in \cdot].$$
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

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

and

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

where \(M_n = \max\{|X_1|, \ldots, |X_n|\}\) and \(M_{m_n}^* = \max\{|X_1^*|, \ldots, |X_{m_n}^*|\}\). Observe that \(\tilde{U}_n(u) = U_n(u a_n/M_n)\) and \(\tilde{U}_n^*(u) = U_n^*(u a_{m_n}/M_{m_n}^*)\). 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 \(X_n\) converge in distribution to \(\psi_+ \Gamma_1^{-1/\alpha}\) where \(\psi_+ := \max_{j=0}^{\infty} |\psi_j|\). It follows that \(\tilde{U}_n(\cdot) \overset{d}{\rightarrow} \tilde{U}(\cdot)\) and

\[
P[\tilde{U}_n^* \in \cdot | X_n] \overset{P}{\rightarrow} P[\tilde{U} \in \cdot]
\]

where \(\tilde{U}(u) = U(u \Gamma_1^{-1/\alpha}/\psi_+)\). 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 \(\hat{u}\) a.s., then

\[
M_n(\hat{\phi}_n - \phi) \overset{d}{\rightarrow} \hat{u}
\]

and

\[
P[M_{m_n}^*(\hat{\phi}_{m_n} - \hat{\phi}_n) \in \cdot | X_n] \overset{P}{\rightarrow} P[\tilde{u} \in \cdot]
\]

where \(\hat{u} = \hat{u}_+ \psi_+ \Gamma_1^{-1/\alpha}\).

4. Extensions

In this section we consider extensions of the preceding sections to multiple regression, unknown location parameter in both regression and autoregression, and LAD estimation.

**Multiple Regression:** Here the model becomes

\[
Y_i = X_i' \beta + Z_i, \quad i = 1, \ldots, n,
\]
where $\beta = (\beta_1, \cdots, \beta_d)'$, and $X_i = (X_{i1}, \cdots, X_{id})'$ are iid random vectors satisfying a $d$-variate regular variation condition (see Assumptions 1 and 2 of Davis and Wu (1994)). Specifically, we assume that there exists a sequence $a_n \to \infty$ and a Lévy measure $\mu$ on $(\mathbb{R}^d, B(\mathbb{R}^d))$, such that

$$nP(a_n^{-1}X_1 \in \cdot) \xrightarrow{\nu_{n \to \infty}} \mu(\cdot),$$

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

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

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

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

so that the minimizer of $W_n(u)$ is $u_n = a_n(\hat{\beta} - \beta)$. In Davis and Wu (1994), it was shown that if the loss function $\rho(x)$ satisfies conditions (a)–(c) of Theorem 2.1, then $\hat{u}_n \xrightarrow{d} \hat{u}$ where $\hat{u}$ is the minimizer of the limit stochastic process of $W_n(\cdot)$.

The bootstrap implementation for the multiple regression case follows the development given in Section 2. A bootstrap replicate $\{(Y_1^*, X_1^*), \ldots, (Y_m^*, X_m^*)\}$ of the data is generated from the equations

$$Y_j^* = X_j^T \hat{\beta} + Z_j^*$$

where $\{X_j^*\}_{j=1}^m$ and $\{Z_j^*\}_{j=1}^m$ are two independent samples drawn from the empirical distributions based on $X_1, \ldots, X_n$, and the estimated residuals $\hat{Z}_1, \ldots, \hat{Z}_n$, respectively. The bootstrap replicate $\hat{u}_m^* = a_m(\beta^* - \hat{\beta})$ is then found by minimizing

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

Using the argument given in Section 2, if $m/n \to 0$ then the bootstrap replicate of $\hat{u}_m$ also converges in distribution to $\hat{u}$ in following sense

$$P[\hat{u}_m^* \in \cdot | X_1, \ldots, X_n, Y_1, \ldots, Y_n] \xrightarrow{p} P[\hat{u} \in \cdot].$$

**UNKNOWN LOCATION PARAMETER:** The bootstrap implementation described in Sections 2 and 3 can also be used for the case when a location parameter is included in the model. While we only describe the results for the AR($p$) case, they also remain valid for the multiple regression situation just considered. 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 $\beta_0$ and $\beta$. Provided $\rho$ is convex with a Lipschitz continuous derivative $\psi(\cdot)$, $a_n(\hat{\phi} - \phi)$ 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 $\hat{\phi}^*$ 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 $a_m(\hat{\phi}^* - \hat{\phi})$ 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 the theorems in Sections 2 and 3. Nevertheless, the bootstrap paradigm as described in Sections 2 and 3 still holds for the LAD estimate. We state a version of this result for the AR($p$) case only.

**Theorem 4.1.** Let $\{X_t\}$ be an $A(p)$ 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 $W(u)$ has a unique minimum $\hat{u}$ a.s., where $W(u) = \sum_{i=1}^\infty \sum_{j=1}^\infty [|Z_{i,j} - (\psi_{i-1} u_1 + \cdots + \psi_{i-p} u_p) \delta_{i-1}^{-1/\alpha} - |Z_{i,j}|$. If $m \to \infty$ and $m/n \to 0$, then

$$
P[a_m^{-1}(\hat{\phi}_m - \hat{\phi}) \in \cdot | X_n] \overset{p}{\to} P[\hat{u} \in \cdot]
$$

where $\hat{\phi}$ is the LAD estimate based on the original data $X_1, \ldots, X_n$ and $\hat{\phi}_m$ is the LAD estimator based on the bootstrap replicate $X_1^*, \ldots, X_m^*$.

**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, Knight and Liu (1992). For Lemmas 1–4 below, the assumptions of Theorem 2.1 are assumed to be met.
Lemma 1. Let $\mu_n(dx, dx)$ and $\hat{\mu}_n(dx, dx)$ be the random measures defined on rectangles of the form $E = A \times B \subset \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ as

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

and

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

Then if $m_n/n \to 0$,

$$
(A.1) \quad \mu_n(E) \xrightarrow{\mathbb{P}} \nu(E)
$$

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

$$
(A.2) \quad \hat{\mu}_n(E) - \mu_n(E) \xrightarrow{\mathbb{P}} 0.
$$

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

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

which by assumptions (2.2) converges to $e^{-i\lambda(B)}$ as $n \to \infty$. Thus $\mu_n(E) \xrightarrow{\mathbb{P}} E I(Z_1 \in B) \lambda(B) = \nu(E)$.

As for the second statement, we have

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

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). \qed

Lemma 2. Let $\mu_n^*$ be the random point measure

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

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

$$
P[(\mu_n^*(E_1), \ldots, \mu_n^*(E_d)) \in \cdot | X_\infty, Y_\infty] \xrightarrow{\mathbb{P}} P[(\mu(E_1), \ldots, \mu(E_d)) \in \cdot],
$$
where $\mu(\cdot)$ is the Poisson process $\sum_{j=1}^{\infty} \epsilon_{Z_j, g_j} \gamma_{j-1}(\cdot)$.

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

$$P[\mu_n^*(E_1) = r_1, \ldots, \mu_n^*(E_d) = r_d \mid X_\infty, Y_\infty] \overset{P}{\rightarrow} P[\mu(E_1) = r_1, \ldots, \mu(E_d) = r_d].$$

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

$$\frac{m_n!}{r_1! \cdots r_d! (m_n - r_1 - \cdots - r_d)!} \left( \frac{\mu_n(E_1)}{m_n} \right)^{r_1} \cdots \left( \frac{\mu_n(E_d)}{m_n} \right)^{r_d} \left( 1 - \frac{\hat{\mu}_n(\bigcup_{j=1}^{d} E_j)}{m_n} \right)^{m_n - r_1 - \cdots - r_d}$$

$$\overset{P}{\rightarrow} \frac{1}{r_1! \cdots r_d!} \nu^{r_1}(E_1) \cdots \nu^{r_d}(E_d) \exp\{-\nu(E_1) + \cdots + \nu(E_d)\}$$

$$= P[\mu(E_1) = r_1, \ldots, \mu(E_d) = r_d],$$

where the limit follows from Lemma 1. \qed

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

(A.3) 

$$P[\mu_n^*(g) \in \cdot \mid X_\infty, Y_\infty] \overset{P}{\rightarrow} P[\mu(g) \in \cdot]$$

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 $K_\epsilon$ and a step function $g_\epsilon$ with support $E$ such that $P[\mu(E) > K_\epsilon] < \epsilon$ and $|g(z, x) - g_\epsilon(z, x)| \leq \epsilon/K_\epsilon$ for all $z$ and $x$. Since

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

it follows that

$$P[\mu_n^*(g) \leq x \mid X_\infty, Y_\infty] \leq P[\mu_n^*(g_\epsilon) \leq x + \epsilon \mu_n^*(E)/K_\epsilon \mid X_\infty, Y_\infty]$$

$$\leq P[\mu_n^*(g_\epsilon) \leq x + \epsilon \mid X_\infty, Y_\infty] + P[\mu_n^*(E) > K_\epsilon \mid X_\infty, Y_\infty]$$

$$\overset{P}{\rightarrow} P[\mu(g_\epsilon) \leq x + \epsilon] + P[\mu(E) > K_\epsilon]$$

$$\leq P[\mu(g) \leq x + 2\epsilon] + 2P[\mu(E) > K_\epsilon].$$

A similar lower bound can be obtained in exactly the same fashion. Letting $\epsilon \rightarrow 0$, we find that

$$P[\mu_n^*(g) \leq x \mid X_\infty, Y_\infty] \overset{P}{\rightarrow} P[\mu(g) \leq x]$$

from which (A.3) follows using a routine weak convergence argument. \qed
Lemma 4. For any \( u_1, \ldots, u_d \in \mathbb{R} \),

\[
P[(W_{n}^{*}(u_1), \ldots, W_{n}^{*}(u_d)) \in \cdot | X_{\infty}, Y_{\infty}] \xrightarrow{p} P[(W(u_1), \ldots, W(u_d)) \in \cdot].
\]

Proof: We just provide the proof for \( d = 1 \); the case \( d > 1 \) being similar using the Cramér-Wold device. First note that

\[
W_{n}^{*}(u) = \int g \, d\mu_{n}^{*}
\]

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

\[
(A.4) \quad P[\mu_{n}^{*}(g_K) \in \cdot | X_{\infty}, Y_{\infty}] \xrightarrow{p} P[\mu(g_K) \in \cdot]
\]

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

For any \( \epsilon > 0 \)

\[
P[|\mu_{n}^{*}(g - g_K)| > 3\epsilon | X_{\infty}, Y_{\infty}] \leq P[|\mu_{n}^{*}(gI(|z| \leq K_2^{-1}))| > \epsilon | X_{\infty}, Y_{\infty}]
+ P[|\mu_{n}^{*}(gI(|z| > K_2))| > \epsilon | X_{\infty}, Y_{\infty}]
+ P[|\mu_{n}^{*}(gI(|z| > K_1, K_2^{-1} < |z| < K_2))| > \epsilon | X_{\infty}, Y_{\infty}]
\]

\[
(A.5) \quad =: I + II + III.
\]

We handle each of the three terms separately. We have

\[
II = P[(\sum_{i=1}^{m_{n}} g(Z_{i}^{*}, a_{m_{n}}^{-1}X_{i}^{*})I(a_{m_{n}}^{-1}|X_{i}^{*}| > K_2)) > \epsilon | X_{\infty}, Y_{\infty}]
\leq P[\bigcup_{i=1}^{m_{n}} \{a_{m_{n}}^{-1}|X_{i}^{*}| > K_2\} | X_{\infty}, Y_{\infty}]
\leq m_{n}P[a_{m_{n}}^{-1}|X_{i}^{*}| > K_2 | X_{\infty}, Y_{\infty}]
= m_{n}n^{-1} \sum_{i=1}^{n} I(a_{m_{n}}^{-1}|X_{i}| > K_2)
\xrightarrow{p} \lambda(K_2, \infty) \quad (\text{as } n \to \infty \text{ by Lemma 1})
\to 0
\]
as \( K_2 \to \infty \). Next, by Lemma 1,

\[
III \leq P[\bigcup_{i=1}^{m_{n}} \{|Z_{i}^{*}| > K_1, K_2^{-1} < a_{m_{n}}^{-1}|X_{i}^{*}| \leq K_2\} | X_{\infty}, Y_{\infty}]
\leq m_{n}P[|Z_{i}^{*}| > K_1, K_2^{-1} < a_{m_{n}}^{-1}|X_{i}^{*}| \leq K_2 | X_{\infty}, Y_{\infty}]
= \bar{\mu}_{n}(\{|z| > K_1\} \times \{K_2^{-1} < |z| \leq K_2\})
\xrightarrow{p} P[|Z_1| > K_1]\lambda(K_2^{-1} < |z| \leq K_2) \quad (\text{as } n \to \infty)
\to 0
\]
as $K_1 \to \infty$ and then $K_2 \to \infty$. As for the first term in (A.5), we have, using the bound $|g(x, z)| \leq |ux\psi(z)| + C|ux|^{1+\tau_1}$, that

$$I \leq P]\sum_{i=1}^{m_n} X_i^n \psi(Z_i^n) I(a_{m_n}^{-1} X_i^n \leq K_2^{-1})| > \varepsilon/2 \mid X_\infty, Y_\infty\]

(A.6) \quad + P|u(a_{m_n}^{-1} X_i^n \psi(Z_i^n) I(a_{m_n}^{-1} X_i^n \leq K_2^{-1})| > \varepsilon/2 \mid X_\infty, Y_\infty\].

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

$$C|u|2\varepsilon^{-1} m_n a_{m_n}^{-\tau_1} \sum_{i=1}^{m_n} \{X_i^n \psi(Z_i^n) I(a_{m_n}^{-1} X_i^n \leq K_2^{-1})| X_{\infty}, Y_{\infty}\} = (\text{const}) m_n a_{m_n}^{-\tau_1} \sum_{i=1}^{m_n} \{X_i^n \psi(Z_i^n) I(a_{m_n}^{-1} X_i^n \leq K_2^{-1})| X_{\infty}, Y_{\infty}\} \sim (\text{const}) \alpha(1 + \tau_1 - \alpha)^{-1} m_n a_{m_n}^{-\tau_1} (K_2^{-1} a_{m_n})^{1+\tau_1} P(|X_1^n| > K_2^{-1} a_{m_n}) \rightarrow (\text{const}) \alpha(1 + \tau_1 - \alpha)^{-1} K_2^{\alpha-\tau_1} \text{ (as } n \rightarrow \infty) \rightarrow 0

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 $\gamma > 0$

(A.7) \quad m_n P\{a_{m_n}^{-1} X_i^n X_i^n \psi(Z_i^n) I(a_{m_n}^{-1} X_i^n \leq \delta)| X_{\infty}, Y_{\infty}\} \overset{P}{\to} 0,

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

(A.8) \quad m_n E\{a_{m_n}^{-1} X_i^n X_i^n \psi(Z_i^n) I(a_{m_n}^{-1} X_i^n \leq \delta)I(|X_i^n \psi(Z_i^n)| \leq a_{m_n} \gamma) | X_{\infty}, Y_{\infty}\} \overset{P}{\to} 0

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

(A.9) \quad m_n E\{a_{m_n}^{-2} (X_i^n X_i^n \psi(Z_i^n)^2) I(a_{m_n}^{-1} X_i^n \leq \delta)I(a_{m_n}^{-1} X_i^n \psi(Z_i^n) | \leq \gamma) | X_{\infty}, Y_{\infty}\} \overset{P}{\to} 0

as $n \rightarrow \infty$, $\delta \rightarrow 0$, and $\gamma \rightarrow 0$.

Applying Markov's inequality and using the relation

(A.10) \quad |\psi(\hat{Z}_j) - \psi(Z_j)| \leq C|(|\hat{\beta} - \beta| X_j|^{\tau_1} = o_p(1)a_{m_n}^{-\tau_1} X_j|^{\tau_1},

where $o_p(1) = Ca_{m_n}^{-\tau_1} |\hat{\beta} - \beta|^{\tau_1}$, (A.7) is bounded by

$$m_n n^{-2} \sum_{1 \leq i,j \leq n} I(a_{m_n}^{-1} X_i \psi(\hat{Z}_j) > \gamma) I(a_{m_n}^{-1} X_i \leq \delta)

\leq m_n n^{-2} \sum_{1 \leq i,j \leq n} I(a_{m_n}^{-1} X_i \psi(Z_j) > \gamma/2) I(a_{m_n}^{-1} X_i \leq \delta)

+ m_n n^{-2} \sum_{1 \leq i,j \leq n} I(o_p(1)a_{m_n}^{-1} X_i a_{m_n}^{-\tau_1} X_j|^{\tau_1} > \gamma/2) I(a_{m_n}^{-1} X_i \leq \delta)].$$
On the set where the $o_p(1)$ term is bounded by 1, the right hand side has expectation bounded by

\[ m_n E \left[ |X_1|^{\tau_1} (\gamma a_{m_n})^{-\tau_1} I(a_{m_n}^{-1} |X_i| \leq \delta) \right] E[\psi(Z_1)] \]

\[ + m_n P[|X_1|^{\tau_1} > a_{m_n}^{\tau_1} \gamma/(2\delta)] + 2m_n n^{-1} P[|X_1| \leq a_{m_n} \delta] \]

\[ \to (\text{const}) \delta^\alpha + \left( \frac{\gamma}{2\delta} \right)^{-\alpha/\tau_1} \] (as $n \to \infty$)

\[ \to 0 \] (as $\delta \to 0$).

As for (A.8), it may be written as

\[ m_n n^{-2} \sum_{1 \leq i, j \leq n} a_{m_n}^{-1} X_i \psi(Z_j) I(a_{m_n}^{-1} |X_i| \leq \delta) I(a_{m_n}^{-1} |X_i \psi(Z_j)| \leq \gamma) \]

\[ = m_n n^{-2} \sum_{1 \leq i, j \leq n} a_{m_n}^{-1} X_i \psi(Z_j) I(a_{m_n}^{-1} |X_i| \leq \delta) I(a_{m_n}^{-1} |X_i \psi(Z_j)| \leq \gamma) \]

\[ + o_p(1) m_n n^{-2} \sum_{1 \leq i, j \leq n} a_{m_n}^{-(1+\tau_1)} X_i |X_j|^{\tau_1} I(a_{m_n}^{-1} |X_i| \leq \delta). \]

Using an argument similar to that given above, the second term in (A.11) converges to 0 in probability as $n \to \infty$ and then $\delta \to 0$. To handle the first term we consider the three cases $\alpha \in (0, 1)$, $\alpha = 1$ and $\alpha \in (1, 2)$. For $\alpha \in (0, 1)$ we have the bound,

\[ E[m_n n^{-2} \sum_{1 \leq i, j \leq n} a_{m_n}^{-1} X_i \psi(Z_j) I(a_{m_n}^{-1} |X_i| \leq \delta) I(a_{m_n}^{-1} |X_i \psi(Z_j)| \leq \gamma)] \]

\[ \leq m_n a_{m_n}^{-1} E \left[ |X_1 \psi(Z_1)| I(a_{m_n}^{-1} |X_1 \psi(Z_1)| \leq \gamma) \right] \]

which, by Karamata’s theorem,

\[ \to (\text{const}) \gamma^{1-\alpha} \] (as $n \to \infty$)

\[ \to 0 \] (as $\gamma \to 0$).

This takes care of (A.8) for $\alpha \in (0, 1)$. For the case $\alpha \in (1, 2)$, first note that

\[ \frac{m_n}{n^2} \sum_{1 \leq i, j \leq n} a_{m_n}^{-1} X_i \psi(Z_j) I(a_{m_n}^{-1} |X_i| \leq \delta) = \frac{m_n}{a_{m_n}} \left( \frac{1}{n} \sum_{i=1}^{n} X_i I(a_{m_n}^{-1} |X_i| \leq \delta) \right) \left( \frac{1}{n} \sum_{j=1}^{n} \psi(Z_j) \right). \]

Now assumptions (b) and (c) of Theorem 2.1 imply \( n^{-1} \sum_{j=1}^{n} \psi(Z_j) = n^{1/\tau_2} (\ln n)^{2/\tau_2} o_p(1) \) and since $E|X_1| < \infty$, \( n^{-1} \sum_{i=1}^{n} X_i I(a_{m_n}^{-1} |X_i| \leq \delta) = O_p(1) \). Writing $a_{m_n}^{-1} = m_n^{-1/\alpha} L(m_n)$ where $L(\cdot)$ is a slowly varying function, it follows that (A.13) is of order

\[ m_n^{1-1/\alpha} n^{-(1-1/\tau_2)} L(m_n) (\ln n)^{2/\tau_2} o_p(1) \to 0 \]

since $\tau_2 > \alpha$ and $m_n/n \to 0$. Moreover, the difference between the first term in (A.11) and (A.13)
has expectation bounded by

\[
m_n a_{m_n}^{-1} E[|X_1 \psi(Z_1)| I(a_{m_n}^{-1} |X_1| \leq \delta) I(a_{m_n}^{-1} |X_1 \psi(Z_1)| > \gamma)] \\
\sim m_n a_{m_n}^{-1} \alpha(\alpha - 1)^{-1} a_{m_n} \gamma P[|X_1 \psi(Z_1)| I(|\psi(Z_1)| > \gamma/\delta) > a_{m_n} \gamma] \\
\rightarrow (\text{const}) \gamma^{1-\alpha} E[|\psi(Z_1)|^\alpha I(|\psi(Z_1)| > \gamma/\delta)] \quad (\text{as } n \rightarrow \infty) \\
\rightarrow 0 \quad (\text{as } \delta \rightarrow 0).
\]

Finally, for the case \( \alpha = 1 \), the mean of (A.13) is zero and, using the symmetry of the distribution of \( X_1 \), it can be shown that the variance of (A.13) converges to 0 as \( n \rightarrow \infty \) and \( \delta \rightarrow 0 \).

Turning to (A.9), we have

\[
E[m_n n^{-2} a_{m_n}^{-2} \sum_{1 \leq i, j \leq n} (X_i \psi(Z_j))^2 I(a_{m_n}^{-1} |X_i| \leq \delta) I(a_{m_n}^{-1} |X_i \psi(Z_j)| \leq \gamma)] \\
\leq m_n a_{m_n}^{-2} E[(X_1 \psi(Z_1))^2 I(a_{m_n}^{-1} |X_1 \psi(Z_1)| \leq \gamma)] \\
\sim \alpha(2 - \alpha)^{-1} (\gamma a_{m_n})^2 P[a_{m_n}^{-1} |X_1 \psi(Z_1)| > \gamma] \\
\rightarrow (\text{const}) \gamma^{2-\alpha} \quad (\text{as } n \rightarrow \infty) \\
\rightarrow 0 \quad (\text{as } \gamma \rightarrow 0)
\]

which establishes (A.9).

To finish off the proof of the lemma, write for any \( \gamma > 0 \)

\[
a_{m_n}^{-1} \sum_{i=1}^{m_n} X_i^* \psi(Z_i^*) I(a_{m_n}^{-1} |X_i^*| \leq K_2^{-1}) \\
= a_{m_n}^{-1} \sum_{i=1}^{m_n} X_i^* \psi(Z_i^*) I(a_{m_n}^{-1} |X_i^*| \leq K_2^{-1}) I(a_{m_n}^{-1} |X_i^* \psi(Z_i^*)| \leq \gamma) \\
+ a_{m_n}^{-1} \sum_{i=1}^{m_n} X_i^* \psi(Z_i^*) I(a_{m_n}^{-1} |X_i^*| \leq K_2^{-1}) I(a_{m_n}^{-1} |X_i^* \psi(Z_i^*)| > \gamma).
\]

By (A.8) and (A.9) the conditional variance and conditional mean of the first term converges to 0 in probability as \( n \rightarrow \infty, K_2 \rightarrow \infty \) and \( \gamma \rightarrow 0 \). 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 \( n \rightarrow \infty, K_2 \rightarrow \infty \) as claimed. \( \square \)

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 \( \{X_i\} \) is an AR\((p)\) process satisfying (3.1) where \( \{Z_i\} \overset{iid}{\sim} F \) with \( F \) satisfying the regular variation condition (2.2), are assumed to be met. The corresponding sequence of point processes is now given by

\[
\mu_n^* (\cdot) = \sum_{i=1}^{m} \delta_{(Z_i^*, a_{m_n}^{-1} Y_i^*) (\cdot)}
\]
where

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

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

(A.14) \[ P_n[\mu_n^* \in \cdot] \xrightarrow{d} P[\mu \in \cdot], \]

where \( \mu \) is the random measure

\[ \mu(\cdot) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \epsilon(\bar{Z}_{i,j}, (u_1 \psi_{i-1} + \cdots + u_p \psi_{i-p}) \delta_j \gamma_j^{-1/\alpha}); \]

the \( \psi_j \)'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 3.1. The proof of (A.14) follows the sequence of steps (see Theorem 2.4 in Davis and Resnick (1985)) used for establishing the result

\[ \mu_n(\cdot) := \sum_{i=1}^{n} \epsilon(Z_i, a_n^{-1} \gamma_i) \xrightarrow{d} \mu(\cdot). \]

We break up the proof of (A.14) 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 \( I_{m,k} = \sum_{i=1}^{m} \epsilon(Z_i, a_n^{-1} \gamma_i), Z_t^* = (Z_{t-1}^*, \ldots, Z_t^*), I_k = \sum_{j=1}^{k} \sum_{i=1}^{\infty} \epsilon(Z_{i,j}, \delta_j \gamma_j^{-1/\alpha} e_i) \) and \( e_i \) is the basis element of \( \mathbb{R}^k \) with \( i \)-th component equal to one and the rest zero. (The relevant state space for the point processes is \( \mathbb{R} \times (\mathbb{R}^k \setminus (0, \ldots, 0)). \))

**Proof:** As in the Davis and Resnick (1985), let \( S \) be the collection of all sets \( B \) of the form

\[ B = (b_0, c_0) \times (b_1, c_1) \times \cdots \times (b_k, c_k), \]

where \( b_0 \) and \( c_0 \) are continuity points of \( F \) and the \( k \)-dimensional rectangle \( (b_1, c_1) \times \cdots \times (b_k, c_k) \) is bounded away from \( (0, \ldots, 0) \).

For future use, we record the Laplace functional of the limit point process as

(A.15) \[ E[\exp\{-I_k(f)\}] = \exp\{- \int (1 - \prod_{i=1}^{k} \exp(-f(z_i, a \epsilon_i))) F(dz_1) \cdots F(dz_k) \lambda(dx) \]

which is easily calculated by first conditioning on the Poisson points, \( \delta_j \gamma_j^{-1/\alpha} \).
Note that \( \hat{Z}_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p} = Z_t + (\phi - \hat{\phi})' X_{t-1} \), where \( X_{t-1} = (X_{t-1}, \ldots, X_{t-p})' \), so that for a fixed interval \((c, d)\) with \(c > 0\), we have for \(\delta > 0\) sufficiently small

\[
|mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} Z_t \in (c, d]) - mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} \hat{Z}_t \in (c, d])|
\]

\[
\leq mn^{-1} \sum_{t=1}^{n} \left( I(c < a_m^{-1} Z_t \leq c + \delta, a_m^{-1}|(\phi - \hat{\phi})' X_{t-1}| \leq \delta) + I(d - \delta < a_m^{-1} Z_t \leq d, a_m^{-1}|(\phi - \hat{\phi})' X_{t-1}| \leq \delta) \right)
\]

\[
+ mn^{-1} \sum_{t=1}^{n} I(a_m^{-1}(\phi - \hat{\phi})' X_{t-1}| > \delta)
\]

Taking expectations on the set \(I(|\phi - \hat{\phi}| < \delta_1)\) for \(\delta_1\) small, the above bound has mean bounded by

\[
mP[c < a_m^{-1} Z_1 \leq c + \delta] + mP[d - \delta < a_m^{-1} Z_1 \leq d] + mP[a_m^{-1}|X_1| > \delta/\delta_1]
\]

\[
\rightarrow (\text{const}) (c^{-\alpha} - (c + \delta)^{-\alpha} + (d - \delta)^{-\alpha} - (d)^{-\alpha} + (\delta/\delta_1)^{-\alpha})
\]

\[
\rightarrow (\text{const}) (c^{-\alpha} - (c + \delta)^{-\alpha} + (d - \delta)^{-\alpha} - (d)^{-\alpha}) \quad (\text{as } \delta_1 \rightarrow 0)
\]

\[
\rightarrow 0 \quad (\text{as } \delta \rightarrow 0).
\]

It follows that

\[(A.16) \quad mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} Z_t \in (c, d]) - mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} \hat{Z}_t \in (c, d]) = o_p(1).\]

On the other hand, if \(c < 0 < d\), then

\[(A.17) \quad n^{-1} \sum_{t=1}^{n} I(a_m^{-1} \hat{Z}_t \in (c, d]) \overset{P}{\rightarrow} 1\]

and for any continuity points \(c < d\) of \(F\),

\[(A.18) \quad n^{-1} \sum_{t=1}^{n} I(\hat{Z}_t \in (c, d]) \overset{P}{\rightarrow} F(d) - F(c)\]

Next we show for any \(B \in S\),

\[E_n I_{m,k}(B) \overset{P}{\rightarrow} EI_k(B).\]

First consider the case when \(c_1 > 0\) or \(d_1 < 0\). Then

\[
E_n I_{m,k}(B) = \left(n^{-1} \sum_{t=1}^{n} I(\hat{Z}_t \in (c_0, d_0])\right) \left(mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} \hat{Z}_{t-1} \in (c_1, d_1])\right)
\]

\[
\cdot \left(n^{-1} \sum_{t=1}^{n} I(a_m^{-1} \hat{Z}_{t-2} \in (c_2, d_2])\right) \cdots \left(n^{-1} \sum_{t=1}^{n} I(a_m^{-1} \hat{Z}_{t-k} \in (c_k, d_k])\right)
\]

which by \((A.16)-(A.18))
\[= (F(d_0) - F(c_0)) \left( mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} Z_{t-1} \in (c_1, d_1]) \right) + o_p(1) \]

\[\rightarrow (F(d_0) - F(c_0))\lambda(c_1, d_1]) \quad \text{(see Lemma 1)} \]

\[= EI_k(B). \]

A similar result holds if \( c_i > 0 \) or \( d_i > 0 \). On the other hand, if \( B \) satisfies (C2) of Davis and Resnick (1985), i.e. if \( \min_{i=1}^{k} c_i > 0 \) or \( \max_{i=1}^{k} d_i < 0 \), then it is easy to show that \( E_n I_{m,k}(B) \rightarrow 0 \) which, combined with the foregoing, establishes (A.18).

To complete the proof of Lemma 5, it is enough to show (see Theorem 4.7 in Kallenberg (1983)) that

\[ P_n[I_{m,k}(R) = 0] \rightarrow P[I_k(R) = 0] \]

where \( R \) is a finite union of disjoint sets in \( S \). We only give full details of this argument for the special case when \( k = 2 \) and \( R = B \cup C \); the other cases being handled in a similar fashion. Write \( B = B_1 \times B_2 \times B_3 \) and \( C_1 \times C_2 \times C_3 \) where we assume that the intervals \( B_2 \) and \( C_3 \) are bounded away from 0 while \( B_3 \) and \( C_2 \) contain 0. It follows from (A.16)--(A.18) that

\[ mP_n[(Z_1^*, a_m^{-1}(Z_0^*, Z_{-1}^*)) \in B] \]

\[= \left( n^{-1} \sum_{t=1}^{n} I(\tilde{Z}_t \in B_1) \right) \left( mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} \tilde{Z}_t \in B_2) \right) \left( n^{-1} \sum_{t=1}^{n} I(a_m^{-1} \tilde{Z}_t \in B_3) \right) \]

\[\rightarrow P[Z_1 \in B_1]\lambda(B_2). \]

Moreover, for \( m \) large

\[ mP_n[(Z_1^*, a_m^{-1}(Z_0^*, Z_{-1}^*)) \in B, (Z_2^*, a_m^{-1}(Z_1^*, Z_0^*)) \in C] \]

\[= \left( n^{-1} \sum_{t=1}^{n} I(\tilde{Z}_t \in B_1) \right) \left( mn^{-1} \sum_{t=1}^{n} I(a_m^{-1} \tilde{Z}_t \in B_2 \cap C_3) \right) \left( n^{-1} \sum_{t=1}^{n} I(a_m^{-1} \tilde{Z}_t \in B_3) \right) \cdot \left( n^{-1} \sum_{t=1}^{n} I(\tilde{Z}_t \in C_1) \right) \]

\[\rightarrow P[Z_1 \in B_1]P[Z_1 \in C_1]\lambda(B_2 \cap C_3). \]

On the other hand,

\[ \limsup_{m} m^2 P_n[(Z_1^*, a_m^{-1}(Z_0^*, Z_{-1}^*)) \in C, (Z_2^*, a_m^{-1}(Z_1^*, Z_0^*)) \in B] = 0 \]

and for all \( j \geq 3 \)

\[ \limsup_{m} m^2 P_n[(Z_1^*, a_m^{-1}(Z_0^*, Z_{-1}^*)) \in R, (Z_j^*, a_m^{-1}(Z_{j-1}^*, Z_{j-2}^*)) \in R] = O_p(1). \]
Combining (A.20)–(A.22) and writing \( A_j = \{(Z^*_j, a^{-1}_m(Z^*_{j-1}, Z^*_{j-2})) \in R\} \) we obtain

\[
\sum_{j=2}^{[m/r]} ([m/r] - j)P_n[A_1 \cap A_j] = r^{-1}P[Z_1 \in B_1]P[Z_1 \in C_1]\lambda(B_2 \cap C_3) + o_p(r^{-1})
\]

as \( m \to \infty \) and then \( r \to \infty \). Similarly,

\[
\sum_{2 \leq j_1 < j_2 \leq [m/r]} P_n[A_1 \cap A_{j_1} \cap A_{j_2}] = o_p(r^{-1})
\]

as \( m \to \infty \) and then \( r \to \infty \). The remainder of the proof is based on a standard ‘big’ block, ‘small’ block style argument for mixing sequences. In this case the underlying process \( \{Z^*_t, Z^*_{t-1}, Z^*_{t-2}\} \) is 2-dependent given \( X_1, X_2, \ldots \). In particular, using a standard argument for mixing sequences (see Chapter 3 of Leadbetter, Lindgren and Rootzén (1983)),

\[
\limsup_m \left| P[I_{m, k}(R) = 0] - P_n^*[\sum_{i=1}^{[m/r]} \epsilon_{Z^*_i, a^{-1}_m(Z^*_{i-1}, Z^*_{i-2})}(R) = 0]\right| \overset{P}{\to} 0
\]

as \( r \to \infty \).

For \( r \) a fixed integer, we have using the inclusion-exclusion principle

\[
1 - \frac{m}{r} P_n[A_1] + \sum_{j=2}^{[m/r]} ([m/r] - j)P_n[A_1 \cap A_j] - \sum_{2 \leq j_1 < j_2 \leq [m/r]} [m/r]P_n[A_1 \cap A_{j_1} \cap A_{j_2}]
\]

\[
\leq P_n[\sum_{i=1}^{[m/r]} \epsilon_{Z^*_i, a^{-1}_m(Z^*_{i-1}, Z^*_{i-2})}(R) = 0]
\]

\[
\leq 1 - \frac{m}{r} P_n[A_1] + \sum_{j=2}^{[m/r]} ([m/r] - j)P_n[A_1 \cap A_j].
\]

(A.24)

The outside two terms, after taking the limit on \( m \to \infty \) and using (A.19)–(A.22), are equal to

\[
1 - r^{-1} (P(Z_1 \in B_1)\lambda(B_2) + (P(Z_1 \in C_1)\lambda(C_3) + P[Z_1 \in B_1]P[Z_1 \in C_1]\lambda(B_2 \cap C_3)) + o_p(r^{-1}).
\]

Now raising all 3 sides of (A.24) to the \( r \)th power and letting \( m \to \infty \) and then \( r \to \infty \) and applying (A.23), we obtain

\[
P_n[I_{m, 3}(R) = 0]
\]

\[
\overset{P}{\to} \exp\{-P(Z_1 \in B_1)\lambda(B_2) - (P(Z_1 \in C_1)\lambda(C_3) + P[Z_1 \in B_1]P[Z_1 \in C_1]\lambda(B_2 \cap C_3)\}.
\]

Setting \( f = tI_R \) in (A.15), one calculates that the limit is equal to \( P[I_3(R) = 0] \) as claimed. For other choices of the initial sets \( B \) and \( C \) in \( S \), the argument is pretty much the same and is omitted. □
From the convergence in Lemma 5 and the weak consistency of \( \hat{\phi} \), it follows that the points in \( I_{m,k} \) may be summed up. Specifically, let \( \{ \hat{\psi}_j, j = 0, 1, \ldots \} \) be the coefficients in the power series expansion of \( 1 - \hat{\phi}_1 z - \cdots - \hat{\phi}_p z^p \) so that \( X_t^* = \sum_{j=0}^{\infty} \hat{\psi}_j Z_{t-j}^* \). Since \( \hat{\psi}_j \rightarrow \psi_j \), it follows that
\[
\begin{align*}
\mathbb{P} \left( \sum_{i=1}^{m} \epsilon(Z_i, e^{-1} \sum_{j=0}^{k} c_j Z_{t-j}) \in \mathcal{L} \right) \leq \sum_{i=1}^{\infty} \epsilon(Z_i, e^{-1} \sum_{j=0}^{k} c_j Z_{t-j}^*) \in \mathcal{L},
\end{align*}
\]
where \( \hat{c}_j = u_1 \hat{\psi}_j + \cdots + u_p \hat{\psi}_j + \cdots + u_{p+1} \hat{\psi}_j \), and \( c_j = u_1 \psi_j + \cdots + u_p \psi_j + \cdots + u_{p+1} \psi_j \). Thus to complete the proof of (A.14), we need to extend (A.25) to the case \( k = \infty \). Before doing so, however, we first require the following two lemmas with the latter paralleling Lemma 2.3 of Davis and Resnick (1985).

**Lemma 6.** Set \( Y^* = \sum_{j=0}^{\infty} |c_j Z_j^*| \) where the coefficients \( \{ c_j \} \) decrease to 0 at an exponential rate. Then
\[
\begin{align*}
\limsup_n m\mathbb{E}[\alpha(\phi - \hat{\phi}) < \delta] & \leq (\text{const}) x^{-\alpha} \sum_{j=0}^{\infty} |c_j|^\alpha, \\
\end{align*}
\]
and
\[
\begin{align*}
\limsup_n m a^{-\alpha} \mathbb{E}[E_n[(Y^*)^\gamma I(\alpha(\phi - \hat{\phi}) < \delta)] & \leq (\text{const}) \delta^{-\alpha}
\end{align*}
\]
for all \( \gamma > \alpha \).

**Proof:** (i) Following the argument given on p.228–230 of Resnick (1987), we have
\[
\begin{align*}
m P_n[Y^* > a_m x] & = m P_n[Y^* > a_m x, \sqrt{\sum_j |c_j Z_j^*| > a_m x}] + m P_n[Y^* > a_m x, \sqrt{\sum_j |c_j Z_j^*| \leq a_m x}] \\
& \leq \sum_j m P_n[|c_j Z_j^*| > a_m x] + m P_n[|c_j Z_j^*| \leq a_m x, \sqrt{|c_j Z_j^*| > a_m x}] \\
& \leq \sum_j \left( m x^{-1} \sum_{i=1}^{n} I(|c_j Z_i| > a_m x) \right) \\
& \quad + \sum_j \left( m x^{-1} \sum_{i=1}^{n} |c_j Z_j| I(|c_j Z_i| > a_m x) \right)
\end{align*}
\]

Also,
\[
\hat{Z}_t = Z_t + (\phi - \hat{\phi})X_{t-1}
\]
which, on the set \( |\phi - \hat{\phi}| < \delta \) for \( \delta \) 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
\[
\begin{align*}
\mathbb{E}[\alpha(\phi - \hat{\phi}) < \delta] & \leq \sum_j m P[|c_j| (|Z_1| + \delta |U_0|) > a_m x] \\
& \leq \sum_j (|c_j|^\alpha (1 + \delta^\alpha d_+) x^{-\alpha}) \\
& \leq (\text{const}) x^{-\alpha} \sum_j |c_j|^\alpha,
\end{align*}
\]
where \( d_+ := \sum_j |d_j|^\alpha \). As for the second term, assume that \( 0 < \alpha < 1 \). Then, by Karamata’s theorem,

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

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

(ii) We have

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

After taking expectations on the set \(|\phi - \hat{\phi}| < \delta\) for \( \delta \) 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. \( \square \)

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

\[
\lim_{k \to \infty} \limsup_{m \to \infty} P_n[|a_m^{-1} \sum_{i=1}^m |\hat{\psi}_j Z^*_{i-j} - Y^*_{i-1}| > \gamma] > \epsilon] = 0.
\]

**Proof:** Since \( \hat{\phi} \) is weakly consistent, it suffices to consider the outside probability on the set \(|\phi - \hat{\phi}| < \delta\) for some small \( \delta \). If \( \delta \) 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 \(|\phi - \hat{\phi}| < \delta\), is bounded by

\[
mP_n[\sum_{j>\delta} |\hat{\psi}_j||Z^*_j| > a_m \gamma]I(|\phi - \hat{\phi}| < \delta) \leq mP_n[\sum_{j>\delta} |c_j||Z^*_j| > a_m \gamma]I(|\phi - \hat{\phi}| < \delta)
\]

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

**Lemma 8.** We have

\[
P_n[\mu_n^* \in .] \overset{\delta}{\to} P[\mu \in .].
\]

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

\[
P_n((U_n^*(u_1), \ldots, U_n^*(u_d)) \in \cdot) \Rightarrow P((U(u_1), \ldots, U(u_d)) \in \cdot).
\]

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 \( \epsilon, \eta > 0 \)

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

\[
a_m^{-(1+\gamma)} \sum_{t=1}^{m} |Y_{t-1}^*|^{1+\gamma} I(|Y_{t-1}^*| \leq a_m \delta) \xrightarrow{p} 0,
\]

as \( m \to \infty \) and then \( \delta \to 0 \), and

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

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


