Model Identification for Infinite Variance Autoregressive Processes

Beth Andrews*
Northwestern University

Richard A. Davis
Columbia University

June 17, 2011

Abstract

We consider model identification for infinite variance autoregressive time series processes. It is shown that a consistent estimate of autoregressive model order can be obtained by minimizing Akaike's information criterion, and we use all-pass models to identify noncausal autoregressive processes and estimate the order of noncausality (the number of roots of the autoregressive polynomial inside the unit circle in the complex plane). We examine the performance of the order selection procedures for finite samples via simulation, and use the techniques to fit a noncausal autoregressive model to stock market trading volume data.

* *Corresponding author.* Department of Statistics, Northwestern University, 2006 Sheridan Road, Evanston, IL 60208, USA. Telephone: 1 847 467 4533. E-mail: bandrews@northwestern.edu.

JEL classification codes. C13, C22.

Keywords. Akaike's information criterion, all-pass models, autoregressive processes, infinite variance, noncausal.

1 Introduction

The autoregressive (AR) model is often used to describe observed, heavy-tailed time series processes which appear to have infinite variance. Note that, in the infinite variance case, causal (all roots of the AR polynomial are outside the unit circle in the complex plane) and noncausal AR processes are distinguishable, and can be used to describe different types of time series behavior. Specific applications for causal, heavy-tailed AR models include network interarrival times (Resnick, 1997), sea surface temperatures (Gallagher, 2001), and stock market log-returns (Ling, 2005), while noncausal models have appeared, for example, for modeling trading volume data (Andrews et al., 2009) and in deconvolution problems (Blass and Halsey, 1981; Donoho, 1981; Scargle, 1981). Since every Gaussian AR process has a causal representation, causal and noncausal models cannot be distinguished using autocorrelations (as seen in Davis and Resnick, 1985a, the autocorrelation function is consistently estimated in the infinite variance case). As a result, while traditional second-order moment techniques, such as least squares and Yule-Walker estimation, can be used for infinite variance AR estimation when the model is known to be causal (Davis and Resnick, 1985a), second-order methods are not sufficient for AR model identification in the general infinite variance setting where causality does not necessarily hold. An alternative to using second-order moment techniques is to consider modeling the AR processes as α-stable, since the non-Gaussian stable distributions are a large class of infinite variance distributions which can be asymmetric and have varying degrees of tail heaviness, and use stable maximum likelihood (ML) for AR model selection and parameter estimation. Properties of stable ML estimators and bootstrap confidence intervals for the AR parameter values are developed in Andrews et al. (2009). However, given no prior model information, one may need to maximize the stable likelihood function for different AR model orders and various configurations of roots inside and outside the unit circle, which can be computationally prohibitive for large order models. Therefore, in this paper, we develop a "toolbox" for model identification in the case of an infinite variance AR process. In particular, we show that a consistent estimate of AR model order can be obtained by minimizing Akaike's information criterion (AIC), and we use all-pass models to identify noncausal AR processes and estimate the order of noncausality (i.e., the number of roots inside the unit circle). Once an appropriate AR model has been identified, ML can be used to estimate the parameter values.

For infinite variance, causal AR processes with noise distributions in the domain of attraction of a non-Gaussian stable law, the AIC statistic, computed using Gaussian likelihood, is a consistent AR order selection criterion (Knight, 1989). In this paper, we show that, even though the Yule-Walker method cannot be used to estimate the true AR parameters when the model is not necessarily causal, it can be used to consistently estimate a causal AR model with the same number of parameters and all-pass innovations. We can, therefore, extend results in Knight (1989) and show that minimizing the Gaussian-based AIC statistic is a consistent order selection procedure for noncausal, infinite variance AR processes. In contrast, for Gaussian and other finite variance AR processes, AIC is not a consistent order selection criterion (Shibata, 1976; Hannan, 1980).

All-pass models are autoregressive-moving average (ARMA) models in which the roots of the AR polynomial are reciprocals of roots of the MA polynomial and vice versa. These models generate series that are dependent in the non-Gaussian case. When a noncausal AR process is modeled as causal, the innovations follow an all-pass model of order s, where s is the number of roots of the true AR polynomial inside the unit circle. Consequently, by identifying the all-pass order of the innovations, one can determine the order of noncausality for the AR process. In addition, a preliminary AR model estimate can be obtained from fitted causal AR and all-pass models. While all-pass parameter estimation has already been considered in the literature for finite variance processes (Giannakis and Swami, 1990, Chi and Kung, 1995, Chien et al., 1997, cumulant-based estimation using cumulants of order greater than two; Breidt et al., 2001, least absolute deviations estimation; Andrews et al., 2006, MLE; Andrews et al., 2007, rank-based estimation), the infinite variance case has yet to be addressed. In this paper, we focus on ML estimation for AR processes with non-Gaussian, stable noise, and give the limiting distribution for estimators of the causal AR parameters, the all-pass parameters, and parameters of the noise distribution. The ML estimators of the causal AR parameters have a faster rate of convergence than the Yule-Walker estimators, and we show that the causal AR and all-pass ML estimators converge in distribution to the maximizer of a random function. The form of this limiting distribution is intractable, but the bootstrap procedure can be used to examine the shape of the distribution and obtain confidence intervals for the parameter values. Confidence intervals for the all-pass parameters can be used to identify an appropriate all-pass model order, which equals the order of noncausality for the AR process. We show the bootstrap is asymptotically valid under general conditions.

ML estimators of parameters of the noise distribution are asymptotically independent of the AR and all-pass estimators, and have a multivariate normal limiting distribution.

Heavy-tailed AR processes with infinite variance are discussed in Section 2, and we give steps that can be taken in practice for AR model identification. In Section 3.1, we look at limiting behavior of the Yule-Walker estimators and show that a consistent estimate of AR model order can be obtained by minimizing AIC. In Section 3.2, we give limits for sample correlations of the Yule-Walker residuals, and absolute values and squares of the residuals. The sample correlations of the residuals converge in probability to zero but, in the case of a noncausal AR process, sample correlations for absolute values and squares have nonzero limits. Hence, these sample correlations can be used to detect all-pass dependence in the Yule-Walker residuals and, consequently, to identify noncausal AR processes. In Section 3.3, we consider simultaneous stable ML estimation for the causal AR parameters, all-pass model parameters, and parameters of the noise distribution, and we develop bootstrap confidence intervals which can be used for all-pass order selection. Proofs of the lemmas used to establish results of Sections 3.1–3.3 can be found in the Appendix. In Section 4.1, we examine the performance of the order selection procedures for finite samples via simulation and, in Section 4.2, the model identification techniques are used to fit a noncausal AR model to the natural logarithms of volumes of Wal-Mart stock traded daily on the New York Stock Exchange, a series also modeled as noncausal AR in Andrews et al. (2009).

2 Preliminaries

Let $\{X_t\}$ be the AR process which satisfies the difference equations

$$\phi_0(B)X_t = Z_t, \tag{1}$$

where the AR polynomial $\phi_0(z) := 1 - \phi_{01}z - \cdots - \phi_{0p_0}z^{p_0} \neq 0$ for $|z| = 1$, B is the backshift operator ($B^k X_t = X_{t-k}$, $k = 0, \pm 1, \pm 2, \ldots$), and $\{Z_t\}$ is a sequence of independent and identically distributed (iid) random variables. We will assume for now that the distribution for Z_t is in the domain of attraction of a non-Gaussian stable law with exponent $\alpha_0 \in (0, 2)$, which is less stringent than the assumption that Z_t is α_0-stable. Hence, $\mathrm{P}(|Z_t| > x) = x^{-\alpha_0}L(x)$ for some function $L(\cdot)$ which is slowly varying at ∞ (i.e.,

$\lim_{x\to\infty} L(sx)/L(x) = 1\ \forall s > 0$), and $\lim_{x\to\infty} \mathrm{P}(Z_t > x)/\mathrm{P}(|Z_t| > x) = p$ for some $p \in [0,1]$ (Feller, 1971, page 312). It follows that $\mathrm{E}|Z_1|^\delta < \infty$ for all $\delta \in [0, \alpha_0)$ and $\mathrm{E}|Z_1|^\delta = \infty$ for all $\delta > \alpha_0$, and so the noise process $\{Z_t\}$ has infinite variance. We further suppose that $\phi_{0p_0} \neq 0$, so p_0 represents the AR model order.

Since $\{Z_t\}$ is non-Gaussian, there are no alternative AR representations for $\{X_t\}$ with iid noise (Breidt and Davis, 1992). In addition, because $\phi_0(z) \neq 0$ for $|z| = 1$, the Laurent series expansion for $1/\phi_0(z)$, $1/\phi_0(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$, exists on some annulus $\{z : a^{-1} < |z| < a\}$, $a > 1$, and the unique strictly stationary solution to (1) is given by $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, where the ψ_j's are geometrically decaying as $j \to \pm\infty$ (Brockwell and Davis, 1991, Chapter 3). If $\phi_0(z) \neq 0$ for $|z| \leq 1$, then $\psi_j = 0$ for $j < 0$, and so $\{X_t\}$ is said to be *causal* since $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, a function of only the past and present $\{Z_t\}$. On the other hand, if $\phi_0(z) \neq 0$ for $|z| \geq 1$, then $X_t = \sum_{j=0}^{\infty} \psi_{-j} Z_{t+j}$ and $\{X_t\}$ is a *purely noncausal* process. In the purely noncausal case, the coefficients $\{\psi_j\}$ satisfy $(1 - \phi_{01} z - \cdots - \phi_{0p_0} z^{p_0})(\psi_0 + \psi_{-1} z^{-1} + \cdots) = 1$, which implies that $\psi_0 = \psi_{-1} = \cdots = \psi_{1-p_0} = 0$ and $\psi_{-p_0} = -\phi_{0p_0}^{-1}$. From Cline (1983, page 12), $\lim_{x\to\infty} \mathrm{P}(|X_t| > x)/\mathrm{P}(|Z_t| > x) = \sum_{j=-\infty}^{\infty} |\psi_j|^{\alpha_0}$, and so it is also the case that $\mathrm{E}|X_t|^\delta < \infty$ for $\delta \in [0, \alpha_0)$ and $\mathrm{E}|X_t|^\delta = \infty$ for $\delta > \alpha_0$.

Let $r_0 \geq 0$ represent the number of roots of the AR polynomial $\phi_0(z) = 1 - \phi_{01} z - \cdots - \phi_{0p_0} z^{p_0}$ that lie outside the unit circle in the complex plane, and let $s_0 \geq 0$ represent the number of roots of $\phi_0(z)$ inside the unit circle. Since $\phi_0(z) \neq 0$ for $|z| = 1$, it must be the case that $r_0 + s_0 = p_0$, and there exist a causal AR polynomial $\theta_0^\dagger(z)$ of order r_0 and a purely noncausal polynomial $\theta_0^*(z)$ of order s_0 for which $\phi_0(z) = \theta_0^\dagger(z)\theta_0^*(z)$. Now suppose $\theta_0^c(z) := 1 - \theta_{01} z - \cdots - \theta_{0s_0} z^{s_0}$ denotes the causal s_0th-order AR polynomial whose roots are the reciprocals of the roots of the noncausal polynomial $\theta_0^*(z)$, so $\theta_0^*(z) = -\theta_{0s_0}^{-1} z^{s_0} \theta_0^c(z^{-1})$ (if $s_0 = 0$, $\theta_{00} := -1$). In addition, we let $\eta_0(z) := 1 - \eta_{01} z - \cdots - \eta_{0p_0} z^{p_0}$ denote the causal p_0th-order AR polynomial $\theta_0^\dagger(z)\theta_0^c(z)$. Since $\phi_0(z) = \theta_0^\dagger(z)\theta_0^c(z)\{\theta_0^*(z)/\theta_0^c(z)\} = \eta_0(z)\{(-\theta_{0s_0}^{-1}) z^{s_0} \theta_0^c(z^{-1})/\theta_0^c(z)\}$, the AR model equation (1) can be expressed as

$$\eta_0(B) \frac{-\theta_{0s_0}^{-1} B^{s_0} \theta_0^c(B^{-1})}{\theta_0^c(B)} X_t = Z_t \tag{2}$$

or

$$\eta_0(B) X_t = U_t, \quad \text{with} \quad U_t = \frac{\theta_0^c(B)}{-\theta_{0s_0}^{-1} B^{s_0} \theta_0^c(B^{-1})} Z_t. \tag{3}$$

Since the series $\{U_t\}$ is an ARMA process for which all s_0 roots of the AR polynomial are reciprocals of roots of the MA polynomial and vice versa, it is an all-pass process. It, therefore, follows that $\{X_t\}$ has a p_0th-order causal AR representation with innovations satisfying an all-pass model of order $s_0 \geq 0$.

Non-Gaussian all-pass processes of order greater than zero are known to be dependent, but, when the second-order moments are finite, all-pass series are uncorrelated (Breidt et al., 2001; Andrews et al., 2006; Andrews et al., 2007). Correlations for $\{U_t\}$ do not exist in this infinite variance case. However, if $U_t = \sum_{j=0}^{\infty} \pi_j Z_{t+j}$ is the infinite-order moving average representation for $\{U_t\}$, then from Davis and Resnick (1985a), the sample correlations

$$\begin{aligned} \tilde{\rho}_U(h) &:= \frac{\sum_{t=h+1}^{n} \left(U_t - [n^{-1}\sum_{t=1}^{n} U_t]\right)\left(U_{t-h} - [n^{-1}\sum_{t=1}^{n} U_t]\right)}{\sum_{t=1}^{n}(U_t - [n^{-1}\sum_{t=1}^{n} U_t])^2} \\ &\xrightarrow{P} \frac{\sum_{j=0}^{\infty} \pi_j \pi_{j+h}}{\sum_{j=0}^{\infty} \pi_j^2} \qquad \text{as } n \to \infty \end{aligned}$$

for any non-negative integer h. Following (3), the coefficients $\{\pi_j\}$ can be obtained from the Laurent series expansion for $\theta_0^c(z)/[-\theta_{0s_0}^{-1} z^{s_0} \theta_0^c(z^{-1})]$. Since $\theta_0^*(z) = -\theta_{0s_0}^{-1} z^{s_0} \theta_0^c(z^{-1}) \neq 0$ for $|z| \geq 1$, the moving average representation for $\{U_t\}$ is purely noncausal and, from Brockwell and Davis (1991, Chapter 3), the values of $\{\pi_j\}$ are geometrically decaying as $j \to \infty$. Note also that $\sum_{j=0}^{\infty} \pi_j \pi_{j+h} = 0$ for all $h > 0$, since, if the iid noise $\{Z_t\}$ were $\mathrm{N}(0, \sigma^2)$ (instead of in the domain of attraction of a non-Gaussian stable law), then the all-pass process $\{U_t\}$ would be uncorrelated with $\mathrm{Cov}\{U_t, U_{t+h}\} = \mathrm{E}\{U_t U_{t+h}\} = \sigma^2 \sum_{j=0}^{\infty} \pi_j \pi_{j+h} = 0$. It follows that $\tilde{\rho}_U(h) \xrightarrow{P} 0$ for $h > 0$, and so the infinite variance all-pass series $\{U_t\}$ in (3) might also be described as "uncorrelated."

Given a realization of length n from (1), $\{X_t\}_{t=1}^{n}$, we recommend the following steps for AR model selection:

1. Estimate AR model order p_0 by minimizing the AIC statistic, computed using Gaussian likelihood, and using Yule-Walker, estimate the parameters of the p_0th-order causal AR model $\eta_0(B)X_t = U_t$.

2. Look at sample correlations of the Yule-Walker residuals, and absolute values and squares of the residuals. If the residuals and their absolute values and squares appear uncorrelated, this suggests $\{U_t\}$ is iid noise, and so $\{X_t\}$ is a causal AR process. On the other hand, if the residuals appear uncorrelated yet dependent (i.e., absolute values and squares appear correlated), this indicates $\{U_t\}$ is an all-pass process of order $s_0 > 0$.

3. If s_0 appears positive, simultaneously estimate the causal AR parameters, the all-pass parameters, and parameters of the iid noise via stable ML, and obtain bootstrap confidence intervals for the all-pass parameter values. The confidence intervals can be used to estimate the all-pass order s_0, which equals the order of noncausality for the AR process $\{X_t\}$.

Once an appropriate AR model order and an appropriate order of noncausality have been identified for $\{X_t\}$, estimates of the parameter values can be obtained via ML.

3 Asymptotic Results

The three steps for AR model identification in Section 2 are discussed in further detail in Sections 3.1–3.3, along with corresponding asymptotic theory. We consider Yule-Walker estimation and AR order selection in Section 3.1 and, in Section 3.2, we show that, for noncausal AR processes, sample correlations for absolute values and squares of the Yule-Walker residuals have nonzero limits, so these sample correlations can be used to identify noncausal series. Results in Sections 3.1 and 3.2 are obtained under the assumption that the distribution for the iid noise $\{Z_t\}$ is in the domain of attraction of a non-Gaussian stable law with exponent $\alpha_0 \in (0, 2)$. In Section 3.3, we consider stable MLE and make the more stringent assumption that the $\{Z_t\}$ are non-Gaussian α_0-stable.

3.1 Yule-Walker Estimation and Autoregressive Order Selection

In this section, we give a limiting result for Yule-Walker estimators of the parameters $\eta_{01}, \ldots, \eta_{0p_0}$ in the causal AR equation $\eta_0(B)X_t = U_t$ in (3), and show that a consistent estimate of the AR model order p_0 can be obtained by minimizing Gaussian-based AIC. From Section 8.1 in Brockwell and Davis (1991), given observed values of $\{X_t\}_{t=1}^n$, for $k \geq \max\{p_0, 1\}$, the Yule-Walker estimate of $\boldsymbol{\eta}_0(k) := (\eta_{01}, \ldots, \eta_{0p_0}, 0, \ldots, 0)' \in \mathbb{R}^k$ is $\hat{\boldsymbol{\eta}}_{YW}(k) = \hat{\mathbf{C}}_k^{-1}\hat{\mathbf{r}}_k$, with $\hat{\mathbf{C}}_k := [\hat{\gamma}(|i - j|)]_{i,j\in\{1,\ldots,k\}}$, $\hat{\mathbf{r}}_k := [\hat{\gamma}(j)]_{j\in\{1,\ldots,k\}}$, $\hat{\gamma}(j) := n^{-1}\sum_{t=j+1}^n (X_t - \overline{X})(X_{t-j} - \overline{X})$, and $\overline{X} := n^{-1}\sum_{t=1}^n X_t$. The following theorem shows that these estimators are $n^{1/2}$-consistent and can converge uniformly over k. When the AR process $\{X_t\}$ is causal (i.e., when $s_0 = 0$), this result holds by Corollary 6 in Knight (1989), but here it is not necessarily the case that $s_0 = 0$. Note that $\|\cdot\|$ represents the Euclidean norm.

Theorem 1. *If $K(n) = O(n^{\delta})$, with $0 \le \delta < \min\{1/2, 1-\alpha_0/2\}$, then*

$$\sqrt{n} \max_{p_0 \le k \le K(n)} \|\hat{\boldsymbol{\eta}}_{YW}(k) - \boldsymbol{\eta}_0(k)\| \xrightarrow{P} 0 \tag{4}$$

as $n \to \infty$.

Proof. The proof of this result is similar to that of Corollary 6 in Knight (1989). For $k \ge \max\{p_0, 1\}$, let $\hat{\boldsymbol{\eta}}_{LS}(k) = \tilde{\mathbf{C}}_k^{-1}\tilde{\mathbf{r}}_k$, with $\tilde{\mathbf{C}}_k := [\sum_{t=k+1}^{n}(X_{t-i} - \overline{X})(X_{t-j} - \overline{X})]_{i,j \in \{1,\ldots,k\}}$ and $\tilde{\mathbf{r}}_k := [\sum_{t=k+1}^{n}(X_t - \overline{X})(X_{t-j} - \overline{X})]_{j \in \{1,\ldots,k\}}$, denote the least-squares estimate of $\boldsymbol{\eta}_0(k)$. From (3), $X_t = \eta_{01}X_{t-1} + \cdots + \eta_{0p_0}X_{t-p_0} + U_t$ for all t, so

$$\tilde{\mathbf{C}}_k\left(\hat{\boldsymbol{\eta}}_{LS}(k) - \boldsymbol{\eta}_0(k)\right) = \left[\sum_{t=k+1}^{n}\left(U_t - \overline{X}\left(1 - \sum_{i=1}^{p_0}\eta_{0i}\right)\right)\left(X_{t-j} - \overline{X}\right)\right]_{j \in \{1,\ldots,k\}}$$

and, by Lemmas 1–3 in the Appendix, for some sufficiently large $\kappa < 2/\alpha_0$, $\max_{p_0 \le k \le K(n)} n^{1/2-\kappa}\|\tilde{\mathbf{C}}_k(\hat{\boldsymbol{\eta}}_{LS}(k) - \boldsymbol{\eta}_0(k))\| \xrightarrow{P} 0$. Using an argument similar to the one used for the proof of Theorem 5a(ii) in Knight (1989), it can be shown that $\min_{p_0 \le k \le K(n)} \min_{\|\mathbf{v}\|=1} n^{-\kappa}\mathbf{v}'\tilde{\mathbf{C}}_k\mathbf{v} \xrightarrow{P} \infty$, and therefore $\sqrt{n}\max_{p_0 \le k \le K(n)} \|\hat{\boldsymbol{\eta}}_{LS}(k) - \boldsymbol{\eta}_0(k)\| \xrightarrow{P} 0$. Since, by the Corollary in Davis and Resnick (1985a, page 193), $n^{1-\kappa}\sum_{t=1}^{K(n)}(X_t - \overline{X})^2 \xrightarrow{P} 0$ for large $\kappa < 2/\alpha_0$, it follows from the proof of Theorem 5(b) in Knight (1989) that $\sqrt{n}\max_{p_0 \le k \le K(n)} \|\hat{\boldsymbol{\eta}}_{YW}(k) - \hat{\boldsymbol{\eta}}_{LS}(k)\| \xrightarrow{P} 0$, and so the result of this theorem holds. $\square$

When $\alpha_0 \le 1$, and when $\alpha_0 > 1$ and $\mathrm{E}X_t = 0$, it can be shown that (4) holds for Yule-Walker estimators computed using the unadjusted sample covariances $\tilde{\gamma}(j) = n^{-1}\sum_{t=j+1}^{n} X_t X_{t-j}$, instead of the mean-corrected sample covariances $\hat{\gamma}(j)$. The proof is similar to that of Theorem 1. It is also possible to obtain the rate of convergence of $\hat{\boldsymbol{\eta}}_{YW}(k)$ for fixed k, since, following Davis and Resnick (1986, Section 5.4), $\hat{\boldsymbol{\eta}}_{YW}(k) - \boldsymbol{\eta}_0(k) = \mathbf{D}_k\left(\hat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\right) + o_p\left(\hat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\right)$, where $\mathbf{D}_k$ is a nonzero, $k \times k$ matrix of constants, $\hat{\boldsymbol{\rho}}_k := [\hat{\gamma}(j)/\hat{\gamma}(0)]_{j \in \{1,\ldots,k\}}$, $\boldsymbol{\rho}_k := [(\sum_{\ell=-\infty}^{\infty}\psi_\ell\psi_{\ell-j})/(\sum_{\ell=-\infty}^{\infty}\psi_\ell^2)]_{j \in \{1,\ldots,k\}}$, and the coefficients $\{\psi_\ell\}$ are from the expansion $X_t = \sum_{j=-\infty}^{\infty}\psi_j Z_{t-j}$. By Davis and Resnick (1985a), $\hat{\boldsymbol{\rho}}_k \xrightarrow{P} \boldsymbol{\rho}_k$, so $\hat{\boldsymbol{\eta}}_{YW}(k)$ has the same rate of convergence as the vector of sample correlations $\hat{\boldsymbol{\rho}}_k$. For fixed k, Davis and Resnick (1985b) give the limiting distribution for $\hat{\boldsymbol{\rho}}_k$ when $\mathrm{E}|Z_t|^{\alpha_0} < \infty$, and the case when $\mathrm{E}|Z_t|^{\alpha_0} = \infty$ is considered in Davis and Resnick (1986). For instance, when Z_t has an α_0-stable distribution (in which case $\mathrm{E}|Z_t|^{\alpha_0} = \infty$), it follows that $n(\hat{\boldsymbol{\eta}}_{YW}(k) - \boldsymbol{\eta}_0(k)) = O_p(1)$ when $\alpha_0 < 1$, $(n/\ln n)(\hat{\boldsymbol{\eta}}_{YW}(k) - \boldsymbol{\eta}_0(k)) = O_p(1)$ when $\alpha_0 = 1$ and the distribution for Z_t is symmetric, and $(n/\ln n)^{1/\alpha_0}(\hat{\boldsymbol{\eta}}_{YW}(k) - \boldsymbol{\eta}_0(k)) = O_p(1)$ when $\alpha_0 > 1$.

Following Knight (1989), for integer-valued $k \geq 0$, we compute Gaussian-based AIC via $\text{AIC}(k) = n \ln \hat{\sigma}_k^2 + 2k$, where $\hat{\sigma}_k^2 = \hat{\gamma}(0) - \hat{\mathbf{r}}_k' \hat{\boldsymbol{\eta}}_{YW}(k) I\{k > 0\}$ is the Yule-Walker estimate of innovations variance and $I\{\cdot\}$ represents the indicator function. By Theorem 2, the minimum AIC estimate of AR model order is consistent for p_0. Note that this estimate, $\hat{p}$, is obtained by minimizing AIC over the integers $0, \ldots, K(n)$, where it is possible for $K(n) \to \infty$ as $n \to \infty$.

Theorem 2. *If $K(n) = O(n^\delta)$, with $0 \leq \delta < \min\{1/2, 1 - \alpha_0/2\}$, and $\hat{p} = \text{arg min}_{0 \leq k \leq K(n)} AIC(k)$, then $\hat{p} \xrightarrow{P} p_0$ as $n \to \infty$.*

Proof. This result follows directly from Theorem 1 and the proof of Theorem 7 in Knight (1989). See Knight (1989, pages 835–836) for details. $\square$

Theorem 1 and results in Knight (1989) can be used to show, more generally, that $\tilde{p} := \text{arg min}_{0 \leq k \leq K(n)} (n \ln \hat{\sigma}_k^2 + h_n k)$ is consistent for p_0 if $h_n/n \to 0$ as $n \to \infty$ and $h_n \geq$ *constant* > 0 for all large n. When $h_n = 2$, $\tilde{p}$ equals $\hat{p}$ and, when $h_n = \ln n$, $\tilde{p}$ minimizes the Bayesian information criterion (BIC). We, however, recommend that in practice one use $\hat{p}$ instead of $\tilde{p}$ with $h_n \to \infty$, to reduce the risk of underestimating p_0 for finite samples. In addition, note that one can also use the sample partial autocorrelation function to identify an appropriate AR model order, since, in the AR case, the underlying theoretical partial autocorrelations are zero at lags greater than p_0 (see, for example, Brockwell and Davis, 1991, page 100). However, in the infinite variance case, quantiles of the limiting distribution for sample partial autocorrelations cannot be computed theoretically, but only via simulation or numerical integration (Adler et al., 1998). Hence, minimizing AIC can be a simpler way to estimate p_0.

3.2 Identifying a Noncausal Autoregressive Process

If the AR process $\{X_t\}$ is causal (i.e., $s_0 = 0$), then the causal polynomial $\eta_0(z)$ in (3) equals the AR polynomial $\phi_0(z)$ in (1), and the uncorrelated all-pass series $\{U_t\}$ is equivalent to the iid noise process $\{Z_t\}$. Hence, in the causal case, estimates of the AR model parameters $\boldsymbol{\phi}_0 := (\phi_{01}, \ldots, \phi_{0p_0})'$ can be obtained using Yule-Walker estimation, and the corresponding residuals $\hat{U}_t := X_t - \hat{\eta}_{1,YW}(\hat{p}) X_{t-1} - \cdots - \hat{\eta}_{\hat{p},YW}(\hat{p}) X_{t-\hat{p}}$, $t = \hat{p} + 1, \ldots, n$, appear iid. On the other hand, if $\{X_t\}$ is noncausal, $\{U_t\}$ is an all-pass process of order

$s_0 > 0$ and, therefore, the Yule-Walker residuals $\{\hat{U}_t\}_{t=\hat{p}+1}^n$ appear uncorrelated but dependent. In this section, we give limits in probability for sample correlations of the Yule-Walker residuals and absolute values and squares of the residuals. We show that, while the residuals are in general uncorrelated, dependence in the noncausal case can be detected in practice by looking at correlations of the absolute values and squares.

Let $\hat{\rho}_U(\cdot)$ denote the sample autocorrelation function for the Yule-Walker residuals $\{\hat{U}_t\}_{t=\hat{p}+1}^n$. So, for any non-negative integer h,

$$\hat{\rho}_U(h) = \frac{\sum_{t=h+\hat{p}+1}^n (\hat{U}_t - \overline{U}_n)(\hat{U}_{t-h} - \overline{U}_n)}{\sum_{t=\hat{p}+1}^n (\hat{U}_t - \overline{U}_n)^2},$$

with $\overline{U}_n := (n-\hat{p})^{-1} \sum_{t=\hat{p}+1}^n \hat{U}_t$. And let $\hat{\rho}_{U^\dagger}(\cdot)$ and $\hat{\rho}_{U^*}(\cdot)$ represent the sample autocorrelation functions for $\{U_t^\dagger\}_{t=\hat{p}+1}^n := \{|\hat{U}_t - \overline{U}_n|\}_{t=\hat{p}+1}^n$ and $\{U_t^*\}_{t=\hat{p}+1}^n := \{(\hat{U}_t - \overline{U}_n)^2\}_{t=\hat{p}+1}^n$, the absolute values and squares of the mean-corrected residuals. Limits for these sample correlations are given in the following theorem. Recall, from Section 2, that $\{\pi_j\}_{j=0}^\infty$ are the coefficients in the infinite-order moving average representation $U_t = \sum_{j=0}^\infty \pi_j Z_{t+j}$ for $\{U_t\}$, and that $\sum_{j=0}^\infty \pi_j \pi_{j+h} = 0$ for all $h > 0$.

Theorem 3. *For any positive integer h, as $n \to \infty$,*

(i) $\hat{\rho}_U(h) \xrightarrow{P} 0$,

(ii) $\hat{\rho}_{U^\dagger}(h) \xrightarrow{P} \sum_{j=0}^\infty |\pi_j \pi_{j+h}| / \sum_{j=0}^\infty \pi_j^2$, *and*

(iii) $\hat{\rho}_{U^*}(h) \xrightarrow{P} \sum_{j=0}^\infty \pi_j^2 \pi_{j+h}^2 / \sum_{j=0}^\infty \pi_j^4$.

Proof. (i) For any $\epsilon > 0$, $\mathrm{P}(|\hat{\rho}_U(h)| > \epsilon) \leq \mathrm{P}(\{|\hat{\rho}_U(h)| > \epsilon\} \cap \{\hat{p} = p_0\}) + \mathrm{P}(\hat{p} \neq p_0)$. Since, by Theorem 2, $\mathrm{P}(\hat{p} \neq p_0) \to 0$, we can, therefore, establish result (i) by showing that

$$\frac{\sum_{t=h+p_0+1}^n [U_t(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n(\hat{\boldsymbol{\eta}}_{YW}(p_0))][U_{t-h}(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n(\hat{\boldsymbol{\eta}}_{YW}(p_0))]}{\sum_{t=p_0+1}^n [U_t(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n(\hat{\boldsymbol{\eta}}_{YW}(p_0))]^2} \xrightarrow{P} 0, \tag{5}$$

where, for $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{p_0})' \in \mathrm{I\!R}^{p_0}$,

$$U_t(\boldsymbol{\eta}) := X_t - \eta_1 X_{t-1} - \cdots - \eta_{p_0} X_{t-p_0} \tag{6}$$

and $\overline{U}_n(\boldsymbol{\eta}) := (n-p_0)^{-1} \sum_{t=p_0+1}^n U_t(\boldsymbol{\eta})$. Note that $U_t = U_t(\boldsymbol{\eta}_0(p_0))$, and let $a_n = \inf\{x : \mathrm{P}(|Z_t| > x) \leq$

$n^{-1}\}$. For any non-negative integer ℓ, we consider

$$a_n^{-2} \sum_{t=\ell+p_0+1}^{n} \left\{ \left[U_t(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right] \left[U_{t-\ell}(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right] - U_t U_{t-\ell} \right\} \tag{7}$$

$$= a_n^{-2} \sum_{t=\ell+p_0+1}^{n} \left[U_t(\hat{\boldsymbol{\eta}}_{YW}(p_0)) U_{t-\ell}(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - U_t(\boldsymbol{\eta}_0(p_0)) U_{t-\ell}(\boldsymbol{\eta}_0(p_0))\right] \tag{8}$$

$$- a_n^{-2} \overline{U}_n(\hat{\boldsymbol{\eta}}_{YW}(p_0)) \sum_{t=\ell+p_0+1}^{n} \left[U_t(\hat{\boldsymbol{\eta}}_{YW}(p_0)) + U_{t-\ell}(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right] \tag{9}$$

$$+ a_n^{-2} (n - \ell - p_0) \left[\overline{U}_n(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right]^2. \tag{10}$$

Using the mean value theorem, (8) equals

$$(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \boldsymbol{\eta}_0(p_0))' \; a_n^{-2} \sum_{t=\ell+p_0+1}^{n} \frac{\partial [U_t(\boldsymbol{\eta}_n^*(p_0)) U_{t-\ell}(\boldsymbol{\eta}_n^*(p_0))]}{\partial \boldsymbol{\eta}},$$

where $\boldsymbol{\eta}_n^*(p_0)$ lies between $\hat{\boldsymbol{\eta}}_{YW}(p_0)$ and $\boldsymbol{\eta}_0(p_0)$ and, following (6),

$$\frac{\partial [U_t(\boldsymbol{\eta}) U_{t-\ell}(\boldsymbol{\eta})]}{\partial \eta_j} = -X_{t-j}(X_{t-\ell} - \eta_1 X_{t-\ell-1} - \cdots - \eta_{p_0} X_{t-\ell-p_0}) - X_{t-\ell-j}(X_t - \eta_1 X_{t-1} - \cdots - \eta_{p_0} X_{t-p_0})$$

for $j \in \{1, \ldots, p_0\}$. By Theorem 1, $\hat{\boldsymbol{\eta}}_{YW}(p_0) \xrightarrow{P} \boldsymbol{\eta}_0(p_0)$ and, by Theorem 4.2 in Davis and Resnick (1985a), $a_n^{-2} \sum_{\ell+p_0+1}^{n} X_{t-j} X_{t-k} = O_p(1)$ for any integers j, k. It, therefore, follows that equation (8) is $o_p(1)$. From the proof of the Corollary in Davis and Resnick (1985a, page 193), $a_n^{-1} \, n^{-1/2} \sum_{t=1}^{n} |X_t| \xrightarrow{P} 0$, so

$$\sum_{t=p_0+1}^{n} U_t(\hat{\boldsymbol{\eta}}_{YW}(p_0)) = \sum_{t=p_0+1}^{n} [X_t - \hat{\eta}_{1,YW}(p_0) X_{t-1} - \cdots - \hat{\eta}_{p_0,YW}(p_0) X_{t-p_0}] = o_p(a_n n^{1/2}),$$

and hence (9) and (10) are $o_p(1)$. Therefore, equation (7) is $o_p(1)$ for any non-negative integer ℓ. By Theorem 4.2 in Davis and Resnick (1985a), we also have $a_n^{-2}(\sum_{t=p_0+1}^{n} U_t^2, \sum_{t=h+p_0+1}^{n} U_t U_{t-h}) \xrightarrow{\mathcal{L}} V(\sum_{j=0}^{\infty} \pi_j^2, \sum_{j=0}^{\infty} \pi_j \pi_{j+h})$, with the random variable $V \in (0, \infty)$ almost surely, and so the left-hand side of (5) converges in probability to $\sum_{j=0}^{\infty} \pi_j \pi_{j+h} / \sum_{j=0}^{\infty} \pi_j^2 = 0$.

(ii) Using a proof similar to that of (i), it can be shown that

$$a_n^{-2} \sum_{t=\ell+p_0+1}^{n} \left\{ \left[U_t^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right] \left[U_{t-\ell}^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right] - |U_t U_{t-\ell}| \right\} \xrightarrow{P} 0$$

for any non-negative integer ℓ, where for $\boldsymbol{\eta} \in \mathbb{R}^{p_0}$, $U_t^\dagger(\boldsymbol{\eta}) := |U_t(\boldsymbol{\eta}) - \overline{U}_n(\boldsymbol{\eta})|$ and $\overline{U}_n^\dagger(\boldsymbol{\eta}) := (n - p_0)^{-1} \sum_{t=p_0+1}^{n} U_t^\dagger(\boldsymbol{\eta})$. In addition, using an argument similar to the proof of Theorem 4.2 in Davis and Resnick (1985a), $a_n^{-2}(\sum_{t=p_0+1}^{n} U_t^2, \sum_{t=h+p_0+1}^{n} |U_t U_{t-h}|) \xrightarrow{\mathcal{L}} V^\dagger(\sum_{j=0}^{\infty} \pi_j^2, \sum_{j=0}^{\infty} |\pi_j \pi_{j+h}|)$, with

$V^\dagger \in (0,\infty)$ almost surely. Hence,

$$\frac{\sum_{t=h+p_0+1}^{n}\left[U_t^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right]\left[U_{t-h}^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right]}{\sum_{t=p_0+1}^{n}\left[U_t^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n^\dagger(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right]^2} \xrightarrow{P} \frac{\sum_{j=0}^{\infty}|\pi_j\pi_{j+h}|}{\sum_{j=0}^{\infty}\pi_j^2},$$

and so the result holds.

(iii) Following the proof of Theorem 4.2 in Davis and Resnick (1985a), it can be shown that, for any integers j, k l, and m, $a_n^{-2}\sum_{t=1}^{n}|X_{t-j}X_{t-k}|$, $a_n^{-3}\sum_{t=1}^{n}|X_{t-j}X_{t-k}X_{t-l}|$, and $a_n^{-4}\sum_{t=1}^{n}|X_{t-j}X_{t-k}X_{t-l}X_{t-m}|$ are all $O_p(1)$. It follows that

$$a_n^{-4}\sum_{t=\ell+p_0+1}^{n}\left\{\left[U_t^*(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n^*(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right]\left[U_{t-\ell}^*(\hat{\boldsymbol{\eta}}_{YW}(p_0)) - \overline{U}_n^*(\hat{\boldsymbol{\eta}}_{YW}(p_0))\right] - U_t^2U_{t-\ell}^2\right\} \xrightarrow{P} 0$$

for any non-negative integer ℓ, where $U_t^*(\boldsymbol{\eta}) := [U_t(\boldsymbol{\eta}) - \overline{U}_n(\boldsymbol{\eta})]^2$ and $\overline{U}_n^*(\boldsymbol{\eta}) := (n-p_0)^{-1}\sum_{t=p_0+1}^{n}U_t^*(\boldsymbol{\eta})$. Using the proof of Theorem 4.2 in Davis and Resnick (1985a), it can also be shown that $a_n^{-4}(\sum_{t=p_0+1}^{n}U_t^4, \sum_{t=h+p_0+1}^{n}U_t^2U_{t-h}^2) \xrightarrow{d} V^*(\sum_{j=0}^{\infty}\pi_j^4, \sum_{j=0}^{\infty}\pi_j^2\pi_{j+h}^2)$, where the random variable $V^* \in (0,\infty)$ almost surely, so the result (iii) holds. $\square$

If the AR process $\{X_t\}$ is causal, and so the all-pass process $\{U_t\}$ is equivalent to $\{Z_t\}$, it must be the case that $\pi_0 = 1$ and $\pi_j = 0$ for $j > 0$. Therefore, when $\{X_t\}$ is causal, the limits in Theorem 3 are all zero for $h > 0$. However, if $\{X_t\}$ is noncausal, $\{U_t\}$ is dependent, which implies that multiple values of $\{\pi_j\}$ are nonzero, and so the limits in (ii) and (iii) must be positive for some $h > 0$. Since the $\{\pi_j\}$ are geometrically decaying as $j \to \infty$, these limits are roughly geometrically decaying as $h \to \infty$. So, in practice, to identify a noncausal AR series, one can look at sample correlations for $\{U_t^\dagger\}_{t=\hat{p}+1}^{n}$ and $\{U_t^*\}_{t=\hat{p}+1}^{n}$, and compare them to confidence bounds for the sample correlations computed under the assumption that $\{U_t\}$ is independent. These confidence bounds could be obtained by generating multiple series containing $n-\hat{p}$ independent values from the empirical distribution of $\{\hat{U}_t\}_{t=\hat{p}+1}^{n}$, and then computing sample correlations for the absolute values and squares of the mean-corrected series. Or, if a stable distribution appears appropriate for the $\{\hat{U}_t\}$, one could model the Yule-Walker residuals as iid stable (ML estimation for the parameters of iid stable random variables is discussed in DuMouchel, 1973), and then simulate sample correlations for absolute values and squares of $n-\hat{p}$ mean-corrected iid stable random variables with the estimated stable parameter values.

Using a proof similar to that of Theorem 3, it can be shown that sample correlations for $\{|\hat{U}_t|\}$ and $\{\hat{U}_t^2\}$ have the same limits as those given in Theorem 3(ii)–(iii) for sample correlations of $\{U_t^\dagger\} = \{|\hat{U}_t - \overline{U}_n|\}$ and

$\{U_t^*\} = \{(\hat{U}_t - \overline{U}_n)^2\}$. We are focusing on absolute values and squares of the mean-corrected residuals since, in practice, for fixed sample size n, dependence in the residuals can often more easily be detected using $\{U_t^\dagger\}$ and $\{U_t^*\}$. For instance, if the observed values of $\{X_t\}_{t=1}^n$ are all positive or all negative, as is the case with the Wal-Mart log-volume series discussed in Section 4.2, then the corresponding Yule-Walker residuals $\{\hat{U}_t\}$ can also be all positive or all negative, which means that the sample correlations for $\{|\hat{U}_t|\}$ and $\{\hat{U}_t\}$ are identical.

3.3 Maximum Likelihood Estimation

In this section, we consider ML parameter estimation. To obtain the limiting distribution for the estimators, we impose further restrictions on the noise, and assume Z_t is non-Gaussian stable with exponent $\alpha_0 \in (0, 2)$, parameter of symmetry $\beta_0 \in (-1, 1)$, scale parameter $\sigma_0 \in (0, \infty)$, and location parameter $\mu_0 \in \mathbb{R}$. When the $\{Z_t\}$ are iid α_0-stable, the AR random variables $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ also have a stable distribution with exponent α_0 (Samorodnitsky and Taqqu, 1994, Properties 1.2.1 and 1.2.3). We use ML to simultaneously estimate the causal AR parameters $\eta_{01}, \ldots, \eta_{0p_0}$, the all-pass model parameters $\theta_{01}, \ldots, \theta_{0s_0}$, and parameters of the stable noise distribution.

The stable characteristic function for Z_t is given by

$$\varphi_0(s) := \mathrm{E}\{\exp(isZ_t)\} = \begin{cases} \exp\left\{-\sigma_0^{\alpha_0}|s|^{\alpha_0}\left[1 + i\beta_0(\text{sign } s)\tan\left(\frac{\pi\alpha_0}{2}\right)\left(|\sigma_0 s|^{1-\alpha_0} - 1\right)\right] + i\mu_0 s\right\}, & \alpha_0 \neq 1, \\ \exp\left\{-\sigma_0|s|\left[1 + i\beta_0\frac{2}{\pi}(\text{sign } s)\ln(\sigma_0|s|)\right] + i\mu_0 s\right\}, & \alpha_0 = 1, \end{cases} \tag{11}$$

and so, if $\boldsymbol{\vartheta}_0 := (\alpha_0, \beta_0, \sigma_0, \mu_0)'$, the density function for the noise can be expressed as $f(z; \boldsymbol{\vartheta}_0) = (2\pi)^{-1}\int_{-\infty}^{\infty} \exp(-izs)\,\varphi_0(s)\,ds$. No general, closed-form expression is known for f, however; although computational formulas exist that can be used to evaluate f (Nolan, 1997; McCulloch, 1998). It can be shown that $f(z; \boldsymbol{\vartheta}_0) = \sigma_0^{-1} f(\sigma_0^{-1}(z - \mu_0); (\alpha_0, \beta_0, 1, 0)')$, $f(z; (\alpha_0, \beta_0, 1, 0)') = f(-z; (\alpha_0, -\beta_0, 1, 0)')$, $f(\cdot; (\alpha_0, \beta_0, 1, 0)')$ is unimodal on $\mathbb{R}$ (Yamazato, 1978), and $f(z; (\alpha, \beta, 1, 0)')$ is infinitely differentiable with respect to (z, α, β) on $\mathbb{R} \times (0, 2) \times (-1, 1)$. There are alternative parameterizations for the stable characteristic function φ_0 (Zolotarev, 1986), but we use (11) so that the noise density function is continuous and differentiable with respect to not only z on $\mathbb{R}$ but also $(\alpha, \beta, \sigma, \mu)'$ on $(0, 2) \times (-1, 1) \times (0, \infty) \times (-\infty, \infty)$. For additional properties of stable distributions/densities see Gnedenko and Kolmogorov (1968), Feller (1971), Zolotarev (1986), and Samorodnitsky and Taqqu (1994).

To obtain the log-likelihood function, we consider model equation (2), which can be expressed as

$$\eta_0(B)\left[-\theta_{0s_0}^{-1}B^{s_0}\theta_0^c(B^{-1})\right]X_t = \theta_0^c(B)Z_t, \tag{12}$$

where $\eta_0(z) = 1 - \eta_{01}z - \cdots - \eta_{0p_0}z^{p_0}$ and $\theta_0^c(z) = 1 - \theta_{01}z - \cdots - \theta_{0s_0}z^{s_0}$. Letting $z_t = -\theta_{0s_0}Z_{t-p_0+s_0}$, which is stable with parameter vector $\boldsymbol{\tau}_0 := (\alpha_0, -(\text{sign }\theta_{0s_0})\beta_0, |\theta_{0s_0}|\sigma_0, -\theta_{0s_0}\mu_0)'$, and rearranging (12), we have the recursion

$$z_t = \theta_{01}z_{t-1} + \cdots + \theta_{0p_0}z_{t-p_0} + \eta_0(B)B^{p_0}\theta_0^c(B^{-1})X_t, \tag{13}$$

where $\theta_{0j} := 0$ for $j > s_0$. For arbitrary causal AR polynomials $\eta(z) = 1 - \eta_1 z - \cdots - \eta_p z^p$ and $\theta(z) = 1 - \theta_1 z - \cdots - \theta_p z^p$, an analogous recursion can be defined as

$$z_t(\boldsymbol{\eta},\boldsymbol{\theta},p) = \begin{cases} 0, & t \le 2p, \\ \theta_1 z_{t-1}(\boldsymbol{\eta},\boldsymbol{\theta},p) + \cdots + \theta_p z_{t-p}(\boldsymbol{\eta},\boldsymbol{\theta},p) + \eta(B)B^p\theta(B^{-1})X_t, & t = 2p+1,\ldots,n, \end{cases} \tag{14}$$

with $\boldsymbol{\eta} := (\eta_1,\ldots,\eta_p)'$ and $\boldsymbol{\theta} := (\theta_1,\ldots,\theta_p)'$. If $\boldsymbol{\eta}_0 := (\eta_{01},\ldots,\eta_{0p_0})'$ and $\boldsymbol{\theta}_0 := (\theta_{01},\ldots,\theta_{0p_0})'$, note that $\{z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)\}_{t=2p_0+1}^n$ closely approximates $\{z_t\}_{t=2p_0+1}^n$; the error is due to the initialization with zeros. Now, if $V_t = \theta_0^*(B)X_t$ (in which case, $V_t = Z_t/\theta_0^\dagger(B)$ also, see Section 2), then from Breidt et al. (1991), the joint density function for $(V_1,\ldots,V_{s_0},X_1,\ldots,X_n)$ equals

$$\begin{aligned} & h_1(V_1,\ldots,V_{r_0})\left[C_n\prod_{t=r_0+1}^n f(X_t - \phi_{01}X_{t-1} - \cdots - \phi_{0p_0}X_{t-p_0};\boldsymbol{\vartheta}_0)\right]h_2(\theta_0^\dagger(B)X_{n-s_0+1},\ldots,\theta_0^\dagger(B)X_n) \\ = \quad & h_1(V_1,\ldots,V_{r_0})\left[C_n\prod_{t=r_0+1}^n f(Z_t;\boldsymbol{\vartheta}_0)\right]h_2(\theta_0^\dagger(B)X_{n-s_0+1},\ldots,\theta_0^\dagger(B)X_n), \end{aligned}$$

where h_1 and h_2 are the joint densities for $(V_1,\ldots,V_{r_0})$ and $(\theta_0^\dagger(B)X_1,\ldots,\theta_0^\dagger(B)X_{s_0})$ respectively and do not depend on n, and $\ln C_n \sim -n\ln|\theta_{0s_0}|$ ($\sim$ indicates the ratio of the two sides converges to one as $n \to \infty$). Letting $\boldsymbol{\zeta} = (\eta_1,\ldots,\eta_p,\theta_1,\ldots,\theta_p,\tau_1,\ldots,\tau_4)'$, with $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_4)' = (\alpha, -(\text{sign }\theta_s)\beta, |\theta_s|\sigma, -\theta_s\mu)'$, $s = \max\{0 \le j \le p : \theta_j \ne 0\}$, and $\theta_0 = -1$ (since $\theta_{00} = -1$), it follows that for large n, the log-likelihood of

$\boldsymbol{\zeta}$ can be approximated by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\zeta}, p) &= \sum_{t=2p+1}^{n} \left[\ln f\left(-z_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p)/\theta_s; (\alpha, \beta, \sigma, \mu)'\right) - \ln|\theta_s|\right] \\ &= \sum_{t=2p+1}^{n} \ln\left[\frac{1}{|\theta_s|\sigma} f\left(\frac{z_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p) + \theta_s\mu}{-\theta_s\sigma}; (\alpha, \beta, 1, 0)'\right)\right] \\ &= \sum_{t=2p+1}^{n} \ln f(z_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p); \boldsymbol{\tau}), \end{aligned} \tag{15}$$

where the values of $\{z_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p)\}_{t=2p+1}^{n}$ are computed recursively via (14).

Given observed values $\{X_t\}_{t=1}^{n}$ and the minimum AIC estimate of AR model order $\hat{p}$, we can estimate $\boldsymbol{\zeta}_0 = (\boldsymbol{\eta}_0', \boldsymbol{\theta}_0', \boldsymbol{\tau}_0')'$ by maximizing $\mathcal{L}$ with respect to $\boldsymbol{\zeta}$ at $p = \hat{p}$, using the Yule-Walker estimate $\hat{\boldsymbol{\eta}}_{YW}(\hat{p})$ of $\boldsymbol{\eta}_0$ as a starting value for $\boldsymbol{\eta}$ when doing the optimization. The order of noncausality for $\{X_t\}$, s_0, can then be estimated by computing confidence intervals for the all-pass parameters $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0p_0})' = (\theta_{01}, \ldots, \theta_{0s_0}, 0, \ldots, 0)'$. In addition, since $\phi_0(z) = 1 - \phi_{01}z - \ldots - \phi_{0p_0}z^{p_0} = \eta_0(z)(-\theta_{0s_0}^{-1})z^{s_0}\theta_0^{c}(z^{-1})/\theta_0^{c}(z)$, given estimates $\hat{p}$, $\hat{s}$, and $\hat{\boldsymbol{\zeta}}$ of p_0, s_0, and $\boldsymbol{\zeta}_0$, a preliminary estimate of $\boldsymbol{\phi}_0 = (\phi_{01}, \ldots, \phi_{0p_0})'$ can be found by canceling roots which are in both the numerator and denominator of

$$\frac{(1 - \hat{\eta}_1 z - \cdots - \hat{\eta}_{\hat{p}} z^{\hat{p}})(-\hat{\theta}_{\hat{s}}^{-1})z^{\hat{s}}(1 - \hat{\theta}_1 z^{-1} - \cdots - \hat{\theta}_{\hat{s}} z^{-\hat{s}})}{(1 - \hat{\theta}_1 z - \cdots - \hat{\theta}_{\hat{s}} z^{\hat{s}})}. \tag{16}$$

For further model accuracy, ML can then be used to directly estimate $\boldsymbol{\phi}_0$ (see Andrews et al., 2009, for details), with the preliminary estimate used as an initial value for the optimizer. The limiting distribution for ML estimators of $\boldsymbol{\zeta}_0 = (\boldsymbol{\eta}_0', \boldsymbol{\theta}_0', \boldsymbol{\tau}_0')'$ is given in Theorem 4, and afterwards we address confidence interval calculation. But first, we need to introduce some notation and define a random function W. MLEs of $(\boldsymbol{\eta}_0', \boldsymbol{\theta}_0')'$ converge in distribution to the maximizer of W.

For $\mathbf{u} = (u_1, \ldots, u_{p_0})' \in \mathbb{R}^{p_0}$ and $\mathbf{v} = (v_1, \ldots, v_{p_0})' \in \mathbb{R}^{p_0}$, define the sequence $\{c_j(\mathbf{u}, \mathbf{v})\}_{j \neq 0}$ so that

$$\sum_{j \neq 0} c_j(\mathbf{u}, \mathbf{v}) z_{t-j} = -\mathbf{u}'[(1/\eta_0(B))z_{t-k}]_{k \in \{1,\ldots,p_0\}} + \mathbf{v}'[(1/\theta_0^{c}(B))z_{t-k} - (1/\theta_0^{c}(B^{-1}))z_{t+k}]_{k \in \{1,\ldots,p_0\}}. \tag{17}$$

Therefore, if the Laurent series expansions for $1/\eta_0(z)$ and $1/\theta_0^{c}(z)$ are given by $1/\eta_0(z) = \sum_{j=0}^{\infty} \gamma_j z^j$ and $1/\theta_0^{c}(z) = \sum_{j=0}^{\infty} \chi_j z^j$, then

$$\begin{aligned}\sum_{j \neq 0} c_j(\mathbf{u}, \mathbf{v}) z_{t-j} &= -u_1 \sum_{j=0}^{\infty} \gamma_j z_{t-1-j} - \cdots - u_{p_0} \sum_{j=0}^{\infty} \gamma_j z_{t-p_0-j} + v_1 \left(\sum_{j=0}^{\infty} \chi_j z_{t-1-j} - \sum_{j=0}^{\infty} \chi_j z_{t+1+j}\right) \\ &\quad + \cdots + v_{p_0} \left(\sum_{j=0}^{\infty} \chi_j z_{t-p_0-j} - \sum_{j=0}^{\infty} \chi_j z_{t+p_0+j}\right),\end{aligned}$$

and so $c_1(\mathbf{u},\mathbf{v}) = -u_1\gamma_0 + v_1\chi_0$, $c_{-1}(\mathbf{u},\mathbf{v}) = -v_1\chi_0$, $c_2(\mathbf{u},\mathbf{v}) = -u_1\gamma_1 - u_2\gamma_0 + v_1\chi_1 + v_2\chi_0$, $c_{-2}(\mathbf{u},\mathbf{v}) = -v_1\chi_1 - v_2\chi_0$, etc. From Brockwell and Davis (1991, Chapter 3), $\{\gamma_j\}_{j=0}^{\infty}$ and $\{\chi_j\}_{j=0}^{\infty}$ decay at geometric rates, so for any $\mathbf{u},\mathbf{v} \in \mathrm{I\!R}^{p_0}$, $\{c_j(\mathbf{u},\mathbf{v})\}_{j\neq 0}$ is also geometrically decaying as $j \to \pm\infty$. We now introduce

$$W(\mathbf{u},\mathbf{v}) = \sum_{k=1}^{\infty}\sum_{j\neq 0}\left\{\ln f\left(z_{k,j} + [\tilde{c}(\alpha_0)]^{1/\alpha_0}\sigma_0|\theta_{0s_0}|c_j(\mathbf{u},\mathbf{v})\delta_k\Gamma_k^{-1/\alpha_0};\boldsymbol{\tau}_0\right) - \ln f\left(z_{k,j};\boldsymbol{\tau}_0\right)\right\},$$

where

- $\{z_{k,j}\}_{k,j}$ is iid with $z_{k,j} \stackrel{\mathcal{L}}{=} z_t$,
- $\tilde{c}(\alpha_0) := \left(\int_0^{\infty} t^{-\alpha_0}\sin(t)\,dt\right)^{-1}$,
- $\{\delta_k\}$ is iid with $\mathrm{P}(\delta_k = 1) = [1 - (\text{sign }\theta_{0s_0})\beta_0]/2$ and $\mathrm{P}(\delta_k = -1) = 1 - \mathrm{P}(\delta_k = 1)$,
- $\Gamma_k = E_1 + \cdots + E_k$, where $\{E_k\}$ is an iid series of exponential random variables with mean one, and
- $\{z_{k,j}\}$, $\{\delta_k\}$, and $\{E_k\}$ are mutually independent.

Note that $\tilde{c}(\alpha_0) = \lim_{x\to\infty} x^{\alpha_0}\mathrm{P}(|z_t| > x)/(|\theta_{0s_0}|\sigma_0)^{\alpha_0}$ and $[1-(\text{sign }\theta_{0s_0})\beta_0]/2 = \lim_{x\to\infty}[\mathrm{P}(z_t > x)/\mathrm{P}(|z_t| > x)]$ (Samorodnitsky and Taqqu, 1994, Property 1.2.15). Also, from the proof of Theorem 3.1 in Andrews et al. (2009), where a function similar to W is considered, $W(\mathbf{u},\mathbf{v})$ is finite for all $\mathbf{u},\mathbf{v} \in \mathrm{I\!R}^{p_0}$ and has a unique maximum almost surely.

In the following theorem, we give the nondegenerate limiting distribution for ML estimators of $\boldsymbol{\zeta}_0 = (\boldsymbol{\eta}_0', \boldsymbol{\theta}_0', \boldsymbol{\tau}_0')'$.

Theorem 4. *There exists a sequence of maximizers $\hat{\boldsymbol{\zeta}}_{ML} = (\hat{\boldsymbol{\eta}}_{ML}', \hat{\boldsymbol{\theta}}_{ML}', \hat{\boldsymbol{\tau}}_{ML}')'$ of $\mathcal{L}(\cdot, p_0)$ in (15) such that, as $n \to \infty$,*

$$n^{1/\alpha_0}(\hat{\boldsymbol{\eta}}_{ML} - \boldsymbol{\eta}_0) \xrightarrow{\mathcal{L}} \boldsymbol{\xi}_1, \quad n^{1/\alpha_0}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \boldsymbol{\xi}_2, \quad \text{and} \quad n^{1/2}(\hat{\boldsymbol{\tau}}_{ML} - \boldsymbol{\tau}_0) \xrightarrow{\mathcal{L}} \mathbf{Y} \sim N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\tau}_0)), \tag{18}$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ is the unique maximizer of $W(\cdot,\cdot)$, $\boldsymbol{\xi}$ and $\mathbf{Y}$ are independent, and $\mathbf{I}(\boldsymbol{\tau}) := -\left[E\{\partial^2 \ln f(z_t;\boldsymbol{\tau})/(\partial\tau_i\partial\tau_j)\}\right]_{i,j\in\{1,\ldots,4\}}$.

Proof. For $\mathbf{u},\mathbf{v} \in \mathrm{I\!R}^{p_0}$ and $\mathbf{w} \in \mathrm{I\!R}^4$, let

$$W_n(\mathbf{u},\mathbf{v},\mathbf{w}) = \mathcal{L}(\boldsymbol{\zeta}_0 + (n^{-1/\alpha_0}\mathbf{u}', n^{-1/\alpha_0}\mathbf{v}', n^{-1/2}\mathbf{w}')', p_0) - \mathcal{L}(\boldsymbol{\zeta}_0, p_0), \tag{19}$$

$$W_n^{\dagger}(\mathbf{u},\mathbf{v},\mathbf{w}) = \sum_{t=2p_0+1}^{n}\left[\ln f\left(z_t + n^{-1/\alpha_0}\sum_{j\neq 0} c_j(\mathbf{u},\mathbf{v})z_{t-j};\boldsymbol{\tau}_0 + \frac{\mathbf{w}}{\sqrt{n}}\right) - \ln f\left(z_t;\boldsymbol{\tau}_0\right)\right], \tag{20}$$

and

$$W_n^{*}(\mathbf{u},\mathbf{v},\mathbf{w}) = \sum_{t=2p_0+1}^{n}\left[\ln f\left(z_t + n^{-1/\alpha_0}\sum_{j\neq 0} c_j(\mathbf{u},\mathbf{v})z_{t-j};\boldsymbol{\tau}_0\right) - \ln f\left(z_t;\boldsymbol{\tau}_0\right) + \frac{\mathbf{w}'}{\sqrt{n}}\frac{\partial \ln f(z_t;\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}}\right].$$

By Lemma 4 in the Appendix, $W_n(\cdot,\cdot,\cdot) - W_n^{\dagger}(\cdot,\cdot,\cdot) = o_p(1)$ on $C(\mathbb{R}^{2p_0+4})$, the space of continuous functions on $\mathbb{R}^{2p_0+4}$ where convergence is equivalent to uniform convergence on every compact subset, and, by Lemma A.5 in Andrews et al. (2009), $W_n^{\dagger}(\mathbf{u},\mathbf{v},\mathbf{w}) - W_n^{*}(\mathbf{u},\mathbf{v},\mathbf{w}) + \mathbf{w}'\mathbf{I}(\boldsymbol{\tau}_0)\mathbf{w}/2 = o_p(1)$ on $C(\mathbb{R}^{2p_0+4})$. In addition, following the proof of Theorem 3.3 in Andrews et al. (2009), where a similar result is established, it can be shown that $W_n^{*}(\mathbf{u},\mathbf{v},\mathbf{w}) \xrightarrow{\mathcal{L}} W(\mathbf{u},\mathbf{v}) + \mathbf{w}'\mathbf{N}$ on $C(\mathbb{R}^{2p_0+4})$, with $\mathbf{N} \sim \mathrm{N}(\mathbf{0},\mathbf{I}(\boldsymbol{\tau}_0))$ independent of $W(\cdot,\cdot)$. Therefore, $W_n(\mathbf{u},\mathbf{v},\mathbf{w}) \xrightarrow{\mathcal{L}} W(\mathbf{u},\mathbf{v}) + \mathbf{w}'\mathbf{N} - \mathbf{w}'\mathbf{I}(\boldsymbol{\tau}_0)\mathbf{w}/2$ on $C(\mathbb{R}^{2p_0+4})$. Since $\boldsymbol{\xi} = (\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)$ uniquely maximizes $W(\cdot,\cdot)$ almost surely, and $\mathbf{Y} = \mathbf{I}^{-1}(\boldsymbol{\tau}_0)\mathbf{N} \sim \mathrm{N}(\mathbf{0},\mathbf{I}^{-1}(\boldsymbol{\tau}_0))$, which is independent of W, uniquely maximizes $\mathbf{w}'\mathbf{N} - \mathbf{w}'\mathbf{I}(\boldsymbol{\tau}_0)\mathbf{w}/2$, by Remark 1 in Davis et al. (1992), there exists a sequence of maximizers of $W_n(\cdot,\cdot,\cdot)$ which converges in distribution to $(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{Y})$. Because $\mathcal{L}(\boldsymbol{\zeta},p_0) - \mathcal{L}(\boldsymbol{\zeta}_0,p_0) = W_n(n^{1/\alpha_0}(\boldsymbol{\eta}-\boldsymbol{\eta}_0), n^{1/\alpha_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0), n^{1/2}(\boldsymbol{\tau}-\boldsymbol{\tau}_0))$, the result of this theorem holds. □

Although the MLEs $\hat{\boldsymbol{\zeta}}_{ML}$ maximize $\mathcal{L}(\cdot,p_0)$, note that $\mathrm{P}(\hat{\boldsymbol{\zeta}}_{ML}$ maximizes $\mathcal{L}(\cdot,\hat{p})) \to 1$ as $n \to \infty$, since, by Theorem 2, $\mathrm{P}(\hat{p} = p_0) \to 1$. Note, also, that the estimators $\hat{\boldsymbol{\tau}}_{ML}$ of $\boldsymbol{\tau}_0$ have the same limiting normal distribution as ML estimators in the case of observed iid stable noise $\{z_t\}_{t=1}^{n}$ (DuMouchel, 1973). Values of the limiting covariance matrix $\mathbf{I}^{-1}(\cdot)$ can be found in Nolan (2001) for different stable parameter vectors.

Since the forms of the limiting distributions for $\hat{\boldsymbol{\eta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}$ in (18) are intractable, we recommend using the bootstrap to examine the distributions for these estimators. Andrews et al. (2009) give a bootstrap procedure for examining the distribution for MLEs of the AR parameters $\boldsymbol{\phi}_0 = (\phi_{01},\ldots,\phi_{0p_0})'$; we consider a similar procedure here. Given observations $\{X_t\}_{t=1}^{n}$ from (1), $\hat{\boldsymbol{\eta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}$ from (18), and corresponding residuals $\{z_t(\hat{\boldsymbol{\eta}}_{ML},\hat{\boldsymbol{\theta}}_{ML},p_0)\}_{t=2p_0+1}^{n}$ obtained via (14), the procedure is implemented by first generating an iid sequence $\{z_t^*\}_{t=1}^{m_n}$ from the empirical distribution for $\{z_t(\hat{\boldsymbol{\eta}}_{ML},\hat{\boldsymbol{\theta}}_{ML},p_0)\}_{t=2p_0+1}^{n}$. A bootstrap replicate $X_1^*,\ldots,X_{m_n}^*$ is then obtained from the estimate of model equation (13)

$$\hat{\eta}_{ML}(B)B^{p_0}\hat{\theta}_{ML}^{c}(B^{-1})X_t^* = \hat{\theta}_{ML}^{c}(B)z_t^*,$$

with $\hat{\eta}_{ML}(z) := 1 - \hat{\eta}_{1,ML}z - \cdots - \hat{\eta}_{p_0,ML}z^{p_0}$ and $\hat{\theta}^c_{ML}(z) := 1 - \hat{\theta}_{1,ML}z - \cdots - \hat{\theta}_{p_0,ML}z^{p_0}$ (let $z^*_t = 0$ for $t \notin \{1, \ldots, m_n\}$). Finally, with

$$z^*_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p) := \begin{cases} 0, & t \leq 2p, \\ \theta_1 z^*_{t-1}(\boldsymbol{\eta}, \boldsymbol{\theta}, p) + \cdots + \theta_p z^*_{t-p}(\boldsymbol{\eta}, \boldsymbol{\theta}, p) + \eta(B)B^p\theta(B^{-1})X^*_t, & t = 2p+1, \ldots, m_n, \end{cases}$$

for $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)' \in \mathbb{R}^p$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)' \in \mathbb{R}^p$, bootstrap replicates $\hat{\boldsymbol{\eta}}^*_{m_n}$ and $\hat{\boldsymbol{\theta}}^*_{m_n}$ of $\hat{\boldsymbol{\eta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}$ can be found by maximizing $\mathcal{L}^*_{m_n}(\boldsymbol{\eta}, \boldsymbol{\theta}, p_0) := \sum_{t=2p_0+1}^{m_n} \ln f(z^*_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p_0); \hat{\boldsymbol{\tau}}_{ML})$ with respect to $(\boldsymbol{\eta}, \boldsymbol{\theta})$. The limiting behavior of $\hat{\boldsymbol{\eta}}^*_{m_n}$ and $\hat{\boldsymbol{\theta}}^*_{m_n}$ is addressed in Theorem 5. To give a precise statement of the results, we let $\mathcal{M}_p(\mathbb{R}^{p_0})$ represent the space of probability measures on $\mathbb{R}^{p_0}$ and we use the metric d_{p_0} from Davis and Wu (1997, page 1139) to metrize the topology of weak convergence on $\mathcal{M}_p(\mathbb{R}^{p_0})$. For random elements Q_n, Q of $\mathcal{M}_p(\mathbb{R}^{p_0})$, $Q_n \xrightarrow{P} Q$ if and only if $d_{p_0}(Q_n, Q) \xrightarrow{P} 0$ on $\mathbb{R}$, which is equivalent to $\int_{\mathbb{R}^{p_0}} h_j\, dQ_n \xrightarrow{P} \int_{\mathbb{R}^{p_0}} h_j\, dQ$ on $\mathbb{R}$ for all $j \in \{1, 2, \ldots\}$, where $\{h_j\}_{j=1}^{\infty}$ is a dense sequence of bounded, uniformly continuous functions on $\mathbb{R}^{p_0}$.

Theorem 5. *If, as $n \to \infty$, $m_n \to \infty$ with $m_n/n \to 0$, then there exists a sequence of maximizers $(\hat{\boldsymbol{\eta}}^*_{m_n}, \hat{\boldsymbol{\theta}}^*_{m_n})$ of $\mathcal{L}^*_{m_n}(\cdot, \cdot, p_0)$ such that $P(m_n^{1/\hat{\alpha}_{ML}}(\hat{\boldsymbol{\eta}}^*_{m_n} - \hat{\boldsymbol{\eta}}_{ML}) \in \cdot | X_1, \ldots, X_n) \xrightarrow{P} P(\boldsymbol{\xi}_1 \in \cdot)$ and $P(m_n^{1/\hat{\alpha}_{ML}}(\hat{\boldsymbol{\theta}}^*_{m_n} - \hat{\boldsymbol{\theta}}_{ML}) \in \cdot | X_1, \ldots, X_n) \xrightarrow{P} P(\boldsymbol{\xi}_2 \in \cdot)$ on $\mathcal{M}_p(\mathbb{R}^{p_0})$.*

Proof. The proof of this result is nearly the same as the proof of Theorem 3.4 in Andrews et al. (2009), so we omit the details. □

Thus, $m_n^{1/\hat{\alpha}_{ML}}(\hat{\boldsymbol{\eta}}^*_{m_n} - \hat{\boldsymbol{\eta}}_{ML})$ and $m_n^{1/\hat{\alpha}_{ML}}(\hat{\boldsymbol{\theta}}^*_{m_n} - \hat{\boldsymbol{\theta}}_{ML})$, conditioned on $\{X_t\}_{t=1}^n$, have the same limiting distributions as $n^{1/\alpha_0}(\hat{\boldsymbol{\eta}}_{ML} - \boldsymbol{\eta}_0)$ and $n^{1/\alpha_0}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0)$ respectively. If n is large, these limiting distributions can, therefore, be approximated by simulating bootstrap values of $\hat{\boldsymbol{\eta}}^*_{m_n}$ and $\hat{\boldsymbol{\theta}}^*_{m_n}$, and looking at the distributions for $m_n^{1/\hat{\alpha}_{ML}}(\hat{\boldsymbol{\eta}}^*_{m_n} - \hat{\boldsymbol{\eta}}_{ML})$ and $m_n^{1/\hat{\alpha}_{ML}}(\hat{\boldsymbol{\theta}}^*_{m_n} - \hat{\boldsymbol{\theta}}_{ML})$. In principle, one could also examine the limiting distributions for $n^{1/\alpha_0}(\hat{\boldsymbol{\eta}}_{ML} - \boldsymbol{\eta}_0)$ and $n^{1/\alpha_0}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0)$ by simulating realizations of $W(\cdot, \cdot)$, with the true parameter values $\boldsymbol{\eta}_0$, $\boldsymbol{\theta}_0$, and $\boldsymbol{\tau}_0$ replaced by estimates, and by finding the corresponding values of the maximizer $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$, but this procedure is more laborious than the bootstrap. Confidence intervals for the elements of $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}_0$ can be obtained using the limiting results for $\hat{\boldsymbol{\eta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML}$ in (18), bootstrap estimates of quantiles for the limiting distributions, and the estimate $\hat{\alpha}_{ML}$ of α_0. For the elements of $\boldsymbol{\tau}_0$, confidence intervals can be directly obtained from the limiting normal result for $\hat{\boldsymbol{\tau}}_{ML}$ in (18).

4 Numerical Results

4.1 Simulation Study

In this section, we describe a simulation experiment to study the accuracy of the order selection procedures for finite samples. We did these simulations in R (http://www.r-project.org), using the *fBasics: Rmetrics–Markets and Basic Statistics* package (http://www.rmetrics.org) to generate stable noise and evaluate stable densities.

For each of 100 replicates, we simulated an AR series of length $n = 500$ with stable noise and found $\hat{p} = \arg\min_{0 \le k \le 5} \text{AIC}(k)$, the minimum AIC estimate of AR model order over the integers $0, 1, \ldots, 5$. We then found the MLE $\hat{\boldsymbol{\zeta}}_{ML} = (\hat{\boldsymbol{\eta}}'_{ML}, \hat{\boldsymbol{\theta}}'_{ML}, \hat{\boldsymbol{\tau}}'_{ML})'$ of $\boldsymbol{\zeta}_0$ by maximizing the log-likelihood $\mathcal{L}(\boldsymbol{\zeta}, p)$ in (15) with respect to $\boldsymbol{\zeta}$ at $p = \hat{p}$. For the likelihood maximization, the Yule-Walker estimate $\hat{\boldsymbol{\eta}}_{YW}(\hat{p})$ was used as the starting value for $\boldsymbol{\eta}$, and we used 100 randomly chosen starting values for $(\boldsymbol{\theta}', \boldsymbol{\tau}')'$. The log-likelihood was evaluated at each of the candidate values, and then we reduced the collection of initial values to the eight with the highest likelihoods. Optimized values were found using the Nelder-Mead algorithm (Nelder and Mead, 1965) and the eight initial values as starting points. The optimized value for which the likelihood was highest was chosen to be $\hat{\boldsymbol{\zeta}}_{ML}$. Lastly, the bootstrap procedure described in Section 3.3 was repeated 1000 times, with $m_n = 150$, in order to estimate the 2.5% and 97.5% quantiles for the distributions of the elements of $\boldsymbol{\xi}_2$ in (18). We used the estimated quantiles to compute 95% confidence intervals for the elements of $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0s_0}, 0, \ldots, 0)'$, and s_0, which corresponds to the order of noncausality for the AR process, was estimated via $\hat{s} = \min\{0 \le j \le \hat{p} : \text{the C.I.s for } \theta_{0k}, k > j, \text{ all contain zero}\}$.

We obtained simulation results for the causal AR(1) model with parameter $\boldsymbol{\phi}_0 = 0.5$, the purely noncausal AR(1) model with parameter $\boldsymbol{\phi}_0 = 2.0$, and the AR(2) model with parameter $\boldsymbol{\phi}_0 = (-1.2, 1.6)'$. The AR(2) polynomial $\phi_0(z) = 1 + 1.2z - 1.6z^2$ equals $(1 - 0.8z)(1 + 2.0z)$, so it has one root inside and the other outside the unit circle. For the stable parameter values $\boldsymbol{\vartheta}_0 = (\alpha_0, \beta_0, \sigma_0, \mu_0)'$, we considered $\alpha_0 \in \{0.8, 1.5\}$,

Model Parameters	α_0	β_0	σ_0	μ_0	$\hat{p}$ 0	1	2	3	4	5	$\hat{s}$ 0	1	2	3	4	5
$\boldsymbol{\phi}_0 = 0.5$	0.8	0.0	1	0	0	93	2	1	0	4	91	4	1	0	1	3
($p_0 = 1$, $s_0 = 0$)	0.8	0.5	1	0	0	94	1	1	1	3	95	0	0	1	1	3
	1.5	0.0	1	0	0	88	2	6	1	3	90	4	1	3	1	1
	1.5	0.5	1	0	0	78	7	7	3	5	93	5	1	0	1	0
$\boldsymbol{\phi}_0 = 2.0$	0.8	0.0	1	0	0	93	3	1	1	2	0	94	2	1	1	2
($p_0 = 1$, $s_0 = 1$)	0.8	0.5	1	0	0	96	2	2	0	0	0	97	1	2	0	0
	1.5	0.0	1	0	0	76	10	3	7	4	0	94	3	1	1	1
	1.5	0.5	1	0	0	84	9	5	1	1	0	98	2	0	0	0
$\boldsymbol{\phi}_0 = (-1.2, 1.6)'$	0.8	0.0	1	0	0	1	82	7	4	6	0	61	25	5	4	5
($p_0 = 2$, $s_0 = 1$)	0.8	0.5	1	0	0	0	95	3	0	2	0	64	32	2	0	2
	1.5	0.0	1	0	0	0	84	4	5	7	0	93	3	0	2	2
	1.5	0.5	1	0	0	1	84	10	2	3	0	95	2	1	1	1

Table 1: *The frequencies for estimates of the AR model order p_0 and the order of noncausality s_0.*

$\beta_0 \in \{0, 0.5\}$, $\sigma_0 = 1$, and $\mu_0 = 0$. Simulation results appear in Table 1, where we give the frequencies for values of $\hat{p}$ and $\hat{s}$. Note that, for all models, p_0 and s_0 were correctly identified most of the time, and underestimation was rare.

4.2 Autoregressive Model Fitting

Figure 1 shows the natural logarithms of the volumes of Wal-Mart stock traded daily on the New York Stock Exchange from December 1, 2003 to December 31, 2004. In Andrews et al. (2009, Section 4.2), the noncausal AR(2) model

$$(1 - 0.7380B)(1 + 2.8146B)X_t = Z_t, \tag{21}$$

with $\{Z_t\}$ iid stable with parameter vector $(\alpha, \beta, \sigma, \mu)' = (1.8335, 0.5650, 0.4559, 16.0030)'$, was fit to this log-volume series $\{X_t\}_{t=1}^{274}$. Andrews et al. used the Gaussian AIC statistic to determine that two is an appropriate AR model order, and then maximized the log-likelihood of a stable AR(2) series to obtain the parameter estimates, considering AR(2) polynomials with all combinations of roots inside and outside the unit circle. Since the residuals from model (21) appeared approximately iid stable with parameter $(1.8335, 0.5650, 0.4559, 16.0030)'$, they concluded that (21) is a satisfactory fitted model for the series. By Theorem 2, minimizing AIC is a consistent AR order selection procedure in the case of a noncausal, infinite variance AR process, supporting the use of the AIC statistic for AR order determination in this example. In this section, we demonstrate that the AR likelihood did not need to be maximized with respect to all

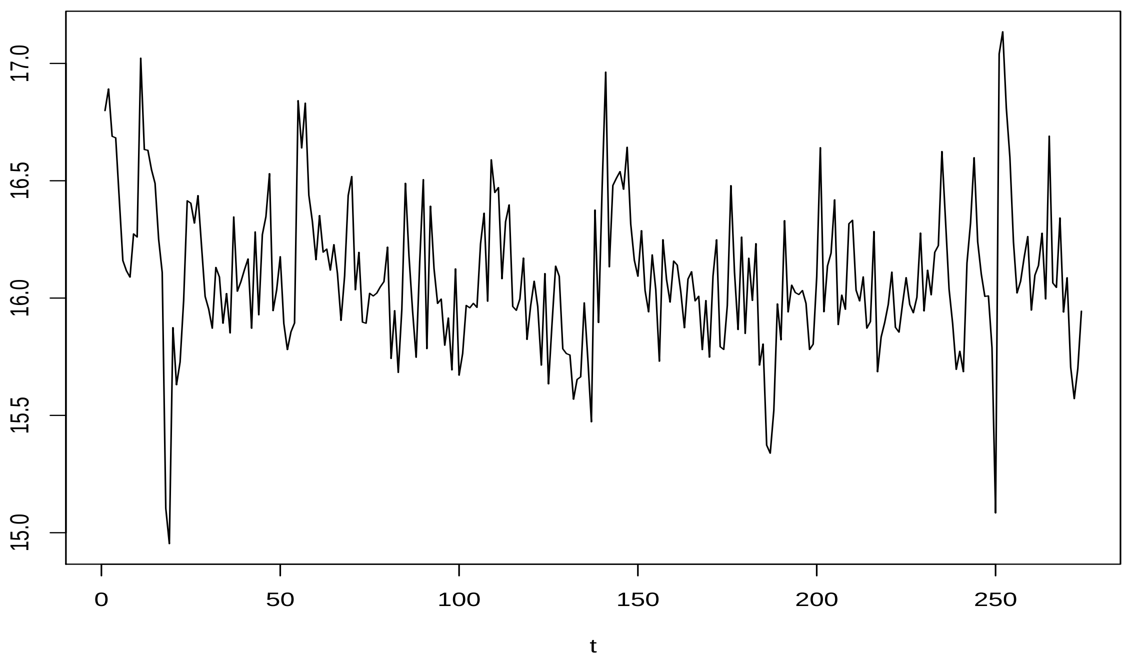

Figure 1: *The natural logarithms of the volumes of Wal-Mart stock traded daily on the New York Stock Exchange from December 1, 2003 to December 31, 2004.*

combinations of two roots inside and outside the unit circle; all-pass models could have been used to determine that one is an appropriate order of noncausality for the fitted AR(2) model. All-pass models could also have been used for preliminary AR estimation.

First of all, the Yule-Walker estimate $\hat{\boldsymbol{\eta}}_{YW}(2)$ equals $(0.4425, 0.0903)'$, so the causal AR residuals are given by $\hat{U}_t = (1 - 0.4425B - 0.0903B^2)X_t$, $t = 3, \ldots, 274$. These residuals $\{\hat{U}_t\}$ are shown in Figure 2, along with sample autocorrelation functions for the residuals and absolute values and squares of the mean-corrected residuals. The bounds in Figure 2(b)–(d) are approximate 95% confidence bounds which we obtained by simulating 100,000 independent sample correlations for the values, absolute values, and squares of 272 mean-corrected iid values from the empirical distribution of $\{\hat{U}_t\}_{t=3}^{274}$. Based on the graphs in Figure 2, $\{\hat{U}_t\}$ does not appear iid, but rather uncorrelated yet dependent, with sample correlations for the absolute values and squares that are roughly geometrically decaying. Following Theorem 3, this suggests that a noncausal AR(2) model is appropriate for $\{X_t\}$.

To identify the appropriate order of noncausality, we maximized the log-likelihood $\mathcal{L}(\boldsymbol{\zeta}, p)$ in (15) with respect to $\boldsymbol{\zeta}$ at $p = 2$. The ML estimates are

$$\begin{aligned} \hat{\boldsymbol{\zeta}}_{ML} &= (\hat{\eta}_1, \hat{\eta}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{\tau}_4)' \\ &= (0.4178, 0.1326, -0.2553, -0.0351, 1.7942, 0.6754, 0.1610, 7.1993)' \end{aligned} \tag{22}$$

and, from 1000 iterations of the bootstrap procedure described in Section 3.3 with $m_n = 135$, approximate 95% bootstrap confidence intervals for the all-pass parameters θ_{01} and θ_{02} are $(-0.3651, -0.2253)$ and $(-0.0705, 0.0039)$. Since the second interval overlaps zero while the first does not, the all-pass order is one, and so the appropriate order of noncausality for $\{X_t\}$ also appears to be one. Given $\hat{p} = 2$, $\hat{s} = 1$, and the parameter estimates in (22), it follows from (16) that a preliminary estimate of the AR(2) polynomial $\phi_0(z) = 1 - \phi_{01}z - \phi_{02}z^2$ is

$$\begin{aligned} \frac{(1 - 0.4178z - 0.1326z^2)(0.2553^{-1})z(1 + 0.2553z^{-1})}{(1 + 0.2553z)} &= \frac{(1 + 0.2109z)(1 - 0.6287z)(1 + 3.9170z)}{(1 + 0.2553z)} \\ &\approx (1 - 0.6287z)(1 + 3.9170z). \end{aligned}$$

The corresponding parameters could have been used as initial values when finding (21).

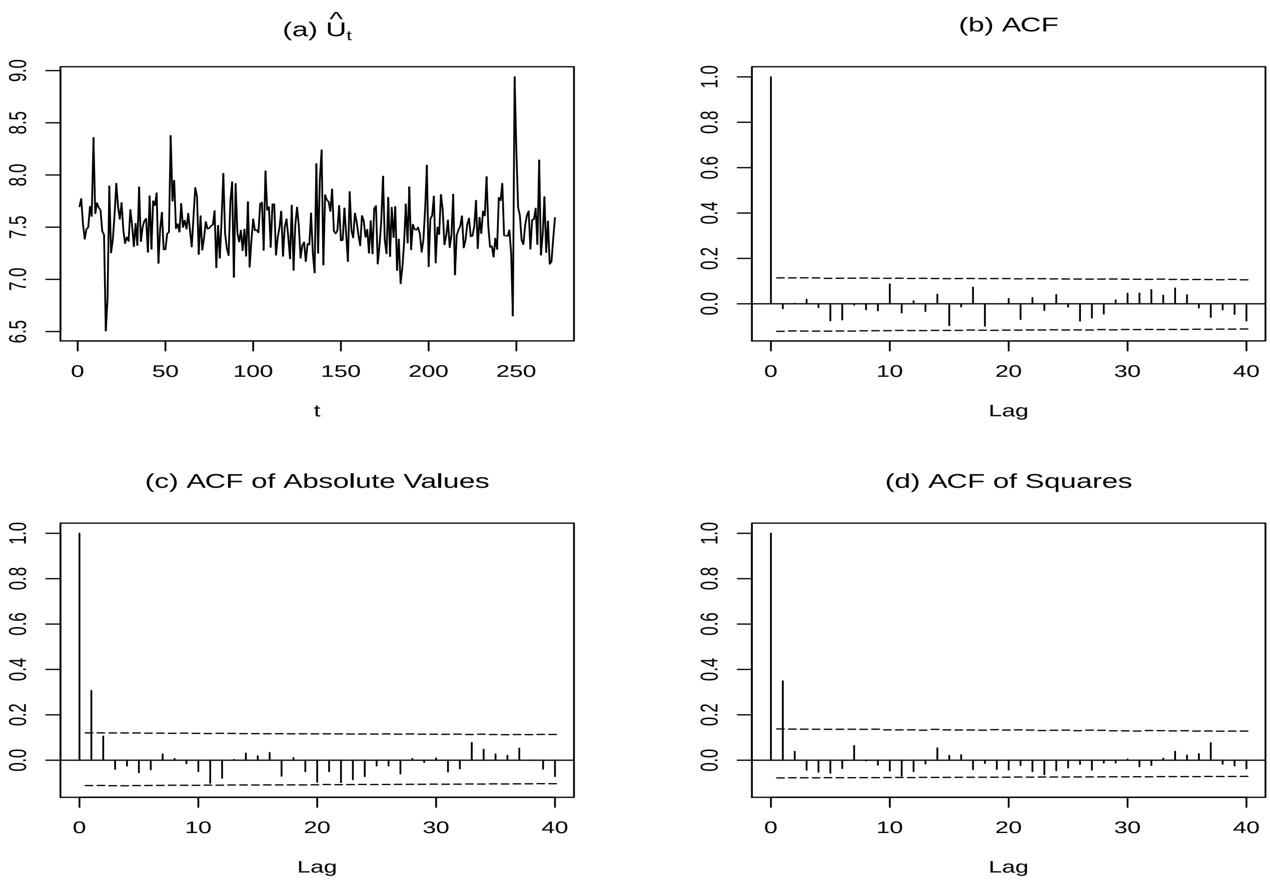

Figure 2: *(a) The causal AR residuals* $\{\hat{U}_t\}$*, and sample autocorrelation functions for (b)* $\{\hat{U}_t\}$*, (c) the absolute values of mean-corrected* $\{\hat{U}_t\}$*, and (d) the squares of mean-corrected* $\{\hat{U}_t\}$*.*

Appendix

In this section, we give proofs of the lemmas used to establish results in Section 3. We begin with Lemmas 1–3, which were used in the proof of Theorem 1. As in the proof of Theorem 1, we assume the distribution for the iid noise $\{Z_t\}$ is in the domain of attraction of a stable law with exponent $\alpha_0 \in (0,2)$, and that $K(n) = O(n^{\delta})$ with $0 \le \delta < \min\{1/2, 1-\alpha_0/2\}$. Additionally, when $\mathrm{E}|Z_t| < \infty$, we let $\mu_X = \mathrm{E}X_t$ and $\mu_U = \mathrm{E}U_t$. Since, by (3), $U_t = X_t - \eta_{01}X_{t-1} - \cdots - \eta_{0p_0}X_{t-p_0}$, it follows that $\mu_U = (1-\sum_{j=1}^{p_0}\eta_{0j})\mu_X$.

Lemma 1. *(i) If* $\alpha_0 \in (0,1]$*, then* $\max_{p_0\le k\le K(n)} n^{1/2-\kappa}\|[\sum_{t=k+1}^{n} U_tX_{t-j}]_{j\in\{1,\ldots,k\}}\| \xrightarrow{P} 0$ *for sufficiently large* $\kappa < 2/\alpha_0$ *and, (ii) if* $\alpha_0 \in (1,2)$*, then* $\max_{p_0\le k\le K(n)} n^{1/2-\kappa}\|[\sum_{t=k+1}^{n}(U_t-\mu_U)(X_{t-j}-\mu_X)]_{j\in\{1,\ldots,k\}}\| \xrightarrow{P} 0$ *for large* $\kappa < 2/\alpha_0$.

Proof. (i) First, recall that $X_t = \sum_{j=-\infty}^{\infty}\psi_j Z_{t-j}$ and $U_t = \sum_{j=0}^{\infty}\pi_j Z_{t+j}$, where the coefficients $\{\psi_j\}$ and $\{\pi_j\}$ are geometrically decaying as $j\to\pm\infty$. For $j\le 0$, we let $\bar\pi_j = \pi_{-j}$, so that $U_t = \sum_{j=-\infty}^{0}\bar\pi_j Z_{t-j}$. Now suppose $\alpha_0 \le 1$ and consider

$$\begin{aligned}
&\max_{p_0\le k\le K(n)} n^{1/2-\kappa}\left\|\left[\sum_{t=k+1}^{n} U_tX_{t-j}\right]_{j\in\{1,\ldots,k\}}\right\| \\
&= \max_{p_0\le k\le K(n)} n^{1/2-\kappa}\left\|\left[\sum_{t=k+1}^{n}\sum_{l=-\infty}^{0}\sum_{m=-\infty}^{\infty}\bar\pi_l\psi_m Z_{t-l}Z_{t-j-m}\right]_{j\in\{1,\ldots,k\}}\right\| \\
&\le \max_{p_0\le k\le K(n)} n^{1/2-\kappa}\left\|\left[\sum_{t=k+1}^{n}\sum_{l=-\infty}^{0}\bar\pi_l\psi_{l-j}Z_{t-l}^2\right]_{j\in\{1,\ldots,k\}}\right\| \qquad (23)\\
&\quad + \max_{p_0\le k\le K(n)} n^{1/2-\kappa}\left\|\left[\sum_{t=k+1}^{n}\sum_{l=-\infty}^{0}\sum_{m\ne l-j}\bar\pi_l\psi_m Z_{t-l}Z_{t-j-m}\right]_{j\in\{1,\ldots,k\}}\right\|. \qquad (24)
\end{aligned}$$

We complete the proof of (i) by showing that (23) and (24) are $o_p(1)$ for sufficiently large $\kappa < 2/\alpha_0$.

Since the Laurent series expansion for $1/\eta_0(z)$ is given by $\sum_{j=0}^{\infty}\gamma_j z^j$, following (3), $X_t = \sum_{j=0}^{\infty}\gamma_j U_{t-j}$ for all t. Therefore, because the all-pass process $\{U_t\}$ is uncorrelated, if the iid noise $\{Z_t\}$ were $\mathrm{N}(0,\sigma^2)$ (instead of in the domain of attraction of a non-Gaussian stable law), then for any $j>0$, it would be the case that

$$0 = E\{U_tX_{t-j}\} = \mathrm{E}\left\{\sum_{l=-\infty}^{0}\sum_{m=-\infty}^{\infty}\bar\pi_l\psi_m Z_{t-l}Z_{t-j-m}\right\} = \sigma^2\sum_{l=-\infty}^{0}\bar\pi_l\psi_{l-j}.$$

It follows that, for any $j>0$, $\sum_{l=-\infty}^{0}\bar\pi_l\psi_{l-j}$ must equal zero, and so, for $\{Z_t\}$ in a stable domain of attraction,

$$\begin{aligned}\sum_{t=k+1}^{n}\sum_{l=-\infty}^{0}\bar{\pi}_l\psi_{l-j}Z_{t-l}^2 &= \sum_{t=k+1}^{n}\sum_{l=-t+k+1}^{0}\bar{\pi}_l\psi_{l-j}Z_t^2+\sum_{t=n+1}^{\infty}\sum_{l=-t+k+1}^{-t+n}\bar{\pi}_l\psi_{l-j}Z_t^2\\ &= -\sum_{t=k+1}^{n}\sum_{l=-\infty}^{-t+k}\bar{\pi}_l\psi_{l-j}Z_t^2+\sum_{t=n+1}^{\infty}\sum_{l=-t+k+1}^{-t+n}\bar{\pi}_l\psi_{l-j}Z_t^2. \end{aligned}\tag{25}$$

Since $\{\bar{\pi}_j\}$ and $\{\psi_j\}$ are geometrically decaying, there exist constants $C_1>0$ and $0<D_1<1$ such that $|\bar{\pi}_j|,|\psi_j|\le C_1D_1^{|j|}$ for all j, so the absolute value of (25) is bounded above by

$$\sum_{t=k+1}^{n}\sum_{l=t-k}^{\infty}C_1^2D_1^{2l+j}Z_t^2+\sum_{t=n+1}^{\infty}\sum_{l=t-n}^{\infty}C_1^2D_1^{2l+j}Z_t^2=\frac{C_1^2D_1^j}{1-D_1^2}\left(\sum_{t=k+1}^{n}D_1^{2(t-k)}Z_t^2+\sum_{t=n+1}^{\infty}D_1^{2(t-n)}Z_t^2\right).$$

Hence, equation (23) is bounded above by

$$\begin{aligned}&n^{1/2-\kappa}\left(D_1^1+\cdots+D_1^{K(n)}\right)\frac{C_1^2}{1-D_1^2}\left(D_1^2\sum_{t=p_0+1}^{K(n)}Z_t^2+\sum_{t=K(n)+1}^{n}D_1^{2(t-K(n))}Z_t^2+\sum_{t=n+1}^{\infty}D_1^{2(t-n)}Z_t^2\right)\\ &\quad\le\quad n^{1/2-\kappa}\frac{C_1^2D_1}{(1-D_1)(1-D_1^2)}\left(D_1^2\sum_{t=p_0+1}^{K(n)}Z_t^2+\sum_{t=K(n)+1}^{n}D_1^{2(t-K(n))}Z_t^2+\sum_{t=n+1}^{\infty}D_1^{2(t-n)}Z_t^2\right).\end{aligned}$$

Now, choose $\kappa_1<2/\alpha_0$ and $\lambda_1<\alpha_0/2$ so that $\lambda_1(\kappa_1-1/2)$ is sufficiently close to $(\alpha_0/2)(2/\alpha_0-1/2)=1-\alpha_0/4$ that we have $\lambda_1(\kappa_1-1/2)>1/2$, and let $\epsilon>0$. Since $\lambda_1<1$, using the Markov inequality, we have

$$\begin{aligned}&\mathrm{P}\left(\left[n^{1/2-\kappa_1}\left(D_1^2\sum_{t=p_0+1}^{K(n)}Z_t^2+\sum_{t=K(n)+1}^{n}D_1^{2(t-K(n))}Z_t^2+\sum_{t=n+1}^{\infty}D_1^{2(t-n)}Z_t^2\right)\right]^{\lambda_1}>\epsilon^{\lambda_1}\right)\\ &\quad\le\quad \epsilon^{-\lambda_1}n^{\lambda_1(1/2-\kappa_1)}\left(D_1^{2\lambda_1}\sum_{t=p_0+1}^{K(n)}\mathrm{E}|Z_1|^{2\lambda_1}+\sum_{t=K(n)+1}^{n}D_1^{2\lambda_1(t-K(n))}\mathrm{E}|Z_1|^{2\lambda_1}+\sum_{t=n+1}^{\infty}D_1^{2\lambda_1(t-n)}\mathrm{E}|Z_1|^{2\lambda_1}\right)\\ &\quad\le\quad \epsilon^{-\lambda_1}n^{\lambda_1(1/2-\kappa_1)}\mathrm{E}|Z_1|^{2\lambda_1}\left(K(n)D_1^{2\lambda_1}+\frac{2D_1^{2\lambda_1}}{1-D_1^{2\lambda_1}}\right),\end{aligned}$$

which is $o(1)$ because $\mathrm{E}|Z_1|^{2\lambda_1}<\infty$, $K(n)=O(n^\delta)$, and $n^{\lambda_1(1/2-\kappa_1)+\delta}\le n^{\lambda_1(1/2-\kappa_1)+1/2}\to 0$. Therefore, (23) is $o_p(1)$ for some sufficiently large $\kappa<2/\alpha_0$ when $\alpha_0\le 1$.

Now consider (24), which is bounded above by $n^{1/2-\kappa}\sum_{j=1}^{K(n)}\sum_{t=1}^{n}\sum_{l=-\infty}^{0}\sum_{m\neq l-j}|\bar{\pi}_l\psi_mZ_{t-l}Z_{t-j-m}|$. Since $\alpha_0(2/\alpha_0-1/2)\ge 3/2$ when $\alpha_0\le 1$, we can choose $\kappa_2<2/\alpha_0$ and $\lambda_2<\alpha_0$ so that $\lambda_2(\kappa_2-1/2)$ is sufficiently close to $\alpha_0(2/\alpha_0-1/2)$ that $1+\delta<\lambda_2(\kappa_2-1/2)<\alpha_0(2/\alpha_0-1/2)$. It follows that

$$n^{\lambda_2(1/2-\kappa_2)}K(n)\,n\mathrm{E}|Z_1Z_2|^{\lambda_2}\sum_{l=-\infty}^{0}\sum_{m=-\infty}^{\infty}|\bar{\pi}_l\psi_m|^{\lambda_2}\to 0,$$

since $n^{\lambda_2(1/2-\kappa_2)+\delta+1}\to 0$, $\mathrm{E}|Z_1Z_2|^{\lambda_2}<\infty$, and $\sum_{l=-\infty}^{0}\sum_{m=-\infty}^{\infty}|\bar{\pi}_l\psi_m|^{\lambda_2}<\infty$. Consequently, (24) is also $o_p(1)$ for sufficiently large κ when $\alpha_0\le 1$.

(ii) We now consider the case $\alpha_0 > 1$, and let $\tilde{Z}_t = Z_t - \mathrm{E}Z_t$, so $U_t - \mu_U = \sum_{j=-\infty}^{0} \bar{\pi}_j \tilde{Z}_{t-j}$ and $X_t - \mu_X = \sum_{j=-\infty}^{\infty} \psi_j \tilde{Z}_{t-j}$. It follows that $\max_{p_0 \le k \le K(n)} n^{1/2-\kappa} \|[\sum_{t=k+1}^{n} (U_t - \mu_U)(X_{t-j} - \mu_X)]_{j\in\{1,\ldots,k\}}\|$ is bounded above by

$$\max_{p_0 \le k \le K(n)} n^{1/2-\kappa} \left\| \left[\sum_{t=k+1}^{n} \sum_{l=-\infty}^{0} \bar{\pi}_l \psi_{l-j} \tilde{Z}_{t-l}^2 \right]_{j\in\{1,\ldots,k\}} \right\| \tag{26}$$

$$+ \max_{p_0 \le k \le K(n)} n^{1/2-\kappa} \left\| \left[\sum_{t=1}^{n} \sum_{l=-\infty}^{0} \sum_{m \ne l-j} \bar{\pi}_l \psi_m \left(\tilde{Z}_{t-l} \tilde{Z}_{t-j-m} I\{|\tilde{Z}_{t-l}\tilde{Z}_{t-j-m}| \le n^{1/\alpha_0}\} - \mathrm{E}\left\{ \tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\}\right\}\right) \right]_{j\in\{1,\ldots,k\}} \right\| \tag{27}$$

$$+ \max_{p_0 \le k \le K(n)} n^{1/2-\kappa} \left\| \left[\sum_{t=1}^{n} \sum_{l=-\infty}^{0} \sum_{m \ne l-j} \bar{\pi}_l \psi_m \left(\tilde{Z}_{t-l} \tilde{Z}_{t-j-m} I\{|\tilde{Z}_{t-l}\tilde{Z}_{t-j-m}| > n^{1/\alpha_0}\} - \mathrm{E}\left\{ \tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| > n^{1/\alpha_0}\}\right\}\right) \right]_{j\in\{1,\ldots,k\}} \right\| \tag{28}$$

$$+ \max_{p_0 \le k \le K(n)} n^{1/2-\kappa} \left\| \left[\sum_{t=1}^{k} \sum_{l=-\infty}^{0} \sum_{m \ne l-j} \bar{\pi}_l \psi_m \tilde{Z}_{t-l} \tilde{Z}_{t-j-m} \right]_{j\in\{1,\ldots,k\}} \right\|. \tag{29}$$

Following the same proof used to show that (23) is $o_p(1)$ for large $\kappa < 2/\alpha_0$, one can also show that (26) is $o_p(1)$ for large $\kappa < 2/\alpha_0$ when $\alpha_0 > 1$. So, to complete the proof of (ii), we show that, when $\alpha_0 > 1$, (27), (28), and (29) are $o_p(1)$ for sufficiently large $\kappa < 2/\alpha_0$.

The expected value of the square of (27) is bounded above by

$$n^{1-2\kappa} \sum_{j=1}^{K(n)} \mathrm{E} \left\{ \sum_{t=1}^{n} \sum_{l=-\infty}^{0} \sum_{m \ne l-j} \bar{\pi}_l \psi_m \left(\tilde{Z}_{t-l} \tilde{Z}_{t-j-m} I\{|\tilde{Z}_{t-l}\tilde{Z}_{t-j-m}| \le n^{1/\alpha_0}\} - \mathrm{E}\left\{ \tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\}\right\}\right) \right\}^2 \tag{30}$$

and, since $\{\bar{\pi}_j\}$ and $\{\psi_j\}$ are absolutely summable, $\{\tilde{Z}_t\}$ is iid, and

$$\mathrm{E}\left(\tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\} - \mathrm{E}\left\{ \tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\}\right\}\right) = 0,$$

(30) is bounded above by

$$\begin{aligned}(constant) n^{1-2\kappa} K(n)\, n \Big[&\mathrm{E}\left(\tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\} - \mathrm{E}\left\{ \tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\}\right\}\right)^2 \\ &+ \mathrm{E}\Big| \left(\tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\} - \mathrm{E}\left\{ \tilde{Z}_1 \tilde{Z}_2 I\{|\tilde{Z}_1 \tilde{Z}_2| \le n^{1/\alpha_0}\}\right\}\right) \\ &\times \left(\tilde{Z}_2 \tilde{Z}_3 I\{|\tilde{Z}_2 \tilde{Z}_3| \le n^{1/\alpha_0}\} - \mathrm{E}\left\{ \tilde{Z}_2 \tilde{Z}_3 I\{|\tilde{Z}_2 \tilde{Z}_3| \le n^{1/\alpha_0}\}\right\}\right) \Big| \Big]\end{aligned}$$

$$\leq \quad 2(constant)n^{1-2\kappa}K(n)\, n\mathrm{E}\left(\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| \leq n^{1/\alpha_0}\} - \mathrm{E}\left\{\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| \leq n^{1/\alpha_0}\}\right\}\right)^2$$

$$\leq \quad 2(constant)n^{1-2\kappa}K(n)\, n\mathrm{E}\left(\tilde{Z}_1^2\tilde{Z}_2^2 I\{|\tilde{Z}_1\tilde{Z}_2| \leq n^{1/\alpha_0}\}\right). \tag{31}$$

By Theorem 3.3(ii) in Cline (1983, page 80), the distribution for $|\tilde{Z}_1\tilde{Z}_2|$ is in the domain of attraction of a stable law with exponent α_0, so, for $x > 0$, $\mathrm{P}(|\tilde{Z}_1\tilde{Z}_2| > x) = x^{-\alpha_0}L_1(x)$, where L_1 is slowly varying at ∞. Therefore, by Karamata's Theorem (see, for example, Feller, 1971, page 283), $\mathrm{E}(\tilde{Z}_1^2\tilde{Z}_2^2 I\{|\tilde{Z}_1\tilde{Z}_2| \leq n^{1/\alpha_0}\}) \sim n^{2/\alpha_0 - 1}L_2(n)$, for some slowly varying function L_2. Now, choose $\kappa_3 < 2/\alpha_0$ sufficiently large so that $2\kappa_3 - 1 > \delta + 2/\alpha_0$, which is possible because $\delta + 2/\alpha_0 < 1 - \alpha_0/2 + 2/\alpha_0 < 4/\alpha_0 - 1$. Since $n^{1-2\kappa_3+\delta+2/\alpha_0} \to 0$ and $L_2(n)n^{-\epsilon} \to 0$ for any $\epsilon > 0$ (Feller, 1971, page 277, Lemma 2), when $\kappa = \kappa_3$, (31) is $o(1)$. Thus, for large $\kappa < 2/\alpha_0$, (27) is $o_p(1)$.

Now let $V_{t_1,t_2,n} = \tilde{Z}_{t_1}\tilde{Z}_{t_2} I\{|\tilde{Z}_{t_1}\tilde{Z}_{t_2}| > n^{1/\alpha_0}\} - \mathrm{E}\{\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| > n^{1/\alpha_0}\}\}$. In order to prove that (28) is $o_p(1)$ for large κ, we consider

$$n^{1-2\kappa}\sum_{j=1}^{K(n)}\left[\sum_{t=1}^{n}\sum_{l=-\infty}^{0}\sum_{m\neq l-j}\bar{\pi}_l\psi_m V_{t-l,t-j-m,n}\right]^2$$

$$= \quad n^{1-2\kappa} \sum_{\substack{j=1 \\ [t_1-l_1\notin\{t_2-l_2,t_2-j-m_2\}]\cap[t_1-j-m_1\notin\{t_2-l_2,t_2-j-m_2\}]}}^{K(n)} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{l_1=-\infty}^{0}\sum_{l_2=-\infty}^{0}\sum_{m_1\neq l_1-j}\sum_{m_2\neq l_2-j} \bar{\pi}_{l_1}\bar{\pi}_{l_2}\psi_{m_1}\psi_{m_2}V_{t_1-l_1,t_1-j-m_1,n}V_{t_2-l_2,t_2-j-m_2,n} \tag{32}$$

$$+\, n^{1-2\kappa} \sum_{\substack{j=1 \\ [t_1-l_1\in\{t_2-l_2,t_2-j-m_2\}]\cup[t_1-j-m_1\in\{t_2-l_2,t_2-j-m_2\}]}}^{K(n)} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{l_1=-\infty}^{0}\sum_{l_2=-\infty}^{0}\sum_{m_1\neq l_1-j}\sum_{m_2\neq l_2-j} \bar{\pi}_{l_1}\bar{\pi}_{l_2}\psi_{m_1}\psi_{m_2}V_{t_1-l_1,t_1-j-m_1,n}V_{t_2-l_2,t_2-j-m_2,n} \tag{33}$$

and show there exist values of $\kappa < 2/\alpha_0$ for which (32) and (33) are $o_p(1)$. First, observe that the expected value of the absolute value of (32) is bounded above by $(constant)n^{1-2\kappa}K(n)\,n^2(\mathrm{E}|\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| > n^{1/\alpha_0}\}|)^2$. By Karamata's Theorem, $\mathrm{E}|\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| > n^{1/\alpha_0}\}| \sim n^{1/\alpha_0 - 1}L_3(n)$ for some slowly varying function L_3. Therefore, since $n^{1-2\kappa_3+\delta+2/\alpha_0} \to 0$, we have $n^{1-2\kappa_3}K(n)\,n^2(\mathrm{E}|\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| > n^{1/\alpha_0}\}|)^2 \to 0$, and so (32) is $o_p(1)$ for large $\kappa < 2/\alpha_0$. Next, choose $\kappa_4 < 2/\alpha_0$ and $\lambda_3 < \alpha_0/2 < 1$ sufficiently large so that $\lambda_3(2\kappa_4 - 1) > \delta + 1$. This is possible because $(\alpha_0/2)(4/\alpha_0 - 1) = 2 - \alpha_0/2 > \delta + 1$. Since $n^{\lambda_3(1-2\kappa_4)+\delta+1} \to 0$,

$\mathrm{E}|\tilde{Z}_1|^{2\lambda_3} < \infty$, and $\mathrm{E}|\tilde{Z}_1| < \infty$,

$$\begin{aligned}
& n^{\lambda_3(1-2\kappa_4)} K(n)\, n\mathrm{E}|V_{1,2,n}|^{2\lambda_3} \\
= \;& n^{\lambda_3(1-2\kappa_4)} K(n)\, n\mathrm{E}\Big|\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| > n^{1/\alpha_0}\} - \mathrm{E}\left\{\tilde{Z}_1\tilde{Z}_2 I\{|\tilde{Z}_1\tilde{Z}_2| > n^{1/\alpha_0}\}\right\}\Big|^{2\lambda_3} \\
\leq \;& 2n^{\lambda_3(1-2\kappa_4)} K(n)\, n\left[\left(\mathrm{E}|\tilde{Z}_1|^{2\lambda_3}\right)^2 + \left(\mathrm{E}|\tilde{Z}_1|\right)^{4\lambda_3}\right] \\
\to \;& 0.
\end{aligned}$$

It follows that (33) is $o_p(1)$ for large $\kappa < 2/\alpha_0$.

Finally, we consider equation (29), which is bounded above by

$$n^{1/2-\kappa} \sum_{j=1}^{K(n)} \sum_{t=1}^{K(n)} \sum_{l=-\infty}^{0} \sum_{m \neq l-j} |\bar{\pi}_l \psi_m \tilde{Z}_{t-l} \tilde{Z}_{t-j-m}|, \tag{34}$$

and we choose $\kappa_5 < 2/\alpha_0$ so that $\kappa_5 - 1/2 > 2 - \alpha_0 > 2\delta$. Since $n^{1/2-\kappa_5+2\delta} \to 0$ and $\mathrm{E}|\tilde{Z}_1\tilde{Z}_2| < \infty$, when $\kappa = \kappa_5$, the expected value of (34) is $o(1)$, and therefore (29) is $o_p(1)$ for sufficiently large κ. □

Lemma 2. *(i) If $\alpha_0 \leq 1$, then $\max_{p_0 \leq k \leq K(n)} n^{1/2-\kappa}|\overline{X}|\|[\sum_{t=k+1}^{n} U_t]_{j\in\{1,\ldots,k\}}\|$ and $\max_{p_0 \leq k \leq K(n)} n^{1/2-\kappa}|\overline{X}|\|[\sum_{t=k+1}^{n} X_{t-j}]_{j\in\{1,\ldots,k\}}\|$ converge in probability to zero for sufficiently large $\kappa < 2/\alpha_0$ and, (ii) if $\alpha_0 > 1$, then $\max_{p_0 \leq k \leq K(n)} n^{1/2-\kappa}|\overline{X} - \mu_X|\|[\sum_{t=k+1}^{n} (U_t - \mu_U)]_{j\in\{1,\ldots,k\}}\|$ and $\max_{p_0 \leq k \leq K(n)} n^{1/2-\kappa}|\overline{X} - \mu_X|\|[\sum_{t=k+1}^{n} (X_{t-j} - \mu_X)]_{j\in\{1,\ldots,k\}}\|$ converge in probability to zero for large $\kappa < 2/\alpha_0$.*

Proof. (i) When $\alpha_0 \leq 1$, $\sum_{t=1}^{n} |U_t|$ and $\sum_{t=1}^{n} |X_t|$ are $o_p(n^{1/\alpha_0+\epsilon})$ for any $\epsilon > 0$ (Davis and Resnick, 1985a, Section 4). Therefore, if we choose $\kappa_6 < 2/\alpha_0$ sufficiently large so that $\kappa_6 + 1/2 > \delta + 2/\alpha_0$, then

$$\max_{p_0 \leq k \leq K(n)} n^{1/2-\kappa_6} \,|\overline{X}| \left\| \left[\sum_{t=k+1}^{n} U_t\right]_{j\in\{1,\ldots,k\}} \right\| \leq n^{-\kappa_6 - 1/2} K(n) \left(\sum_{t=1}^{n} |X_t|\right)\left(\sum_{t=1}^{n} |U_t|\right) \xrightarrow{P} 0.$$

It can be shown similarly that $\max_{p_0 \leq k \leq K(n)} n^{1/2-\kappa}|\overline{X}|\|[\sum_{t=k+1}^{n} X_{t-j}]_{j\in\{1,\ldots,k\}}\| \xrightarrow{P} 0$ for large $\kappa < 2/\alpha_0$.

(ii) This result can be established using the fact that, when $\alpha_0 > 1$, $\overline{X} - \mu_X = o_p(n^{1/\alpha_0 - 1+\epsilon})$ for any $\epsilon > 0$ (Davis and Resnick, 1985a, Section 4). We omit the details. □

Lemma 3. *When $\alpha_0 \leq 1$, $n^{1/2-\kappa}K(n)n\overline{X}^2 \xrightarrow{P} 0$ for sufficiently large $\kappa < 2/\alpha_0$ and, when $\alpha_0 > 1$, $n^{1/2-\kappa}K(n)n(\overline{X} - \mu_X)^2 \xrightarrow{P} 0$ for large $\kappa < 2/\alpha_0$.*

Proof. When $\alpha_0 \leq 1$, $\overline{X} = o_p(n^{1/\alpha_0 - 1 + \epsilon})$ for any $\epsilon > 0$ and, when $\alpha_0 > 1$, $\overline{X} - \mu_X = o_p(n^{1/\alpha_0 - 1 + \epsilon})$ for any $\epsilon > 0$. The results of this lemma follow. □

Finally, we give Lemma 4, which was used in the proof of Theorem 4. For the proof of this lemma, we assume the distribution for the iid noise $\{Z_t\}$ is stable with parameter vector $\boldsymbol{\vartheta}_0 = (\alpha_0, \beta_0, \sigma_0, \mu_0)'$, since that is also assumed in the proof of Theorem 4. It follows that the scale-transformed series $\{z_t\} = \{-\theta_{0s_0} Z_{t-p_0+s_0}\}$ is iid stable with parameter $\boldsymbol{\tau}_0 = (\alpha_0, -(\text{sign } \theta_{0s_0})\beta_0, |\theta_{0s_0}|\sigma_0, -\theta_{0s_0}\mu_0)'$.

Lemma 4. *For W_n and $W_n^\dagger$ defined in equations (19) and (20) respectively, $W_n(\cdot,\cdot,\cdot) - W_n^\dagger(\cdot,\cdot,\cdot) \xrightarrow{P} 0$ on $C(I\!R^{2p_0+4})$ as $n \to \infty$.*

But before proving this result, we look at partial and mixed partial derivatives for the residuals $\{z_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p)\}$ in (14), which are used in the proof. First, for $i \in \{1, \ldots, p\}$, note that

$$\frac{\partial z_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p)}{\partial \eta_i} = \begin{cases} 0, & t \leq 2p, \\ \theta_1 \frac{\partial z_{t-1}(\boldsymbol{\eta}, \boldsymbol{\theta}, p)}{\partial \eta_i} + \cdots + \theta_p \frac{\partial z_{t-p}(\boldsymbol{\eta}, \boldsymbol{\theta}, p)}{\partial \eta_i} - B^p \theta(B^{-1}) X_{t-i}, & t = 2p+1, \ldots, n, \end{cases} \tag{35}$$

and

$$\frac{\partial z_t(\boldsymbol{\eta}, \boldsymbol{\theta}, p)}{\partial \theta_i} = \begin{cases} 0, & t \leq 2p, \\ \theta_1 \frac{\partial z_{t-1}(\boldsymbol{\eta}, \boldsymbol{\theta}, p)}{\partial \theta_i} + \cdots + \theta_p \frac{\partial z_{t-p}(\boldsymbol{\eta}, \boldsymbol{\theta}, p)}{\partial \theta_i} + z_{t-i}(\boldsymbol{\eta}, \boldsymbol{\theta}, p) - \eta(B) B^p X_{t+i}, & t = 2p+1, \ldots, n. \end{cases} \tag{36}$$

Since $1/\theta_0^c(z) = 1/(1 - \theta_{01} z - \cdots - \theta_{0p_0} z^{p_0}) = \sum_{j=0}^{\infty} \chi_j z^j$ and, from (13), $\eta_0(B) B^{p_0} \theta_0^c(B^{-1}) X_t = \theta_0^c(B) z_t$, if we evaluate (35) and (36) at the true parameter values and then ignore the recursion initialization, for $t \in \{2p_0+1, \ldots, n\}$, we have

$$\frac{\partial z_t(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0, p_0)}{\partial \eta_i} = -\sum_{j=0}^{t-2p_0-1} \chi_j B^{p_0} \theta_0^c(B^{-1}) X_{t-i-j} = -\sum_{j=0}^{t-2p_0-1} \chi_j \frac{\theta_0^c(B)}{\eta_0(B)} z_{t-i-j} \approx -\frac{z_{t-i}}{\eta_0(B)}$$

and

$$\begin{aligned} \frac{\partial z_t(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0, p_0)}{\partial \theta_i} &= \sum_{j=0}^{t-2p_0-1} \chi_j \left[z_{t-i-j}(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0, p_0) - \eta_0(B) B^{p_0} X_{t+i-j} \right] \\ &= \sum_{j=0}^{t-2p_0-1} \chi_j z_{t-i-j}(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0, p_0) - \sum_{j=0}^{t-2p_0-1} \chi_j \frac{\theta_0^c(B)}{\theta_0^c(B^{-1})} z_{t+i-j} \\ &\approx \frac{z_{t-i}}{\theta_0^c(B)} - \frac{z_{t+i}}{\theta_0^c(B^{-1})}. \end{aligned}$$

Hence, following (17), for $\mathbf{u},\mathbf{v}\in\mathbb{R}^{p_0}$, $\mathbf{u}'[\partial z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)/\partial\boldsymbol{\eta}]+\mathbf{v}'[\partial z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)/\partial\boldsymbol{\theta}]\approx\sum_{j\neq 0}c_j(\mathbf{u},\mathbf{v})z_{t-j}$. In particular, since $\{\chi_j\}$ is geometrically decaying, it can be shown that, for any $T>0$, there exist constants $C_2>0$ and $D_2\in(0,1)$ such that

$$\sup_{\mathbf{u},\mathbf{v}\in[-T,T]^{p_0}}\left|\mathbf{u}'\frac{\partial z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)}{\partial\boldsymbol{\eta}}+\mathbf{v}'\frac{\partial z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)}{\partial\boldsymbol{\theta}}-\sum_{j\neq 0}c_j(\mathbf{u},\mathbf{v})z_{t-j}\right|\leq C_2D_2^t\sum_{j=0}^{\infty}D_2^j\left(|X_{2p_0-j}|+|z_{2p_0-j}|\right) \tag{37}$$

for all $t\geq 2p_0+1$. Next, for $i,j\in\{1,\ldots,p\}$, note that $\partial^2z_t(\boldsymbol{\eta},\boldsymbol{\theta},p)/(\partial\eta_i\partial\eta_j)=0\ \forall t$. Additionally, for $t\in\{2p+1,\ldots,n\}$, we have the recursions

$$\frac{\partial^2z_t(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\eta_i\partial\theta_j}=\theta_1\frac{\partial^2z_{t-1}(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\eta_i\partial\theta_j}+\cdots+\theta_p\frac{\partial^2z_{t-p}(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\eta_i\partial\theta_j}+\frac{\partial z_{t-j}(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\eta_i}+B^pX_{t-i+j}$$

and

$$\frac{\partial^2z_t(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\theta_i\partial\theta_j}=\theta_1\frac{\partial^2z_{t-1}(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\theta_i\partial\theta_j}+\cdots+\theta_p\frac{\partial^2z_{t-p}(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\theta_i\partial\theta_j}+\frac{\partial z_{t-j}(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\theta_i}+\frac{\partial z_{t-i}(\boldsymbol{\eta},\boldsymbol{\theta},p)}{\partial\theta_j},$$

with $\partial^2z_t(\boldsymbol{\eta},\boldsymbol{\theta},p)/(\partial\eta_i\partial\theta_j)=\partial^2z_t(\boldsymbol{\eta},\boldsymbol{\theta},p)/(\partial\theta_i\partial\theta_j)=0$ for $t\leq 2p$. It can therefore be shown that there exists an $\epsilon>0$ and constants $C_3>0$ and $D_3\in(0,1)$ such that

$$\sup_{\|\boldsymbol{\eta}-\boldsymbol{\eta}_0\|,\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\epsilon}\left|\frac{\partial^2z_t(\boldsymbol{\eta},\boldsymbol{\theta},p_0)}{\partial\eta_i\partial\theta_j}\right|+\sup_{\|\boldsymbol{\eta}-\boldsymbol{\eta}_0\|,\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\epsilon}\left|\frac{\partial^2z_t(\boldsymbol{\eta},\boldsymbol{\theta},p_0)}{\partial\theta_i\partial\theta_j}\right|\leq C_3\sum_{j=0}^{\infty}D_3^j|X_{t-j}| \tag{38}$$

for all $i,j\in\{1,\ldots,p_0\}$ and $t\geq 2p_0+1$.

Proof of Lemma 4. For $\mathbf{u},\mathbf{v}\in\mathbb{R}^{p_0}$ and $\mathbf{w}\in\mathbb{R}^4$, note that $W_n(\mathbf{u},\mathbf{v},\mathbf{w})-W_n^{\dagger}(\mathbf{u},\mathbf{v},\mathbf{w})$ equals

$$\sum_{t=2p_0+1}^{n}\left[\ln f\left(z_t\left(\boldsymbol{\eta}_0+\frac{\mathbf{u}}{n^{1/\alpha_0}},\boldsymbol{\theta}_0+\frac{\mathbf{v}}{n^{1/\alpha_0}},p_0\right);\boldsymbol{\tau}_0+\frac{\mathbf{w}}{\sqrt{n}}\right)\right.$$
$$\left.-\ln f\left(z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)+\frac{1}{n^{1/\alpha_0}}\sum_{j\neq 0}c_j(\mathbf{u},\mathbf{v})z_{t-j};\boldsymbol{\tau}_0+\frac{\mathbf{w}}{\sqrt{n}}\right)\right] \tag{39}$$
$$+\sum_{t=2p_0+1}^{n}\left[\ln f\left(z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)+\frac{1}{n^{1/\alpha_0}}\sum_{j\neq 0}c_j(\mathbf{u},\mathbf{v})z_{t-j};\boldsymbol{\tau}_0+\frac{\mathbf{w}}{\sqrt{n}}\right)-\ln f(z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0);\boldsymbol{\tau}_0)\right.$$
$$\left.-\ln f\left(z_t+\frac{1}{n^{1/\alpha_0}}\sum_{j\neq 0}c_j(\mathbf{u},\mathbf{v})z_{t-j};\boldsymbol{\tau}_0+\frac{\mathbf{w}}{\sqrt{n}}\right)+\ln f(z_t;\boldsymbol{\tau}_0)\right]. \tag{40}$$

We first consider equation (39), which can be expressed as

$$\sum_{t=2p_0+1}^{n}\left\{\frac{\partial\ln f(z_{t,n}^*(\mathbf{u},\mathbf{v},\mathbf{w});\boldsymbol{\tau}_0+\mathbf{w}/\sqrt{n})}{\partial z}\right.$$
$$\left.\times\left[z_t\left(\boldsymbol{\eta}_0+\frac{\mathbf{u}}{n^{1/\alpha_0}},\boldsymbol{\theta}_0+\frac{\mathbf{v}}{n^{1/\alpha_0}},p_0\right)-z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)-\frac{1}{n^{1/\alpha_0}}\sum_{j\neq 0}c_j(\mathbf{u},\mathbf{v})z_{t-j}\right]\right\}, \tag{41}$$

with $z^*_{t,n}(\mathbf{u},\mathbf{v},\mathbf{w})$ between $z_t(\boldsymbol{\eta}_0+n^{-1/\alpha_0}\mathbf{u},\boldsymbol{\theta}_0+n^{-1/\alpha_0}\mathbf{v},p_0)$ and $z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)+n^{-1/\alpha_0}\sum_{j\neq 0}c_j(\mathbf{u},\mathbf{v})z_{t-j}$, and we note that $z_t(\boldsymbol{\eta}_0+n^{-1/\alpha_0}\mathbf{u},\boldsymbol{\theta}_0+n^{-1/\alpha_0}\mathbf{v},p_0)$ also equals

$$z_t\left(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0\right)+\frac{1}{n^{1/\alpha_0}}\left(\mathbf{u}',\mathbf{v}'\right)\begin{bmatrix}\frac{\partial z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)}{\partial\boldsymbol{\eta}}\\ \frac{\partial z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)}{\partial\boldsymbol{\theta}}\end{bmatrix}$$
$$+\quad\frac{1}{2n^{2/\alpha_0}}(\mathbf{u}',\mathbf{v}')\begin{bmatrix}\frac{\partial^2 z_t(\boldsymbol{\eta}^*_{t,n}(\mathbf{u},\mathbf{v}),\boldsymbol{\theta}^*_{t,n}(\mathbf{u},\mathbf{v}),p_0)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'} & \frac{\partial^2 z_t(\boldsymbol{\eta}^*_{t,n}(\mathbf{u},\mathbf{v}),\boldsymbol{\theta}^*_{t,n}(\mathbf{u},\mathbf{v}),p_0)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\theta}'}\\ \frac{\partial^2 z_t(\boldsymbol{\eta}^*_{t,n}(\mathbf{u},\mathbf{v}),\boldsymbol{\theta}^*_{t,n}(\mathbf{u},\mathbf{v}),p_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\eta}'} & \frac{\partial^2 z_t(\boldsymbol{\eta}^*_{t,n}(\mathbf{u},\mathbf{v}),\boldsymbol{\theta}^*_{t,n}(\mathbf{u},\mathbf{v}),p_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\end{bmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix},$$

where $\boldsymbol{\eta}^*_{t,n}(\mathbf{u},\mathbf{v})$ is between $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_0+n^{-1/\alpha_0}\mathbf{u}$, and $\boldsymbol{\theta}^*_{t,n}(\mathbf{u},\mathbf{v})$ is between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0+n^{-1/\alpha_0}\mathbf{v}$. Now, let $T>0$. Since $\sup_{z\in\mathbb{R},\,\mathbf{w}\in[-T,T]^4}|\partial\ln f(z;\boldsymbol{\tau}_0+\mathbf{w}/\sqrt{n})/\partial z|=O(1)$ as $n\to\infty$ (Andrews et al., 2009, Section 2), it follows from (37) and (38) that, for all n sufficiently large, $\sup_{(\mathbf{u}',\mathbf{v}',\mathbf{w}')'\in[-T,T]^{2p_0+4}}|\cdot|$ of (41) is bounded above by

$$(constant)\left[\frac{1}{n^{1/\alpha_0}}\sum_{t=2p_0+1}^{n}D_2^t\sum_{j=0}^{\infty}D_2^j\left(|X_{2p_0-j}|+|z_{2p_0-j}|\right)+\frac{1}{n^{2/\alpha_0}}\sum_{t=2p_0+1}^{n}\sum_{j=0}^{\infty}D_3^j|X_{t-j}|\right]$$

for some $D_2,D_3\in(0,1)$. Therefore, because $\sum_{t=2p_0+1}^{n}D_2^t\sum_{j=0}^{\infty}D_2^j(|X_{2p_0-j}|+|z_{2p_0-j}|)=O_p(1)$ and $n^{-2/\alpha_0}\sum_{t=2p_0+1}^{n}\sum_{j=0}^{\infty}D_3^j|X_{t-j}|\xrightarrow{P}0$ (Davis and Resnick, 1985a, Section 4), (41) and hence also (39) must be $o_p(1)$ on $C([-T,T]^{2p_0+4})$. Using the fact that, for some $C_4>0$ and $D_4\in(0,1)$, $|z_t(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0,p_0)-z_t|\le C_4D_4^t\sum_{j=0}^{\infty}D_4^j|X_{2p_0-j}|\ \forall t\ge 2p_0+1$, it can also be shown that (40) is $o_p(1)$ on $C([-T,T]^{2p_0+4})$. Since $T>0$ is arbitrary, it follows that (39) and (40) are $o_p(1)$ on $C(K)$ for any compact set $K\subset\mathbb{R}^{2p_0+4}$, and thus $W_n(\cdot,\cdot,\cdot)-W_n^{\dagger}(\cdot,\cdot,\cdot)\xrightarrow{P}0$ on $C(\mathbb{R}^{2p_0+4})$. □

Acknowledgements We gratefully acknowledge the support of NSF Grants DMS0806104 (Andrews) and DMS0743459 (Davis), and thank two anonymous reviewers for their helpful comments.

References

Adler, R.J., Feldman, R.E., and Gallagher, C., 1998, Analysing stable time series. In: R.J. Adler, R.E. Feldman, and M.S. Taqqu (Eds.), A practical guide to heavy tails. Birkhäuser, Boston, pp. 133–158.

Andrews, B., Calder, M., and Davis, R.A., 2009, Maximum likelihood estimation for α-stable autoregressive processes. Annals of Statistics 37, 1946–1982.

Andrews, B., Davis, R.A., and Breidt, F.J., 2006, Maximum likelihood estimation for all-pass time series models. Journal of Multivariate Analysis 97, 1638–1659.

Andrews, B., Davis, R.A., and Breidt, F.J., 2007, Rank-based estimation for all-pass time series models. Annals of Statistics 35, 844–869.

Blass, W.E. and Halsey, G.W., 1981, Deconvolution of absorption spectra. Academic Press, New York.

Breidt, F.J. and Davis, R.A., 1992, Time-reversibility, identifiability and independence of innovations for stationary time series. Journal of Time Series Analysis 13, 377–390.

Breidt, F.J., Davis, R.A., Lii, K.-S., and Rosenblatt, M., 1991, Maximum likelihood estimation for noncausal autoregressive processes. Journal of Multivariate Analysis 36, 175–198.

Breidt, F.J., Davis, R.A., and Trindade, A.A., 2001, Least absolute deviation estimation for all-pass time series models. Annals of Statistics 29, 919–946.

Brockwell, P.J. and Davis, R.A., 1991, Time series: theory and methods, 2nd ed. Springer-Verlag, New York.

Chi, C.-Y. and Kung, J.-Y., 1995, A new identification algorithm for allpass systems by higher-order statistics. Signal Processing 41, 239–256.

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 Transactions on Signal Processing 45, 1742–1762.

Cline, D.B.H., 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, Colorado.

Davis, R.A., Knight, K., and Liu, J., 1992, M-estimation for autoregressions with infinite variance. Stochastic Processes and their Applications 40, 145–180.

Davis, R. and Resnick, S., 1985a, Limit theory for moving averages of random variables with regularly varying tail probabilities. Annals of Probability 13, 179–195.

Davis, R. and Resnick, S., 1985b, More limit theory for the sample correlation function of moving averages. Stochastic Processes and their Applications 20, 257–279.

Davis, R. and Resnick, S., 1986, Limit theory for the sample covariance and correlation functions of moving averages. Annals of Statistics 14, 533–558.

Davis, R.A. and Wu, W., 1997, Bootstrapping M-estimates in regression and autoregression with infinite variance. Statistica Sinica 7, 1135–1154.

Donoho, D., 1981, On minimum entropy deconvolution. In: D.R. Findley (Ed.), Applied time series analysis II. Academic Press, New York, pp. 565–608.

DuMouchel, W.H., 1973, On the asymptotic normality of the maximum likelihood estimate when sampling from a stable distribution. Annals of Statistics 1, 948–957.

Feller, W., 1971, An introduction to probability theory and its applications, vol. 2, 2nd ed. Wiley, New York.

Gallagher, C., 2001, A method for fitting stable autoregressive models using the autocovariation function. Statistics and Probability Letters 53, 381–390.

Giannakis, G.B. and Swami, A., 1990, On estimating noncausal nonminimum phase ARMA models of non-Gaussian processes. IEEE Transactions on Acoustics, Speech, and Signal Processing 38, 478–495.

Gnedenko, B.V. and Kolmogorov, A.N., 1968, Limit distributions for sums of independent random variables. (Translated by K.L. Chung.) Addison-Wesley, Reading, MA.

Hannan, E.J., 1980, The estimation of the order of an ARMA process. Annals of Statistics 8, 1071–1081.

Knight, K., 1989, Consistency of Akaike's information criterion for infinite variance autoregressive processes. Annals of Statistics 17, 824–840.

Ling, S., 2005, Self-weighted least absolute deviation estimation for infinite variance autoregressive models. Journal of the Royal Statistical Society: Series B 67, 381–393.

McCulloch, J.H., 1998, Numerical approximation of the symmetric stable distribution and density. In: R.J. Adler, R.E. Feldman, and M.S. Taqqu (Eds.), A practical guide to heavy tails. Birkhäuser, Boston, pp. 489–499.

Nelder, J. A. and Mead, R., 1965, A simplex method for function minimization. Computer Journal 7, 308–313.

Nolan, J.P., 1997, Numerical calculation of stable densities and distribution functions. Communications in Statistics, Stochastic Models 13, 759–774.

Nolan, J.P., 2001, Maximum likelihood estimation and diagnostics for stable distributions. In: O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick (Eds.), Lévy processes: theory and applications. Birkhäuser, Boston, pp. 379–400.

Resnick, S.I., 1997, Heavy tail modeling and teletraffic data. Annals of Statistics 25, 1805–1869.

Samorodnitsky, G. and Taqqu, M.S., 1994, Stable non-Gaussian random processes. Chapman & Hall, New York.

Scargle, J.D., 1981, Phase-sensitive deconvolution to model random processes, with special reference to astronomical data. In: D.R. Findley (Ed.), Applied time series analysis II. Academic Press, New York, pp. 549–564.

Shibata, R., 1976, Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 63, 117–126.

Yamazato, M., 1978, Unimodality of infinitely divisible distribution functions of class L. Annals of Probability 6, 523–531.

Zolotarev, V.M., 1986, One-dimensional stable distributions. American Mathematical Society, Providence, RI.