Least Absolute Deviation Estimation for All-Pass Time Series Models

F. Jay Breidt
Iowa State University

Richard A. Davis\(^1\)
Colorado State University

Alex Trindade\(^1\)
Colorado State University

September 23, 1999

Abstract

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 Laplace (two-sided exponential) noise yields a modified absolute deviations criterion, which can be used even if the underlying noise is not Laplace. 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 "non-linear" behavior, and is applied to stock market volume data as part of a two-step procedure for fitting noncausal autoregressions.

*Keywords.* Laplace density, noncausal, noninvertible, white noise.

\(^1\) This research supported in part by NSF DMS Grant No. DMS-9972015.
1 Introduction

In the analysis of returns on financial assets such as stocks, it is common to observe the following characteristics:

- lack of serial correlation
- heavy-tailed marginal distributions
- 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) and the stochastic volatility models (Clark, 1973; Jacquier, Polson, and Rossi, 1994) are suggested for such time series. It is perhaps less well known that linear, non-Gaussian models can display exactly this behavior.

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

All-pass models are widely used in the engineering literature, and usually 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 could then be modeled 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, method of moments estimators using moments of order greater than two are often used to estimate such models (Giannakis and Swami, 1990; Chi and Kung, 1995).

In this paper 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 Laplace (two-sided exponential) noise is derived, yielding a modified absolute deviations criterion. This criterion can be used even if the underlying noise is not Laplace. Asymptotic normality for least absolute deviation estimators of the model parameters is established under general conditions in Section 3, and order selection is considered. Behavior of the estimators in finite samples is studied via simulation in Section 4.1. 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, a noncausal AR(1) model is shown to provide a reasonable fit to the time series of daily log volumes of Microsoft stock. In contrast, 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 $(B^k X_t = X_{t-k}, k = 0, \pm 1, \pm 2, \ldots)$ and let

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

be a $p$th-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 polynomial $\phi(B^{-1})$ is purely noncausal in the sense that all of its roots are inside the unit circle, hence

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

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

Let

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

where $\phi_0(z) \neq 0$ for $|z| \leq 1$. Define $\phi_0 = 1$ and suppose that $\phi_{0r} \neq 0$ for some $r = 0, 1, \ldots, p$ and $\phi_{0j} = 0$ for $j = r+1, \ldots, p$. Then a causal all-pass time series is the autoregressive-moving average (ARMA) \( \{X_t\} \) which satisfies the difference equations

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

(1)

where \( \{Z_t\} \) is an independent and identically distributed (iid) sequence of random variables with mean 0, variance $\sigma^2$, and common distribution function $F_\sigma$. We assume that $F_\sigma$ 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 $F_\sigma$, where $\sigma$ is a scale parameter.

Note that the spectral density of \( \{X_t\} \) in (1) is

$$\frac{|e^{-i \omega t}|^2 |\phi_0(e^{i \omega})|^2}{\hat{\phi}_0^2} \left( \frac{\sigma^2}{2 \pi} \right) = \frac{\sigma^2}{\hat{\phi}_{0r}^2 2 \pi},$$

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 $N(0, \sigma^2 \hat{\phi}_{0r}^2)$, but independence does not hold in the non-Gaussian case (e.g., Breidt and Davis, 1991).

Rearranging (1), we have the backward recursion

$$z_{t-p} = \phi_{01} z_{t-p+1} + \cdots + \phi_{0p} z_t - \left( X_t - \phi_{01} X_{t-1} - \cdots - \phi_{0p} X_{t-p} \right),$$

where $z_t := Z_t \phi_{0r}^{-1}$. We define the analogous recursion for an arbitrary, causal autoregressive polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| \leq 1$ as follows:

$$z_{t-p}(\phi) = \begin{cases} 0, & t = n+p, \ldots, n+1, \\ \phi_{1z_{t-p+1}}(\phi) + \cdots + \phi_{pz_t}(\phi) - \phi(B)X_t, & t = n, \ldots, p+1, \end{cases}$$

(2)
where \( \phi = (\phi_1, \ldots, \phi_p)' \). (We do not require \( \phi_p \neq 0 \).) Let \( \phi_0 = (\phi_{01}, \ldots, \phi_{0p})' = (\phi_0, 0, \ldots, 0)' \). Note that \( \{x_t(\phi_0)\} \) is a close approximation to \( \{x_t\} \), in which the error is due to the initialization with zeros. Though \( \{x_t\} \) is iid, \( \{x_t(\phi)\} \), 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), and write

\[
-\phi(B^{-1})B^p z_t(\phi) = \phi(B) x_t.
\]

We then approximate the likelihood of a realization of length \( n \), \((X_1, \ldots, X_n)\), from the model (1) using techniques similar to those in Breidt, Davis, Lii, and Rosenblatt (1991) and Lii and Rosenblatt (1992, 1996).

Consider

\[
x := (X_{-p}, \ldots, X_0, X_1, \ldots, X_n, z_{n-p+1}(\phi), \ldots, z_n(\phi))',
\]

\[
y := (X_{-p}, \ldots, X_0, -\phi(B^{-1})B^p z_1(\phi), \ldots, -\phi(B^{-1})B^p z_n(\phi), z_{n-p+1}(\phi), \ldots, z_n(\phi))',
\]

and

\[
z := (X_{-p}, \ldots, X_0, z_1(\phi), \ldots, z_{n-p}(\phi), z_1(\phi), \ldots, z_{n-p+1}(\phi), \ldots, z_n(\phi))'.
\]

Note that if \( \phi \) is the true parameter vector, then the first \( 2p \) terms of \( z \) are independent of the last \( n \) terms by causality of \( \phi(\cdot) \).

From (3), it follows that

\[
y = Ax
\]

where \( |A| = 1 \). Also, since

\[
-\phi(B^{-1})B^p z_t(\phi) = -(z_{t-p}(\phi) - \phi_1 z_{t-p+1}(\phi) - \cdots - \phi_p z_t(\phi)),
\]

\[
y = \begin{bmatrix}
X_{-p} \\
\vdots \\
X_0 \\
-\phi(B^{-1}) z_{1-p}(\phi) \\
\vdots \\
-\phi(B^{-1}) z_{n-p}(\phi) \\
z_{n}(\phi)
\end{bmatrix}
= \begin{bmatrix}
\begin{bmatrix}
1 \\
-\phi_1 \\
-\phi_1 \\
\vdots \\
-\phi_1 \\
\phi_1
\end{bmatrix}
\end{bmatrix}
\begin{bmatrix}
X_{-p} \\
\vdots \\
X_0 \\
\phi_{-p+1}(\phi) \\
\vdots \\
\phi_{n-p+1}(\phi)
\end{bmatrix}
= \begin{bmatrix}
X_{-p} \\
\vdots \\
X_0 \\
z_{1-p}(\phi) \\
\vdots \\
z_{n}(\phi)
\end{bmatrix}
\]
\[ \begin{align*}
\mathbf{z} &= B\mathbf{z} \\
\text{where } |B| &= 1.
\end{align*} \]

Now the joint distribution of \( \mathbf{z} \) under \( \phi \) is given by

\[ h(\mathbf{z}) = h_1(X_{1-p}, \ldots, X_0, z_{1-p}(\phi), \ldots, z_0(\phi)) \left( \prod_{t=1}^{n-p} f_\sigma(\phi_t z_t(\phi)) \mid \phi_q \right) h_2(z_{n-p+1}(\phi), \ldots, z_n(\phi)), \]

where \( q = \max\{0 \leq j \leq p : \phi_j \neq 0\} \), so the joint distribution of \( \mathbf{x} \) under \( \phi \) is given by

\[ h(\mathbf{x}) = h_1 \left( \prod_{t=1}^{n-p} f_\sigma(\phi_t z_t(\phi)) \mid \phi_q \right) h_2, \]

where \( h_1 \) and \( h_2 \) do not depend on \( n \).

This suggests approximating the log-likelihood of \( (\phi, \sigma) \) given the data as

\[ \mathcal{L}(\phi, \sigma) = \sum_{t=1}^{n-p} \ln f_\sigma(\phi_t z_t(\phi)) + (n-p) \ln |\phi_q| \]

\[ = -(n-p) \ln \sigma + \sum_{t=1}^{n-p} \ln f(\sigma^{-1} \phi_t z_t(\phi)) + (n-p) \ln |\phi_q|, \]

where the \( \{z_t(\phi)\} \) can be computed recursively from (2).

### 2.3 Least Absolute Deviations

If the noise distribution is Laplacian, or two-sided exponential, with mean 0, variance \( \sigma^2 \), and density

\[ f_\sigma(x) = \frac{1}{\sigma} f \left( \frac{x}{\sigma} \right) = \frac{1}{\sqrt{2\sigma}} \exp \left( -\frac{\sqrt{2}|x|}{\sigma} \right), \]

then the log-likelihood is given by

\[ \text{constant} - (n-p) \ln \kappa - \sum_{t=1}^{n-p} \frac{\sqrt{2}|z_t(\phi)|}{\kappa}, \]

where \( \kappa = \sigma |\phi_q|^{-1} \). Setting the partial derivative of (8) with respect to \( \kappa \) equal to zero, we obtain the least absolute deviations (LAD) estimator of \( \kappa \),

\[ \hat{\kappa} = \frac{\sqrt{2}}{n-p} \sum_{t=1}^{n-p} |z_t(\phi)|, \]

where the \( \{z_t(\phi)\} \) are computed from (2) using the LAD estimates of \( \phi \). Substituting \( \hat{\kappa} \) for \( \kappa \) in (8), we obtain the concentrated Laplacian likelihood

\[ \ell(\phi) = \text{constant} - (n-p) \ln \sum_{t=1}^{n-p} |z_t(\phi)|. \]

Maximizing \( \ell(\phi) \) is equivalent to minimizing the absolute deviations criterion,

\[ m_n(\phi) = \sum_{t=1}^{n-p} |z_t(\phi)|. \]
3 Asymptotic Results

3.1 Parameter Estimation

The following preliminary result parallels Theorem 1 of Davis and Dunsmuir (1997), which covers one-sided linear processes.

**Theorem 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

\[
S_n := \sum_{t=1}^{n-p} \left( |z_t - n^{-1/2}Y_t| - |z_t| \right) \overset{p}{\to} \text{Var}(Y_t) g(0) + N,
\]

where

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

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

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

**Proof:** Using the identity for \( z \neq 0, \)

\[
|z - y| - |z| = -y \text{sgn}(z) + 2(y - z) \left\{ 1_{0 < z < y} - 1_{y < z < 0} \right\},
\]

we have

\[
S_n = -n^{-1/2} \sum_{t=1}^{n-p} Y_t \text{sgn}(z_t)
\]

\[
+ 2 \sum_{t=1}^{n-p} \left( n^{-1/2} Y_t - z_t \right) \left\{ 1_{0 < z_t < n^{-1/2}Y_t} - 1_{n^{-1/2}Y_t < z_t < 0} \right\}
\]

\[
=: A_n + B_n.
\]

A standard truncation argument, truncating \( Y_t \) to create the \( 2M \)-dependent sequence \( \{Y_t^M, \text{sgn}(z_t)\} = \{\sum_{j=-M}^{M} c_j z_{t-j}, \text{sgn}(z_t)\} \), allows application of a central limit theorem (Brockwell and Davis, 1991, Theorem 6.4.2) for each \( M \), from which asymptotic normality of \( A_n \) can be deduced.

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

\[
\limsup_{n \to \infty} n \mathbb{E} \left[ W_{nt}^2 \right]
\]

\[
= \limsup_{n \to \infty} \left[ n \int_0^{cn^{1/2}} \int_0^{n^{-1/2}y} (n^{-1/2}y - b^{-1}z)^2 G(dz) F_Y(dy) \right]
\]
\[ n \int_{e^{-1/2}}^{\infty} \int_{0}^{n^{-1/2}y} \left( n^{-1/2}y - b^{-1}z \right)^2 G(dz)F_Y(dy) \]
\[ \leq \limsup_{n \to \infty} \left[ n \int_{0}^{n^{-1/2}y} \left( n^{-1/2}y - b^{-1}z \right)^2 (g(0) + \delta) \, dz \, F_Y(dy) \right] \]
\[ = \limsup_{n \to \infty} (\text{const}) n \int_{0}^{en^{-1/2}} n^{-1}y^2 G(dz)F_Y(dy) \]
\[ \leq \limsup_{n \to \infty} (\text{const}) \epsilon \mathbb{E} \left[ Y_1^2 1_{\{ Y_1 > 0 \}} \right], \]  
\[ \text{(11)} \]

and since \( \epsilon > 0 \) 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 \( \{ Y_t > 0 \} \),
\[ \mathbb{E}[W_{nt} \mid z_{t-1}, z_{t-2}, \ldots] \]
\[ = \mathbb{E} \left[ (n^{-1/2} Y_t - z_t) 1_{\{ 0 < z_t < n^{-1/2} Y_t \}} \mid z_{t-1}, z_{t-2}, \ldots \right] \]
\[ = \int_{-Y_t^-}^{\infty} \int_{0}^{n^{-1/2}(Y_t^- + y)} \left\{ n^{-1/2}(Y_t^- + y) - z \right\} G(dz)F_Y(dy) \]
\[ = \int_{-Y_t^-}^{\infty} n^{-1/2}(Y_t^- + y) \left\{ G(n^{-1/2}(Y_t^- + y)) - G(0) \right\} F_Y(dy) \]
\[ - \int_{-Y_t^-}^{\infty} n^{-1/2}(Y_t^- + y) zG(dz)F_Y(dy) \]
\[ = \int_{-Y_t^-}^{\infty} n^{-1}(Y_t^- + y)^2 g(0)F_Y(dy) \]
\[ - \int_{-Y_t^-}^{\infty} g(0) \frac{n^{-1}(Y_t^- + y)^2}{2} F_Y(dy) \]
\[ = \frac{g(0)}{2n} \int_{-Y_t^-}^{\infty} (Y_t^- + y)^2 F_Y(dy), \]

where the approximation holds on the set \( \{ n^{-1/2} Y_t < \epsilon \} \), for \( \epsilon > 0 \) small. Since
\[ \Pr \left\{ n^{-1/2} \max(|Y_1|, \ldots, |Y_n|) > \epsilon \right\} \leq \Pr \left\{ \bigcup_{t=1}^{\infty} \{ |Y_t| > \epsilon n^{1/2} \} \right\} \]
\[ \leq n \Pr \left\{ |Y_1| > \epsilon n^{1/2} \right\} \]
\[ \leq \epsilon^{-2} \mathbb{E} \left[ Y_1^2 1_{\{ Y_1^2 > \epsilon^2 n \}} \right] \to 0 \]

as \( n \to \infty \), it follows from the ergodic theorem that
\[ \sum_{t=1}^{n-p} \mathbb{E}[W_{nt} \mid z_{t-1}, z_{t-2}, \ldots] \overset{p}{\to} \frac{g(0)}{2} \mathbb{E} \left[ \int_{-Y_t^-}^{\infty} (Y_t^- + y)^2 F_Y(dy) \right]. \]  
\[ \text{(12)} \]
By (11),
\[
\text{Var} \left( \sum_{t=1}^{n-p} (W_{nt} - E[W_{nt} | z_{t-1}, z_{t-2}, \ldots]) \right) = \sum_{t=1}^{n-p} \text{Var} \left( W_{nt} - E[W_{nt} | z_{t-1}, z_{t-2}, \ldots] \right) \\
\leq \sum_{t=1}^{n-p} E \left[ W_{nt}^2 \right] \to 0,
\]
so that from (12) we have
\[
\sum_{t=1}^{n-p} W_{nt} \xrightarrow{P} \frac{g(0)}{2} E \left[ \int_{-\infty}^{\infty} (Y_t^- + y)^2 F_{Y_+}(dy) \right].
\]
Using the same argument for the second indicator in \( B_n \), we obtain
\[
B_n \xrightarrow{P} \frac{g(0)}{2} E \left[ \int_{-\infty}^{\infty} (Y_t^- + y)^2 F_{Y_+}(dy) \right] = \frac{g(0)}{2} \text{Var} (Y_t),
\]
which concludes the proof. \( \square \)

Define \( \varphi(z) = \phi_1 z + \cdots + \phi_p z^p = 1 - \phi(z) \) and \( \varphi_0(z) = 1 - \phi_0(z) \). In what follows, we linearize \( \varphi(B^{-1})z_t(\phi) \) around \( \phi_0 \) within the criterion function \( m_n \); that is, \( \varphi(B^{-1})z_t(\phi) \) is approximated by
\[
\varphi_0(B^{-1})z_t(\phi_0) + \sum_{j=1}^{p} \frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1})z_t(\phi) \right\} \bigg|_{\phi=\phi_0} (\phi_j - \phi_{0j}).
\]
By (2), the criterion function (10) can be written as
\[
m_n = \sum_{t=1}^{n-p} \left| \varphi(B^{-1})z_t(\phi) - \phi(B)X_{t+p} \right|
\]
\[
= \sum_{t=1}^{n-p} \left| \varphi(B^{-1})B^p z_{t+p}(\phi) - \phi_0(B)X_{t+p} + (\phi_0(B) - \phi(B))X_{t+p} \right|
\]
\[
= \sum_{t=1}^{n-p} \left| \varphi_0(B^{-1})B^p z_{t+p}(\phi_0) - B^p z_{t+p}(\phi_0) + z_t(\phi_0) \right|
\]
\[
- \sum_{j=1}^{p} \frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1})z_t(\phi) \right\} \bigg|_{\phi=\phi_0} (\phi_j - \phi_{0j})
\]
\[
- \phi_0(B)X_{t+p} + n^{1/2}(\phi - \phi_0)'n^{-1/2}(X_{t+p-1}, \ldots, X_t)'
\]
\[
= \sum_{t=1}^{n-p} z_t(\phi_0) + n^{-1/2} \mathbf{u}' \left[ \frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1})z_t(\phi) \right\} \bigg|_{\phi=\phi_0 + X_{t+p-j}} \right]_{j=1}^{p}, \quad (13)
\]
where \( \mathbf{u} = n^{1/2}(\phi - \phi_0) \).

Now
\[
\phi(B)X_{t+p} = -z_t(\phi) + \varphi(B^{-1})z_t(\phi),
\]
so
\[
\frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1})z_t(\phi) \right\} = -X_{t+p-j} + \frac{\partial}{\partial \phi_j} z_t(\phi). \quad (14)
\]
Also,
\[
\frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1})z_t(\phi) \right\} = \varphi(B^{-1}) \frac{\partial}{\partial \phi_j} z_t(\phi) + z_{t+j}(\phi).
\]  

(15)

Equating (14) and (15) and solving for \( \frac{\partial z_t(\phi) / \partial \phi_j} \), we obtain
\[
\frac{\partial}{\partial \phi_j} z_t(\phi) = \frac{1}{\varphi(B^{-1})} \{X_{t+p-j} + z_{t+j}(\phi)\}.
\]  

(16)

Substituting (16) in (14), we have
\[
\frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1})z_t(\phi) \right\} \bigg|_{\phi=\phi_0}
= \left\{ -X_{t+p-j} + \frac{1}{\varphi(B^{-1})} \{X_{t+p-j} + z_{t+j}(\phi)\} \right\} \bigg|_{\phi=\phi_0}
= \left\{ -X_{t+p-j} + \frac{\phi_0(B^{-1})B^pZ_{t+p-j} + z_{t+j}(\phi)}{-\phi_0\phi(B^{-1})\phi_0(B)} \right\} \bigg|_{\phi=\phi_0}
= -X_{t+p-j} - \frac{z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}(\phi_0)}{\phi_0(B^{-1})}.
\]  

(17)

Finally, note that (17) implies that the coefficient of \( n^{-1/2} \) in (13) is
\[
\mathbf{u}^T \left[ \frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1})z_t(\phi) \right\} \bigg|_{\phi=\phi_0} + X_{t+p-j} \right]^p_{j=1}
= \mathbf{u}^T \left[ \frac{z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}(\phi_0)}{\phi_0(B^{-1})} \right]^p_{j=1}
\approx \mathbf{u}^T \left[ \frac{z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}}{\phi_0(B^{-1})} \right]^p_{j=1}
= -Y_t^- - Y_t^+ = -Y_t,
\]  

(18)

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 \( z_t := Z_t\phi_{0r} \).

Note that
\[
\text{Var} (Y_t) = \phi_0^{-2} \mathbf{u}^T \left[ \text{Cov} \left( \frac{-Z_{t-j}}{\phi_0(B)} + \frac{Z_{t+j}}{\phi_0(B^{-1})}, -\frac{Z_{t-k}}{\phi_0(B)} + \frac{Z_{t+k}}{\phi_0(B^{-1})} \right) \right]^p_{j,k=1} \mathbf{u}
= \phi_0^{-2} \mathbf{u}^T [2\gamma(j-k)]^p_{j,k=1} \mathbf{u}
= 2\phi_0^{-2} \mathbf{u}^T \mathbf{\Gamma}_p \mathbf{u},
\]  

(19)

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

We now compute the autocovariance function \( \gamma^*(h) \) of the stationary process \( \{Y_t \text{ sgn} (z_t)\} \):
\[
\gamma^*(h) = \mathbb{E} [Y_t \text{ sgn} (z_t) Y_{t+h} \text{ sgn} (z_{t+h})]
= \mathbf{u}^T \mathbb{E} \left[ \left( \frac{-z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}}{\phi_0(B^{-1})} \right) \text{ sgn} (z_t)
- \frac{z_{t+k}}{\phi_0(B)} + \frac{z_{t+k}}{\phi_0(B^{-1})} \right] \text{ sgn} (z_{t+h})^p_{j,k=1} \mathbf{u}
\]

\[
= \mathbf{u}^T \mathbf{\Gamma}_p \mathbf{u},
\]  

where \( \gamma(\cdot) \) is the autocovariance function of the causal AR\((r)\) \( \{Z_t/\phi_0(B)\} \) and \( \mathbf{\Gamma}_p = [\gamma(j-k)]^p_{j,k=1} \).
\[
\begin{align*}
&= u' E \left[ \left( - \sum_{\ell=0}^{\infty} \psi_{\ell} z_{t-\ell} + \sum_{\ell=0}^{\infty} \psi_{\ell} z_{t+\ell} \right) \operatorname{sgn}(z_t) \right. \\
&\quad \left. \left( - \sum_{m=0}^{\infty} \psi_{m} z_{t+h-k-m} + \sum_{m=0}^{\infty} \psi_{m} z_{t+h+k+m} \right) \operatorname{sgn}(z_{t+h}) \right]_{j,k=1}^p u \\
&= u' \left[ \nu_{jk}(h) \right]_{j,k=1}^p u,
\end{align*}
\]

where
\[
\nu_{jk}(h) = \begin{cases} 
\frac{2\gamma(j-k)}{\phi_0}, & h = 0 \\
-\psi_{|h-j|} \psi_{|h-k|} E^2 |Z_1|, & h \neq 0,
\end{cases}
\]

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

Thus,
\[
\gamma^*(0) + 2 \sum_{h=1}^{\infty} \gamma^*(h) = u' \left\{ 2\phi_0^{-2} \gamma(j-k) \left[ j, k \right]_{j,k=1}^p - 2\phi_0^{-2} E^2 |Z_1| \left[ \sum_{h=1}^{\infty} \psi_{h-j} \psi_{h-k} \right]_{j,k=1}^p \right\} u \\
= u' \left\{ \frac{2}{\phi_0^2} \Gamma_p - \frac{2E^2 |Z_1| \Gamma_p}{\phi_0^2 \sigma^2} \right\} u \\
= \frac{2 \text{Var}(|Z_1|)}{\phi_0^2 \sigma^2} u' \Gamma_p u.
\]

**Theorem 2** For \( u \in \mathbb{R}^p \), let
\[
S_n(u) = m_n(\phi_0 + n^{-1/2} u) - \sum_{t=1}^{n-p} |z_t(\phi_0)|
\]

and define
\[
S^*_n(u) = \sum_{t=1}^{n-p} \left\{ z_t(\phi_0) + n^{-1/2} u' \left[ \frac{\partial}{\partial \phi_j} \{ \varphi(B^{-1}) z_t(\phi) \} \right]_{\phi=\phi_0 + X_{t+p-j}}^{p} \right\} - |z_t(\phi_0)|.
\]

Then
1. \( S^*_n \xrightarrow{L} S \) on \( C(\mathbb{R}^p) \) where
\[
S(u) = \frac{f_\varphi(0)}{\phi_0} u' \Gamma_p u + u' N
\]

and
\[
N \sim N \left( 0, \frac{2 \text{Var}(|Z_1|) \Gamma_p}{\phi_0^2 \sigma^2} \right).
\]

2. \( S_n \xrightarrow{L} S \).

**Proof:** (1) Define
\[
S^t_n(u) = \sum_{t=1}^{n-p} \left\{ |z_t - n^{-1/2} Y_t| - |z_t| \right\},
\]
where $Y_t$ is given in equation (18). By Theorem 1 and (19),

$$S_n^1(u) = -n^{-1/2} \sum_{t=1}^{n-p} Y_t \text{sgn}(z_t) + \frac{f_\sigma(0)}{\phi_0} u \mathbf{T}_p u + o_p(1).$$

Thus, using (21), we have that the finite dimensional distributions of $S_n^1$ converge to those of $S$. But since $S_n^1$ has convex sample paths, this implies that the convergence is in fact on $C(\mathbb{R}^p)$. (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 $S_n^1$ can be established.) It follows that $S_n^1 \overset{p}{\to} S$ on $C(\mathbb{R}^p)$.

In order to transfer the convergence of $S_n^1$ onto $S_n^*$, we first note that

$$z_{n-t-p} = \sum_{j=0}^{\infty} \psi_j U_{n-t+j} \quad \text{and} \quad z_{n-t-p}(\phi_0) = \sum_{j=0}^{t} \psi_j U_{n-t+j}$$

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

$$|z_{n-t-p} - z_{n-t-p}(\phi_0)| = \left| \sum_{j=t+1}^{\infty} \psi_j U_{n-t+j} \right|$$

and hence

$$\limsup_{n \to \infty} \sum_{t=M}^{n-p+1} \sum_{j=t+1}^{\infty} |\psi_j| \leq C \sum_{t=M}^{\infty} \sum_{j=t+1}^{\infty} |\psi_j| \to 0,$$

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

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

**Corollary 1** There exists a sequence of local minimizers $\hat{\phi}_{LAD}$ of $S_n$ such that

$$\frac{n^{1/2}(\hat{\phi}_{LAD} - \phi_0)}{\sqrt{\frac{\text{Var}(Z_1)}{2f_\sigma(0)}}} \sim N(0, \frac{\text{Var}(Z_1)}{2\sigma^4 f_\sigma^2(0)} \mathbf{T}_p^{-1}). \quad (22)$$

**Proof:** The minimizer of the limit process in Theorem 2 is $-|\phi_0|/(2f_\sigma(0))\mathbf{T}_p^{-1}$. (See Davis and Dunsmuir (1997), Corollary 1.)

**Remark:** 1. The sequence of local minimizers in the corollary depends on the unknown $\phi_0$, which may not be the unique global minimizer of $\mathbf{E}|z_1(\phi)|$. If $\phi_0$ and $\phi_1$ are both local minimizers of $\mathbf{E}|z_1(\phi)|$, then there may exist a sequence of local minimizers of the LAD criterion which converges to $\phi_0$ and another sequence of local minimizers which converges to $\phi_1$. Without prior knowledge of the unknown $\phi_0$, we do not know in practice which local minimizer to choose.

In the Gaussian case, for example, any choice of $\phi_0$ (with $\phi_0 \neq 0$) together with $\sigma_0^2 := \phi_0^2 \text{Var}(X_t)$ satisfies model (1) with innovations $\{Z_t\}$ iid $N(0, \sigma_0^2)$ and $\{X_t\}$ iid $N(0, \sigma_0^2 \phi_0^{-2})$. Choose any $\phi_1 \neq \phi_0$ with $\phi_1p \neq 0$ and set $\sigma_1^2 := \phi_1^2 \text{Var}(X_t)$. Then

$$\mathbf{E}|z_1(\phi_1)| = \mathbf{E}\left|\frac{Z_1 \sigma_1}{\phi_1 \sigma_0}\right| = \mathbf{E}\left|\frac{Z_1 \text{Var}^{1/2}(X_t)}{\sigma_0}\right| = \mathbf{E}|z_1(\phi_0)|$$
so that $E|z_1(\phi)|$ is not uniquely minimized at $\phi_0$.

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

$$E \left| \sum_{j=-\infty}^{\infty} c_j Z_{t-j} \right| > E|Z_1|$$

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

$$E|z_1(\phi)| = E \left| \frac{\phi_0(B^{-1})\phi(B)}{\phi_0(B^{-1})\phi_0(B)} Z_t \right| > E|z_1(\phi_0)|,$$

so that $\phi_0$ is the unique global minimizer. Jian and Pawitan (1998) give sufficient conditions for (23) and show that it is satisfied by the Laplace, Student's $t$, contaminated normal, and other standard heavy-tailed distributions.

2. Note that the asymptotic covariance matrix from (22) is a scalar multiple of the asymptotic covariance matrix for the vector of Gaussian likelihood estimators of the corresponding $p$th-order autoregressive process.

**Examples:** For the Laplace density, $E|Z_1| = \sigma/\sqrt{2}$ and $f_\sigma(0) = 1/(\sqrt{2}\sigma)$, so that the constant factor appearing in the limiting covariance matrix in (22) is

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

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

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

and

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

so that the constant factor in (22) is

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

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 2 is useful in order selection.

**Corollary 2** Assume the conditions of Theorem 2. If the true all-pass model order is $r$ and the fitted model order is $p > r$ then

$$n^{1/2} \hat{\phi}_{p,LAD} \xrightarrow{d} \mathcal{N} \left( 0, \frac{\text{Var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} \right),$$

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

Proof: By Problem 8.15 of Brockwell and Davis (1991), the pth diagonal element of $\Gamma_p^{-1}$ is $\sigma^{-2}$ for $p > r$, so the result follows from (22). □

A practical approach to order determination in large samples then proceeds as follows:

1. Fit a high order (say $P$th order) all-pass model and obtain residuals $\{z_t(\hat{\phi})\}$.

   (a) Estimate $\text{Var}(|Z_1|) \phi_0^{-2}$ consistently by $\hat{\nu}_1$, the empirical variance of $\{|z_t(\hat{\phi})|\}$.

   (b) Estimate $\text{Var}(Z_1) \phi_0^{-2} = \sigma^2 \phi_0^{-2}$ consistently by $\hat{\nu}_2$, the empirical variance of $\{z_t(\hat{\phi})\}$.

   (c) Estimate $|\phi_0| f_\sigma(0)$ consistently by $\hat{d}$, a kernel estimator of the density at zero based on $\{z_t(\hat{\phi})\}$.

   (d) Compute

   $$\hat{\sigma}^2 := \frac{\hat{\nu}_1}{2d_2 \hat{d}^2} \frac{P}{2} \frac{\text{Var}(|Z_1|)}{2d_2 f_\sigma^2(0)}$$

2. Fit all-pass models of order $p = 1, 2, \ldots, P$ via LAD and obtain the $p$th 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,

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

A more formal 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. To make fair comparisons across different model orders, we consider the Laplace likelihood computed on the basis of $n - P$ observations, where the candidate model order $p$ is no greater than $P$. Let $X_1^*, \ldots, X_n^*$ be a realization from the model $(\phi_0', \kappa_0')$, independent of $X_1, \ldots, X_n$. Then, from (7),

$$-2L_{X^*}(\hat{\phi}, \hat{\kappa}) = -2L_X(\hat{\phi}, \hat{\kappa}) - 2\sqrt{2} \sum_{t=1}^{n-P} |z_t(\hat{\phi})| - 2\sqrt{2} \sum_{t=1}^{n-P} |z_t^*(\hat{\phi})|$$

$$= -2L_X(\hat{\phi}, \hat{\kappa}) - 2(n - P) + 2\sqrt{2} \sum_{t=1}^{n-P} |z_t^*(\hat{\phi})| - \sum_{t=1}^{n-P} |z_t^*(\phi_0)|$$

$$+ 2\sqrt{2} \sum_{t=1}^{n-P} |z_t^*(\phi_0)|.$$

Using Theorem 2, (22), and the ergodic theorem, we have that

$$\frac{\sum_{t=1}^{n-P} |z_t^*(\hat{\phi})| - \sum_{t=1}^{n-P} |z_t^*(\phi_0)|}{\hat{\kappa}} \leq \frac{u'N^*}{\sqrt{2E|Z_1||\phi_0||^{-1}}} + \frac{f_\sigma(0)}{\sqrt{2E|Z_1|}} \frac{u'T_P u}{\sqrt{2E|Z_1||\phi_0||^{-1}}},$$

where $u' = -|\phi_0|/(2f_\sigma(0))\Gamma_p^{-1}N$ and $N, N^*$ are iid $N(0, 2\text{Var}|Z_1|\phi_0^{-2}\sigma^{-2}|Z_1|)$. It follows that

$$E \left[ \frac{\sum_{t=1}^{n-P} |z_t^*(\hat{\phi})| - \sum_{t=1}^{n-P} |z_t^*(\phi_0)|}{\hat{\kappa}} \right] \approx \frac{f_\sigma(0)}{\sqrt{2E|Z_1|}} \text{trace} (T_P E[uu'])$$

$$= \frac{\text{Var}|Z_1|}{2\sqrt{2E|Z_1|}\sigma^2 f_\sigma(0)^p}.$$
Further,
\[
E \left[ \frac{\sum_{t=1}^{n-P} |z_t^*(\phi_0)|}{\hat{\kappa}} \right] = E \left[ \frac{\sum_{t=1}^{n-P} |z_t^*(\phi_0)|}{\hat{\kappa}} \right] E \left[ \frac{1}{\hat{\kappa}} \right] 
\approx \frac{(n-P)E|Z_1|}{|\phi_0|} \frac{|\phi_0|}{\sqrt{2E|Z_1|}} = \frac{n-P}{\sqrt{2}}.
\]

Therefore the quantity
\[
AIC(p) := -2 \mathcal{L}_X(\hat{\phi}, \hat{\kappa}) + \frac{\text{Var}|Z_1|}{E|Z_1|\sigma^2 f_\sigma(0)^p}
\] (26)

is approximately unbiased for (25). The model order \( p \in \{0, 1, \ldots, P\} \) which minimizes \( AIC(p) \) is selected. Note that in the Laplace case, the penalty term in (26) is
\[
\frac{\text{Var}|Z_1|}{E|Z_1|\sigma^2 f_\sigma(0)^p} = \frac{\sigma^2 / 2}{(\sigma/\sqrt{2})\sigma^2 (1/\sqrt{2})^p} = \sigma^2 / 2.
\]

unlike the \( 2p \) penalty associated with a Gaussian likelihood. The penalty term can be estimated consistently with
\[
\frac{\hat{\phi}_1}{\hat{\epsilon}_1 \hat{\phi}_2^2},
\]
where \( \hat{\epsilon}_1 \) is the sample mean of the \( |z_t(\hat{\phi})| \) from the \( P \)th order fit, and the remaining terms are defined above.

4 Empirical Results

4.1 Simulation Results

In this section we describe a small simulation study undertaken to evaluate the asymptotic theory. We considered all-pass model orders one and two 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 algorithm to minimize the LAD criterion for each replicate.

Figure 1(a) shows a realization of length 500 from a causal all-pass process of order 2 with parameter values \( \phi_1 = .3, \phi_2 = .4 \) and noise that is distributed as \( t \) with 3 degrees of freedom. The ACFs of the process, its squares, and its absolute values are displayed in Figure 1(b)-(d). As is evident from these graphs, the data are uncorrelated, the squares and absolute values are correlated and the data display some stochastic volatility. For this particular realization, we applied the estimation and identification methods described in Section 3. The estimates of \( \phi_1 \) and \( \phi_2 \) were .297 and .374 with an estimated standard error of .0381. The latter is computed as \( \hat{\epsilon} \sqrt{(1 - \hat{\phi}_2^2)/500} \) where \( \hat{\epsilon} \) is given by (24). The estimates of \( \hat{\phi}_{pp} \) are given in Table 1. With \( P = 10 \), the value of \( \hat{\epsilon} \) in (24) is .908 so that the cut-off value in step 3 of the first order selection procedure described in Section 3.2 is 1.96 * .908/\( \sqrt{500} = .0796 \). As seen from Table 1, this method correctly identifies the order for this particular realization.

The AIC values are also displayed in Table 1. Here we took the maximum order \( P = 10 \) and the estimate of the coefficient of \( p \) in (26) was 1.8955. These AIC values show three competitive models at the correct order \( p = 2 \) and at orders \( p = 6 \) and 9.

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
Figure 1: (a) Realization of an all-pass model of order 2; (b) ACF of the data; (c) ACF of the squares; (d) ACF of the absolute values.

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 \( k \)th starting value \( (\phi_{p1}^{(k)}, \ldots, \phi_{pp}^{(k)})' \) was computed recursively as follows:

1. Draw \( \phi_{11}^{(k)}, \phi_{22}^{(k)}, \ldots, \phi_{pp}^{(k)} \) iid uniform\((-1,1)\).

2. For \( j = 2, \ldots, p \), compute

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

<table>
<thead>
<tr>
<th>Order</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\phi}_{pp} )</td>
<td>.289</td>
<td>.374</td>
<td>.009</td>
<td>.011</td>
<td>.010</td>
<td>.047</td>
<td>.034</td>
<td>-.054</td>
<td>.083</td>
<td>.021</td>
</tr>
<tr>
<td>AIC((p))</td>
<td>2450.6</td>
<td>2345.8</td>
<td>2347.2</td>
<td>2348.2</td>
<td>2349.7</td>
<td>2347.6</td>
<td>2348.5</td>
<td>2345.1</td>
<td>2343.0</td>
<td>2344.6</td>
</tr>
</tbody>
</table>

Table 1: Estimates \( \hat{\phi}_{pp} \) and AIC\((p)\) for \( p = 1, \ldots, 10 \).
Table 2: Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for LAD estimates of all-pass model of order one. Empirical confidence intervals (c.i.'s) are based on standard asymptotic theory for 1000 iid replicates at each sample size, n. Asymptotic means and standard deviations are from (22). Noise distribution is $t$ with 3 degrees of freedom.

<table>
<thead>
<tr>
<th>n</th>
<th>mean std.dev.</th>
<th>mean std.dev.</th>
<th>% coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 0.5</td>
<td>(c.i.)</td>
<td>(c.i.)</td>
</tr>
<tr>
<td>500</td>
<td>0.4979</td>
<td>0.0397</td>
<td>94.2</td>
</tr>
<tr>
<td></td>
<td>(0.4954,0.5004)</td>
<td>(0.0379,0.0414)</td>
<td>(92.8,95.6)</td>
</tr>
<tr>
<td>5000</td>
<td>0.4998</td>
<td>0.0109</td>
<td>95.4</td>
</tr>
<tr>
<td></td>
<td>(0.4991,0.5005)</td>
<td>(0.0105,0.0112)</td>
<td>(94.1,96.7)</td>
</tr>
</tbody>
</table>

Table 3: Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for LAD estimates of all-pass model of order two. Empirical confidence intervals (c.i.'s) are based on standard asymptotic theory for 1000 iid replicates at each sample size, n. Asymptotic means and standard deviations are from (22). Noise distribution is $t$ with 3 degrees of freedom.

<table>
<thead>
<tr>
<th>n</th>
<th>mean std.dev.</th>
<th>mean std.dev.</th>
<th>% coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 0.3</td>
<td>(c.i.)</td>
<td>(c.i.)</td>
</tr>
<tr>
<td>500</td>
<td>0.2990</td>
<td>0.04557</td>
<td>92.5</td>
</tr>
<tr>
<td></td>
<td>(0.2962,0.3018)</td>
<td>(0.0435,0.0475)</td>
<td>(90.9,94.1)</td>
</tr>
<tr>
<td></td>
<td>0.3965</td>
<td>0.0447</td>
<td>92.1</td>
</tr>
<tr>
<td></td>
<td>(0.3937,0.3993)</td>
<td>(0.0427,0.0467)</td>
<td>(90.4,93.8)</td>
</tr>
<tr>
<td>5000</td>
<td>0.3003</td>
<td>0.0118</td>
<td>95.5</td>
</tr>
<tr>
<td></td>
<td>(0.2996,0.3010)</td>
<td>(0.0113,0.0123)</td>
<td>(94.2,96.8)</td>
</tr>
<tr>
<td></td>
<td>0.3990</td>
<td>0.0117</td>
<td>94.7</td>
</tr>
<tr>
<td></td>
<td>(0.3983,0.3997)</td>
<td>(0.0112,0.0122)</td>
<td>(93.3,96.1)</td>
</tr>
</tbody>
</table>

The initial 250 candidate starting values was 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.

Results appear in Tables 2 and 3. 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.

4.2 Linear Time Series with “Nonlinear” Behavior

We now turn to some examples with real data. Figure 2(a)–(d) shows 500 daily log returns of the New Zealand/U.S. exchange rate together with autocorrelations for the returns, their squares, and
Least Absolute Deviations for All-Pass Models

Figure 2: (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.

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 as a linear alternative. 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 3(a) and (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 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, one use of all-pass models is in noncausal autoregressive model fitting. Suppose that $\{X_t\}$ satisfies the difference equations

$$\phi_c(B)\phi_{nc}(B)X_t = Z_t,$$
Figure 3: Diagnostics for fitted all-pass model of order six for New Zealand/U.S. exchange rate returns: (a) ACF of residuals; (b) ACF for squares of residuals.

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 \( r \)-th 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

\[
U_t = \frac{\phi_c(B)\phi_{nc}^{(c)}(B)}{\phi_c(B)\phi_{nc}(B)}Z_t
= \frac{\phi_{nc}^{(c)}(B)}{-\phi_{nc,r}B^r\phi_{nd}^{(c)}(B^{-1})}Z_t,
\]  \hspace{1cm} (27)

where \( \phi_{nc,r} \) is the coefficient of \(-B^r\) in \( \phi_{nc}(B) \). Thus, by (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 \( p \)-th order autoregressive model to \( \{X_t\} \) and obtain residuals \( \{\hat{U}_t\} \). Select a model order \( r \) and fit a purely noncausal \( r \)-th order all-pass model to \( \{\hat{U}_t\} \). The fitted model can be evaluated by residual diagnostics, looking for 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\( (p) \) 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 \( 2^p \) configurations of roots inside and outside the unit circle.

Example: Microsoft Trading Volume. The data in Figure 4 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 one or three 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 $\hat{\phi}_{nc} = 0.5834$ with standard error 0.0296. The resulting residuals $\{\hat{U}_t \}$ show little evidence of correlation, but both $\{\hat{U}_t^2\}$ and $\{|\hat{U}_t|\}$ have significant lag one autocorrelations as shown in Figures 5 (a) and (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 one to $\{\hat{U}_t\}$, we obtain the estimate $\tilde{\phi}_{nc} = 1.7522$, with standard error 0.0989. From (27),

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

so that the all-pass residuals are obtained from

$$\hat{Z}_t = \frac{(1 - 1.7522B)(1 - 0.5834B)}{1 - (1.7522)^{-1}B}X_t$$

$$= \frac{(1 - 1.7522B)(1 - 0.5834B)}{1 - 0.5707B}X_t.$$  \hspace{1cm} (28)
Figure 5: Diagnostics for causal and noncausal autoregressive models fitted to log Microsoft volume: (a) ACF of squares of residuals \( \{\hat{U}_t\} \) from causal AR(1) fit; (b) ACF of absolute values of \( \{\hat{U}_t\} \); (c) ACF of squares of residuals \( \{\hat{Z}_t\} \) from noncausal all-pass fit; (d) ACF of absolute values of \( \{\hat{Z}_t\} \).

In Figures 5 (c) and (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.

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 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 non-causal 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 have fitted all possible causal/noncausal models, and compared diagnostics, rather than employing 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 (28). Diagnostics for the residuals from the noncausal AR(1) fit are virtually identical to those for the \( \{\hat{Z}_t\} \) above. Note that for higher-order models it may not be possible to fit and assess all \(2^p\) possible models.
5 Discussion

This paper 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 Laplace noise yielded a modified absolute deviations criterion, which can be used even if the underlying noise is not Laplace. 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. The methodology was applied to exchange rate returns to show that linear all-pass models can mimic "non-linear" 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 compare our methodology to methods based on higher-order moments. We are also currently looking at maximum likelihood estimation for the same problem.

References


