The Annals of Statistics
2001, Vol. 29, No. 4, 919–946

LEAST ABSOLUTE DEVIATION ESTIMATION FOR ALL-PASS TIME SERIES MODELS[1]

BY F. JAY BREIDT, RICHARD A. DAVIS AND A. ALEXANDRE TRINDADE

Colorado State University, Colorado State University and University of Florida

An autoregressive moving average model in which all of the roots of the autoregressive polynomial are reciprocals of roots of the moving average polynomial and vice versa is called an all-pass time series model. All-pass models generate uncorrelated (white noise) time series, but these series are not independent in the non-Gaussian case. An approximation to the likelihood of the model in the case of Laplacian (two-sided exponential) noise yields a modified absolute deviations criterion, which can be used even if the underlying noise is not Laplacian. Asymptotic normality for least absolute deviation estimators of the model parameters is established under general conditions. Behavior of the estimators in finite samples is studied via simulation. The methodology is applied to exchange rate returns to show that linear all-pass models can mimic "nonlinear" behavior, and is applied to stock market volume data to illustrate a two-step procedure for fitting noncausal autoregressions.

1. Introduction. In the analysis of returns on financial assets such as stocks, it is common to observe lack of serial correlation, heavy-tailed marginal distributions, and volatility clustering. Volatility clustering is the name given to the phenomenon noticed by Mandelbrot (1963), in which small observations tend to be followed by small observations, and large observations by large observations. This kind of dependence is not reflected in the second-order properties of the series, which is serially uncorrelated, but can be detected through the analysis of higher-order moments, such as in the autocorrelations of the squared returns.

Typically, nonlinear models with time-dependent conditional variances, such as the autoregressive conditionally heteroskedastic (ARCH) models [Engle (1982), Bollerslev, Chou and Kroner (1992)] or the stochastic volatility models [Clark (1973), Jacquier, Polson and Rossi (1994)] are suggested for such time series. In this article we consider a class of *linear* models which can also mimic this behavior. Data from these models are serially uncorrelated and can have heavy-tailed marginals. The data are dependent and can display volatility clustering. This class is a particularly striking illustration of a known result that linear, non-Gaussian models can display "nonlinear" behavior [Bickel and Bühlmann (1996)].

The linear models which we will consider are all-pass models: autoregressive moving average models in which all of the roots of the autoregressive

Received September 1999; revised April 2001.
[1]Supported in part by NSF Grant DMS-99-72015.
AMS 2000 *subject classifications*. Primary 62M10; secondary 62E20, 62F10.
Key words and phrases. Laplacian density, noncausal, noninvertible, nonminimum phase, white noise.

polynomial are reciprocals of roots of the moving average polynomial and vice versa. All-pass models generate uncorrelated (white noise) time series, but these series are not independent in the non-Gaussian case.

While all-pass models can generate examples of linear time series with "nonlinear" behavior, their dependence structure is highly constrained, limiting their ability to compete with ARCH. A far more important application of all-pass models is in the fitting of noncausal autoregressions. Noncausal models are important tools in a number of applications, including deconvolution of absorption spectra [Blass and Halsey (1981)], design of communication systems [Benveniste, Goursat and Roget (1980)], processing of blurry images [Donoho (1981), Chien, Yang and Chi (1997)], deconvolution of seismic signals [Wiggins (1978), Ooe and Ulrych (1979), Donoho (1981), Godfrey and Rocca (1981), Hsueh and Mendel (1985)], modeling of vocal tract filters [Rabiner and Schafer (1978), Chien, Yang and Chi (1997)] and analysis of astronomical data [Scargle (1981)].

In many of these applications, the models are essentially one-dimensional random fields, in which the direction of "time" is irrelevant and prediction is not of interest. The form of the predictive density of a future observation given a sample of n consecutive observations is fairly difficult to compute, as it depends on the underlying noise density and on the marginal density of the series (an infinite convolution). We do not discuss prediction for all-pass or noncausal models further in this article. Rosenblatt (2000) is a monograph which covers identification, estimation and prediction aspects of noncausal models.

All-pass models are widely used in the fitting of noncausal models, where they arise as the result of whitening a series with a causal filter (all of the roots of the autoregressive polynomial outside the unit circle) when in fact the true model is noncausal. The whitened series in this case can then be represented as an all-pass of order r, where r is the number of roots of the true autoregressive polynomial which lie inside the unit circle.

Estimation methods based on Gaussian likelihood, least-squares, or related second-order moment techniques are unable to identify all-pass models. Instead, cumulant-based estimators using cumulants of order greater than two are often used to estimate such models [Wiggins (1978), Donoho (1981), Lii and Rosenblatt (1982), Giannakis and Swami (1990), Chi and Kung (1995), Chien, Yang and Chi (1997)].

In this article we consider estimation based on a quasi-likelihood approach. In Section 2, an approximation to the likelihood of an all-pass model in the case of Laplacian (two-sided exponential) noise is derived, yielding a modified absolute deviations criterion. This criterion can be used even if the underlying noise is not Laplacian. Asymptotic normality for least absolute deviation estimators of the model parameters is established under general conditions in Section 3 and order selection is considered. This asymptotic theory relies on two preliminary results stated and proved in the Appendix. The first result extends a theorem of Davis and Dunsmuir (1997) to the case of two-sided linear processes, and the second result uses the first in establishing a functional convergence theorem for the modified absolute deviations criterion.

Behavior of the estimators in finite samples is studied via simulation in Section 4.1. For illustration purposes, the estimation procedure is applied to exchange rate data in Section 4.2 and to noncausal autoregressive modeling in Section 4.3. In the latter, the two-step procedure for fitting noncausal models is applied not to a standard engineering deconvolution problem but to a nonstandard example: time series of daily log volumes of Microsoft stock. A noncausal AR(1) model is shown to provide a reasonable fit to these data. Though the purpose of this example is purely illustrative, it is interesting to note that causal AR models are found to provide better fits for the log volumes of Atmel and Microchip, two smaller companies with considerably less public exposure. A brief discussion follows in Section 5.

2. Preliminaries.

2.1. *All-pass models.* Let B denote the backshift operator ($B^k X_t = X_{t-k}$, $k = 0, \pm 1, \pm 2, \dots$) and let

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_s z^s$$

be an sth-order autoregressive polynomial, where $\phi(z) \neq 0$ for $|z| = 1$. The polynomial is said to be *causal* if all its roots are outside the unit circle in the complex plane. In this case, for a sequence $\{W_t\}$,

$$\phi^{-1}(B) W_t = \left(\sum_{j=0}^{\infty} \psi_j B^j \right) W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j},$$

a function of only the past and present of the $\{W_t\}$. Note that the filter $\phi(B^{-1})$ is *purely noncausal* in the sense that

$$\phi^{-1}(B^{-1}) W_t = \left(\sum_{j=0}^{\infty} \psi_j B^{-j} \right) W_t = \sum_{j=0}^{\infty} \psi_j W_{t+j},$$

a function of only the present and future of the $\{W_t\}$. See, for example, Chapter 3 of Brockwell and Davis (1991).

We introduce notation which will be useful in our later discussion of order selection. Consider the sth-order autoregressive polynomial

$$\phi_0(z) = 1 - \phi_{01} z - \cdots - \phi_{0s} z^s,$$

where $\phi_0(z) \neq 0$ for $|z| \leq 1$ and s is known. Define $\phi_{00} = 1$ and assume

A1. $\phi_{0r} \neq 0$ *for some* $r \in \{0, 1, \dots, s\}$ *and* $\phi_{0j} = 0$ *for* $j = r+1, \dots, s$.

That is, r is the unknown, *real* model order, while s is a known, *sufficiently large* model order. Then a causal all-pass time series is the autoregressive moving average (ARMA) $\{X_t\}$ which satisfies the difference equations

$$\phi_0(B)X_t = \frac{B^s\phi_0(B^{-1})}{-\phi_{0r}} Z_t, \tag{2.1}$$

where $\{Z_t\}$ is an independent and identically distributed (iid) sequence of random variables. In principle, it is possible to consider all-pass models with both causal and noncausal factors. We restrict attention to causal all-pass models because they suffice for our main application: the fitting of noncausal autoregressive models.

We assume

A2. *$\{Z_t\}$ is iid with mean* 0, *finite variance* $\sigma^2 > 0$, *and common distribution function* F_σ.

A3. *F_σ has median zero and is continuously differentiable in a neighborhood of zero. Let $f_\sigma(z) = \sigma^{-1} f(\sigma^{-1} z)$ denote the density function corresponding to F_σ, where σ is a scale parameter.*

A4. $f_\sigma(0) > 0$.

A2 implies that the mean of $\{X_t\}$ in (2.1) is zero. This suffices for the applications we consider, in which $\{X_t\}$ is a zero-mean white noise sequence. In the case of nonzero mean, it is possible to center by subtracting off the sample mean, which is $n^{1/2}$-consistent and asymptotically equivalent to the best linear unbiased estimator [Brockwell and Davis (1991), Section 7.1]. Another possibility is to include the mean when constructing the approximate likelihood. A comparison of these alternatives is beyond the scope of this article.

Note that the spectral density of $\{X_t\}$ in (2.1) is

$$\frac{|e^{-is\omega}|^2|\phi_0(e^{i\omega})|^2}{\phi_{0r}^2|\phi_0(e^{-i\omega})|^2}\frac{\sigma^2}{2\pi} = \frac{\sigma^2}{\phi_{0r}^2 2\pi},$$

which is constant for $\omega \in [-\pi, \pi]$, hence $\{X_t\}$ is an uncorrelated sequence. In the case of Gaussian $\{Z_t\}$, this implies that $\{X_t\}$ is iid $N(0, \sigma^2\phi_{0r}^{-2})$, but independence does not hold in the non-Gaussian case [e.g., Breidt and Davis (1991)].

Rearranging (2.1), we have the backward recursion

$$z_{t-s} = \phi_{01}z_{t-s+1} + \cdots + \phi_{0s}z_t - (X_t - \phi_{01}X_{t-1} - \cdots - \phi_{0s}X_{t-s}), \tag{2.2}$$

where $z_t := Z_t\phi_{0r}^{-1}$. In practice, the model order r is unknown. We *propose* a model order $p \le s$ and a corresponding causal autoregressive polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| \le 1$, where $\phi_p \neq 0$. The analogous recursion to (2.2) is then

$$z_{t-s}(\boldsymbol{\phi}) = \begin{cases} 0, & t = n+s, \ldots, n+1, \\ \phi_1 z_{t-s+1}(\boldsymbol{\phi}) + \cdots + \phi_s z_t(\boldsymbol{\phi}) - \phi(B)X_t, & t = n, \ldots, s+1, \end{cases} \tag{2.3}$$

where the $s \times 1$ vector $\boldsymbol{\phi}$ is defined as $(\phi_1, \ldots, \phi_p, 0, \ldots, 0)'$.

Let $\boldsymbol{\phi}_0=(\phi_{01},\dots,\phi_{0s})'=(\phi_{01},\dots,\phi_{0r},0,\dots,0)'$. Note that $\{z_t(\boldsymbol{\phi}_0)\}$ is a close approximation to $\{z_t\}$, in which the error is due to the initialization with zeros. Though $\{z_t\}$ is iid, $\{z_t(\boldsymbol{\phi})\}$, in general, is not iid, even after ignoring the transient behavior due to initialization.

2.2. *Approximating the likelihood.* The modified absolute deviations criterion we consider is motivated by a likelihood approximation. In this subsection, we ignore the effect of recursion initialization in (2.3), and write

$$-\phi(B^{-1})B^s z_t(\boldsymbol{\phi})=\phi(B)X_t. \tag{2.4}$$

We then approximate the likelihood of a realization of length n, $(X_1,\dots,X_n)$, from the model (2.1) using techniques similar to those in Breidt, Davis, Lii and Rosenblatt (1991) and Lii and Rosenblatt (1992, 1996).

Consider the augmented data vector

$$\mathbf{x}:=(X_{1-s},\dots,X_0,X_1,\dots,X_n,z_{n-s+1}(\boldsymbol{\phi}),\dots,z_n(\boldsymbol{\phi}))'$$

and the augmented noise vector

$$\mathbf{z}:=(X_{1-s},\dots,X_0,z_{1-s}(\boldsymbol{\phi}),\dots,z_0(\boldsymbol{\phi}),z_1(\boldsymbol{\phi}),\dots,z_{n-s+1}(\boldsymbol{\phi}),\dots,z_n(\boldsymbol{\phi}))'.$$

Note that when $\boldsymbol{\phi}=\boldsymbol{\phi}_0$, the first $2s$ terms of $\mathbf{z}$ are independent of the last n terms by causality.

From (2.4), it is easy to show that

$$A\mathbf{x}=B\mathbf{z} \tag{2.5}$$

with $|A|=|B|=1$. Now the joint distribution of $\mathbf{z}$ under $\boldsymbol{\phi}$ is given by

$$\begin{aligned}h(\mathbf{z}) &= h_1(X_{1-s},\dots,X_0,z_{1-s}(\boldsymbol{\phi}),\dots,z_0(\boldsymbol{\phi}))\\ &\quad\times\left(\prod_{t=1}^{n-s} f_\sigma(\phi_p z_t(\boldsymbol{\phi}))|\phi_p|\right)h_2(z_{n-s+1}(\boldsymbol{\phi}),\dots,z_n(\boldsymbol{\phi})),\end{aligned}$$

so the joint distribution of $\mathbf{x}$ under $\boldsymbol{\phi}$ is given by

$$h(\mathbf{x})=h_1\left(\prod_{t=1}^{n-s} f_\sigma(\phi_p z_t(\boldsymbol{\phi}))|\phi_p|\right)h_2, \tag{2.6}$$

where h_1 and h_2 do not depend on n. This suggests approximating the log-likelihood of $(\boldsymbol{\phi},\sigma)$ given the data as

$$\begin{aligned}\mathscr{L}(\boldsymbol{\phi},\sigma) &= \sum_{t=1}^{n-s}\ln f_\sigma(\phi_p z_t(\boldsymbol{\phi}))+(n-s)\ln|\phi_p|\\ &= -(n-s)\ln\sigma+\sum_{t=1}^{n-s}\ln f(\sigma^{-1}\phi_p z_t(\boldsymbol{\phi}))+(n-s)\ln|\phi_p|,\end{aligned} \tag{2.7}$$

where the $\{z_t(\boldsymbol{\phi})\}$ can be computed recursively from (2.3).

2.3. *Least absolute deviations.* If the noise distribution is Laplacian, or two-sided exponential, with mean 0, variance σ^2, and density

$$f_\sigma(z)=\frac{1}{\sigma}f\left(\frac{z}{\sigma}\right)=\frac{1}{\sqrt{2}\sigma}\exp\left(-\frac{\sqrt{2}|z|}{\sigma}\right),$$

then the log-likelihood is given by

$$\text{constant}-(n-s)\ln\kappa-\sum_{t=1}^{n-s}\frac{\sqrt{2}|z_t(\boldsymbol{\phi})|}{\kappa}, \tag{2.8}$$

where $\kappa=\sigma|\phi_p|^{-1}$. Setting the partial derivative of (2.8) with respect to κ equal to zero, we obtain

$$\kappa(\boldsymbol{\phi})=\frac{\sqrt{2}}{n-s}\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi})|, \tag{2.9}$$

where the $\{z_t(\boldsymbol{\phi})\}$ are computed from (2.3). Substituting $\kappa(\boldsymbol{\phi})$ for κ in (2.8), we obtain the concentrated Laplacian likelihood

$$\ell(\boldsymbol{\phi})=\text{constant}-(n-s)\ln\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi})|.$$

Maximizing $\ell(\boldsymbol{\phi})$ is equivalent to minimizing the absolute deviations criterion,

$$m_n(\boldsymbol{\phi})=\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi})|. \tag{2.10}$$

The minimizer $\hat{\boldsymbol{\phi}}$ of (2.10) will be referred to as the least absolute deviations (LAD) estimator of $\boldsymbol{\phi}_0$.

3. Asymptotic results.

3.1. *Parameter estimation.* We now state our main result, which parallels Davis and Dunsmuir (1997), Corollary 1.

THEOREM 1. *Assume the all-pass model* (2.1) *holds with* A1–A4. *Then there exists a sequence of local minimizers* $\hat{\boldsymbol{\phi}}_{\text{LAD}}$ *of* (2.10) *such that*

$$n^{1/2}(\hat{\boldsymbol{\phi}}_{\text{LAD}}-\boldsymbol{\phi}_0)\xrightarrow{\mathscr{L}}-\frac{|\phi_{0r}|\boldsymbol{\Gamma}_s^{-1}}{2f_\sigma(0)}\mathbf{N}\sim N\left(\mathbf{0},\frac{\text{var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)}\sigma^2\boldsymbol{\Gamma}_s^{-1}\right), \tag{3.1}$$

where $\boldsymbol{\Gamma}_s=[\gamma(j-k)]_{j,k=1}^s$ *and* $\gamma(\cdot)$ *is the autocovariance function of the causal* $AR(r)$ $\{Z_t/\phi_0(B)\}$.

PROOF. The proof of this theorem relies on two lemmas which are stated and proved in the Appendix. For $\mathbf{u}\in\mathbb{R}^s$, let

$$S_n(\mathbf{u})=m_n(\boldsymbol{\phi}_0+n^{-1/2}\mathbf{u})-\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi}_0)|. \tag{3.2}$$

Then minimizing (2.10) with respect to $\boldsymbol{\phi}$ is equivalent to minimizing (3.2) with respect to $\mathbf{u}=n^{1/2}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)$. Lemma 1 of the Appendix is used to establish a functional convergence theorem in Lemma 2; specifically, $S_n \xrightarrow{\mathscr{L}} S$ on $C(\mathbb{R}^s)$ where

$$S(\mathbf{u})=\frac{f_\sigma(0)}{|\phi_{0r}|}\mathbf{u}'\boldsymbol{\Gamma}_s\mathbf{u}+\mathbf{u}'\mathbf{N}$$

and

$$\mathbf{N}\sim \mathrm{N}\left(\mathbf{0},\frac{2\operatorname{var}(|Z_1|)}{\phi_{0r}^2\sigma^2}\boldsymbol{\Gamma}_s\right).$$

Since the minimizer of the limit process $S(\mathbf{u})$ is $-|\phi_{0r}|/(2f_\sigma(0))\boldsymbol{\Gamma}_s^{-1}\mathbf{N}$, the result (3.1) follows by the continuous mapping theorem. □

REMARK 1. The sequence of local minimizers in the theorem depends on the unknown $\boldsymbol{\phi}_0$, which may not be the unique global minimizer of $\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|$, where $\tilde{z}_1(\boldsymbol{\phi})=-\phi(B)X_{1+s}/\phi(B^{-1})$. If $\boldsymbol{\phi}_0$ is the unique global minimizer of $\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|$, then Proposition 1 in the Appendix establishes strong consistency of the LAD estimators.

Now suppose that $\boldsymbol{\phi}_0$ is *not* the unique global minimizer, and $\boldsymbol{\phi}_0$ and $\boldsymbol{\phi}_1$ are both local minimizers of $\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|$. Then there may exist a sequence of local minimizers of the LAD criterion which converges to $\boldsymbol{\phi}_0$ and another sequence of local minimizers which converges to $\boldsymbol{\phi}_1$. Unless $\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|$ has a unique global minimizer at $\boldsymbol{\phi}=\boldsymbol{\phi}_0$, it is unclear whether the global minimizer of (2.10) satisfies the condition of the theorem.

In the Gaussian case, for example, any choice of $\boldsymbol{\phi}_0$ (with $\phi_{0r}\neq 0$) together with $\sigma_0^2:=\phi_{0r}^2\operatorname{var}(X_t)$ satisfies model (2.1) with innovations $\{Z_t\}$ iid $N(0,\sigma_0^2)$ and $\{X_t\}$ iid $N(0,\sigma_0^2\phi_{0r}^{-2})$. Choose any $\boldsymbol{\phi}_1\neq\boldsymbol{\phi}_0$ with $\phi_{1p}\neq 0$ and set $\sigma_1^2=\phi_{1p}^2\operatorname{var}(X_t)$. Then

$$\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi}_1)|=\mathrm{E}\left|\frac{Z_1\sigma_1}{\sigma_0\phi_{1p}}\right|=\mathrm{E}\left|\frac{Z_1\operatorname{var}^{1/2}(X_t)}{\sigma_0}\right|=\mathrm{E}|z_1(\boldsymbol{\phi}_0)|$$

so that $\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|$ is not uniquely minimized at $\boldsymbol{\phi}_0$.

On the other hand, if Z_t has heavier tails than Gaussian, in the sense that

$$\mathrm{E}\left|\sum_{j=-\infty}^{\infty}c_jZ_{t-j}\right|>\mathrm{E}|Z_1| \tag{3.3}$$

for any $\{c_j\}$ with at least two non-zero elements, $\sum_j|c_j|<\infty$, and $\sum_j c_j^2=1$, then

$$\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|=\mathrm{E}\left|\frac{\phi_0(B^{-1})\phi(B)}{\phi_{0r}\phi(B^{-1})\phi_0(B)}Z_t\right|>\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi}_0)|,$$

so that $\boldsymbol{\phi}_0$ is the unique global minimizer. Huang and Pawitan (2000) give sufficient conditions for (3.3) and show that it is satisfied by the Laplacian, Student's t, contaminated normal, and other standard heavy-tailed distributions. In these cases, $\boldsymbol{\phi}_0$ is the unique global minimizer of $\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|$.

REMARK 2. Note that the asymptotic covariance matrix from (3.1) is a scalar multiple of the asymptotic covariance matrix for the vector of Gaussian likelihood estimators of the corresponding sth-order autoregressive process.

REMARK 3. In practice, computation of $\hat{\boldsymbol{\phi}}_{\mathrm{LAD}}$ requires numerical minimization, in which local minima are of concern. In Section 4.1, we describe our methods for generating initial values and guarding against local minima.

EXAMPLES. For the Laplacian density, $\mathrm{E}|Z_1| = \sigma/\sqrt{2}$ and $f_\sigma(0) = 1/(\sqrt{2}\sigma)$, so that the constant factor appearing in the limiting covariance matrix in (3.1) is

$$\frac{\mathrm{var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} = \frac{1}{2}.$$

For Student's t-distribution with $\nu > 2$ degrees of freedom, $\sigma = (\nu/(\nu-2))^{1/2}$,

$$\mathrm{E}|Z_1| = 2\frac{(\nu-2)^{1/2}}{\nu-1}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi}}\sigma$$

and

$$f_\sigma(0) = \frac{\Gamma((\nu+1)/2)}{\sigma\Gamma(\nu/2)\sqrt{(\nu-2)\pi}},$$

so that the constant factor in (3.1) is

$$\frac{\mathrm{var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} = \frac{\Gamma^2(\nu/2)(\nu-2)\pi}{2\Gamma^2((\nu+1)/2)} - \frac{2(\nu-2)^2}{(\nu-1)^2}.$$

For $\nu = 3$, the value of this expression is 0.7337.

3.2. *Order selection.* In practice the order r of the all-pass model is usually unknown. The following corollary to Theorem 1 is useful in order selection.

COROLLARY 1. *Assume the conditions of Theorem* 1. *If the true all-pass model order is r and the fitted model order is $p > r$ then*

$$n^{1/2}\widehat{\phi}_{p,\mathrm{LAD}} \xrightarrow{\mathscr{L}} N\left(0, \frac{\mathrm{var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)}\right),$$

where $\hat{\phi}_{p,\mathrm{LAD}}$ is the pth element of $\hat{\boldsymbol{\phi}}_{\mathrm{LAD}}$.

PROOF. By Problem 8.15 of Brockwell and Davis (1991), the pth diagonal element of $\mathbf{\Gamma}_p^{-1}$ is σ^{-2} for $p > r$, so the result follows from (3.1). □

Recall that we have assumed there is a known model order s which is sufficiently large in the sense that $s \geq r$. A practical approach to order determination in large samples then proceeds as follows:

1. Fit an sth-order all-pass model and obtain residuals $\{z_t(\hat{\boldsymbol{\phi}})\}$.
   (a) Estimate $\operatorname{var}(|Z_1|)\phi_{0r}^{-2}$ consistently by $\hat{v}_1$, the empirical variance of $\{|z_t(\hat{\boldsymbol{\phi}})|\}$.
   (b) Estimate $\operatorname{var}(Z_1)\phi_{0r}^{-2} = \sigma^2\phi_{0r}^{-2}$ consistently by $\hat{v}_2$, the empirical variance of $\{z_t(\hat{\boldsymbol{\phi}})\}$.
   (c) Estimate $|\phi_{0r}|f_\sigma(0)$ consistently by $\hat{d}$, a kernel estimator of the density at zero based on $\{z_t(\hat{\boldsymbol{\phi}})\}$.
   (d) Compute

$$\hat{\theta}^2 := \frac{\hat{v}_1}{2\hat{v}_2^2\hat{d}^2}. \tag{3.4}$$

   We conjecture that this estimator converges in probability to

$$\frac{\operatorname{var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)},$$

   using extensions of the results in Kreiss (1987) and Robinson (1987). Those results do not apply directly here because they assume that the $\{Z_t\}$ are the one-step prediction errors in the Wold decomposition.

2. Fit all-pass models of order $p = 1, 2, \ldots, s$ via LAD and obtain the pth coefficient, $\hat{\phi}_{pp}$ for each.
3. Choose the model order r as the smallest order beyond which the estimated coefficients are statistically insignificant; that is,

$$r = \min\{0 \leq p \leq s \colon |\hat{\phi}_{jj}| < 1.96\hat{\theta}n^{-1/2} \text{ for } j > p\}.$$

A more formal order selection procedure is based on a version of AIC, the information criterion of Akaike (1973), which is designed to be an approximately unbiased estimator of the Kullback–Leibler index of the fitted model relative to the true model. We take the same heuristic approach here, using the Laplacian likelihood computed on the basis of $n - s$ observations to make fair comparisons across different model orders. The proposed model order p is no greater than s. Let $X_1^*, \ldots, X_n^*$ be a realization from the model $(\boldsymbol{\phi}_0', \kappa_0)'$, independent of $X_1, \dots, X_n$. Then, from (2.7),

$$
\begin{aligned}
-2\mathscr{L}_{X^*}(\hat{\boldsymbol{\phi}}, \hat{\kappa}) &= -2\mathscr{L}_X(\hat{\boldsymbol{\phi}}, \hat{\kappa}) - 2\frac{\sqrt{2}\sum_{t=1}^{n-s}|z_t(\hat{\boldsymbol{\phi}})|}{\hat{\kappa}} + 2\frac{\sqrt{2}\sum_{t=1}^{n-s}|z_t^*(\hat{\boldsymbol{\phi}})|}{\hat{\kappa}} \\
&= -2\mathscr{L}_X(\hat{\boldsymbol{\phi}}, \hat{\kappa}) - 2(n-s) + 2\sqrt{2}\frac{\sum_{t=1}^{n-s}|z_t^*(\hat{\boldsymbol{\phi}})| - \sum_{t=1}^{n-s}|z_t^*(\boldsymbol{\phi}_0)|}{\hat{\kappa}} \\
&\quad + 2\sqrt{2}\frac{\sum_{t=1}^{n-s}|z_t^*(\boldsymbol{\phi}_0)|}{\hat{\kappa}}.
\end{aligned} \tag{3.5}
$$

Using Lemma 2, (3.1), and the ergodic theorem, we have that

$$
\frac{\sum_{t=1}^{n-s}|z_t^*(\hat{\boldsymbol{\phi}})| - \sum_{t=1}^{n-s}|z_t^*(\boldsymbol{\phi}_0)|}{\hat{\kappa}} \xrightarrow{\mathscr{L}} \frac{\mathbf{u}'\mathbf{N}^*}{\sqrt{2}\mathrm{E}|Z_1||\phi_{0r}|^{-1}} + \frac{f_\sigma(0)}{|\phi_{0r}|}\frac{\mathbf{u}'\boldsymbol{\Gamma}_s\mathbf{u}}{\sqrt{2}\mathrm{E}|Z_1||\phi_{0r}|^{-1}},
$$

where $\mathbf{u}' = -|\phi_{0r}|/(2f_\sigma(0))\boldsymbol{\Gamma}_s^{-1}\mathbf{N}$ and $\mathbf{N}$, $\mathbf{N}^*$ are iid $N(\mathbf{0}, 2\,\mathrm{var}|Z_1|\phi_{0r}^{-2}\sigma^{-2}\boldsymbol{\Gamma}_s)$. Given the existence of the relevant moments, it follows that

$$
\begin{aligned}
\mathrm{E}\left[\frac{\sum_{t=1}^{n-s}|z_t^*(\hat{\boldsymbol{\phi}})| - \sum_{t=1}^{n-s}|z_t^*(\boldsymbol{\phi}_0)|}{\hat{\kappa}}\right] &\simeq \frac{f_\sigma(0)}{\sqrt{2}\mathrm{E}|Z_1|}\ \mathrm{trace}(\boldsymbol{\Gamma}_s\mathrm{E}[\mathbf{u}\mathbf{u}']) \\
&= \frac{\mathrm{var}|Z_1|}{2\sqrt{2}\mathrm{E}|Z_1|\sigma^2 f_\sigma(0)}\,p
\end{aligned}
$$

and

$$
\mathrm{E}\left[\frac{\sum_{t=1}^{n-s}|z_t^*(\boldsymbol{\phi}_0)|}{\hat{\kappa}}\right] = \mathrm{E}\left[\sum_{t=1}^{n-s}|z_t^*(\boldsymbol{\phi}_0)|\right]\mathrm{E}\left[\frac{1}{\hat{\kappa}}\right] \simeq \frac{(n-s)\mathrm{E}|Z_1|}{|\phi_{0r}|}\frac{|\phi_{0r}|}{\sqrt{2}\mathrm{E}|Z_1|} = \frac{n-s}{\sqrt{2}}.
$$

Therefore the quantity

$$
\mathrm{AIC}(p) := -2\mathscr{L}_X(\hat{\boldsymbol{\phi}}, \hat{\kappa}) + \frac{\mathrm{var}|Z_1|}{\mathrm{E}|Z_1|\sigma^2 f_\sigma(0)}\,p \tag{3.6}
$$

is approximately unbiased for (3.5). The model order $p \in \{0, 1, \dots, s\}$ which minimizes $\mathrm{AIC}(p)$ is selected. Note that in the Laplacian case, the penalty term in (3.6) is

$$
\frac{\mathrm{var}|Z_1|}{\mathrm{E}|Z_1|\sigma^2 f_\sigma(0)}\,p = \frac{\sigma^2/2}{(\sigma/\sqrt{2})\sigma^2(1/\sqrt{2}\sigma)}\,p = p,
$$

unlike the $2p$ penalty associated with a Gaussian likelihood. The penalty term can be estimated consistently with

$$
\frac{\hat{v}_1}{\hat{e}_1\hat{v}_2\hat{d}},
$$

where $\hat{e}_1$ is the sample mean of the $|z_t(\hat{\boldsymbol{\phi}})|$ from the sth order fit, and the remaining terms are defined above.

4. Empirical results.

4.1. *Simulation results.* In this section we describe a simulation study undertaken to evaluate the asymptotic theory. We considered all-pass model orders 1 and 2 and sample sizes $n=500$ and 5000. For each case, we simulated 1000 replications of the all-pass model, using as noise Student's t with 3 degrees of freedom. We used the Hooke and Jeeves (1961) algorithm to minimize the LAD criterion for each replicate.

To guard against the possibility of being trapped in local minima, we used a large number (250) of starting values for each replicate. These were distributed uniformly in the space of partial autocorrelations, then mapped to the space of autoregressive coefficients using the Durbin–Levinson algorithm [Brockwell and Davis (1991), Proposition 5.2.1]. That is, for a model of order p, the kth starting value $(\phi_{p1}^{(k)}, \dots, \phi_{pp}^{(k)})'$ was computed recursively as follows:

1. Draw $\phi_{11}^{(k)}, \phi_{22}^{(k)}, \dots, \phi_{pp}^{(k)}$ iid uniform$(-1, 1)$.
2. For $j=2, \dots, p$, compute

$$\begin{bmatrix} \phi_{j1}^{(k)} \\ \vdots \\ \phi_{j,j-1}^{(k)} \end{bmatrix} = \begin{bmatrix} \phi_{j-1,1}^{(k)} \\ \vdots \\ \phi_{j-1,j-1}^{(k)} \end{bmatrix} - \phi_{jj}^{(k)} \begin{bmatrix} \phi_{j-1,j-1}^{(k)} \\ \vdots \\ \phi_{j-1,1}^{(k)} \end{bmatrix}.$$

The initial 250 candidate starting values were pared to the 10 that gave the smallest function evaluations. Optimized values were then found by implementing the Hooke and Jeeves algorithm with each of these 10 candidates as starting values. Among the 10 optimized values, the one that gave the smallest function evaluation was selected as the estimate. Residuals for each realization were obtained, and confidence intervals for $\boldsymbol{\phi}_0$ were constructed using equations (3.1) and (3.4). In computing (3.4), we used a normal kernel density estimator with a normal scale bandwidth selector $\hat{v}_2^{1/2}(3n/4)^{-1/5}$.

Results appear in Tables 1 and 2. In all cases, the LAD estimates are approximately unbiased and the confidence interval coverages are close to the nominal 95% level. The asymptotic standard errors understate the true variability of the LAD estimates for the smaller sample size but are accurate at the larger sample size. Normal probability plots and histograms suggest that this extra variation in the LAD estimates comes from a relatively small number of large outliers, while most of the estimates follow the asymptotic normal law quite closely.

Table 1 shows results for all-pass of order one with $\phi_1=0.1$, 0.5 and 0.9. Asymptotic results are symmetric about zero and empirical results for $\phi_1= -0.1$, -0.5 and -0.9 (not shown) are roughly symmetric. The simulation results show that estimation is more difficult when $\{X_t\}$ has weaker dependence, and convergence to the limiting distribution is slower. Unlike the usual unit root case for autoregressive processes, dependence is weaker for all-pass as $\phi_1 \to \pm 1$, since these boundary cases correspond to iid noise as the AR and

TABLE 1

Empirical means, standard deviations and percent coverages of nominal 95% *confidence intervals for LAD estimates of all-pass model of order* 1*

	Asymptotic		Empirical		
n	**Mean**	**Std. dev.**	**Mean (c.i.)**	**Std. dev. (c.i.)**	**% coverage (c.i.)**
500	$\phi_1 = 0.1$	0.0381	0.1013 (0.0931, 0.1095)	0.1323 (0.1264, 0.1380)	91.1 (89.3, 92.9)
5000	$\phi_1 = 0.1$	0.0121	0.0999 (0.0991, 0.1007)	0.0130 (0.0124, 0.0135)	94.5 (93.1, 95.9)
500	$\phi_1 = 0.5$	0.0332	0.4979 (0.4954, 0.5004)	0.0397 (0.0379, 0.0414)	94.2 (92.8, 95.6)
5000	$\phi_1 = 0.5$	0.0105	0.4998 (0.4991, 0.5005)	0.0109 (0.0105, 0.0112)	95.4 (94.1, 96.7)
500	$\phi_1 = 0.9$	0.0167	0.8834 (0.8770, 0.8898)	0.1027 (0.0981, 0.1071)	91.2 (89.4, 93.0)
5000	$\phi_1 = 0.9$	0.0053	0.8993 (0.8990, 0.8996)	0.0056 (0.0054, 0.0059)	95.7 (94.4, 97.0)

*To quantify simulation uncertainty, empirical confidence intervals (c.i.'s) are computed from standard asymptotic theory for 1000 iid replicates at each sample size, n. Asymptotic means and standard deviations are from (3.1). Noise distribution is t with 3 degrees of freedom.

TABLE 2

Empirical means, standard deviations and percent coverages of nominal 95% *confidence intervals for LAD estimates of all-pass model of order* 2*

	Asymptotic		Empirical		
n	**Mean**	**Std. dev.**	**Mean (c.i.)**	**Std. dev. (c.i.)**	**% coverage (c.i.)**
500	$\phi_1 = 0.3$	0.0351	0.2990 (0.2962, 0.3018)	0.0456 (0.0435, 0.0475)	92.5 (90.9, 94.1)
	$\phi_2 = 0.4$	0.0351	0.3965 (0.3937, 0.3993)	0.0447 (0.0427, 0.0467)	92.1 (90.4, 93.8)
5000	$\phi_1 = 0.3$	0.0111	0.3003 (0.2996, 0.3010)	0.0118 (0.0113, 0.0123)	95.5 (94.2, 96.8)
	$\phi_2 = 0.4$	0.0111	0.3990 (0.3983, 0.3997)	0.0117 (0.0112, 0.0122)	94.7 (93.3, 96.1)

*To quantify simulation uncertainty, empirical confidence intervals (c.i.'s) are computed from standard asymptotic theory for 1000 iid replicates at each sample size, n. Asymptotic means and standard deviations are from (3.1). Noise distribution is t with 3 degrees of freedom.

MA factors $(1-\phi_1 B)$ and $(1-\phi_1^{-1}B)$ cancel. Dependence is also weaker as $\phi_1 \to 0$. To see this, rescale $X_t \sim (0, \sigma^2\phi_1^{-2})$ to have bounded variance as $\phi_1 \to 0$,

$$\phi_1 X_t = \phi_1 Z_t + \phi_1(\phi_1 - \phi_1^{-1}) \sum_{j=0}^{\infty} \phi_1^j Z_{t-1-j}.$$

Now the variance of the $(t-1)$ term is $(1-\phi_1^2)^2\sigma^2 = O(1)$, while the variance of the sum of the remaining terms is $\phi_1^2(2-\phi_1^2)\sigma^2 = O(\phi_1^2)$. Hence $\{X_t\}$ behaves like the iid sequence $\{-\phi_1^{-1} Z_{t-1}\}$ for small ϕ_1.

We also compared the performance of the LAD estimators to the performance of a cumulant-based estimator, which maximizes the absolute residual kurtosis

$$\left| \frac{1}{n-s} \sum_{t=1}^{n-s} \left(\frac{z_t(\boldsymbol{\phi})}{\hat{v}_2^{1/2}} \right)^4 - 3 \right| \tag{4.1}$$

with respect to $\boldsymbol{\phi}$ [see Rosenblatt (2000), Section 8.7, and the references therein]. Results are tabled in Table 3. The cumulant-based estimator suffers from some bias at the smaller sample size, primarily due to a pile-up effect on ± 1. The LAD estimators have much smaller mean squared error (MSE) in most cases. The best case for the cumulant-based estimator is $\phi_1 = 0.9$, $n = 500$, for which the empirical MSE of the cumulant-based estimator is still 20% higher than that of the LAD estimator. For this case, 347 of the 1000 estimates were equal to $+1$, reducing the variability of the estimator, but missing the dependence structure in the data. The performance of the cumulant-based estimators was much worse for second-order all-pass models. We do not report those results here.

4.2. *Linear time series with "nonlinear" behavior.* We now turn to some examples with real data. Figure 1(a)–1(d) shows 500 daily log returns of the New Zealand/U.S. exchange rate together with autocorrelations for the

TABLE 3

*Empirical means, standard deviations and efficiencies relative to LAD for maximum absolute residual kurtosis estimation method**

	True values	**Empirical**		
n	**Mean**	**Mean**	**Std. dev.**	**MSE relative to LAD**
500	$\phi_1 = 0.1$	0.2999	0.4949	16.3
5000	$\phi_1 = 0.1$	0.1180	0.1496	134.3
500	$\phi_1 = 0.5$	0.5254	0.1342	11.8
5000	$\phi_1 = 0.5$	0.5011	0.0333	9.3
500	$\phi_1 = 0.9$	0.9203	0.1114	1.2
5000	$\phi_1 = 0.9$	0.9197	0.0420	67.6

*MSE relative to LAD is empirical mean squared error of cumulant estimator divided by empirical MSE of LAD estimator. Results are based on the same 1000 simulated realizations as in Table 1.

returns, their squares, and their absolute values. These data show many of the stylized facts that would lead to consideration of GARCH or stochastic volatility models: lack of serial correlation, heavy-tailed marginal distribution and volatility clustering. We fit an all-pass model of order 6 to show that a linear model can produce this same behavior. The order was determined using the model selection procedure based on the $\hat{\phi}_{pp}$ as described in Section 3.2. (The AIC had local minima at $p=6$ and 10.) The autoregressive polynomial of the fitted model is

$$1+0.367B+0.75B^2+0.391B^3-0.088B^4+0.193B^5+0.096B^6.$$

Autocorrelations for the residuals and the squares of the residuals from the all-pass fit are shown in Figure 2(a) and 2(b). These diagnostics show that a non-Gaussian linear model can capture many of the features often regarded as characteristic of nonlinearity. Though this example shows that in some cases all-pass models can mimic the behavior of more familiar nonlinear models for

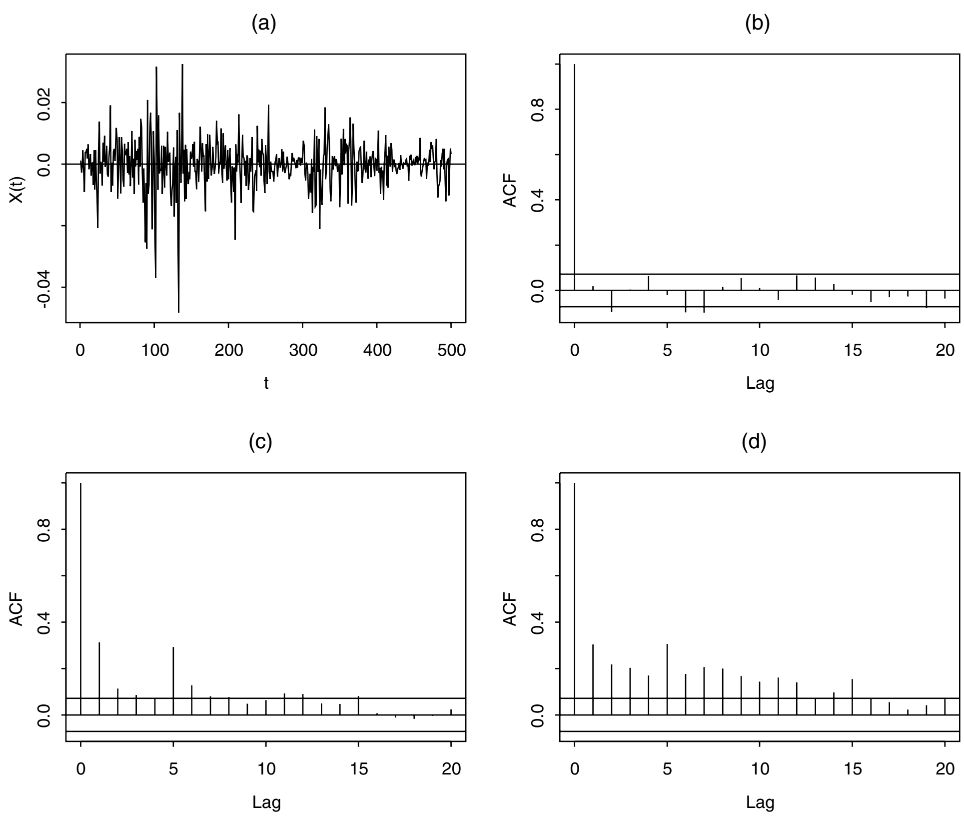

FIG. 1. (a) *Daily log returns of the New Zealand/U.S. exchange rate*. (b) *ACF for the returns*. (c) *ACF for squares of returns*. (d) *ACF for absolute values of returns*.

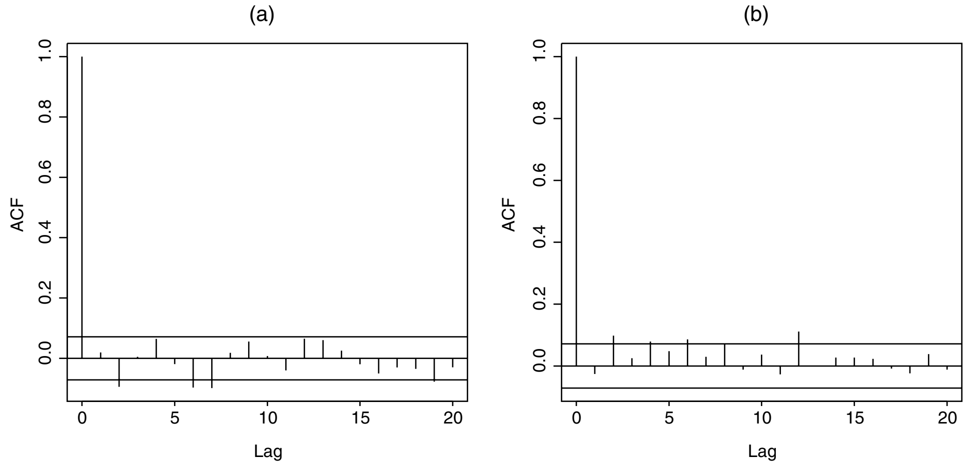

FIG. 2. *Diagnostics for fitted all-pass model of order six for New Zeland/U.S. exchange rate returns*. (a) *ACF of residuals*. (b) *ACF for squares of residuals*.

financial data, the constrained forms of all-pass models limit their usefulness in general for this kind of application. A more natural application of all-pass modeling is illustrated in the next subsection.

4.3. *Noncausal autoregressive modeling.* As mentioned in the Introduction, an important application of all-pass models is in noncausal autoregressive model fitting. Suppose that $\{X_t\}$ satisfies the difference equations

$$\phi_c(B)\phi_{nc}(B)X_t = Z_t,$$

where the q roots of $\phi_c(z)$ are outside the unit circle, the r roots of $\phi_{nc}(z)$ are inside the unit circle and $\{Z_t\}$ is iid. Let $\phi_{nc}^{(c)}(z)$ denote the causal rth order polynomial whose roots are the reciprocals of the roots of $\phi_{nc}(z)$. If $\{X_t\}$ is mistakenly modeled with the second-order equivalent causal representation,

$$\phi_c(B)\phi_{nc}^{(c)}(B)X_t = U_t,$$

then $\{U_t\}$ satisfies the difference equations

$$U_t = \frac{\phi_c(B)\phi_{nc}^{(c)}(B)}{\phi_c(B)\phi_{nc}(B)} Z_t = \frac{\phi_{nc}^{(c)}(B)}{-\phi_{nc,r}B^r\phi_{nc}^{(c)}(B^{-1})} Z_t, \tag{4.2}$$

where $\phi_{nc,r}$ is the coefficient of $-B^r$ in $\phi_{nc}(B)$. Thus, by (2.1), $\{U_t\}$ is a purely noncausal all-pass time series. Equivalently, the reversed-time process $\{U_{-t}\}$ is a causal all-pass time series.

This suggests a two-step procedure for fitting noncausal autoregressive time series models. Using a standard method such as Gaussian maximum likelihood, fit a causal sth order autoregressive model to $\{X_t\}$ and obtain residuals $\{\widehat{U}_t\}$. Select a model order r and fit a purely noncausal rth order all-pass model to $\{\widehat{U}_t\}$. The fitted model can be evaluated by residual diagnostics, looking for

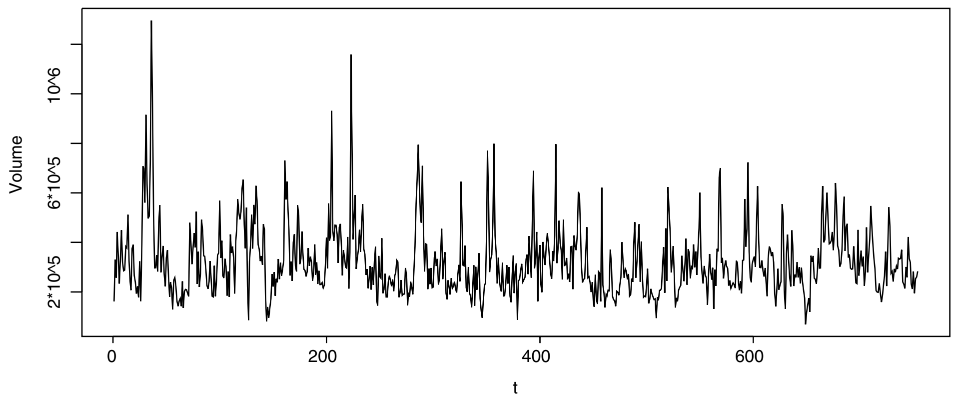

FIG. 3. *Volumes of Microsoft (MSFT) stock traded over* 754 *transaction days from* 06/03/96 *to* 05/27/99.

iid (not merely white) noise. Once a suitable all-pass model is fitted to obtain the purely noncausal AR(r), the appropriate causal AR(q) polynomial can be identified by canceling the roots in the causal AR(s) polynomial which correspond to the inverses of the roots in the purely noncausal AR(r) polynomial. The resulting estimates could be used as preliminary estimates in a more refined estimation procedure as in Breidt, Davis, Lii and Rosenblatt (1991). This two-step procedure avoids the need to study all possible 2^s configurations of roots inside and outside the unit circle. Note that this methodology can be easily adapted to identify the roots of a noninvertible moving average.

EXAMPLE (Microsoft trading volume). The data in Figure 3 are volumes of Microsoft (MSFT) stock traded over 754 transaction days from 06/03/96 to 05/27/99. Because the data are skewed and show some evidence of heteroskedasticity, we transformed with natural logarithms. The autocorrelations and partial autocorrelations of the resulting series suggest that an autoregressive model of order 1 or 3 might be appropriate. To focus on the estimation problem and not on the order selection problem, we fit an AR(1) via Gaussian maximum likelihood, yielding the estimate $\hat{\phi}_{nc}^{(c)} = 0.5834$ with standard error 0.0296. The resulting residuals $\{\widehat{U}_t\}$ show little evidence of correlation, but both $\{\widehat{U}_t^2\}$ and $\{|\widehat{U}_t|\}$ have significant lag 1 autocorrelations, with asymptotic p-values less than 0.001 [McLeod and Li (1983)]; see Figures 4(a) and 4(b). Thus a causal AR(1) model with iid noise is inappropriate for the MSFT data, and we investigate the noncausal alternative.

Fitting a purely noncausal all-pass of order 1 to $\{\widehat{U}_t\}$, we obtain the estimate $\tilde{\phi}_{nc} = 1.7522$, with standard error 0.0989. From (4.2),

$$\widehat{U}_t = \hat{\phi}_c(B)\hat{\phi}_{nc}^{(c)}(B)X_t \simeq \frac{\tilde{\phi}_{nc}^{(c)}(B)}{-\tilde{\phi}_{nc,r}B^r\tilde{\phi}_{nc}^{(c)}(B^{-1})}\widetilde{Z}_t,$$

so that the all-pass residuals are obtained from

$$\widetilde{Z}_t = \frac{(1-1.7522B)(1-0.5834B)}{1-(1.7522)^{-1}B} X_t \tag{4.3}$$
$$= \frac{(1-1.7522B)(1-0.5834B)}{1-0.5707B} X_t.$$

In Figures 4(c) and 4(d), these residuals show no evidence of correlation in their squares or absolute values, suggesting that a noncausal AR(1) is a more appropriate model than a causal AR(1) for these data.

Note that another possible modeling strategy would be to fit a causal AR(1) and then model the non-iid residuals as GARCH. This would require at least two more parameters (intercept and slope in ARCH(1)) than the noncausal AR(1) fitted here.

We also fitted log volumes over the same trading period for two small companies [Atmel Corporation (ATML) and Microchip (MCHP)] in the same sector as

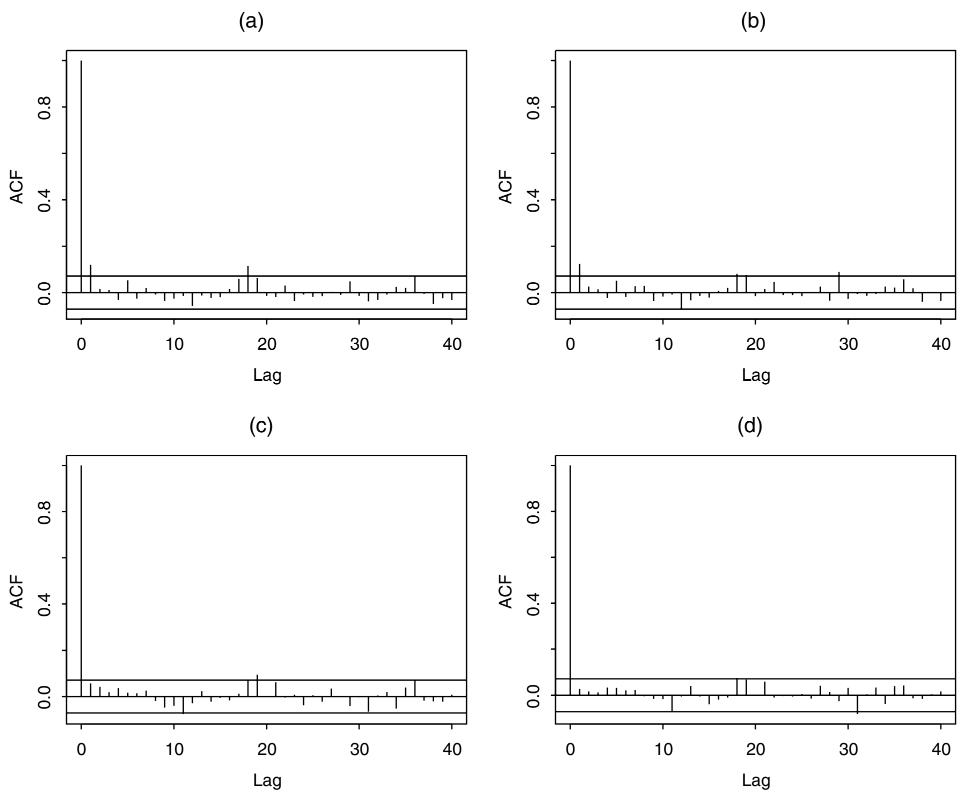

FIG. 4. *Diagnostics for causal and noncausal autoregressive models fitted to log Microsoft volume* (a) *ACF of squares of residuals* $\{\widehat{U}_t\}$ *from causal AR*(1) *fit.* (b) *ACF of absolute values of* $\{\widehat{U}_t\}$. (c) *ACF of squares of residuals* $\{\widetilde{Z}_t\}$ *from noncausal all-pass fit.* (d) *ACF of absolute values of* $\{\widetilde{Z}_t\}$.

Microsoft, but found that causal AR models adequately described their dynamics. A possible explanation for this phenomenon is that forthcoming actions of Microsoft are widely anticipated by the market, so that the effect of shocks precedes their arrival and a noncausal model is appropriate. The actions of smaller companies do not receive as much attention, so causal models are appropriate.

Because the model order is low in the Microsoft example, we could fit all possible causal–noncausal models and compare diagnostics, rather than employ the two-step procedure. If we had fitted a noncausal AR(1) model directly, rather than via the two-step procedure, we would have obtained the estimated model $(1-1.7141B)X_t = Z_t$, which is quite close to the model which would be obtained through cancellation of the common factors in (4.3). Diagnostics for the residuals from the noncausal AR(1) fit are virtually identical to those for the $\{\widetilde{Z}_t\}$ above. Note that for higher-order models it may not be possible to fit and assess all 2^s possible models.

5. Discussion. This article has reviewed all-pass models, which generate uncorrelated but dependent time series in the non-Gaussian case. An approximation to the likelihood of the model in the case of Laplacian noise yielded a modified absolute deviations criterion, which can be used even if the underlying noise is not Laplacian. Asymptotic normality for least absolute deviation estimators of the model parameters was established under general conditions, and order selection methods were developed. Behavior of the LAD estimators in finite samples was studied via simulation, showing agreement with the asymptotic theory and marked superiority over the maximum absolute residual kurtosis technique. The methodology was applied to exchange rate returns to show that linear all-pass models can mimic "nonlinear" behavior often associated with GARCH or stochastic volatility models. The methodology was also applied to Microsoft volume data as part of a two-step procedure for fitting noncausal autoregressions. In this example, a noncausal AR(1) model provides a better fit than does a causal AR(1). Because of the low order of the fitted model, order selection was not an issue in this example.

In future work, we intend to investigate the behavior of the LAD estimates for all-pass models when order selection is required and further compare our methodology to methods based on higher-order moments. We are also currently looking at maximum likelihood estimation for the same problem.

APPENDIX

In this Appendix we derive two preliminary results used in establishing our main theorem, and we prove a strong consistency result for the LAD estimator. The first preliminary result extends Theorem 1 of Davis and Dunsmuir (1997) from one-sided to two-sided linear processes.

LEMMA 1. *Suppose* $\{Y_t\}$ *is the linear process*

$$Y_t = \sum_{j=-\infty}^{\infty} c_j z_{t-j},$$

where $c_0 = 0$, $\sum_{j=-\infty}^{\infty} |c_j| < \infty$, $\{z_t\}$ *is iid with mean* 0, *finite variance, and common distribution function* G *which has median* 0 *and is continuously differentiable in a neighborhood of* 0. *Then*

$$S_n := \sum_{t=1}^{n-s} \left(|z_t - n^{-1/2} Y_t| - |z_t|\right) \xrightarrow{\mathscr{L}} \operatorname{var}(Y_t) g(0) + N,$$

where

$$N \sim N\left(0, \gamma^*(0) + 2\sum_{h=1}^{\infty} \gamma^*(h)\right),$$

$$\gamma^*(h) = E[Y_t \operatorname{sgn}(z_t) Y_{t+h} \operatorname{sgn}(z_{t+h})],$$

and $g(z)$ *is the density corresponding to* G.

PROOF. Using the identity for $z \neq 0$,

$$|z-y| - |z| = -y \operatorname{sgn}(z) + 2(y-z)\{1_{\{0<z<y\}} - 1_{\{y<z<0\}}\},$$

we have

$$\begin{aligned} S_n &= -n^{-1/2} \sum_{t=1}^{n-s} Y_t \operatorname{sgn}(z_t) \\ &\quad + 2\sum_{t=1}^{n-s} (n^{-1/2} Y_t - z_t)\{1_{\{0<z_t<n^{-1/2}Y_t\}} - 1_{\{n^{-1/2}Y_t<z_t<0\}}\} \\ &=: A_n + B_n. \end{aligned}$$

A standard truncation argument, truncating Y_t to create the $2M$-dependent sequence $\{Y_t^M \operatorname{sgn}(z_t)\} = \{\sum_{j=-M}^{M} c_j z_{t-j} \operatorname{sgn}(z_t)\}$, allows application of a central limit theorem [Brockwell and Davis (1991), Theorem 6.4.2] for each M, from which it follows that $A_n \xrightarrow{\mathscr{L}} N$.

Now turning to B_n, let

$$W_{nt} := (n^{-1/2} Y_t - z_t) 1_{\{0<z_t<n^{-1/2}Y_t\}}.$$

Let F_Y denote the distribution of Y_1. Then

$$\begin{aligned}
&\limsup_{n\to\infty} n\mathrm{E}[W_{nt}^2]\\
&\quad=\limsup_{n\to\infty}\left[n\int_0^{\varepsilon n^{1/2}}\int_0^{n^{-1/2}y}(n^{-1/2}y-z)^2 G(dz)F_Y(dy)\right.\\
&\qquad\qquad\left.+n\int_{\varepsilon n^{1/2}}^{\infty}\int_0^{n^{-1/2}y}(n^{-1/2}y-z)^2 G(dz)F_Y(dy)\right]\\
&\quad\le\limsup_{n\to\infty}\left[n\int_0^{\varepsilon n^{1/2}}\int_0^{n^{-1/2}y}(n^{-1/2}y-z)^2(g(0)+\delta)\,dz\,F_Y(dy)\right.\\
&\qquad\qquad\left.+n\int_{\varepsilon n^{1/2}}^{\infty}\int_0^{n^{-1/2}y}n^{-1}y^2 G(dz)F_Y(dy)\right]\\
&\quad\le\limsup_{n\to\infty}(\mathrm{const})\; n\int_0^{\varepsilon n^{1/2}}n^{-3/2}y^3F_Y(dy)\\
&\quad\le\limsup_{n\to\infty}(\mathrm{const})\;\varepsilon\mathrm{E}[Y_1^2 1_{\{Y_1>0\}}],
\end{aligned}\tag{A.1}$$

and since $\varepsilon>0$ is arbitrary, the bound must be zero.

Write

$$Y_t=Y_t^-+Y_t^+=\sum_{j=1}^{\infty}c_j z_{t-j}+\sum_{j=1}^{\infty}c_{-j}z_{t+j}.$$

Then, on the set $\{Y_t>0\}$,

$$\begin{aligned}
&\mathrm{E}[W_{nt}|z_{t-1},z_{t-2},\dots]\\
&\quad=\mathrm{E}[(n^{-1/2}Y_t-z_t)1_{\{0<z_t<n^{-1/2}Y_t\}}|z_{t-1},z_{t-2},\dots]\\
&\quad=\int_{-Y_t^-}^{\infty}\int_0^{n^{-1/2}(Y_t^-+y)}\{n^{-1/2}(Y_t^-+y)-z\}G(dz)F_{Y^+}(dy)\\
&\quad=\int_{-Y_t^-}^{\infty}n^{-1/2}(Y_t^-+y)\{G(n^{-1/2}(Y_t^-+y))-G(0)\}F_{Y^+}(dy)\\
&\qquad-\int_{-Y_t^-}^{\infty}\int_0^{n^{-1/2}(Y_t^-+y)}zG(dz)F_{Y^+}(dy)\\
&\quad\sim\int_{-Y_t^-}^{\infty}n^{-1}(Y_t^-+y)^2g(0)F_{Y^+}(dy)-\int_{-Y_t^-}^{\infty}g(0)\frac{n^{-1}(Y_t^-+y)^2}{2}F_{Y^+}(dy)\\
&\quad=\frac{g(0)}{2n}\int_{-Y_t^-}^{\infty}(Y_t^-+y)^2F_{Y^+}(dy),
\end{aligned}$$

where the approximation holds on the set $|n^{-1/2}Y_t|<\varepsilon$, for $\varepsilon>0$ small. Since

$$\begin{aligned}
\Pr\{n^{-1/2}\max(|Y_1|,\dots,|Y_n|)>\varepsilon\}&\le\Pr\left\{\bigcup_{t=1}^{n}\{|Y_t|>\varepsilon n^{1/2}\}\right\}\\
&\le n\Pr\{|Y_1|>\varepsilon n^{1/2}\}\\
&\le\varepsilon^{-2}\mathrm{E}[Y_1^2 1_{\{Y_1^2>\varepsilon^2 n\}}]\to 0
\end{aligned}$$

as $n \to \infty$, it follows from the ergodic theorem that

$$\sum_{t=1}^{n-s} \mathrm{E}[W_{nt} | z_{t-1}, z_{t-2}, \ldots] \xrightarrow{P} \frac{g(0)}{2} \mathrm{E}\left[\int_{-Y_t^-}^{\infty} (Y_t^- + y)^2 F_{Y^+}(dy)\right]. \tag{A.2}$$

By (A.1),

$$\begin{aligned} \mathrm{var}\left(\sum_{t=1}^{n-s} (W_{nt} - \mathrm{E}[W_{nt} | z_{t-1}, z_{t-2}, \ldots])\right) &= \sum_{t=1}^{n-s} \mathrm{var}(W_{nt} - \mathrm{E}[W_{nt} | z_{t-1}, z_{t-2}, \ldots]) \\ &\le \sum_{t=1}^{n-s} \mathrm{E}[W_{nt}^2] \to 0, \end{aligned}$$

so that from (A.2) we have

$$\sum_{t=1}^{n-s} W_{nt} \xrightarrow{P} \frac{g(0)}{2} \mathrm{E}\left[\int_{-Y_t^-}^{\infty} (Y_t^- + y)^2 F_{Y^+}(dy)\right].$$

Using the same argument for the second indicator in B_n, we obtain

$$\begin{aligned} B_n &\xrightarrow{P} \frac{g(0)}{2} \mathrm{E}\left[\int_{-\infty}^{\infty} (Y_t^- + y)^2 F_{Y^+}(dy)\right] \\ &= \frac{g(0)}{2} \mathrm{var}(Y_t), \end{aligned}$$

which concludes the proof. $\square$

To apply Lemma 1 in the context of LAD for all-pass models, we need to identify an appropriate $\{Y_t\}$ and compute the autocovariance function $\gamma^*(h)$ of the stationary process $\{Y_t \mathrm{sgn}(z_t)\}$. We now undertake these intermediate computations, which are then used in Lemma 2 to establish a functional convergence theorem for the centered absolute deviations criterion.

Define $\varphi(z) = \phi_1 z + \cdots + \phi_s z^s = 1 - \phi(z)$ and $\varphi_0(z) = 1 - \phi_0(z)$. In what follows, we linearize $\varphi(B^{-1}) z_t(\boldsymbol{\phi})$ around $\boldsymbol{\phi}_0$ within the criterion function m_n; that is, $\varphi(B^{-1}) z_t(\boldsymbol{\phi})$ is approximated by

$$\varphi_0(B^{-1}) z_t(\boldsymbol{\phi}_0) + \sum_{j=1}^{s} \frac{\partial}{\partial \phi_j} \{\varphi(B^{-1}) z_t(\boldsymbol{\phi})\}\Big|_{\boldsymbol{\phi} = \boldsymbol{\phi}_0} (\phi_j - \phi_{0j}).$$

By (2.3), the criterion function (2.10) can be written as

$$
\begin{aligned}
m_n &= \sum_{t=1}^{n-s} |\varphi(B^{-1})z_t(\boldsymbol{\phi}) - \phi(B)X_{t+s}| \\
&= \sum_{t=1}^{n-s} |\varphi(B^{-1})B^s z_{t+s}(\boldsymbol{\phi}) - \phi_0(B)X_{t+s} + (\phi_0(B) - \phi(B))X_{t+s}| \\
&\simeq \sum_{t=1}^{n-s} \Big| \varphi_0(B^{-1})B^s z_{t+s}(\boldsymbol{\phi}_0) - B^s z_{t+s}(\boldsymbol{\phi}_0) + z_t(\boldsymbol{\phi}_0) \\
&\qquad + \sum_{j=1}^{s} \frac{\partial}{\partial \phi_j} \{\varphi(B^{-1})z_t(\boldsymbol{\phi})\}\Big|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0} (\phi_j - \phi_{0j}) \\
&\qquad - \phi_0(B)X_{t+s} + n^{1/2}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)' n^{-1/2}(X_{t+s-1}, \dots, X_t)' \Big| \\
&= \sum_{t=1}^{n-s} \left| z_t(\boldsymbol{\phi}_0) + n^{-1/2}\mathbf{u}' \left[\frac{\partial}{\partial \phi_j} \{\varphi(B^{-1})z_t(\boldsymbol{\phi})\}\Big|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0} + X_{t+s-j} \right]_{j=1}^{s} \right|,
\end{aligned}
\tag{A.3}
$$

where $\mathbf{u} = n^{1/2}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)$.

Now

$$\phi(B)X_{t+s} = -z_t(\boldsymbol{\phi}) + \varphi(B^{-1})z_t(\boldsymbol{\phi}),$$

so

$$\frac{\partial}{\partial \phi_j}\{\varphi(B^{-1})z_t(\boldsymbol{\phi})\} = -X_{t+s-j} + \frac{\partial}{\partial \phi_j} z_t(\boldsymbol{\phi}). \tag{A.4}$$

Also,

$$\frac{\partial}{\partial \phi_j}\{\varphi(B^{-1})z_t(\boldsymbol{\phi})\} = \varphi(B^{-1})\frac{\partial}{\partial \phi_j} z_t(\boldsymbol{\phi}) + z_{t+j}(\boldsymbol{\phi}). \tag{A.5}$$

Equating (A.4) and (A.5) and solving for $\partial z_t(\phi)/\partial \phi_j$, we obtain

$$\frac{\partial}{\partial \phi_j} z_t(\boldsymbol{\phi}) = \frac{1}{\phi(B^{-1})}\{X_{t+s-j} + z_{t+j}(\boldsymbol{\phi})\}. \tag{A.6}$$

Substituting (A.6) in (A.4), we have

$$
\begin{aligned}
&\frac{\partial}{\partial \phi_j}\{\varphi(B^{-1})z_t(\boldsymbol{\phi})\}\Big|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0} \\
&\qquad = \left\{ -X_{t+s-j} + \frac{1}{\phi(B^{-1})}(X_{t+s-j} + z_{t+j}(\boldsymbol{\phi})) \right\}_{\boldsymbol{\phi}=\boldsymbol{\phi}_0} \\
&\qquad = \left\{ -X_{t+s-j} + \frac{\phi_0(B^{-1})B^s Z_{t+s-j}}{-\phi_{0r}\phi(B^{-1})\phi_0(B)} + \frac{z_{t+j}(\boldsymbol{\phi})}{\phi(B^{-1})} \right\}_{\boldsymbol{\phi}=\boldsymbol{\phi}_0} \\
&\qquad = -X_{t+s-j} - \frac{z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}(\boldsymbol{\phi}_0)}{\phi_0(B^{-1})}.
\end{aligned}
\tag{A.7}
$$

Finally, note that (A.7) implies that the coefficient of $n^{-1/2}$ in (A.3) is

$$\begin{aligned}
&\mathbf{u}'\left[\frac{\partial}{\partial \phi_j}\{\varphi(B^{-1})z_t(\boldsymbol{\phi})\}\Big|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0}+X_{t+s-j}\right]_{j=1}^{s}\\
&\quad=\mathbf{u}'\left[-\frac{z_{t-j}}{\phi_0(B)}+\frac{z_{t+j}(\boldsymbol{\phi}_0)}{\phi_0(B^{-1})}\right]_{j=1}^{s}\simeq\mathbf{u}'\left[-\frac{z_{t-j}}{\phi_0(B)}+\frac{z_{t+j}}{\phi_0(B^{-1})}\right]_{j=1}^{s}\\
&\quad=:-Y_t^{-}-Y_t^{+}=-Y_t,
\end{aligned}\tag{A.8}$$

where $Y_t^{-}\in\sigma(z_{t-1},z_{t-2},\ldots)$ because $\phi_0(B)$ is a causal operator, and $Y_t^{+}\in\sigma(z_{t+1},z_{t+2},\ldots)$ because $\phi_0(B^{-1})$ is a purely noncausal operator. It follows that Y_t is independent of $z_t:=Z_t\phi_{0r}^{-1}$.

Note that

$$\begin{aligned}
\operatorname{var}(Y_t)&=\phi_{0r}^{-2}\mathbf{u}'\left[\operatorname{cov}\left(-\frac{Z_{t-j}}{\phi_0(B)}+\frac{Z_{t+j}}{\phi_0(B^{-1})},-\frac{Z_{t-k}}{\phi_0(B)}+\frac{Z_{t+k}}{\phi_0(B^{-1})}\right)\right]_{j,k=1}^{s}\mathbf{u}\\
&=\phi_{0r}^{-2}\mathbf{u}'[2\gamma(j-k)]_{j,k=1}^{s}\mathbf{u}\\
&=2\phi_{0r}^{-2}\mathbf{u}'\boldsymbol{\Gamma}_s\mathbf{u},
\end{aligned}\tag{A.9}$$

where $\gamma(\cdot)$ is the autocovariance function of the causal AR(r) $\{Z_t/\phi_0(B)\}$ and $\boldsymbol{\Gamma}_s=[\gamma(j-k)]_{j,k=1}^{s}$.

We now compute the autocovariance function $\gamma^*(h)$ of the stationary process $\{Y_t\operatorname{sgn}(z_t)\}$:

$$\begin{aligned}
\gamma^*(h)&=\mathrm{E}\,[Y_t\operatorname{sgn}(z_t)Y_{t+h}\operatorname{sgn}(z_{t+h})]\\
&=\mathbf{u}'\mathrm{E}\left[\left[\left(-\frac{z_{t-j}}{\phi_0(B)}+\frac{z_{t+j}}{\phi_0(B^{-1})}\right)\operatorname{sgn}(z_t)\right.\right.\\
&\qquad\left.\left.\times\left(-\frac{z_{t+h-k}}{\phi_0(B)}+\frac{z_{t+h+k}}{\phi_0(B^{-1})}\right)\operatorname{sgn}(z_{t+h})\right]_{j,k=1}^{s}\right]\mathbf{u}\\
&=\mathbf{u}'\mathrm{E}\left[\left[\left(-\sum_{\ell=0}^{\infty}\psi_\ell z_{t-j-\ell}+\sum_{\ell=0}^{\infty}\psi_\ell z_{t+j+\ell}\right)\operatorname{sgn}(z_t)\right.\right.\\
&\qquad\left.\left.\times\left(-\sum_{m=0}^{\infty}\psi_m z_{t+h-k-m}+\sum_{m=0}^{\infty}\psi_m z_{t+h+k+m}\right)\operatorname{sgn}(z_{t+h})\right]_{j,k=1}^{s}\right]\mathbf{u}\\
&=\mathbf{u}'[\nu_{jk}(h)]_{j,k=1}^{s}\mathbf{u},
\end{aligned}\tag{A.10}$$

where

$$\nu_{jk}(h)=\begin{cases}\dfrac{2\gamma(j-k)}{\phi_{0r}^2}, & h=0,\\[2ex] \dfrac{-\psi_{|h|-j}\psi_{|h|-k}}{\phi_{0r}^2}\;\mathrm{E}^{\,2}|Z_1|, & h\neq 0,\end{cases}$$

and the $\{\psi_\ell\}$ are given by $\sum_{\ell=0}^{\infty}\psi_\ell z^\ell=1/\phi_0(z)$.

Thus,

$$
\begin{aligned}
&\gamma^*(0)+2\sum_{h=1}^{\infty}\gamma^*(h)\\
&\quad =\mathbf{u}'\left\{2\phi_{0r}^{-2}[\gamma(j-k)]_{j,k=1}^{s}-2\phi_{0r}^{-2}\ \mathrm{E}^{\,2}|Z_1|\left[\sum_{h=1}^{\infty}\psi_{h-j}\psi_{h-k}\right]_{j,k=1}^{s}\right\}\mathbf{u}\\
&\quad =\mathbf{u}'\left\{\frac{2}{\phi_{0r}^2}\mathbf{\Gamma}_s-\frac{2\ \mathrm{E}^{\,2}|Z_1|}{\phi_{0r}^2\sigma^2}\mathbf{\Gamma}_s\right\}\mathbf{u}\\
&\quad =\frac{2\ \mathrm{var}\ |Z_1|}{\phi_{0r}^2\sigma^2}\mathbf{u}'\mathbf{\Gamma}_s\mathbf{u}.
\end{aligned}
\tag{A.11}
$$

LEMMA 2. *For* $\mathbf{u}\in\mathbb{R}^s$, *let*

$$S_n(\mathbf{u})=m_n(\boldsymbol{\phi}_0+n^{-1/2}\mathbf{u})-\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi}_0)|$$

and define

$$
\begin{aligned}
S_n^*(\mathbf{u}) = \sum_{t=1}^{n-s}\Bigg\{\Bigg|&z_t(\boldsymbol{\phi}_0)+n^{-1/2}\mathbf{u}'\\
&\times\left[\frac{\partial}{\partial\phi_j}\{\varphi(B^{-1})z_t(\boldsymbol{\phi})\}\Big|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0}+X_{t+s-j}\right]_{j=1}^{s}\Bigg|-|z_t(\boldsymbol{\phi}_0)|\Bigg\}.
\end{aligned}
$$

Then:

1. $S_n^*\xrightarrow{\mathscr{L}}S$ *on* $C(\mathbb{R}^s)$ *where*

$$S(\mathbf{u})=\frac{f_\sigma(0)}{|\phi_{0r}|}\mathbf{u}'\mathbf{\Gamma}_s\mathbf{u}+\mathbf{u}'\mathbf{N}$$

and

$$\mathbf{N}\sim N\left(\mathbf{0},\frac{2\mathrm{var}(|Z_1|)}{\phi_{0r}^2\sigma^2}\mathbf{\Gamma}_s\right).$$

2. $S_n\xrightarrow{\mathscr{L}}S$.

PROOF. (1) Define

$$S_n^\dagger(\mathbf{u}) = \sum_{t=1}^{n-s}\{|z_t-n^{-1/2}Y_t|-|z_t|\},$$

where Y_t is given in equation (A.8). By Lemma (1) and (A.9),

$$S_n^\dagger(\mathbf{u})=-n^{-1/2}\sum_{t=1}^{n-s}Y_t\,\mathrm{sgn}(z_t)+\frac{f_\sigma(0)}{|\phi_{0r}|}\mathbf{u}'\mathbf{\Gamma}_s\mathbf{u}+o_p(1).$$

Thus, using (A.11), we have that the finite-dimensional distributions of $S_n^\dagger$ converge to those of S. But since $S_n^\dagger$ has convex sample paths, this implies that the convergence is in fact on $C(\mathbb{R}^s)$. [As shown in Theorem 10.8 of Rockafellar (1970), pointwise convergence of convex functions implies uniform convergence on compact sets, from which tightness of the $S_n^\dagger$ can be established.] It follows that $S_n^\dagger \xrightarrow{\mathscr{L}} S$ on $C(\mathbb{R}^s)$.

In order to transfer the convergence of $S_n^\dagger$ onto S_n^*, we first note that

$$z_{n-t-s} = \sum_{j=0}^{\infty} \psi_j U_{n-t+j} \quad \text{and} \quad z_{n-t-s}(\boldsymbol{\phi}_0) = \sum_{j=0}^{t} \psi_j U_{n-t+j}$$

for $t=0,1,\ldots,n-s+1$, where $U_t = -\phi_0(B)X_t$ and $\psi(B) = 1/\phi_0(B)$. Thus,

$$|z_{n-t-s} - z_{n-t-s}(\boldsymbol{\phi}_0)| = \left| \sum_{j=t+1}^{\infty} \psi_j U_{n-t+j} \right|$$

and hence

$$\begin{aligned}\limsup_{n\to\infty} E \sum_{t=M}^{n-s+1} |z_{n-t-s} - z_{n-t-s}(\boldsymbol{\phi}_0)| &\le C \sum_{t=M}^{\infty} \sum_{j=t+1}^{\infty} |\psi_j| \\ &\to 0,\end{aligned}$$

as $M\to\infty$. It now follows simply from these relations and the triangle inequality that $S_n^*(\mathbf{u}) - S_n^\dagger(\mathbf{u}) \xrightarrow{P} 0$ uniformly on compact sets which, combined with the convergence of $S_n^\dagger(\mathbf{u})$, yields (1).

(2) This argument is nearly identical to the one given on page 487 of Davis and Dunsmuir (1997) and is omitted. □

We conclude this Appendix with a result on strong consistency of the LAD estimators under a suitable identifiability condition.

PROPOSITION 1. *Assume the all-pass model* (2.1) *holds with* A1–A4. *Let* $\tilde{z}_1(\boldsymbol{\phi}) = -\phi(B)X_{1+s}/\phi(B^{-1})$. *Given* $\varepsilon > 0$, *let* $\boldsymbol{\Theta}$ *be the compact parameter space consisting of*

$$\{\boldsymbol{\phi} : \phi(z) \neq 0 \textit{ for all } |z| \le 1-\varepsilon\}.$$

If $E|\tilde{z}_1(\boldsymbol{\phi})|$ *has a unique minimum at* $\boldsymbol{\phi} = \boldsymbol{\phi}_0 \in \boldsymbol{\Theta}$, *then*

$$\hat{\boldsymbol{\phi}}_{\mathrm{LAD}} = \operatorname*{argmin}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}} m_n(\boldsymbol{\phi}) \to \boldsymbol{\phi}_0$$

almost surely.

PROOF. By the ergodic theorem, $T_n(\boldsymbol{\phi}) = n^{-1} m_n(\boldsymbol{\phi}) \to \mathrm{E}|\tilde{z}(\boldsymbol{\phi})|$ a.s. It suffices to show that $T_n(\boldsymbol{\phi}) \to \mathrm{E}|\tilde{z}(\boldsymbol{\phi})|$ a.s. uniformly on $\boldsymbol{\phi} \in \boldsymbol{\Theta}$. We begin by showing that $\{T_n(\boldsymbol{\phi})\}$ is uniformly equicontinuous on $\boldsymbol{\Theta}$ a.s.

Using the identity for $z \neq 0$,

$$|y|-|z|=(y-z)\operatorname{sgn}(z)+2y\{1_{\{z<0<y\}}-1_{\{y<0<z\}}\},$$

we have for $\boldsymbol{\phi},\boldsymbol{\theta}\in\boldsymbol{\Theta}$,

$$\begin{aligned}
T_n(\boldsymbol{\phi})-T_n(\boldsymbol{\theta}) &= n^{-1}\sum_{t=1}^{n-s}(|z_t(\boldsymbol{\phi})|-|z_t(\boldsymbol{\theta})|)\\
&= n^{-1}\sum_{t=1}^{n-s}(z_t(\boldsymbol{\phi})-z_t(\boldsymbol{\theta}))\operatorname{sgn}(z_t(\boldsymbol{\theta}))\\
&\quad +2\sum_{t=1}^{n-s}z_t(\boldsymbol{\phi})\{1_{\{z_t(\boldsymbol{\theta})<0<z_t(\boldsymbol{\phi})\}}-1_{\{z_t(\boldsymbol{\phi})<0<z_t(\boldsymbol{\theta})\}}\}\\
&=: I+II.
\end{aligned} \tag{A.12}$$

By the mean value theorem,

$$|I|\le n^{-1}\sum_{t=1}^{n-s}\left|\frac{\partial z_t(\boldsymbol{\phi}^*)}{\partial\boldsymbol{\phi}}\right||\boldsymbol{\phi}-\boldsymbol{\theta}|,$$

where $\boldsymbol{\phi}^*$ is between $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. Using (A.6) and the definition of $z_t(\boldsymbol{\phi})$, it follows that there exist coefficients $\psi_j\ge 0$ decaying at a geometric rate such that

$$\sup_{\boldsymbol{\phi}\in\boldsymbol{\Theta}}\left|\frac{\partial z_t(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right|\le\sum_{j=0}^{\infty}\psi_j|X_{t-s+j}|$$

and

$$\sup_{\boldsymbol{\phi}\in\boldsymbol{\Theta}}|z_t(\boldsymbol{\phi})|\le\sum_{j=0}^{\infty}\psi_j|X_{t-s+j}|.$$

Hence

$$|I|\le|\boldsymbol{\phi}-\boldsymbol{\theta}|n^{-1}\sum_{t=1}^{n-s}\sum_{j=0}^{\infty}\psi_j|X_{t-s+j}|=|\boldsymbol{\phi}-\boldsymbol{\theta}|O(1)\qquad\text{a.s.} \tag{A.13}$$

Turning to the second term in (A.12), we have for a fixed $\delta>0$,

$$\begin{aligned}
|II|&\le 2n^{-1}\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi})|1_{\{|z_t(\boldsymbol{\phi})|\le\delta\}}+2n^{-1}\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi})|1_{\{|z_t(\boldsymbol{\phi})|>\delta\}}1_{\{|z_t(\boldsymbol{\phi})-z_t(\boldsymbol{\theta})|>\delta\}}\\
&\le 2\delta+2n^{-1}\sum_{t=1}^{n-s}|z_t(\boldsymbol{\phi})||z_t(\boldsymbol{\phi})-z_t(\boldsymbol{\theta})|/\delta\\
&\le 2\delta+2n^{-1}\delta^{-1}\sum_{t=1}^{n-s}|\boldsymbol{\phi}-\boldsymbol{\theta}|\left(\sum_{j=0}^{\infty}\psi_j|X_{t-s+j}|\right)^2\\
&=2\delta+\delta^{-1}|\boldsymbol{\phi}-\boldsymbol{\theta}|O(1)\qquad\text{a.s.}
\end{aligned} \tag{A.14}$$

Since the $O(1)$ terms in (A.13) and (A.14) do not depend on $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, or δ, it follows that $\{T_n\}$ is equicontinuous on $\boldsymbol{\theta}$ a.s. It is also easily shown that the sequence $\{T_n\}$ is uniformly bounded a.s. Applying the Arzelà-Ascoli theorem, we conclude that $T_n(\boldsymbol{\phi}) \to \mathrm{E}\,|\tilde{z}_1(\boldsymbol{\phi})|$ a.s. uniformly. The uniqueness of the minimizer of $\mathrm{E}|\tilde{z}_1(\boldsymbol{\phi})|$ ensures that $\hat{\boldsymbol{\phi}}_{\mathrm{LAD}} \to \boldsymbol{\phi}_0$ a.s. $\square$

Acknowledgments. We are grateful to three anonymous referees for numerous constructive suggestions. Some of the work of the first author was conducted while he was Associate Professor of Statistics at Iowa State University.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (B. N. Petrov and F. Csaki, eds.) 267–281. Akademiai Kiado, Budapest.

BENVENISTE, A., GOURSAT, M. and ROGET, G. (1980). Robust identification of a nonminimum phase system: blind adjustment of linear equalizer in data communications. *IEEE Trans. Auto. Control* **AC–25** 385–398.

BICKEL, P. J. and BÜHLMANN, P. (1996). What is a linear process? *Proc Nat. Acad. Sci.* **93** 12128–12131.

BLASS, W. E. and HALSEY, G. W. (1981). *Deconvolution of Absorption Spectra*. Academic Press, New York.

BOLLERSLEV, T., CHOU, R. Y., and KRONER, K. F. (1992). ARCH modeling in finance. *J. Econometrics* **52** 5–59.

BREIDT, F. J. and DAVIS, R. A. (1991). Time-reversibility, identifiability and independence of innovations for stationary time series. *J. Time Ser. Anal.* **13** 377–390.

BREIDT, F. J., DAVIS, R. A., LII, K.-S. and ROSENBLATT, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *J. Multivariate Anal.* **36** 175–198.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods,* 2nd ed. Springer, 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 base phase estimation of 1-d and 2-d nonminimum phase systems by allpass filtering. *IEEE Trans. Signal Processing* **45** 1742–1762.

CLARK, P. K. (1973). A subordinated stochastic process model with finite variances for speculative prices. *Econometrica* **41** 135–156.

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

DONOHO, D. (1981). On minimum entropy deconvolution. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 565–608. Academic Press, New York.

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007.

GIANNAKIS, G. B. and SWAMI, A. (1990). On estimating noncausal nonminimum phase ARMA models of non-Gaussian processes. *IEEE Trans. Acoustics, Speech, and Signal Processing* **38** 478–495.

GODFREY, R. and ROCCA, F. (1981). Zero memory nonlinear deconvolution. *Geophysical Prospecting* **29** 189–228.

HOOKE, R. and JEEVES, T. (1961). A direct search solution of numerical and statistical problems. *J. Assoc. Comput. Mach.* **8** 212–229.

HSUEH, A.-C. and MENDEL, J. M. (1985). Minimum-variance and maximum-likelihood deconvolution for noncausal channel models. *IEEE Trans. Geoscience and Remote Sensing* **23** 797–808.

JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Bus. Econom. Statist.* **12** 371–417.

HUANG, J. and PAWITAN, Y. (2000). Quasi-likelihood estimation of noninvertible moving average processes. *Scandinavian J. Statist.* **27** 689–702.

KREISS, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.

LII, K.-S. and ROSENBLATT, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. *Ann. Statist.* **10** 1195–1208.

LII, K.-S. and ROSENBLATT, M. (1992). An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes. *J. Multivariate Anal.* **43** 272–299.

LII, K.-S. and ROSENBLATT, M. (1996). Maximum likelihood estimation for non-Gaussian nonminimum phase ARMA sequences. *Statist. Sinica* **6** 1–22.

MANDELBROT, B. (1963). The variation of certain speculative prices. *J. Business* **36**, 394–419.

MCLEOD, A. I. and LI, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Ser. Anal.* **4** 269–273.

OOE, M. and ULRYCH, T. J. (1979). Minimum entropy deconvolution with an exponential transformation. *Geophysical Prospecting* **27** 458–473.

RABINER, L. R. and SCHAFER, R. M. (1978). *Digital Processing of Speech Signals*. Prentice-Hall, Englewood Cliffs, NJ.

ROBINSON, P. M. (1987). Time series residuals with application to probability density estimation. *Time Ser. Anal.* **8** 329–344.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton Univ. Press.

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

SCARGLE, J. D. (1981). Phase-sensitive deconvolution to model random processes, with special reference to astronomical data. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 549–564. Academic Press, New York.

WIGGINS, R. A. (1978). Minimum entropy deconvolution. *Geoexploration* **16** 21–35.

F. J. BREIDT
RICHARD A. DAVIS
DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523
E-MAIL: jbreidt@stat.colostate.edu
rdavis@stat.colostate.edu

A. ALEXANDRE TRINDADE
DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611
E-MAIL: trindade@stat.ufl.edu