Statistica Sinica **21** (2011), 779-805

LEAST ABSOLUTE DEVIATION ESTIMATION FOR GENERAL ARMA TIME SERIES MODELS WITH INFINITE VARIANCE

Rongning Wu

Baruch College

Abstract: We study least absolute deviation (LAD) estimation for general autoregressive moving average (ARMA) models with infinite variance. The assumptions of causality and invertibility, which are necessary for Gaussian ARMA models to ensure the identifiability of the model parameters, are removed because they are not required for models with non-Gaussian noise. Following the approach taken by Davis, Knight, and Liu (1992) and Davis (1996), we derive a functional limit theorem for random processes based on the LAD objective function, and establish asymptotic results of the LAD estimator. A simulation study is presented to evaluate the finite sample performance of LAD estimation. An empirical example of financial time series is also provided.

Key words and phrases: ARMA model, infinite variance, LAD estimation, noncausality, noninvertibility, stable distribution, time series.

1. Introduction

Because of its fundamental role in studying stationary processes, the class of autoregressive moving average (ARMA) models has been extensively investigated in the time series literature. Suppose $\{X_t\}$ is an ARMA(p,q) process, i.e., a (strictly) stationary solution of the recursions

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

or, in short,

$$\phi(B)\, X_t = \theta(B)\, Z_t, \tag{1.1}$$

where $\{Z_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables, B is the backward shift operator ($B^i X_t = X_{t-i}$ for all integer i), $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ is the autoregressive (AR) polynomial, and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ is the moving average (MA) polynomial. We assume that polynomials $\phi(z)$ and $\theta(z)$ have neither common roots nor roots on the unit circle in the complex plane. Then, there exists a strictly stationary solution $\{X_t\}$, and such a solution is unique.

Causality and invertibility are two important conditions conventionally assumed when modeling a time series using ARMA models. Causality means that the ARMA(p, q) process X_t can be expressed in terms of present and past Z_t's: $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ for any t, where $\sum_{j=0}^{\infty} |\psi_j| < \infty$. It is equivalent to the condition that $\phi(z) \neq 0$ for $|z| \leq 1$, or all the roots of $\phi(z)$ are outside the unit circle. If there exists any roots inside the unit circle, then $\{X_t\}$ is said to be noncausal; furthermore, if all the roots are inside the unit circle, then the process is purely noncausal. On the other hand, $\{X_t\}$ is said to be invertible if there exists an absolutely summable sequence $\{\pi_j\}$ such that $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ for any t. Like causality, this condition is equivalent to that the zeros of $\theta(z)$ lie outside the unit circle. The process is noninvertible if any roots of $\theta(z)$ are inside the unit circle, and is purely noninvertible if all the roots are inside the unit circle.

When modeling a time series with ARMA models within the classical Gaussian framework, the assumptions of causality and invertibility are necessary to ensure the identifiability of the model parameters. To see this, note that the probability structure of a Gaussian process $\{X_t\}$ is completely determined by its first two moments. It follows that for any noncausal/noninvertible ARMA(p, q) process one can find the equivalent causal-invertible representation in the sense that both processes have the same probability structure (see Brockwell and Davis (1991)). Thus, there is no point to distinguish noncausality (noninvertibility) from causality (invertibility) in the Gaussian framework. Causality and invertibility are commonly assumed in order to remove the nonidentifiability of the model parameters and configure the model uniquely.

In contrast, if the underlying noise $\{Z_t\}$ is non-Gaussian, then a noncausal/noninvertible ARMA(p, q) process will have a different probability structure than its causal-invertible representation; see Breidt and Davis (1992) and Rosenblatt (2000). That is, the model parameters are identifiable for non-Gaussian processes even without being confined to the causal-invertible case. Therefore, a general model with the assumptions of causality and invertibility removed may potentially yield a better description of the observed data when fitting time series in the sense that the residuals appear to be more compatible with the assumption of independence than the residuals produced by its causal-invertible counterpart.

Noncausal/noninvertible ARMA models have a wide variety of applications in real life. For example, noncausal models have been used for the deconvolution of seismic signals (Wiggins (1978); Donoho (1981)), and the modeling of vocal tract filters (Chien, Yang, and Chi (1997)) and daily trading volume of a financial asset (Breidt, Davis, and Trindade (2001); Andrews, Calder, and Davis (2009); Wu and Davis (2010)). Noninvertible models have been applied to the analysis of monthly time series of unemployment in the USA (Huang and Pawitan (2000)) and seismogram deconvolution (Andrews, Davis, and Breidt (2006, 2007)).

In regard to parameter estimation, least absolute deviation (LAD) method is frequently used for time series models in a non-Gaussian setting. LAD estimation does not require specification of the density of the underlying noise, although it can be viewed as a quasi-likelihood procedure assuming Laplacian (or double exponential) noise. This is akin to the connection between least squares (LS) estimation and Gaussian noise. In addition, LAD estimation is robust if the observed data display heavy tails. When the underlying noise has finite variance, Davis and Dunsmuir (1997) proved the asymptotic normality of the LAD estimator for causal-invertible ARMA models. Huang and Pawitan (2000) established the conditions that guarantee the consistency or inconsistency of LAD estimator for general MA processes, and conjectured that similar results hold for general ARMA processes. Breidt, Davis, and Trindade (2001) showed asymptotic normality of the LAD estimator for all-pass models, where the roots of the AR polynomial are reciprocals of the roots of the MA polynomial, and vice versa. Wu and Davis (2010) showed the asymptotic normality of the LAD estimator for general ARMA models. When the underlying noise is heavy-tailed, Davis, Knight, and Liu (1992) studied LAD estimation for causal AR processes with innovations in the domain of attraction of a stable law, and showed that the LAD estimator is $n^{1/\alpha}$-consistent, where α is the index of stable distributions and n represents sample size. The asymptotic distribution of the LAD estimator was derived using point process methods for moving averages. Davis (1996) further extended the results to causal-invertible ARMA processes. Huang and Pawitan (1999) showed the consistency of the LAD estimator for general AR processes when the noise has a stable law distribution with index $\alpha \in (1, 2)$. Pan, Wang, and Yao (2007) proposed weighted LAD estimation for parameters of causal-invertible ARMA models, and showed asymptotic normality of the weighted LAD estimator with the standard $n^{1/2}$ convergence rate. Other researchers also studied weighted LAD estimation, e.g., Ling (2005) and Horvath and Liese (2004).

In this paper we consider general ARMA models where the underlying noise has infinite variance, and study LAD estimation for the model parameters. We extend the asymptotic results of Davis (1996) for the LAD estimator to general ARMA processes. The rest of the paper is organized as follows. In the next section, we deconstruct a general ARMA model into its causal, purely noncausal, invertible, and purely noninvertible components. The deconstruction plays a key role in studying estimation for general ARMA models. In Section 3, we establish a functional limit theorem for random processes, and show asymptotic results of the LAD estimator. A simulation study is presented in Section 4 to evaluate the finite sample performance of LAD estimation via comparison with LS estimation. An empirical example of financial time series is also provided. Technical details can be found in the Appendix.

2. Setup

We assume that the underlying noise $\{Z_t\}$ is a sequence of i.i.d. random variables with infinite variance. Specifically, we assume that $\{Z_t\}$ satisfies the following two conditions: for all $x > 0$,

(C1) $\mathrm{P}(|Z_1| > x) = x^{-\alpha} g(x)$, where $g(x)$ is a slowly varying function at infinity and $\alpha \in (0, 2)$;
(C2) there exists a constant $\rho \in [0, 1]$ such that

$$\lim_{x\to\infty} \frac{\mathrm{P}(Z_1 > x)}{\mathrm{P}(|Z_1| > x)} = \rho.$$

That is, Z_t has a distribution in the domain of attraction of a stable distribution with index $\alpha \in (0, 2)$.

We deconstruct the model (1.1) into its causal, purely noncausal, invertible, and purely noninvertible components. On the one hand, we factor the AR polynomial $\phi(\cdot)$ into its causal component $\phi^+(\cdot)$ and purely noncausal component $\phi^*(\cdot)$; namely we put $\phi(z) = \phi^+(z)\phi^*(z)$ where

$$\begin{aligned} \phi^+(z) &= 1 - \phi_1^+ z - \cdots - \phi_{r'}^+ z^{r'} \neq 0 \qquad && \text{for } |z| \leq 1, \\ \phi^*(z) &= 1 - \phi_1^* z - \cdots - \phi_{s'}^* z^{s'} \neq 0 \qquad && \text{for } |z| \geq 1, \end{aligned}$$

with $r', s' \geq 0$ and $r' + s' = p$. On the other hand, we factor the MA polynomial $\theta(\cdot)$ into its invertible component $\theta^+(\cdot)$ and purely noninvertible component $\theta^*(\cdot)$, putting $\theta(z) = \theta^+(z)\theta^*(z)$ where

$$\begin{aligned} \theta^+(z) &= 1 + \theta_1^+ z + \cdots + \theta_r^+ z^r \neq 0 \qquad && \text{for } |z| \leq 1, \\ \theta^*(z) &= 1 + \theta_1^* z + \cdots + \theta_s^* z^s \neq 0 \qquad && \text{for } |z| \geq 1, \end{aligned}$$

with $r, s \geq 0$ and $r + s = q$. With the factorizations, the model (1.1) can be written as

$$\phi^+(B)\phi^*(B)X_t = \theta^+(B)\theta^*(B)Z_t. \tag{2.1}$$

Moreover, defining

$$U_t^+ = \phi^+(B)X_t, \quad U_t^* = \phi^*(B)X_t, \quad V_t^+ = \theta^+(B)Z_t, \quad V_t^* = \theta^*(B)Z_t,$$

we obtain

$$\begin{aligned} \phi^+(B)U_t^* &= \theta^+(B)V_t^*, \qquad & \phi^*(B)U_t^+ &= \theta^+(B)V_t^*, \\ \phi^+(B)U_t^* &= \theta^*(B)V_t^+, \qquad & \phi^*(B)U_t^+ &= \theta^*(B)V_t^+. \end{aligned} \tag{2.2}$$

If the model is causal and invertible, then $s' = s = 0$. Otherwise, $s' > 0$ in the noncausal case, and $s > 0$ in the noninvertible case. In this paper we assume that

s' and s are fixed. However, it is easily seen later that the asymptotic results still hold when the LAD objective function depends on s' and s; we refer the reader to the discussion in Lii and Rosenblatt (1996).

As to parameter estimation for general ARMA processes, it is convenient to work with the deconstructed model (2.1) using the parameterization

$$\boldsymbol{\kappa} = \left(\phi_1^+, \ldots, \phi_{r'}^+, \phi_1^*, \ldots, \phi_{s'}^*, \theta_1^+, \ldots, \theta_r^+, \theta_1^*, \ldots, \theta_s^*\right)^T$$

for the ARMA coefficients. Let

$$\boldsymbol{\kappa}_0 = \left(\phi_{01}^+, \ldots, \phi_{0r'}^+, \phi_{01}^*, \ldots, \phi_{0s'}^*, \theta_{01}^+, \ldots, \theta_{0r}^+, \theta_{01}^*, \ldots, \theta_{0s}^*\right)^T$$

be the true value of $\boldsymbol{\kappa}$, and suppose $X_1, \ldots, X_n$ are observations from the true model. We are interested in estimating the parameter $\boldsymbol{\kappa}$ using the LAD method. The LAD objective function is

$$l_n(\boldsymbol{\kappa}) = \sum_{t=1}^{n+p-q} \left| \frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa}) \right|, \tag{2.3}$$

whose derivation in the finite variance case can be found in Wu and Davis (2010). The LAD estimator $\widehat{\boldsymbol{\kappa}}_{\text{LAD}}$ is defined as any minimizer of $l_n(\boldsymbol{\kappa})$. Given a $\boldsymbol{\kappa}$ value, the residuals $z_t(\boldsymbol{\kappa})$ in the summands of the LAD objective function can be computed as follows. Take the augmented observations $X_{1-p} = \cdots = X_0 = 0$ and $X_{n+1} = \cdots = X_{n+p} = 0$. It follows from the equation $\phi(B) X_t = \theta^+(B) V_t^*(\boldsymbol{\kappa})$ that

$$V_t^*(\boldsymbol{\kappa}) = \phi(B) X_t - \theta_1^+ V_{t-1}^*(\boldsymbol{\kappa}) - \cdots - \theta_r^+ V_{t-r}^*(\boldsymbol{\kappa}). \tag{2.4}$$

Setting $V_t^*(\boldsymbol{\kappa}) = 0$ for $t \le 0$, we compute $V_t^*(\boldsymbol{\kappa})$ forwards by applying (2.4) recursively for $t = 1, \ldots, n+p$. Then it follows from $V_t^*(\boldsymbol{\kappa}) = \theta^*(B) z_t(\boldsymbol{\kappa})$ that

$$z_t(\boldsymbol{\kappa}) = \frac{1}{\theta_s^*}\left[V_{t+s}^*(\boldsymbol{\kappa}) - z_{t+s}(\boldsymbol{\kappa}) - \theta_1^* z_{t+s-1}(\boldsymbol{\kappa}) - \cdots - \theta_{s-1}^* z_{t+1}(\boldsymbol{\kappa})\right], \tag{2.5}$$

and $z_t(\boldsymbol{\kappa})$ is computed backwards by setting $z_t(\boldsymbol{\kappa}) = 0$ for $t > n+p-s$ and using (2.5) recursively for $t = n+p-s, \ldots, -s+1$.

3. Asymptotic Results

To circumvent the difficulty caused by the non-convexity of the objective function $l_n(\boldsymbol{\kappa})$ in $\boldsymbol{\kappa}$ when studying the asymptotic behavior of $\widehat{\boldsymbol{\kappa}}_{\text{LAD}}$, we adopt the local linearization technique of Davis and Dunsmuir (1997) that was used for studying LAD estimation for causal-invertible ARMA processes. To be specific, for $t = 1, \ldots, n+p-q$, we approximate $(\theta_s^*/\phi_{s'}^*) z_t(\boldsymbol{\kappa})$ by

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*} z_t(\boldsymbol{\kappa}_0) - \mathbf{D}_t^T(\boldsymbol{\kappa}_0)(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0),$$

where

$$\mathbf{D}_t(\boldsymbol{\kappa}) = -\frac{\partial\left((\theta_s^*/\phi_{s'}^*)z_t(\boldsymbol{\kappa})\right)}{\partial\boldsymbol{\kappa}} = \left(D_{t,1}(\boldsymbol{\kappa}), \ldots, D_{t,p+q}(\boldsymbol{\kappa})\right)^T.$$

It can be shown that, for $t = 1, \ldots, n+p-q$, $D_{t,\ell}(\boldsymbol{\kappa})$ satisfies the difference equations

$$\theta(B)D_{t,\ell}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*}U_{t-\ell}^*(\boldsymbol{\phi}^*), \qquad \text{for } \ell = 1, \ldots, r',$$

$$\theta(B)D_{t,\ell}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*}U_{t+r'-\ell}^+(\boldsymbol{\phi}^+), \qquad \text{for } \ell = r'+1, \ldots, p-1,$$

$$\theta(B)D_{t,\ell}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{(\phi_{s'}^*)^2}\phi^*(B)U_t^+(\boldsymbol{\phi}^+) + \frac{\theta_s^*}{\phi_{s'}^*}U_{t-s'}^+(\boldsymbol{\phi}^+), \qquad \text{for } \ell = p,$$

$$\theta(B)D_{t,\ell}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*}V_{t+p-\ell}^*(\boldsymbol{\kappa}), \qquad \text{for } \ell = p+1, \ldots, p+r,$$

$$\theta(B)D_{t,\ell}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*}V_{t+p+r-\ell}^+(\boldsymbol{\kappa}), \qquad \text{for } \ell = p+r+1, \ldots, p+q-1,$$

$$\theta(B)D_{t,\ell}(\boldsymbol{\kappa}) = -\frac{1}{\phi_{s'}^*}\theta^*(B)V_t^+(\boldsymbol{\kappa}) + \frac{\theta_s^*}{\phi_{s'}^*}V_{t-s}^+(\boldsymbol{\kappa}), \qquad \text{for } \ell = p+q.$$

Moreover, $D_{t,\ell}(\boldsymbol{\kappa}_0)$ is well approximated by

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*}W_{1,t-\ell}, \qquad \text{for } \ell = 1, \ldots, r',$$

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*}W_{2,t+r'-\ell}, \qquad \text{for } \ell = r'+1, \ldots, p-1,$$

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*}\left(W_{2,t-s'} + \frac{1}{\phi_{0s'}^*}Z_t\right), \qquad \text{for } \ell = p,$$

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*}W_{3,t+p-\ell}, \qquad \text{for } \ell = p+1, \ldots, p+r,$$

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*}W_{4,t+p+r-\ell}, \qquad \text{for } \ell = p+r+1, \ldots, p+q-1,$$

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*}\left(W_{4,t-s} - \frac{1}{\theta_{0s}^*}Z_t\right), \qquad \text{for } \ell = p+q,$$

where $W_{j,t}$, $j = 1, 2, 3, 4$, are AR processes defined by the recursions

$$\phi_0^+(B)W_{1,t} = Z_t, \quad \phi_0^*(B)W_{2,t} = Z_t, \quad \theta_0^+(B)W_{3,t} = Z_t, \quad \text{and} \quad \theta_0^*(B)W_{4,t} = Z_t,$$

respectively. The polynomials $\phi_0^+(z)$, $\phi_0^*(z)$, $\theta_0^+(z)$, and $\theta_0^*(z)$ correspond to the true parameter $\boldsymbol{\kappa}_0$. The reciprocals of these polynomials can be expressed as

$$\frac{1}{\phi_0^+(z)} = \sum_{i=0}^{\infty}\beta_{0i}^+ z^i, \qquad \frac{1}{\phi_0^*(z)} = \sum_{j=s'}^{\infty}\beta_{0j}^* z^{-j},$$

$$\frac{1}{\theta_0^+(z)} = \sum_{i=0}^{\infty} \alpha_{0i}^+ z^i, \qquad \frac{1}{\theta_0^*(z)} = \sum_{j=s}^{\infty} \alpha_{0j}^* z^{-j},$$

where $\beta_{00}^+ = \alpha_{00}^+ = 1$, $\beta_{0s'}^* = -1/\phi_{0s'}^*$, $\alpha_{0s}^* = 1/\theta_{0s}^*$, and each of the sequences $\{\beta_{0i}^+\}$, $\{\beta_{0j}^*\}$, $\{\alpha_{0i}^+\}$, and $\{\alpha_{0j}^*\}$ is absolutely summable. It follows that

$$\begin{aligned} W_{1,t} &= \sum_{i=0}^{\infty} \beta_{0i}^+ Z_{t-i}, \qquad W_{2,t} = \sum_{j=s'}^{\infty} \beta_{0j}^* Z_{t+j}, \\ W_{3,t} &= \sum_{i=0}^{\infty} \alpha_{0i}^+ Z_{t-i}, \qquad W_{4,t} = \sum_{j=s}^{\infty} \alpha_{0j}^* Z_{t+j}. \end{aligned} \tag{3.1}$$

Now define $\mathbf{Q}_t = \left(\mathbf{W}_1^T, \mathbf{W}_2^T, \mathbf{W}_3^T, \mathbf{W}_4^T\right)^T$, where

$$\begin{aligned} \mathbf{W}_1 &= \left(W_{1,t-1}, \ldots, W_{1,t-r'}\right)^T, \\ \mathbf{W}_2 &= \left(W_{2,t-1}, \ldots, W_{2,t-s'+1}, \left(W_{2,t-s'} + \frac{Z_t}{\phi_{0s'}^*}\right)\right)^T, \\ \mathbf{W}_3 &= \left(W_{3,t-1}, \ldots, W_{3,t-r}\right)^T, \\ \mathbf{W}_4 &= \left(W_{4,t-1}, \ldots, W_{4,t-s+1}, \left(W_{4,t-s} - \frac{Z_t}{\theta_{0s}^*}\right)\right)^T. \end{aligned}$$

Then $\mathbf{D}_t(\boldsymbol{\kappa}_0)$ is well approximated by $(\theta_{0s}^*/\phi_{0s'}^*)\mathbf{Q}_t$. Note that $\mathbf{W}_1, \mathbf{W}_3 \in \mathcal{F}_{-\infty}^{t-1}$ and $\mathbf{W}_2, \mathbf{W}_4 \in \mathcal{F}_{t+1}^{\infty}$, where $\mathcal{F}_{-\infty}^{t}$ and $\mathcal{F}_{t}^{\infty}$ are σ-fields generated by $\{Z_k, k \le t\}$ and $\{Z_k, k \ge t\}$, respectively. Therefore, $\mathbf{W}_1$ and $\mathbf{W}_3$ are independent of $\mathbf{W}_2$ and $\mathbf{W}_4$. Moreover, $\mathbf{Q}_t$ is independent of Z_t.

To examine the asymptotic behavior of $\widehat{\boldsymbol{\kappa}}_{\text{LAD}}$, let $a_n = \inf\{x : \mathrm{P}(|Z_1| > x) \le n^{-1}\}$, and build it into the parameterization $\boldsymbol{\nu} = a_n(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0)$. Under this parameterization, minimizing $l_n(\boldsymbol{\kappa})$ with respect to $\boldsymbol{\kappa}$ is equivalent to minimizing

$$S_n(\boldsymbol{\nu}) \equiv \sum_{t=1}^{n+p-q} \left(\left| \frac{\theta_s^*}{\phi_{s'}^*} z_t\left(\boldsymbol{\kappa}_0 + a_n^{-1}\boldsymbol{\nu}\right) \right| - \left| \frac{\theta_{0s}^*}{\phi_{0s'}^*} z_t(\boldsymbol{\kappa}_0) \right| \right) \tag{3.2}$$

with respect to $\boldsymbol{\nu}$. Let

$$Y_t(\boldsymbol{\nu}) = \mathbf{Q}_t^T \boldsymbol{\nu} = \mathbf{W}_1^T \boldsymbol{\nu}_1 + \mathbf{W}_2^T \boldsymbol{\nu}_2 + \mathbf{W}_3^T \boldsymbol{\nu}_3 + \mathbf{W}_4^T \boldsymbol{\nu}_4, \tag{3.3}$$

where the vectors $\boldsymbol{\nu}_1 = (\nu_1, \ldots, \nu_{r'})^T$, $\boldsymbol{\nu}_2 = (\nu_{r'+1}, \ldots, \nu_p)^T$, $\boldsymbol{\nu}_3 = (\nu_{p+1}, \ldots, \nu_{p+r})^T$, and $\boldsymbol{\nu}_4 = (\nu_{p+r+1}, \ldots, \nu_{p+q})^T$. By (3.1), for any given $\boldsymbol{\nu}$ we can express $Y_t(\boldsymbol{\nu})$ as a two-sided moving average of $\{Z_t\}$:

$$Y_t(\boldsymbol{\nu}) = \sum_{i=-\infty}^{\infty} \left(c_i' + c_i\right) Z_{t-i},$$

where $c_0' = c_0 = 0$ and

$$\begin{aligned}
c_i' &= \beta^+_{0,i-1}\nu_1 + \beta^+_{0,i-2}\nu_2 + \cdots + \beta^+_{0,i-r'}\nu_{r'} && \text{for } i > 0,\\
c_i' &= \beta^*_{0,-i+1}\nu_{r'+1} + \beta^*_{0,-i+2}\nu_{r'+2} + \cdots + \beta^*_{0,-i+s'}\nu_p && \text{for } i < 0,\\
c_i &= \alpha^+_{0,i-1}\nu_{p+1} + \alpha^+_{0,i-2}\nu_{p+2} + \cdots + \alpha^+_{0,i-r}\nu_{p+r} && \text{for } i > 0,\\
c_i &= \alpha^*_{0,-i+1}\nu_{p+r+1} + \alpha^*_{0,-i+2}\nu_{p+r+2} + \cdots + \alpha^*_{0,-i+s}\nu_{p+q} && \text{for } i < 0\ ,
\end{aligned}$$

with the conventions that $\beta^+_{0i} = \alpha^+_{0i} = 0$ for $i < 0$, $\beta^*_{0j} = 0$ for $j < s'$, and $\alpha^*_{0j} = 0$ for $j < s$. Furthermore, for any given $\boldsymbol{\nu}$ let

$$Y_t^-(\boldsymbol{\nu}) = \sum_{i=1}^{\infty}\left(c_i' + c_i\right) Z_{t-i}, \quad \text{and} \quad Y_t^+(\boldsymbol{\nu}) = \sum_{i=1}^{\infty}\left(c_{-i}' + c_{-i}\right) Z_{t+i}.$$

Then, $Y_t(\boldsymbol{\nu}) = Y_t^-(\boldsymbol{\nu}) + Y_t^+(\boldsymbol{\nu})$.

Theorem 1. *Let $\{X_t\}$ be the ARMA(p,q) process* (2.1)*, where the underlying noise $\{Z_t\}$ satisfies conditions* (C1) *and* (C2) *and has median zero if $\alpha \geq 1$. For each of* (a) $\alpha < 1$, (b) $\alpha > 1$ *and $E(|Z_t|^{\beta}) < \infty$ for some $\beta < 1 - \alpha$, and* (c) $\alpha = 1$ *and $E(\ln|Z_t|) > -\infty$, $S_n(\boldsymbol{\nu})$ converges to*

$$\begin{aligned}
S(\boldsymbol{\nu}) = &\left|\frac{\theta^*_{0s}}{\phi^*_{0s'}}\right| \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,i} - \left(c_i' + c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right| - |Z_{k,i}|\right)\\
&+ \left|\frac{\theta^*_{0s}}{\phi^*_{0s'}}\right| \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,-i} - \left(c_{-i}' + c_{-i}\right)\delta_k\Gamma_k^{-1/\alpha}\right| - |Z_{k,-i}|\right)
\end{aligned}$$

in distribution on $C(\mathbb{R}^{p+q})$, where

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

(2) $\{\delta_k\}$ *is i.i.d. with* $P(\delta_k = 1) = \rho$ *and* $P(\delta_k = -1) = 1 - \rho$,

(3) $\Gamma_k = E_1 + \cdots + E_k$ *where $\{E_k\}$ is a sequence of i.i.d. unit exponential random variables,*

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

Proof. The theorem is an immediate consequence of Lemmas 6−8 in the Appendix.

Remark 1. Note that, $S(\boldsymbol{\nu})$ is well-defined in all three cases by Proposition 5 in the Appendix.

Theorem 2. *Under the conditions postulated in Theorem* 1*, if $S(\boldsymbol{\nu})$ has a unique minimizer $\boldsymbol{\nu}_{min}$ almost surely, then there exists a sequence of LAD estimators $\widehat{\boldsymbol{\kappa}}_{LAD}$ such that $a_n(\widehat{\boldsymbol{\kappa}}_{LAD} - \boldsymbol{\kappa}_0) \stackrel{d}{\rightarrow} \boldsymbol{\nu}_{min}$.*

Proof. Note that, the weak convergence of $a_n\left(\widehat{\boldsymbol{\kappa}}_{\text{LAD}}-\boldsymbol{\kappa}_0\right)$ is equivalent to that of $\widehat{\boldsymbol{\nu}}_{\text{LAD}}$, where $\widehat{\boldsymbol{\nu}}_{\text{LAD}}$ is a minimizer of $S_n(\boldsymbol{\nu})$. Since $S_n(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ on $C\left(\mathbb{R}^{p+q}\right)$ and $\boldsymbol{\nu}_{\min}$ is the unique minimizer of $S(\boldsymbol{\nu})$, the result follows from Remark 1 of Davis, Knight, and Liu (1992).

4. Numerical Examples

4.1. Simulation study

In general, the random variable $\boldsymbol{\nu}_{\min}$ cannot be expressed in a closed form and its distribution is intractable. In order to permit statistical inferences such as the construction of tests of hypotheses and confidence intervals, one typically needs to turn to nonparametric methods for an approximation of the limit distribution (see Section 5). In regard to the finite sample performance of LAD estimation, a simulation study was conducted that was limited to the comparison of LAD and LS estimations.

We generated data of size 100 from each of the following three models

1. AR(1) model: $X_t - \phi X_{t-1} = Z_t$,
2. MA(1) model: $X_t = Z_t + \theta Z_{t-1}$,
3. ARMA(1, 1) model: $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$,

where $\{Z_t\}$ was a sequence of i.i.d. symmetric α-stable ($S\alpha S$) random variables, and three values of α were considered: 0.5, 1.0, and 1.5. For each case, we simulated 1,000 replications, estimated the parameters of interest using both LAD and LS methods, and here report the empirical mean and mean absolute deviation of the estimates. It is easy to see that, for general ARMA processes, the correct objective function of LS estimation is

$$l_n^{LS}(\boldsymbol{\kappa}) = \sum_{t=1}^{n+p-q}\left[\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right]^2 .$$

When searching for the minimizer of the LAD objective function, for each replication we used 10 starting values and found the optimized value for each of them such that the chance of being trapped in a local minimum was reduced. Among the 10 optimized values we chose as estimate the one that yielded the smallest evaluation of the LAD objective function. When searching for the minimizer of the LS objective function, however, we had to restrict the range of parameters to be compatible with the (non)causality and (non)invertibility of the true model. This is because, for general ARMA processes, the LS objective function has multiple global minimizers. To see this, we took the AR(1) model $X_t = \phi X_{t-1} + Z_t$

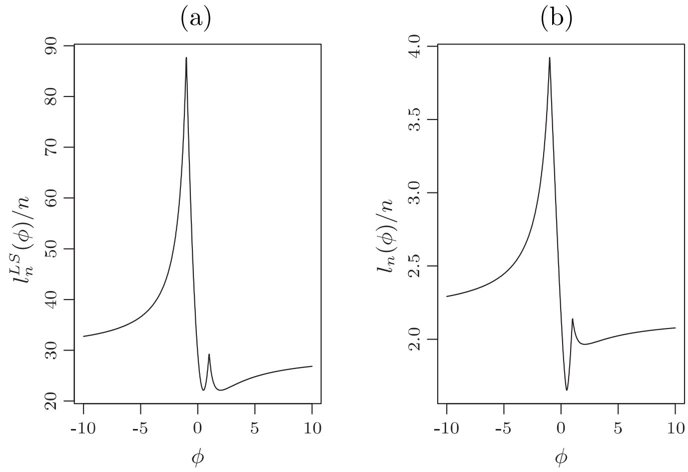

Figure 1. (a) The average LS objective function versus ϕ, and (b) the average LAD objective function versus ϕ.

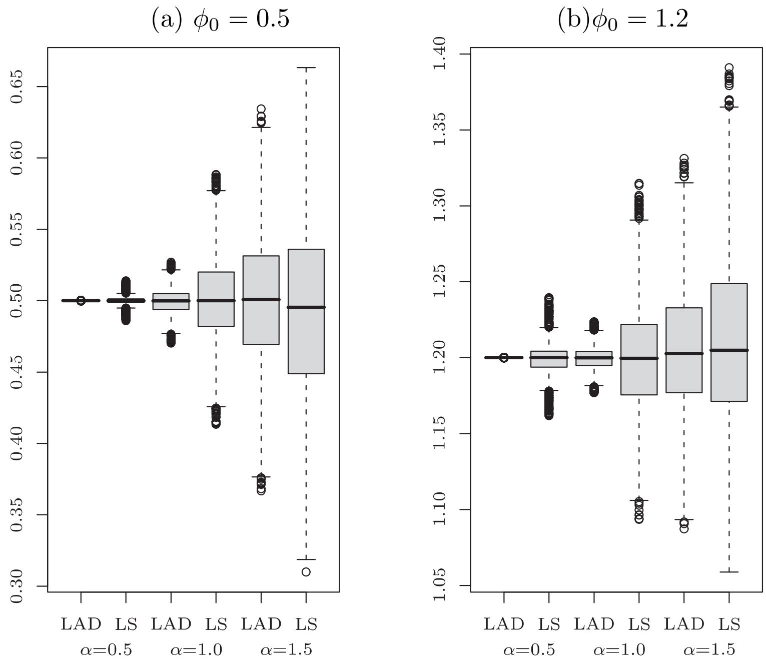

Figure 2. Boxplots of LAD and LS estimates for the AR(1) model: (a) $\phi_0 = 0.5$ and (b) $\phi_0 = 1.2$.

with true parameter value $\phi = 0.5$ and index of the $S\alpha S$ distribution of innovations $\alpha = 1.5$. We generated ten thousand x_t values from the true model and plotted the average LS objective function $l_n^{LS}(\phi)/n$ versus ϕ. From Figure 1(a)

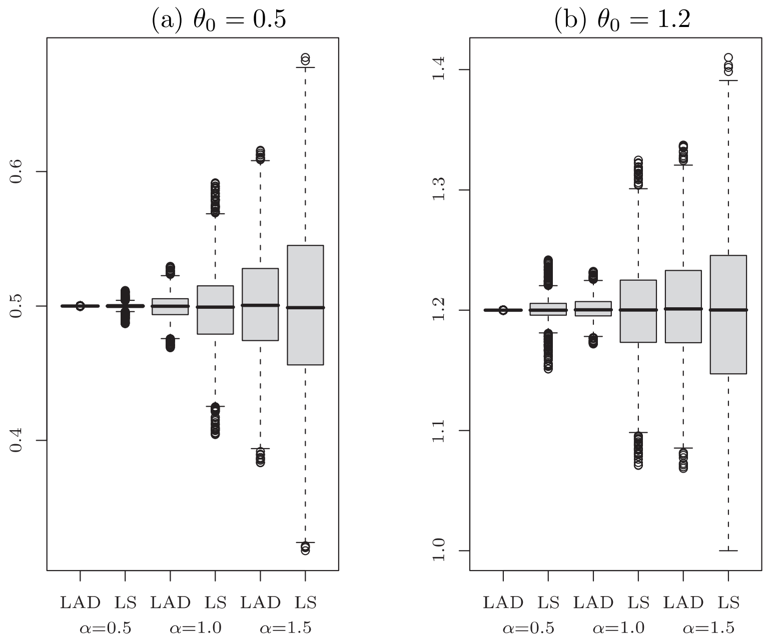

Figure 3. Boxplots of LAD and LS estimates for the MA(1) model: (a) $\theta_0 = 0.5$ and (b) $\theta_0 = 1.2$.

we see that both the true model parameter 0.5 and its reciprocal 2 are global minimizers. In contrast, Figure 1(b) for the LAD estimation case shows clearly that the true parameter 0.5 is the unique global minimizer; its reciprocal 2 is only a local minimizer.

For the first model, the true value of ϕ was taken to be 0.5 and 1.2. The results are reported in Table 1, and side-by-side boxplots are given in Figure 2. The LAD estimates in general outperformed the LS estimates. This is expected since LAD estimation is more efficient than LS method for heavy-tailed data. In addition, there was more improvement in performance of LAD estimation over LS estimation as the tail of the noise distribution grew heavier. On the other hand, for both LAD and LS estimates the performance improved as the tail of the noise distribution grew heavier. From Theorem 2, the convergence rate of $\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}}$ is $n^{1/\alpha}$, which is the same as in the causal-invertible case; and the convergence rate of $\widehat{\boldsymbol{\kappa}}_{\mathrm{LS}}$ should also be in agreement with that in the causal-invertible case, namely $(n/\log n)^{1/\alpha}$.

For the MA(1) model, the true value of θ was taken to be 0.5 and 1.2. The simulation results are reported in Table 2, and side-by-side boxplots are given in Figure 3. The true values of (ϕ, θ) for the ARMA(1,1) model were taken to be $(0.5, 1.2)$ and $(1.2, 0.5)$, and the simulation results are reported in Table 3.

Table 1. Mean and mean absolute deviation for LAD and LS estimates for the AR(1) model.

True value	α	LAD estimates		LS estimates	
$\phi = 0.5$	0.5	0.5000	(5.0648e-5)	0.5000	(0.0030)
	1.0	0.4993	(0.0080)	0.5006	(0.0258)
	1.5	0.5002	(0.0391)	0.4918	(0.0530)
$\phi = 1.2$	0.5	1.2000	(3.4469e-5)	1.1997	(0.0098)
	1.0	1.1998	(0.0066)	1.2007	(0.0310)
	1.5	1.2055	(0.0355)	1.2125	(0.0483)

Table 2. Mean and mean absolute deviation for LAD and LS estimates for the MA(1) model.

True value	α	LAD estimates		LS estimates	
$\theta = 0.5$	0.5	0.5000	(5.4845e-5)	0.4998	(0.0026)
	1.0	0.4993	(0.0083)	0.4983	(0.0266)
	1.5	0.5013	(0.0350)	0.5000	(0.0553)
$\theta = 1.2$	0.5	1.2000	(3.7825e-5)	1.2003	(0.0095)
	1.0	1.2013	(0.0084)	1.1990	(0.0356)
	1.5	1.2042	(0.0382)	1.1942	(0.0644)

Table 3. Mean and mean absolute deviation for LAD and LS estimates for the ARMA(1,1) model.

True values	α	LAD estimates		LS estimates	
$\phi = 0.5$, $\theta = 1.2$	0.5	0.5000	(7.1848e-5)	0.5000	(0.0048)
		1.2000	(5.6999e-5)	1.2012	(0.0110)
	1.0	0.4992	(0.0092)	0.4997	(0.0281)
		1.2028	(0.0088)	1.2030	(0.0375)
	1.5	0.4948	(0.0391)	0.4981	(0.0527)
		1.2186	(0.0471)	1.2189	(0.0696)
$\phi = 1.2$, $\theta = 0.5$	0.5	1.2000	(4.4627e-5)	1.2022	(0.0143)
		0.5000	(7.0291e-5)	0.4962	(0.0109)
	1.0	1.2008	(0.0079)	1.2074	(0.0357)
		0.4970	(0.0094)	0.4944	(0.0384)
	1.5	1.2084	(0.0369)	1.2221	(0.0564)
		0.4905	(0.0393)	0.4898	(0.0602)

The side-by-side boxplots display a similar pattern as in the AR(1) and MA(1) cases, and hence are omitted. In all cases the LAD estimates outperformed the LS estimates, especially when the tail of the noise distribution got heavier.

4.2. Empirical example

Figure 4 shows the natural logarithms of the traded Microsoft (MSFT) stock volumes from June 3, 1996 to May 27, 1999. Based on the normal probability

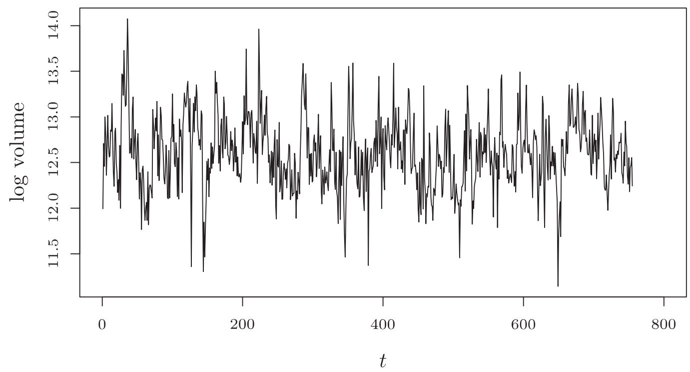

Figure 4. Log-volumes of Microsoft stock transactions.

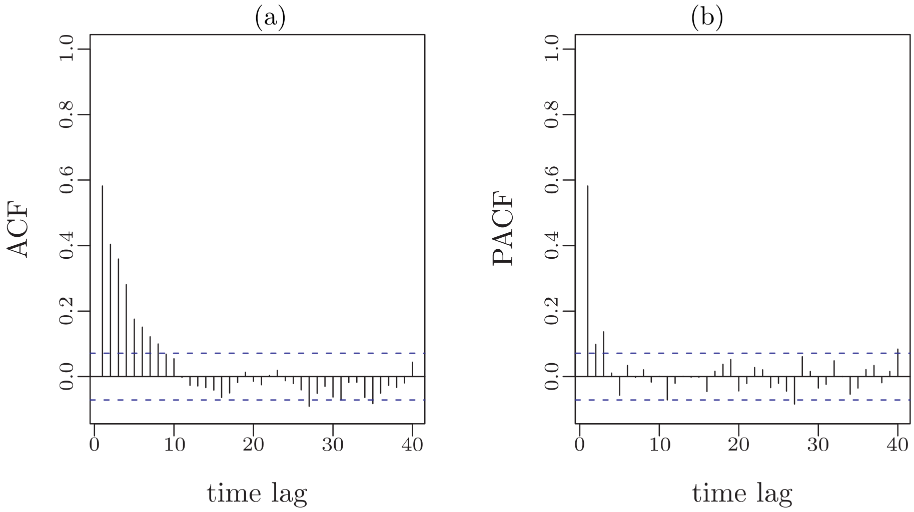

Figure 5. (a) Sample ACF of log-volumes, and (b) sample PACF of log-volumes.

plot, not included, the log-volumes appear heavy-tailed rather than Gaussian. The Jarque-Bera test for normality yields a p-value of 0.0021, while the Shapiro–Wilk test gives a p-value of 0.0149; both indicate strong evidence to reject normality of the data. Even if data are heavy-tailed, a variety of techniques commonly used in the Gaussian framework for exploratory data analysis are still useful to get insight into the underlying data structure. For example, the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) plots can be employed for visualizing dependency and for tentative identification of a suitable ARMA model for the data; see Adler, Feldman, and Gallagher (1998) and Andrews, Calder, and Davis (2009). Based on the sample ACF and PACF plots of the log-volume series shown in Figure 5, it is reasonable to consider fitting an AR(3) model to the data. Breidt, Davis, and Trindade (2001) studied the same data set and fitted an AR(1) model to illustrate noncausal AR model fitting using all-pass models.

Note that, LAD estimation does not require the specification of the density of innovations in order to estimate ARMA coefficients. Instead, the distribution of innovations can be specified and distribution parameters can be estimated based on the resulting residuals from the model-fitting. We applied the LAD method to the mean-corrected series to obtain a purely noncausal AR(3) model

$$X_t = -0.0643X_{t-1} - 3.3954X_{t-2} + 6.4155X_{t-3} + Z_t. \tag{4.1}$$

All the roots of the AR polynomial lie inside the unit circle, namely 0.7910 and $-0.1309 \pm 0.4242i$. The residual sequence appears independent based on the sample ACF plots in Figure 6(b)–(d). The p-values of 3.741e-6 and 0.0003 from the Jarque-Bera and Shapiro–Wilk tests, together with the normal probability plot, support heavy-tailedness of the residuals with strong evidence. On the other hand, under the false assumption of finite variance noise, the best causal AR model-fitting based on AIC is given by

$$X_t = 0.5139X_{t-1} + 0.0237X_{t-2} + 0.1378X_{t-3} + W_t. \tag{4.2}$$

Although the sample ACF plot in Figure 7(b) indicates that the residual sequence is white noise, the sample ACF plots of squared residuals and absolute values of residuals in Figure 7(c),(d) suggest that the residual sequence is dependent because both squared residuals and absolute values of residuals have significant lag 1 sample autocorrelation. Therefore, in regard to the independence assumption of the noise term, the model (4.1) provides a better description for the log-volumes of Microsoft stock transactions. The McLeod–Li test of independence performed on the residuals from both model-fittings also supports the superiority of the model (4.1).

Based on the residuals obtained from the fitted model (4.1), it was plausible to consider a non-Gaussian stable distribution for the innovation sequence $\{Z_t\}$. The distribution parameters were estimated using maximum likelihood method, implemented with R package "fBasics". And the estimates were $\alpha = 1.9200$, $\beta = 0.3069$, $\gamma = 1.3832$, and $\delta = -0.0539$.

5. Discussion

The LAD estimator $\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}}$ is $n^{1/\alpha}$-consistent, and, as $\alpha \in (0, 2)$, it has a higher convergence rate than the standard one of $n^{1/2}$, a desirable property when conducting finite sample studies, especially with small α. However, its limiting distribution is in general intractable. To overcome this hurdle in applying our

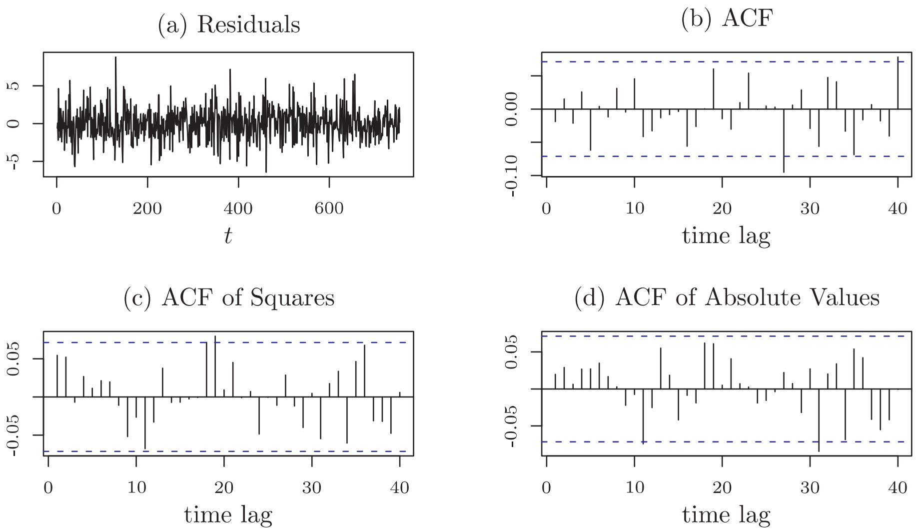

Figure 6. (a) Residuals from the AR model-fitting using the LAD method, (b) sample ACF of the residuals, (c) sample ACF of the squared residuals, and (d) sample ACF of the absolute values of residuals.

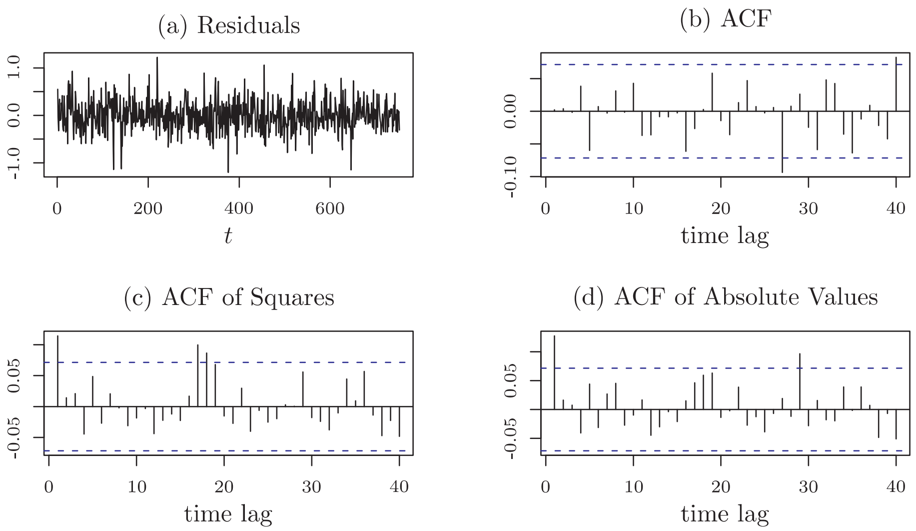

Figure 7. (a) Residuals from the causal AR model-fitting, (b) sample ACF of the residuals, (c) sample ACF of the squared residuals, and (d) sample ACF of the absolute values of residuals.

asymptotic results to statistical inferences with respect to $\boldsymbol{\kappa}$, we can resort to such techniques as the bootstrap to approximate the limiting distribution. To this end, Davis and Wu (1997) explored the bootstrap procedure for causal autoregressive processes with infinite variance, and showed its asymptotic validity. Extension of such a procedure to the present setup is a topic for future research.

Alternatively, one might consider extending the weighted LAD estimation proposed by Pan, Wang, and Yao (2007) to general ARMA models. Although we conjecture, in agreement with two anonymous referees, that the extension ought to be valid, the derivation of asymptotic normality is quite challenging. For example, martingale central limit theorems (see e.g., Hall and Heyde (1980)), on which the derivation of Pan, Wang, and Yao (2007) relies heavily, are no longer applicable. Indeed, with the assumptions of causality and invertibility removed, the ARMA process X_t cannot be expressed in terms of present and past Z_t's, and likewise Z_t cannot be written in terms of present and past X_t's. Therefore, a new device is called for in order to establish the asymptotic normality of the weighted LAD estimator. This is the subject of on-going research. Once the asymptotic properties of the weighted LAD estimator are established for general ARMA models, numerical comparison between the two LAD estimators is in order.

Acknowledgement

I thank Professor Richard Davis, my thesis advisor, for his guidance and support. The comments of two anonymous referees were very helpful, leading to improvement of the paper. This work was supported in part by The City University of New York PSC-CUNY Research Award 60052-38 39.

Appendix

The Appendix contains the lemmas used in establishing Theorem 1, and their proofs. Throughout the Appendix we denote by C an unspecified constant whose value may vary. To facilitate our argument, we introduce some preliminary results that extend the results given in the Appendix of Davis, Knight, and Liu (1992) to two-sided moving average processes.

Suppose $\{Y_t\}$ is the two-sided linear process

$$Y_t = \sum_{i=-\infty}^{\infty} c_i Z_{t-i}, \tag{A.1}$$

where $\{Z_t\}$ is a sequence of i.i.d. random variables satisfying (C1) and (C2), and $\{c_i\}$ is a sequence of constants such that $c_0 = 0$ and $\sum_{i=-\infty}^{\infty} |c_i|^\delta < \infty$ for some $\delta < \min(\alpha, 1)$. Write

$$Y_t^- = \sum_{i=1}^{\infty} c_i Z_{t-i} \qquad \text{and} \qquad Y_t^+ = \sum_{i=1}^{\infty} c_{-i} Z_{t+i}.$$

Then, $Y_t = Y_t^- + Y_t^+$.

Proposition 3. *Suppose* $\{Y_t\}$ *is the process given by* (A.1). *For any continuous function* f *of* Z_t, $a_n^{-1}Y_t^-$ *and* $a_n^{-1}Y_t^+$ *on* $\mathbb{R}\times(\overline{\mathbb{R}}^2\setminus\{\mathbf{0}\})$ *with compact support,*

$$\sum_{t=1}^{n} f\left(Z_t, a_n^{-1}Y_t^-, a_n^{-1}Y_t^+\right) \xrightarrow{d} \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left[f\left(Z_{k,i}, c_i\delta_k\Gamma_k^{-1/\alpha}, 0\right) + f\left(Z_{k,-i}, 0, c_{-i}\delta_k\Gamma_k^{-1/\alpha}\right)\right]. \qquad \text{(A.2)}$$

Proof. See Calder (1998).

Proposition 4. *Suppose* $\{Y_t\}$ *is the process given by* (A.1). *Let* $\{V_t\}$ *be a sequence of i.i.d. random variables with finite mean such that, for every* t, $\{V_t\}$ *and* $\{Y_t\}$ *are independent. Then for all* $\delta>0$ *and* $\eta>0$,

(a) $$\limsup_{n\to\infty} P\left(\sum_{t=1}^{n}|V_t|\left|a_n^{-1}Y_t\right|^{\gamma} 1_{\{|a_n^{-1}Y_t|\le\delta\}} > \eta\right) \le \eta^{-1}CE|V_1|\,\delta^{\gamma-\alpha}$$

for all $\gamma>\alpha$.

(b) $$\limsup_{n\to\infty} P\left(\sum_{t=1}^{n}|V_t|\left|a_n^{-1}Y_t\right|^{\gamma} 1_{\{|a_n^{-1}Y_t|>\delta\}} > \eta\right) \le C\delta^{-\alpha}P\left(|V_1|>0\right)$$

for all $\gamma>0$.

If in addition $\{V_1\}$ *has zero mean and finite variance* σ^2 *and* $\alpha\in[1,2)$, *then*

(c) $$Var\left(\sum_{t=1}^{n}V_t a_n^{-1}Y_t 1_{\{|a_n^{-1}Y_t|\le\delta\}}\right) = na_n^{-2}E\left(Y_1^2 1_{\{|a_n^{-1}Y_1|\le\delta\}}\right)EV_1^2 \to 0$$

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

Proof. (a) and (c) are straightforward extensions of Proposition A.2 of Davis, Knight, and Liu (1992). Turning to (b), note that by Theorem 4.1 of Cline (1983), the two-sided linear process $\{Y_t\}$ is absolutely convergent almost surely and has regularly varying tails equivalent to those of $\{Z_t\}$; that is,

$$\lim_{x\to\infty}\frac{\mathrm{P}\left(|Y_1|>x\right)}{\mathrm{P}\left(|Z_1|>x\right)} = \sum_{i=-\infty}^{\infty}|c_i|^{\alpha}.$$

It follows that

$$\lim_{n\to\infty} n\mathrm{P}\left(|Y_1|>a_n x\right) = \sum_{i=-\infty}^{\infty}|c_i|^{\alpha}x^{-\alpha}$$

for all $x>0$. Then, the argument follows the same lines in Davis, Knight, and Liu (1992).

Proposition 5. *Let* $\{Z_{k,i}\}, k = 1, 2, \ldots, i = \pm 1, \pm 2, \ldots,$ *be an array of i.i.d. symmetric random variables. Suppose* $\{Z_{k,i}\}$ *is independent of* $\{\delta_k\}$ *and* $\{\Gamma_k^{-1/\alpha}\}$*. Let*

$$V = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left[\left(\left|Z_{k,i} - c_i\delta_k\Gamma_k^{-1/\alpha}\right| - |Z_{k,i}|\right) + \left(\left|Z_{k,-i} - c_{-i}\delta_k\Gamma_k^{-1/\alpha}\right| - |Z_{k,-i}|\right)\right],$$

where $\sum_{i\neq 0}|c_i| < \infty$*. Then* (a) *for* $\alpha < 1$*,* V *is finite with probability* 1*,* (b) *for* $\alpha > 1$*,* V *is finite with probability* 1 *if and only if* $E(|Z_{1,1}|^{1-\alpha}) < \infty$*, and (c) for* $\alpha = 1$*,* V *is finite with probability* 1 *if and only if* $E(\ln|Z_{1,1}|) > -\infty$*.*

Proof. See Davis, Knight, and Liu (1992, pp.175-176)

Remark 2. As pointed out by Davis, Knight, and Liu (1992), for $\alpha \geq 1$, the sufficiency still holds if the symmetry of $Z_{k,i}$ is weakened to the condition that $Z_{k,i}$ has median 0.

Lemma 6. *For* $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$*, let*

$$S_n^{\ddagger}(\boldsymbol{\nu}) = \left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right| \sum_{t=1}^{n+p-q}\left(\left|Z_t - a_n^{-1}Y_t(\boldsymbol{\nu})\right| - |Z_t|\right).$$

If the conditions in Theorem 1 are satisfied, then $S_n^{\ddagger}(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ *on* $C(\mathbb{R}^{p+q})$*.*

Proof. We first show pointwise weak convergence. Recalling that $Y_t(\boldsymbol{\nu}) = Y_t^-(\boldsymbol{\nu}) + Y_t^+(\boldsymbol{\nu})$, we rewrite $S_n^{\ddagger}(\boldsymbol{\nu})$ as

$$S_n^{\ddagger}(\boldsymbol{\nu}) = \left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right| \sum_{t=1}^{n+p-q}\left(\left|Z_t - a_n^{-1}Y_t^-(\boldsymbol{\nu}) - a_n^{-1}Y_t^+(\boldsymbol{\nu})\right| - |Z_t|\right)$$

so that we can apply Proposition 3 with $f(x, y, z) = |x - y - z| - |x|$. However, the convergence holds only if the function f is restricted to compact sets. In this regard, we define

$$S_n^{\ddagger}(\boldsymbol{\nu}; \delta, M) = \left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right| \sum_{t=1}^{n+p-q}\left(\left|Z_t - a_n^{-1}Y_t^-(\boldsymbol{\nu}) - a_n^{-1}Y_t^+(\boldsymbol{\nu})\right| - |Z_t|\right) I_{nt}^{\delta,M},$$

where $I_{nt}^{\delta,M} = 1_{\{|Z_t|\leq M\}}1_{\{|a_n^{-1}(Y_t^-(\boldsymbol{\nu})+Y_t^+(\boldsymbol{\nu}))|>\delta\}}$ for large $M > 0$ and small $\delta > 0$. Then, by (A.2), $S_n^{\ddagger}(\boldsymbol{\nu}; \delta, M)$ converges in distribution to

$$S(\boldsymbol{\nu};\delta,M)$$
$$=\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right|\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,i}-\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,i}\right|\right)1_{\left\{\left|Z_{k,i}\right|\le M\right\}}1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|>\delta\}}$$
$$+\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right|\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,-i}-\left(c_{-i}'+c_{-i}\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,-i}\right|\right)1_{\left\{\left|Z_{k,-i}\right|\le M\right\}}$$
$$\times 1_{\{|(c_{-i}'+c_{-i})\delta_k\Gamma_k^{-1/\alpha}|>\delta\}}.$$

Therefore, in order to show $S_n^{\ddagger}(\boldsymbol{\nu})\xrightarrow{d}S(\boldsymbol{\nu})$, by Theorem 3.2 of Billingsley (1999) it suffices to show that, for all $\epsilon>0$,

$$\lim_{\delta\to 0}\lim_{M\to\infty}\limsup_{n\to\infty}\mathrm{P}\left(\left|\sum_{t=1}^{n+p-q}\left(\left|Z_t-a_n^{-1}Y_t^-(\boldsymbol{\nu})-a_n^{-1}Y_t^+(\boldsymbol{\nu})\right|-|Z_t|\right)(1-I_{nt}^{\delta,M})\right|>\epsilon\right)=0, \tag{A.3}$$

and

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,i}-\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,i}\right|\right)1_{\{|Z_{k,i}|\le M\}}1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|>\delta\}}$$
$$\xrightarrow{P}\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,i}-\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,i}\right|\right), \tag{A.4}$$

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,-i}-\left(c_{-i}'+c_{-i}\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,-i}\right|\right)1_{\{|Z_{k,-i}|\le M\}}1_{\{|(c_{-i}'+c_{-i})\delta_k\Gamma_k^{-1/\alpha}|>\delta\}}$$
$$\xrightarrow{P}\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,-i}-\left(c_{-i}'+c_{-i}\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,-i}\right|\right), \tag{A.5}$$

as $M\to\infty$ and $\delta\to 0$.

Proof of (A.3)**.** For the case of $\alpha<1$, since

$$\left|\sum_{t=1}^{n+p-q}\left(\left|Z_t-a_n^{-1}Y_t^-(\boldsymbol{\nu})-a_n^{-1}Y_t^+(\boldsymbol{\nu})\right|-|Z_t|\right)1_{\{|Z_t|>M\}}1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|>\delta\}}\right|$$
$$\le\sum_{t=1}^{n+p-q}\left|a_n^{-1}Y_t(\boldsymbol{\nu})\right|1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|>\delta\}}1_{\{|Z_t|>M\}}$$

and, by applying Proposition 4(b) with $V_t=1_{\{|Z_t|>M\}}$,

$$\limsup_{n\to\infty}\mathrm{P}\left(\sum_{t=1}^{n+p-q}\left|a_n^{-1}Y_t(\boldsymbol{\nu})\right|1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|>\delta\}}1_{\{|Z_t|>M\}}>\epsilon\right)$$

$$\leq C\delta^{-\alpha}\mathrm{P}\left(1_{\{|Z_t|>M\}}>0\right)=C\delta^{-\alpha}\mathrm{P}\left(|Z_t|>M\right)\to 0 \quad \text{as } M\to\infty,$$

we have

$$\lim_{M\to\infty}\limsup_{n\to\infty}\mathrm{P}\left(\left|\sum_{t=1}^{n+p-q}\left(\left|Z_t-a_n^{-1}Y_t^{-}(\boldsymbol{\nu})-a_n^{-1}Y_t^{+}(\boldsymbol{\nu})\right|-|Z_t|\right)1_{\{|Z_t|>M\}}\right.\right.$$
$$\left.\left.\times 1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|>\delta\}}\right|>\epsilon\right)=0. \tag{A.6}$$

Similarly, since

$$\left|\sum_{t=1}^{n+p-q}\left(\left|Z_t-a_n^{-1}Y_t^{-}(\boldsymbol{\nu})-a_n^{-1}Y_t^{+}(\boldsymbol{\nu})\right|-|Z_t|\right)1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|\leq\delta\}}\right|$$
$$\leq\sum_{t=1}^{n+p-q}\left|a_n^{-1}Y_t(\boldsymbol{\nu})\right|1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|\leq\delta\}}$$

and, on the other hand by Proposition 4(a),

$$\limsup_{n\to\infty}\mathrm{P}\left(\sum_{t=1}^{n+p-q}\left|a_n^{-1}Y_t(\boldsymbol{\nu})\right|1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|\leq\delta\}}>\epsilon\right)\leq\frac{C\delta^{1-\alpha}}{\epsilon}\to 0 \quad \text{as } \delta\to 0,$$

we obtain

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathrm{P}\left(\left|\sum_{t=1}^{n+p-q}\left(\left|Z_t-a_n^{-1}Y_t^{-}(\boldsymbol{\nu})-a_n^{-1}Y_t^{+}(\boldsymbol{\nu})\right|-|Z_t|\right)1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|\leq\delta\}}\right|\right.$$
$$\left.>\epsilon\right)=0. \tag{A.7}$$

Note that $1-I_{nt}^{\delta,M}=1_{\{|Z_t|>M\}}1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|>\delta\}}+1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|\leq\delta\}}$. Therefore, (A.3) follows from (A.6) and (A.7).

To show (A.3) for the case $\alpha\geq 1$, we use the identity

$$\sum_{t=1}^{n+p-q}\left(\left|Z_t-a_n^{-1}Y_t^{-}(\boldsymbol{\nu})-a_n^{-1}Y_t^{+}(\boldsymbol{\nu})\right|-|Z_t|\right)$$
$$=-\sum_{t=1}^{n+p-q}a_n^{-1}Y_t(\boldsymbol{\nu})\,\mathrm{sgn}(Z_t)$$
$$+2\sum_{t=1}^{n+p-q}\left(a_n^{-1}Y_t(\boldsymbol{\nu})-Z_t\right)\left(1_{\left\{a_n^{-1}Y_t(\boldsymbol{\nu})>Z_t>0\right\}}-1_{\left\{a_n^{-1}Y_t(\boldsymbol{\nu})<Z_t<0\right\}}\right). \tag{A.8}$$

By an application of Proposition 4(b) with $V_t = \operatorname{sgn}(Z_t)\, 1_{\{|Z_t|>M\}}$, it is easy to show

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\sum_{t=1}^{n+p-q} a_n^{-1} Y_t(\boldsymbol{\nu}) \operatorname{sgn}(Z_t)\, 1_{\{|Z_t|>M\}} 1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|>\delta\}}\right| > \epsilon\right) = 0.$$

Likewise, an application of Proposition 4(c) with $V_t = \operatorname{sgn}(Z_t)$, together with Chebyshev's inequality, yields

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathrm{P}\left(\left|\sum_{t=1}^{n+p-q} a_n^{-1} Y_t(\boldsymbol{\nu}) \operatorname{sgn}(Z_t)\, 1_{\{|a_n^{-1}Y_t(\boldsymbol{\nu})|\le\delta\}}\right| > \epsilon\right) = 0.$$

It follows that

$$\lim_{\delta\to 0} \lim_{M\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\sum_{t=1}^{n+p-q} a_n^{-1} Y_t(\boldsymbol{\nu}) \operatorname{sgn}(Z_t)\, (1 - I_{nt}^{\delta,M})\right| > \epsilon\right) = 0. \tag{A.9}$$

Hence, (A.3) holds provided that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\sum_{t=1}^{n+p-q} \left(a_n^{-1} Y_t(\boldsymbol{\nu}) - Z_t\right) 1_{\left\{a_n^{-1}Y_t(\boldsymbol{\nu})>Z_t>M\right\}} > \epsilon\right) = 0,$$

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\sum_{t=1}^{n+p-q} \left(a_n^{-1} Y_t(\boldsymbol{\nu}) - Z_t\right) 1_{\left\{a_n^{-1}Y_t(\boldsymbol{\nu})<Z_t<-M\right\}} < -\epsilon\right) = 0,$$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathrm{P}\left(\sum_{t=1}^{n+p-q} \left(a_n^{-1} Y_t(\boldsymbol{\nu}) - Z_t\right) 1_{\left\{\delta\ge a_n^{-1}Y_t(\boldsymbol{\nu})>Z_t>0\right\}} > \epsilon\right) = 0,$$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathrm{P}\left(\sum_{t=1}^{n+p-q} \left(a_n^{-1} Y_t(\boldsymbol{\nu}) - Z_t\right) 1_{\left\{-\delta\le a_n^{-1}Y_t(\boldsymbol{\nu})<Z_t<0\right\}} < -\epsilon\right) = 0,$$

which can be established using similar arguments in Davis, Knight, and Liu (1992, pp.158-159)

Proof of (A.4). When $\alpha < 1$, it is equivalent to showing that

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty} \left(\left|Z_{k,i} - \left(c_i' + c_i\right)\delta_k \Gamma_k^{-1/\alpha}\right| - |Z_{k,i}|\right) 1_{\left\{|Z_{k,i}|>M\right\}} 1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|>\delta\}} \xrightarrow{P} 0 \tag{A.10}$$

as $M \to \infty$ and $\delta \to 0$, and

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty} \left(\left|Z_{k,i} - \left(c_i' + c_i\right)\delta_k \Gamma_k^{-1/\alpha}\right| - |Z_{k,i}|\right) 1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|\le\delta\}} \xrightarrow{P} 0 \tag{A.11}$$

as $\delta \to 0$. Note that

$$\left|\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,i}-\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,i}\right|\right)1_{\{|Z_{k,i}|>M\}}1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|>\delta\}}\right|$$
$$\leq \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left|\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|1_{\{|Z_{k,i}|>M\}}.$$

On the other hand,

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left|\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|1_{\{|Z_{k,i}|\leq M\}}\overset{\text{a.s.}}{\to}\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left|\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|$$

as $M \to \infty$, because the left-hand side is non-decreasing as $M \to \infty$ and the right-hand side is almost surely finite. Therefore

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left|\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|1_{\{|Z_{k,i}|>M\}}\overset{\text{a.s.}}{\to}0.$$

and hence (A.10) follows. Similarly, (A.11) holds because

$$\left|\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(\left|Z_{k,i}-\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,i}\right|\right)1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|\leq\delta\}}\right|$$
$$\leq \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left|\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|\leq\delta\}}\overset{\text{a.s.}}{\to}0 \qquad \text{as } \delta\to 0.$$

To show (A.4) when $\alpha \geq 1$, we write

$$\left|Z_{k,i}-\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\right|-\left|Z_{k,i}\right|$$
$$=-\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\operatorname{sgn}\left(Z_{k,i}\right)+2\left[\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}-Z_{k,i}\right]$$
$$\times\left(1_{\{(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}>Z_{k,i}>0\}}-1_{\{(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}<Z_{k,i}<0\}}\right).$$

Each of the terms on the right-hand side, when summed over i and k, is finite almost surely. After the fashion of proving (A.9), we can show

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}\operatorname{sgn}\left(Z_{k,i}\right)\left(1-1_{\{|Z_{k,i}|\leq M\}}1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|>\delta\}}\right)\overset{P}{\to}0$$

as $M \to \infty$ and $\delta \to 0$. Moreover, it is easy to show that

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left[\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}-Z_{k,i}\right]\Big(1_{\{(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}>Z_{k,i}>0\}}$$
$$-1_{\{(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}<Z_{k,i}<0\}}\Big)1_{\{|Z_{k,i}|\le M\}}1_{\{|(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}|>\delta\}}$$
$$\overset{\text{a.s.}}{\to}\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\left[\left(c_i'+c_i\right)\delta_k\Gamma_k^{-1/\alpha}-Z_{k,i}\right]\Big(1_{\{(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}>Z_{k,i}>0\}}$$
$$-1_{\{(c_i'+c_i)\delta_k\Gamma_k^{-1/\alpha}<Z_{k,i}<0\}}\Big)$$

as $M \to \infty$ and $\delta \to 0$, by noting that the term on the left-hand side is non-negative and non-decreasing. This completes the proof of (A.4).

Proof of (A.5). The proof follows along the same lines as that of (A.4), and hence is omitted.

Now, the weak convergence of finite-dimensional distributions of $S_n^{\ddagger}(\boldsymbol{\nu})$ to those of $S(\boldsymbol{\nu})$ follows by applying the Cramér–Wold device. Moreover, the distributions of $S_n^{\ddagger}(\boldsymbol{\nu})$ are tight on compact sets in $C(\mathbb{R}^{p+q})$ since $S_n^{\ddagger}$ is linear in $\boldsymbol{\nu}$. By Theorem 7.1 of Billingsley (1999), therefore, $S_n^{\ddagger}(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ on $C(\mathbb{R}^{p+q})$. This completes the proof.

Lemma 7. *For $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$, let*

$$S_n^{\dagger}(\boldsymbol{\nu}) = \sum_{t=1}^{n+p-q}\left(\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)-a_n^{-1}\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}\right|-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)\right|\right),$$

then $S_n^{\dagger}(\boldsymbol{\nu}) - S_n^{\ddagger}(\boldsymbol{\nu}) \to 0$ in probability uniformly on compact sets in $C(\mathbb{R}^{p+q})$.

Proof. Let $m_n = o(n^{1/2})$. We rewrite

$$\begin{aligned}S_n^{\dagger}(\boldsymbol{\nu}) - S_n^{\ddagger}(\boldsymbol{\nu}) =& \sum_{t=1}^{m_n-1}\left(\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)-a_n^{-1}\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}\right|-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)\right|\right)\\&-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right|\sum_{t=1}^{m_n-1}\left(\left|Z_t-a_n^{-1}\mathbf{Q}_t^T\boldsymbol{\nu}\right|-|Z_t|\right)\\&+\sum_{t=n-m_n+1}^{n+p-q}\left(\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)-a_n^{-1}\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}\right|-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)\right|\right)\\&-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right|\sum_{t=n-m_n+1}^{n+p-q}\left(\left|Z_t-a_n^{-1}\mathbf{Q}_t^T\boldsymbol{\nu}\right|-|Z_t|\right)\end{aligned}$$

$$+\sum_{t=m_n}^{n-m_n}\left(\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)-a_n^{-1}\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}\right|-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)\right|\right)$$
$$-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right|\sum_{t=m_n}^{n-m_n}\left(\left|Z_t-a_n^{-1}\mathbf{Q}_t^T\boldsymbol{\nu}\right|-|Z_t|\right).$$

Since $a_n = n^{1/\alpha}g_1(n)$ for some slowly varying function g_1, it is easily seen that each of the first four sums on the right-hand side converges to zero almost surely. Hence

$$\begin{aligned}\left|S_n^{\dagger}(\boldsymbol{\nu})-S_n^{\ddagger}(\boldsymbol{\nu})\right| &\sim \left|\sum_{t=m_n}^{n-m_n}\left(\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)-a_n^{-1}\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}\right|-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}z_t(\boldsymbol{\kappa}_0)\right|\right)\right.\\ &\qquad\left.-\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right|\sum_{t=m_n}^{n-m_n}\left(\left|Z_t-a_n^{-1}\mathbf{Q}_t^T\boldsymbol{\nu}\right|-|Z_t|\right)\right|\\ &\le 2\left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right|\sum_{t=m_n}^{n-m_n}\left|z_t(\boldsymbol{\kappa}_0)-Z_t\right|+a_n^{-1}\sum_{t=m_n}^{n-m_n}\left|\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}-\frac{\theta_{0s}^*}{\phi_{0s'}^*}\mathbf{Q}_t^T\boldsymbol{\nu}\right|\\ &\equiv I+II,\end{aligned}$$

where the symbol "$\sim$" means that the difference between the two sides goes to zero in probability as $n\to\infty$.

Now, there exists a constant $u\in(0,1)$ such that

$$|V_t^*|\le C\sum_{i=0}^{\infty}u^i|X_{t-i}| \quad \text{and} \quad |V_t^*(\boldsymbol{\kappa}_0)-V_t^*|\le C\sum_{i=0}^{\infty}u^{i+t}|X_{-i}|$$

for all t (see Davis (1996, p.92)). On the other hand, since

$$\begin{aligned} Z_t &= \sum_{j=s}^{\infty}\alpha_{0j}^*V_{t+j}^*, && \text{for all } t,\\ z_t(\boldsymbol{\kappa}_0) &= \sum_{j=s}^{n+p-t}\alpha_{0j}^*V_{t+j}^*(\boldsymbol{\kappa}_0), && \text{for } t=-s+1,\ldots,n+p-s,\end{aligned}$$

we have, for $t=m_n,\ldots,n-m_n$,

$$\begin{aligned}|z_t(\boldsymbol{\kappa}_0)-Z_t| &\le \sum_{j=s}^{n+p-t}\left|\alpha_{0j}^*\right|\left|V_{t+j}^*(\boldsymbol{\kappa}_0)-V_{t+j}^*\right|+\sum_{j=n+p-t+1}^{\infty}\left|\alpha_{0j}^*\right|\left|V_{t+j}^*\right|\\ &\equiv A_1+A_2.\end{aligned}$$

Note that $|\alpha_{0j}^*| \leq Cu_1^j$ for some constant $u_1 \in (0,1)$. Hence,

$$A_1 \leq C \sum_{j=s}^{n+p-t} \left|\alpha_{0j}^*\right| \sum_{i=0}^{\infty} u^{i+j+t} \left|X_{-i}\right| \leq C \sum_{i=0}^{\infty} u^{i+t} \left(\sum_{j=s}^{n+p-t} (u_1 u)^j \right) \left|X_{-i}\right|$$
$$\leq C \sum_{i=0}^{\infty} u^{i+t} \left|X_{-i}\right|.$$

Moreover,

$$A_2 \leq C \sum_{j=n+p-t+1}^{\infty} \left|\alpha_{0j}^*\right| \sum_{i=0}^{\infty} u^i \left|X_{t+j-i}\right|$$
$$\leq C \sum_{j=n+p+1}^{\infty} u u_1^{j-t} \left|X_j\right| + C \sum_{j=0}^{\infty} u_1^{n+p-t+1} u^{j+2} \left|X_{n+p-j}\right|.$$

Then, the term I converges to zero in probability as $n \to \infty$ because, when $\alpha > 1$,

$$\mathrm{E}\left(\sum_{t=m_n}^{n-m_n} \left|z_t(\boldsymbol{\kappa}_0) - Z_t\right| \right) \leq C \sum_{t=m_n}^{n-m_n} u^t + C \sum_{t=m_n}^{n-m_n} u_1^{n+p+1-t} \to 0,$$

and likewise when $\alpha \leq 1$ we have for $\delta \in (0,\alpha)$,

$$\mathrm{E}\left(\sum_{t=m_n}^{n-m_n} \left|z_t(\boldsymbol{\kappa}_0) - Z_t\right| \right)^{\delta} \leq C \sum_{t=m_n}^{n-m_n} u^{\delta t} + C \sum_{t=m_n}^{n-m_n} u_1^{\delta(n+p+1-t)} \to 0.$$

The convergence of the term II to zero in probability can be shown similarly. Therefore, $|S_n^{\dagger}(\boldsymbol{\nu}) - S_n^{\ddagger}(\boldsymbol{\nu})| \to 0$ in probability; the convergence is in fact uniform on compact sets in $C(\mathbb{R}^{p+q})$ because $S_n^{\dagger}(\boldsymbol{\nu}) - S_n^{\ddagger}(\boldsymbol{\nu})$ is linear in $\boldsymbol{\nu}$.

Lemma 8. *$S_n(\boldsymbol{\nu}) - S_n^{\dagger}(\boldsymbol{\nu}) \to 0$ in probability uniformly on compact sets in $C(\mathbb{R}^{p+q})$.*

Proof. The absolute difference $|S_n(\boldsymbol{\nu}) - S_n^{\dagger}(\boldsymbol{\nu})|$ is bounded by

$$\sum_{t=1}^{n+p-q} \left| \frac{\theta_s^*}{\phi_{s'}^*} z_t\left(\boldsymbol{\kappa}_0 + a_n^{-1}\boldsymbol{\nu}\right) - \left(\frac{\theta_{0s}^*}{\phi_{0s'}^*} z_t(\boldsymbol{\kappa}_0) - a_n^{-1} \mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu} \right) \right|$$
$$= \frac{a_n^{-2}}{2} \sum_{t=1}^{n+p-q} \left|\boldsymbol{\nu}^T \mathbf{H}_t(\boldsymbol{\kappa}_t^*)\boldsymbol{\nu}\right|,$$

where $\mathbf{H}_t(\boldsymbol{\kappa}) = \partial^2((\theta_s^*/\phi_{s'}^*)z_t(\boldsymbol{\kappa}))/\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^T$ and $\boldsymbol{\kappa}_t^*$ is between $\boldsymbol{\kappa}_0$ and $\boldsymbol{\kappa}_0 + a_n^{-1}\boldsymbol{\nu}$. It can be shown that, for all $\boldsymbol{\kappa}$ sufficiently close to $\boldsymbol{\kappa}_0$, the absolute value of each element of $\mathbf{H}_t(\boldsymbol{\kappa})$ is upper bounded by $C\sum_{j=-\infty}^{\infty} u_2^{|j|}|X_j|$ for a constant $u_2 \in (0,1)$. It follows immediately that the right-hand side converges to zero in probability uniformly on compact sets in $C(\mathbb{R}^{p+q})$. This finishes the proof of the lemma.

References

Adler, R. J., Feldman, R. E. and Gallagher, C. (1998). Analysing stable time series. In *A Practical Guide to Heavy Tails* (Edited by R. J. Adler, R. E. Feldman and M. S. Taqqu), 133-158. Birkhäuser, Boston.

Andrews, B., Calder, M. and Davis, R. A. (2009). Maximum likelihood estimation for α-stable autoregressive processes. *Ann. Statist.* **37**, 1946-1982.

Andrews, B., Davis, R. A. and Breidt, F. J. (2006). Maximum likelihood estimation for all-pass time series models. *J. Multivariate Anal.* **97**, 1638-1659.

Andrews, B., Davis, R. A. and Breidt, F. J. (2007). Rank-based estimation for all-pass time series models. *Ann. Statist.* **35**, 844-869.

Billingsley, P. (1999). *Convergence of Probability Measures*. 2nd edition. Wiley, New York.

Breidt, F. J. and Davis, R. A. (1992). Time-reversibility, identifiability and independence of innovations for stationary time series. *J. Time Ser. Anal.* **13**, 377-390.

Breidt, F. J., Davis, R. A. and Trindade, A. A. (2001). Least absolute deviation estimation for all-pass time series models. *Ann. Statist.* **29**, 919-946.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd edition. Springer-Verlag, New York.

Calder, M. (1998). Parameter estimation for noncausal and heavy tailed autoregressive processes. Ph.D. Dissertation, Department of Statistics, Colorado State University, Fort Collins, CO.

Chien, H.-M., Yang, H.-L. and Chi, C.-Y. (1997). Parametric cumulant based phase estimation of 1-d and 2-d nonminimum phase systems by allpass filtering. *IEEE Trans. Signal Process* **45**, 1742-1762.

Cline, D. (1983). Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data. Ph.D. Dissertation, Department of Statistics, Colorado State University, Fort Collins, CO.

Davis, R. A. (1996). Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Process. Appl.* **63**, 75-95.

Davis, R. A. and Dunsmuir, W. T. M. (1997). Least absolute deviation estimation for regression with ARMA errors. *J. Theoret. Probab.* **10**, 481-497.

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

Davis, R. A. and Wu, W. (1997). Bootstrapping M-estimates in regression and autoregression with infinite variance. *Statist. Sinica* **7**, 1135-1154.

Donoho, D. (1981). On minimum entropy deconvolution. In *Applied Time Series Analysis II* (Edited by D. F. Findley), 565-608. Academic Press, New York.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press.

Horvath, L. and Liese, F. (2004). L_p-estimators in ARCH models. *J. Statist. Plann. Inference* **119**, 227-309.

Huang, J. and Pawitan, Y. (1999). Consistent estimation for non-Gaussian non-causal autoregressive processes. *J. Time Ser. Anal.* **20**, 417-423.

Huang, J. and Pawitan, Y. (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scand. J. Statist.* **27**, 689-702.

Lii, K. S. and Rosenblatt, M. (1996). Maximum likelihood estimation for nonGaussian nonminimum phase ARMA sequences. *Statist. Sinica* **6**, 1-22.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J. Roy. Statist. Soc. Ser. B* **67**, 381-393.

Pan, J., Wang, H. and Yao, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory* **23**, 852-879.

Rosenblatt, M. (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields.* Springer-Verlag, New York.

Wiggins, R. A. (1978). Minimum entropy deconvolution. *Geoexploration* **16**, 21-35.

Wu, R. and Davis, R. A. (2010). Least absolute deviation estimation for general autoregressive moving average time series models. *J. Time Ser.Anal.* **31**, 98-112.

Department of Statistics and Computer Information Systems, Baruch College, The City University of New York, New York, New York 10010, U.S.A.

E-mail: rongning.wu@baruch.cuny.edu

(Received April 2009; accepted October 2009)