Maximum Likelihood Estimation for All-Pass Time Series Models

Beth Andrews[1]
Northwestern University

Richard A. Davis[1,2] and F. Jay Breidt[2]
Colorado State University

September 21, 2004

Abstract

An autoregressive-moving average model in which all 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 approximate likelihood for a causal all-pass model is given and used to establish asymptotic normality for maximum likelihood estimators under general conditions. Behavior of the estimators for finite samples is studied via simulation. A two-step procedure using all-pass models to identify and estimate noninvertible autoregressive-moving average models is developed and used in the deconvolution of a simulated water gun seismogram.

[1]Supported in part by NSF Grants DMS9972015 and DMS0308109.
[2]Supported in part by EPA STAR grant CR-829095.
AMS 2000 subject classifications. Primary 62M10; secondary 62E20 62F10.
Key words and phrases. Gaussian mixture, non-Gaussian, noninvertible moving average, white noise.

1 Introduction

All-pass models are autoregressive-moving average (ARMA) models in which the roots of the autoregressive polynomial are reciprocals of roots of the moving average polynomial and vice versa. They generate uncorrelated (white noise) time series, but these series are not independent in the non-Gaussian case. An all-pass series can be obtained by fitting a causal, invertible ARMA model (all the roots of the autoregressive and moving average polynomials are outside the unit circle) to a series generated by a causal, noninvertible ARMA model (all the roots of the autoregressive polynomial are outside the unit circle and at least one root of the moving average polynomial is inside the unit circle). The residuals follow an all-pass model of order r, where r is the number of roots of the true moving average polynomial inside the unit circle. Therefore, by identifying the all-pass order of the residuals, the order of noninvertibility of the ARMA can be determined without considering all possible configurations of roots inside and outside the unit circle, which is computationally prohibitive for large order models. As discussed in Breidt, Davis, and Trindade [6], all-pass models can also be used to fit noncausal autoregressive models.

Noninvertible ARMA models have appeared, for example, in vocal tract filters (Chi and Kung [8], Chien, Yang, and Chi [9]) and in the analysis of unemployment rates (Huang and Pawitan [16]). We use them in this paper in the deconvolution of a simulated water gun seismogram. Other deconvolution approaches are discussed in Blass and Halsey [3], Donoho [11], Godfrey and Rocca [13], Hsueh and Mendel [15], Lii and Rosenblatt [17], Ooe and Ulrych [20], and Wiggins [21].

Estimation methods based on Gaussian likelihood, least-squares, or related second-order moment techniques cannot identify all-pass models. Hence, cumulant-based estimators, using cumulants of order greater than two, are often used to estimate such models (Chi and Kung [8], Chien, Yang, and Chi [9], Giannakis and Swami [12]). Breidt, Davis, and Trindade [6] consider a least absolute deviations (LAD) approach which is motivated by the approximate likelihood of an all-pass model with Laplace (two-sided exponential) noise. Under general conditions, the LAD estimators are consistent and asymptotically normal.

Breidt, Davis, and Trindade [6] compare LAD estimates of all-pass model parameters with estimates obtained using a fourth-order moment technique and, in these simulation results, the LAD estimates are considerably more efficient. LAD estimation, however, is relatively efficient only when the noise distribution is Laplace. Therefore, in this paper, we use a maximum likelihood (ML) approach to estimate all-pass model parameters. Although the ML estimation procedure is limited by the assumption that the probability density function for the noise is known to within some parameter values, the residuals from a fitted all-pass model obtained using LAD might indicate an appropriate noise density when it is unknown. The ML estimators are consistent and asymptotically normal under general conditions. Related likelihood approaches are considered in Breidt, Davis, Lii, and Rosenblatt [5] for noncausal autoregressive processes, in Lii and Rosenblatt [18] for noninvertible moving average processes, and in Lii and Rosenblatt [19] for general ARMA processes. Although all-pass models are ARMA models, their special parameterization makes the results of Lii and Rosenblatt [19] inapplicable.

In Section 2, we give an approximate likelihood for all-pass model parameters. Asymptotic normality and strong consistency are established for ML estimators under general conditions and order selection is considered in Section 3. Proofs of the lemmas used to establish the results of Section 3 can be found in the Appendix. The behavior of the estimators for finite samples is studied via simulation in Section 4.1. A two-step procedure using all-pass models to fit noninvertible ARMA models is developed in Section 4.2 and applied to the deconvolution of a simulated water gun seismogram in Section 4.3.

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 pth-order autoregressive polynomial, where $\phi(z) \neq 0$ for $|z| = 1$. The filter $\phi(B)$ is said to be *causal* if all the roots of $\phi(z)$ 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 $\{W_t\}$. Note that if $\phi(B)$ is causal, the filter $B^p\phi(B^{-1})$ is *purely noncausal*, and hence

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

a function of only the present and future $\{W_t\}$. See, for example, Chapter 3 of Brockwell and Davis [7].

Let

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

where $\phi_0(z) \neq 0$ for $|z| \leq 1$. Define $\phi_{00} = 1$, and suppose $\phi_{0r} \neq 0$ for some $r \in \{0, 1, \ldots, p\}$ and $\phi_{0j} = 0$ for $j = r+1, \ldots, p$. Then, a causal all-pass time series is the ARMA series $\{X_t\}$ which satisfies the difference equations

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

or

$$X_t - \phi_{01}X_{t-1} - \cdots - \phi_{0r}X_{t-r} = Z_t^* + \frac{\phi_{0,r-1}}{\phi_{0r}} Z_{t-1}^* + \cdots + \frac{\phi_{01}}{\phi_{0r}} Z_{t-r+1}^* - \frac{1}{\phi_{0r}} Z_{t-r}^*.$$

The true order of the all-pass model is r, and the series $\{Z_t^*\}$ is an independent and identically distributed (iid) sequence of random variables with mean 0 and variance $\sigma_0^2 \in (0, \infty)$. We assume throughout that Z_1^* has probability density function $f_{\sigma_0}(z; \boldsymbol{\theta}_0) = \sigma_0^{-1} f(\sigma_0^{-1}z; \boldsymbol{\theta}_0)$, where f is a density function symmetric about zero and $\boldsymbol{\theta}$ is a parameter of the density f. We also assume that the true value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0d})'$, lies in the interior of a parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^d$, $d \geq 1$. Note that the roots of the autoregressive polynomial $\phi_0(z)$ are reciprocals of the roots of the moving average polynomial $-\phi_{0r}^{-1} z^r \phi_0(z^{-1})$ and vice versa.

The spectral density for $\{X_t\}$ in (1) is

$$\frac{|e^{-ir\omega}|^2 |\phi_0(e^{i\omega})|^2}{\phi_{0r}^2 |\phi_0(e^{-i\omega})|^2} \frac{\sigma_0^2}{2\pi} = \frac{\sigma_0^2}{\phi_{0r}^2 2\pi},$$

which is constant for $\omega \in [-\pi, \pi]$, and thus $\{X_t\}$ is an uncorrelated sequence. In the case of Gaussian $\{Z_t^*\}$, this implies that $\{X_t\}$ is iid $\mathrm{N}(0, \sigma_0^2 \phi_{0r}^{-2})$, but independence does not hold in the non-Gaussian case if $r \geq 1$ (see Breidt and Davis [4]). The model (1) is called all-pass because the power transfer function of the all-pass filter passes all the power for every frequency in the spectrum. In other words, an all-pass filter does not change the distribution of power over the spectrum.

We can express (1) as

$$\phi_0(B)X_t = \frac{B^p \phi_0(B^{-1})}{-\phi_{0r}} Z_t, \tag{2}$$

where $\{Z_t\} = \{Z_{t+p-r}^*\}$ is an iid sequence of random variables with mean 0, variance σ_0^2, and probability density function $f_{\sigma_0}(z; \boldsymbol{\theta}_0)$. Rearranging (2) and setting $z_t = \phi_{0r}^{-1} Z_t$, we have the backward recursion

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

An analogous recursion for an arbitrary, causal autoregressive polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ can be defined as follows:

$$z_{t-p}(\boldsymbol{\phi}) = \begin{cases} 0, & t = n+p, \ldots, n+1, \\ \phi_1 z_{t-p+1}(\boldsymbol{\phi}) + \cdots + \phi_p z_t(\boldsymbol{\phi}) - \phi(B) X_t, & t = n, \ldots, p+1, \end{cases} \tag{3}$$

where $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_p)'$. If $\boldsymbol{\phi}_0 := (\phi_{01}, \ldots, \phi_{0p})' = (\phi_{01}, \ldots, \phi_{0r}, 0, \ldots, 0)'$, note that $\{z_t(\boldsymbol{\phi}_0)\}_{t=1}^{n-p}$ closely approximates $\{z_t\}_{t=1}^{n-p}$; the error is due to the initialization with zeros. Although $\{z_t\}$ is iid, $\{z_t(\boldsymbol{\phi}_0)\}_{t=1}^{n-p}$ is not iid if $r \geq 1$.

2.2 Approximating the Likelihood

In this subsection, we ignore the effect of the recursion initialization in (3), and write

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

Let $q = \max\{0 \leq j \leq p : \phi_j \neq 0\}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p+d+1})' = (\phi_1, \ldots, \phi_p, \sigma/|\phi_q|, \theta_1, \ldots, \theta_d)'$, and $\boldsymbol{\alpha}_0 = (\alpha_{01}, \ldots, \alpha_{0,p+d+1})' = (\phi_{01}, \ldots, \phi_{0p}, \sigma_0/|\phi_{0r}|, \theta_{01}, \ldots, \theta_{0d})'$. Following equation (2.7) in Breidt, Davis, and Trindade [6], we approximate the log-likelihood of $\boldsymbol{\alpha}$ given a realization of length n from model (1), $\{X_t\}_{t=1}^{n}$, with

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}) &= \sum_{t=1}^{n-p} \ln f_\sigma\left(\phi_q z_t(\boldsymbol{\phi}); \boldsymbol{\theta}\right) + (n-p)\ln|\phi_q| \\ &= \sum_{t=1}^{n-p} \{\ln f(z_t(\boldsymbol{\phi})/\alpha_{p+1}; \boldsymbol{\theta}) - \ln \alpha_{p+1}\} \\ &=: \sum_{t=1}^{n-p} g_t(\boldsymbol{\alpha}), \end{aligned} \tag{4}$$

where $\{z_t(\boldsymbol{\phi})\}$ can be computed recursively from (3).

3 Asymptotic Results

3.1 Parameter Estimation

In order to establish asymptotic normality for the MLE of $\boldsymbol{\alpha}$, we make the following additional, yet still fairly general, assumptions on f:

- **A1** *For all* $s \in I\!R$ *and all* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)' \in \boldsymbol{\Theta}$, $f(s; \boldsymbol{\theta}) > 0$ *and* $f(s; \boldsymbol{\theta})$ *is twice continuously differentiable with respect to* $(s, \theta_1, \ldots, \theta_d)'$.
- **A2** *For all* $\boldsymbol{\theta}$ *in some neighborhood of* $\boldsymbol{\theta}_0$, $\int s f'(s; \boldsymbol{\theta})\, ds = s f(s; \boldsymbol{\theta})|_{-\infty}^{\infty} - \int f(s; \boldsymbol{\theta})\, ds = -1$.
- **A3** $\int f''(s; \boldsymbol{\theta}_0)\, ds = f'(s; \boldsymbol{\theta}_0)|_{-\infty}^{\infty} = 0$.
- **A4** $\int s^2 f''(s; \boldsymbol{\theta}_0)\, ds = s^2 f'(s; \boldsymbol{\theta}_0)|_{-\infty}^{\infty} - 2\int s f'(s; \boldsymbol{\theta}_0)\, ds = 2$.
- **A5** $1 < \int \left(f'(s; \boldsymbol{\theta}_0)\right)^2 / f(s; \boldsymbol{\theta}_0)\, ds$.
- **A6** *If* $\tilde{K} := \alpha_{0,p+1}^{-2} \left\{ \int \frac{s^2 \left(f'(s; \boldsymbol{\theta}_0)\right)^2}{f(s; \boldsymbol{\theta}_0)}\, ds - 1 \right\}$, $L := \left[-\alpha_{0,p+1}^{-1} \int \frac{s f'(s; \boldsymbol{\theta}_0)}{f(s; \boldsymbol{\theta}_0)} \frac{\partial f(s; \boldsymbol{\theta}_0)}{\partial \theta_j}\, ds \right]_{j=1}^{d}$, *and* $I := \left[\int \frac{1}{f(s; \boldsymbol{\theta}_0)} \frac{\partial f(s; \boldsymbol{\theta}_0)}{\partial \theta_j} \frac{\partial f(s; \boldsymbol{\theta}_0)}{\partial \theta_k}\, ds \right]_{j,k=1}^{d}$, *the matrix* $\begin{bmatrix} \tilde{K} & L' \\ L & I \end{bmatrix}$ *is positive definite.*
- **A7** *For* $j, k = 1, \ldots, d$ *and all* $\boldsymbol{\theta}$ *in some neighborhood of* $\boldsymbol{\theta}_0$,

– $f(s;\boldsymbol{\theta})$ *is dominated by some function* $f_1(s)$ *such that* $\int s^2 f_1(s)\,ds < \infty$, *and*

– $s^2\frac{(f'(s;\boldsymbol{\theta}))^2}{f^2(s;\boldsymbol{\theta})}$, $s^2\left|\frac{f''(s;\boldsymbol{\theta})}{f(s;\boldsymbol{\theta})}\right|$, $|s|\frac{1}{f(s;\boldsymbol{\theta})}\left|\frac{\partial}{\partial\theta_j}f'(s;\boldsymbol{\theta})\right|$, $\frac{1}{f^2(s;\boldsymbol{\theta})}\left(\frac{\partial}{\partial\theta_j}f(s;\boldsymbol{\theta})\right)^2$, *and* $\frac{1}{f(s;\boldsymbol{\theta})}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}f(s;\boldsymbol{\theta})\right|$ *are dominated by* $a_1 + a_2|s|^{c_1}$, *where* a_1, a_2, c_1 *are non-negative constants and* $\int |s|^{c_1} f_1(s)\,ds < \infty$.

Theorem 1 *If* f *satisfies A1–A7, then there exists a sequence of maximizers*

$$\hat{\boldsymbol{\alpha}}_{ML} = (\hat{\boldsymbol{\phi}}'_{ML}, \hat{\alpha}_{p+1,ML}, \hat{\boldsymbol{\theta}}'_{ML})'$$

of $\mathcal{L}(\cdot)$ *in (4) such that*

$$n^{1/2}(\hat{\boldsymbol{\alpha}}_{ML} - \boldsymbol{\alpha}_0) \xrightarrow{d} \mathbf{Y} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\right), \tag{5}$$

where

$$\boldsymbol{\Sigma}^{-1} := \begin{bmatrix} \frac{\sigma_0^2}{2(\sigma_0^2\tilde{J}-1)}\boldsymbol{\Gamma}_p^{-1} & \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times d} \\ \mathbf{0}_{1\times p} & (\tilde{K} - L'I^{-1}L)^{-1} & -\tilde{K}^{-1}L'(I - L\tilde{K}^{-1}L')^{-1} \\ \mathbf{0}_{d\times p} & -(I - L\tilde{K}^{-1}L')^{-1}L\tilde{K}^{-1} & (I - L\tilde{K}^{-1}L')^{-1} \end{bmatrix},$$

$\tilde{J} := \sigma_0^{-2}\int (f'(s;\boldsymbol{\theta}_0))^2/f(s;\boldsymbol{\theta}_0)\,ds$, $\boldsymbol{\Gamma}_p := [\gamma(j-k)]_{j,k=1}^p$, *and* $\gamma(\cdot)$ *is the autocovariance function of the autoregressive process* $\{(1/\phi_0(B))Z_t\}$.

Proof: $\mathcal{L}(\boldsymbol{\alpha}) - \mathcal{L}(\boldsymbol{\alpha}_0) = S_n(\sqrt{n}\,(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0))$, where $S_n(\cdot)$ is defined in Lemma 3 of the Appendix. Because $\mathbf{Y} := \boldsymbol{\Sigma}^{-1}\mathbf{N}$ maximizes the limit process $S(\cdot)$ in Lemma 3, the result (5) follows by Remark 1 of Davis, Knight, and Liu [10]. □

Remark 1: Note that $\hat{\boldsymbol{\phi}}_{ML}$ is asymptotically independent of $(\hat{\alpha}_{p+1,ML}, \hat{\boldsymbol{\theta}}'_{ML})'$. Given n observations from $\{Z_t\}$, there exists a sequence of maximizers $(\hat{\sigma}_Z, \hat{\boldsymbol{\theta}}'_Z)'$ of the log-likelihood

$$\sum_{t=1}^{n}\{\ln f(Z_t/\sigma;\boldsymbol{\theta}) - \ln\sigma\}$$

such that

$$n^{1/2}[(\hat{\sigma}_Z, \hat{\boldsymbol{\theta}}'_Z)' - (\sigma_0, \boldsymbol{\theta}'_0)'] \xrightarrow{d} \mathbf{Y}_Z \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_Z^{-1}\right),$$

where

$$\boldsymbol{\Sigma}_Z^{-1} := \begin{bmatrix} \phi_{0r}^2(\tilde{K} - L'I^{-1}L)^{-1} & -|\phi_{0r}|\tilde{K}^{-1}L'(I - L\tilde{K}^{-1}L')^{-1} \\ -|\phi_{0r}|(I - L\tilde{K}^{-1}L')^{-1}L\tilde{K}^{-1} & (I - L\tilde{K}^{-1}L')^{-1} \end{bmatrix}$$

and $\boldsymbol{\Sigma}_Z^{-1}$ does not depend on ϕ_{0r}.

Remark 2: Using A2 and the Cauchy-Schwarz inequality,

$$\begin{aligned} 1 &= \left\{\int s\frac{f'(s;\boldsymbol{\theta}_0)}{f(s;\boldsymbol{\theta}_0)}f(s;\boldsymbol{\theta}_0)\,ds\right\}^2 \\ &\leq \left\{\int s^2 f(s;\boldsymbol{\theta}_0)\,ds\right\}\left\{\int\left(\frac{f'(s;\boldsymbol{\theta}_0)}{f(s;\boldsymbol{\theta}_0)}\right)^2 f(s;\boldsymbol{\theta}_0)\,ds\right\} \qquad (6) \\ &= \sigma_0^2\tilde{J}, \end{aligned}$$

with equality in (6) if and only if f is Gaussian. Thus, A5 holds for non-Gaussian f. Further,

$$\begin{aligned} 1 &= \left\{\int s\frac{f'(s;\boldsymbol{\theta}_0)}{f(s;\boldsymbol{\theta}_0)}f(s;\boldsymbol{\theta}_0)\,ds\right\}^2 \\ &< \left\{\int s^2\left(\frac{f'(s;\boldsymbol{\theta}_0)}{f(s;\boldsymbol{\theta}_0)}\right)^2 f(s;\boldsymbol{\theta}_0)\,ds\right\}\left\{\int f(s;\boldsymbol{\theta}_0)\,ds\right\} \qquad (7) \\ &= \alpha_{0,p+1}^2\tilde{K} + 1, \end{aligned}$$

so that $\tilde{K} > 0$. We do not have equality in (7) because, by Cauchy-Schwarz, there is equality if and only if $sf'(s;\theta_0)/f(s;\theta_0) = -1$ for all $s \in \mathbb{R}$ which cannot ever be the case.

Remark 3: The asymptotic covariance matrix for the estimators of the p autoregressive parameters is a scalar multiple of the asymptotic covariance matrix for the Gaussian likelihood estimators for the corresponding pth-order autoregressive process. The same property holds for LAD estimators of all-pass model parameters, as shown in Breidt, Davis, and Trindade [6]. The LAD estimators are obtained by maximizing the likelihood of an all-pass model with Laplace noise. This yields a modified LAD criterion, which can be used even if the underlying noise distribution is not Laplace. The constant in (5) is

$$\frac{1}{2(\sigma_0^2\tilde{J} - 1)}, \qquad (8)$$

while, in the LAD case, the appropriate constant is

$$\frac{\mathrm{Var}\,(|Z_1|)}{2\,(2\sigma_0^2 f_{\sigma_0}(0;\boldsymbol{\theta}_0) - \mathrm{E}|Z_1|)^2}. \qquad (9)$$

(Breidt, Davis, and Trindade [6] contains an error in the calculation of the asymptotic variance; see Andrews [1] for the correction.) Although the Laplace density,

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

does not meet assumptions A1–A7, $\mathrm{E}|Z_1| = \sigma_0/\sqrt{2}$, $f_{\sigma_0}(0) = 1/(\sqrt{2}\sigma_0)$, and $\tilde{J} = 2\sigma_0^{-2}$, so that (8) and (9) are both 1/2 for this density.

Remark 4: We obtain the asymptotic relative efficiency (ARE) of ML to LAD for the autoregressive parameters by dividing (9) by (8):

$$\mathrm{ARE} = (\sigma_0^2\tilde{J} - 1)\frac{\mathrm{Var}\,(|Z_1|)}{\left(2\sigma_0^2 f_{\sigma_0}(0;\boldsymbol{\theta}_0) - \mathrm{E}|Z_1|\right)^2}.$$

The density function

$$f(s;\theta) = \sqrt{\frac{\theta}{\theta-2}}\,\frac{\Gamma((\theta+1)/2)}{\Gamma(\theta/2)}\,\frac{1}{\sqrt{\theta\pi}}\,\frac{1}{(1+s^2/(\theta-2))^{(\theta+1)/2}} \tag{10}$$

is symmetric about zero, has variance one, and satisfies assumptions A1–A7 with $c_1 = 2$ when $\boldsymbol{\Theta} \subseteq (2,\infty)$. If $\sigma_0 = (\theta_0/(\theta_0-2))^{1/2}$, then $f_{\sigma_0}(s;\theta_0) = \sigma_0^{-1} f(\sigma_0^{-1}s;\theta_0)$ is the Students' t-density with θ_0 degrees of freedom. In this case,

$$\mathrm{E}|Z_1| = 2\sigma_0\frac{\sqrt{\theta_0-2}}{\theta_0-1}\,\frac{\Gamma((\theta_0+1)/2)}{\Gamma(\theta_0/2)\sqrt{\pi}},$$

$$f_{\sigma_0}(0;\theta_0) = \sigma_0^{-1}\frac{\Gamma((\theta_0+1)/2)}{\Gamma(\theta_0/2)\sqrt{(\theta_0-2)\pi}},$$

and

$$\tilde{J} = \sigma_0^{-2}\frac{\theta_0(\theta_0+1)}{(\theta_0-2)(\theta_0+3)},$$

so the ARE of ML to LAD is

$$\mathrm{ARE} = \frac{6}{(\theta_0+3)}\left\{\frac{\pi}{4}(\theta_0-1)^2\left(\frac{\Gamma(\theta_0/2)}{\Gamma((\theta_0+1)/2)}\right)^2 - (\theta_0-2)\right\}. \tag{11}$$

When $\theta_0 = 3$, the value of (11) is $2(0.7337) = 1.4674$, and thus ML is nearly 50% more efficient than LAD for the Students' t-distribution with three degrees of freedom. For values of θ_0 greater than three, (11) is even

θ_{01}	θ_{02}	ARE
0.1	0.2	1.5506
0.1	0.8	2.1506
0.4	0.4	1.4988
0.4	0.6	1.7124
0.6	0.4	1.6012
0.6	0.6	1.8327
0.9	0.2	1.6395
0.9	0.8	2.9997

Table 1: AREs for ML to LAD when f is the Gaussian scale mixture density.

larger than 1.4674. As $\theta_0 \to \infty$, however, the Students' t-distribution approaches the standard Gaussian distribution, and so the autoregressive parameters cannot be consistently estimated.

Remark 5: The Gaussian scale mixture density

$$f(s;(\theta_1,\theta_2)') = \frac{\theta_1}{\theta_2\sqrt{2\pi}} \exp\left(\frac{-s^2}{2\theta_2^2}\right) + \frac{(1-\theta_1)^{3/2}}{\sqrt{1-\theta_1\theta_2^2}\,\sqrt{2\pi}} \exp\left(\frac{-s^2(1-\theta_1)}{2(1-\theta_1\theta_2^2)}\right) \tag{12}$$

is symmetric about 0 and satisfies A1–A7 with $c_1 = 4$ when $\boldsymbol{\Theta} \subseteq (0,1)\times(0,1)$. In this case, Z_1 is $\mathrm{N}(0,\sigma_0^2\theta_{02}^2)$ with probability θ_{01} and $\mathrm{N}(0,\sigma_0^2(1-\theta_{01}\theta_{02}^2)/(1-\theta_{01}))$ with probability $1-\theta_{01}$. Some values of ARE for ML to LAD for the autoregressive parameters are given in Table 1.

Remark 6: By A7 and the dominated convergence theorem,

$$\frac{1}{2(\sigma_0^2\tilde{J}-1)} = \frac{1}{2}\left(\int \frac{(f'(s;\boldsymbol{\theta}_0))^2}{f(s;\boldsymbol{\theta}_0)}\,ds - 1\right)^{-1},$$

$\tilde{K}$, L, and I are continuous with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. Thus, because $\hat{\boldsymbol{\theta}}_{ML} \xrightarrow{P} \boldsymbol{\theta}_0$,

$$\frac{1}{2}\left(\int \frac{(f'(s;\hat{\boldsymbol{\theta}}_{ML}))^2}{f(s;\hat{\boldsymbol{\theta}}_{ML})}\,ds - 1\right)^{-1}, \tag{13}$$

$$\frac{1}{\hat{\alpha}_{p+1,ML}^2}\left(\int \frac{s^2(f'(s;\hat{\boldsymbol{\theta}}_{ML}))^2}{f(s;\hat{\boldsymbol{\theta}}_{ML})}\,ds - 1\right), \tag{14}$$

$$\left[\frac{-1}{\hat{\alpha}_{p+1,ML}} \int \frac{sf'(s;\hat{\boldsymbol{\theta}}_{ML})}{f(s;\hat{\boldsymbol{\theta}}_{ML})} \frac{\partial f(s;\hat{\boldsymbol{\theta}}_{ML})}{\partial\theta_j}\,ds\right]_{j=1}^{d}, \tag{15}$$

and

$$\left[\int \frac{1}{f(s;\hat{\boldsymbol{\theta}}_{ML})} \frac{\partial f(s;\hat{\boldsymbol{\theta}}_{ML})}{\partial \theta_j} \frac{\partial f(s;\hat{\boldsymbol{\theta}}_{ML})}{\partial \theta_k}\, ds\right]_{j,k=1}^{d} \tag{16}$$

are consistent estimators of $[2(\sigma_0^2 \tilde{J} - 1)]^{-1}$, $\tilde{K}$, L, and I respectively.

If we restrict $\boldsymbol{\alpha}$ to a compact, convex space $\Xi \subset \mathbb{R}^{p+d+1}$ with $\boldsymbol{\alpha}_0$ in the interior, then the MLE of $\boldsymbol{\alpha}_0$ is both strongly consistent and asymptotically normal, as shown in the following two theorems.

Theorem 2 *If*

- *the parameter space $\Xi \subset \mathbb{R}^{p+d+1}$ is compact and convex with $\boldsymbol{\alpha}_0$ in the interior,*
- *$\boldsymbol{\phi}$ forms a causal polynomial, $\alpha_{p+1} > 0$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ for all $\boldsymbol{\alpha} = (\boldsymbol{\phi}', \alpha_{p+1}, \boldsymbol{\theta}')' \in \Xi$,*
- *f satisfies A1,*
- *$\int |\ln f(s;\boldsymbol{\theta}_0)| f(s;\boldsymbol{\theta}_0)\, ds < \infty$,*
- *$|sf'(s;\boldsymbol{\theta})|/f(s;\boldsymbol{\theta})$ and $|\frac{\partial}{\partial \theta_j} f(s;\boldsymbol{\theta})|/f(s;\boldsymbol{\theta})$, $j = 1, \ldots, d$, are dominated by $a_3 + a_4|s|^{c_2}$ for all $\boldsymbol{\alpha} \in \Xi$, where a_3, a_4, c_2 are non-negative constants such that $\int |s|^{c_2} f(s;\boldsymbol{\theta}_0) < \infty$, and*
- *the unique maximum of $E\{\ln f(\tilde{z}_1(\boldsymbol{\phi})/\alpha_{p+1};\boldsymbol{\theta}) - \ln \alpha_{p+1}\}$ on Ξ is at $\boldsymbol{\alpha}_0$ if $\tilde{z}_1(\boldsymbol{\phi}) := -(\phi(B)/\phi(B^{-1}))X_{1+p}$,*

then

$$\hat{\boldsymbol{\alpha}} := argmax_{\boldsymbol{\alpha} \in \Xi} \mathcal{L}(\boldsymbol{\alpha}) \stackrel{a.s.}{\to} \boldsymbol{\alpha}_0.$$

Proof: By Lemma 4 in the Appendix,

$$n^{-1}\mathcal{L}(\boldsymbol{\alpha}) \stackrel{a.s.}{\to} \mathrm{E}\{\ln f(\tilde{z}_1(\boldsymbol{\phi})/\alpha_{p+1};\boldsymbol{\theta}) - \ln \alpha_{p+1}\}$$

uniformly on Ξ. Since the limit has a unique maximum at $\boldsymbol{\alpha}_0$, the result follows. □

Theorem 3 *If f satisfies A1–A7 and the conditions of Theorem 2 hold, then*

$$n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} \mathbf{Y} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\right).$$

Proof: Because $\mathcal{L}$ is differentiable on $\boldsymbol{\Xi}$ and $\hat{\boldsymbol{\alpha}}$ maximizes $\mathcal{L}(\cdot)$, $\frac{\partial}{\partial \boldsymbol{\alpha}} \mathcal{L}(\hat{\boldsymbol{\alpha}}) = \mathbf{0}$ for all n sufficiently large almost surely. Thus, we have

$$\begin{aligned} \mathbf{0} &= n^{-1} \sum_{t=1}^{n-p} \frac{\partial g_t(\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\alpha}} \\ &= n^{-1} \sum_{t=1}^{n-p} \frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} + n^{-1} \sum_{t=1}^{n-p} \frac{\partial^2 g_t(\boldsymbol{\alpha}_n^*)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0), \end{aligned}$$

where $\boldsymbol{\alpha}_n^*$ is between $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$, and so

$$n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = -\left(n^{-1} \sum_{t=1}^{n-p} \frac{\partial^2 g_t(\boldsymbol{\alpha}_n^*)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}\right)^{-1} \left(n^{-1/2} \sum_{t=1}^{n-p} \frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}\right).$$

By Lemma 1,

$$n^{-1/2} \sum_{t=1}^{n-p} \frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \xrightarrow{d} \mathbf{N} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

and, by Lemma 2,

$$n^{-1} \sum_{t=1}^{n-p} \frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \xrightarrow{P} -\boldsymbol{\Sigma}.$$

Because $\hat{\boldsymbol{\alpha}} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$ and f satisfies A7,

$$n^{-1} \sum_{t=1}^{n-p} \frac{\partial^2 g_t(\boldsymbol{\alpha}_n^*)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - n^{-1} \sum_{t=1}^{n-p} \frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \xrightarrow{P} \mathbf{0},$$

and the result follows. □

Remark 7: The first five conditions of Theorem 2 are generally relatively straightforward to check, but the final condition may be more of a challenge to verify without resorting to simulation.

3.2 Order Selection

In practice, the order r of an all-pass model is usually unknown. Therefore, we present the following corollary to Theorem 1 for use in order selection.

Corollary 1 *Assume f satisfies A1–A7. If the true order of the all-pass model is r and the order of the fitted model is $p > r$, then*

$$n^{1/2}\hat{\phi}_{p,ML} \xrightarrow{d} N\left(0, \frac{1}{2(\sigma_0^2 \tilde{J} - 1)}\right).$$

Proof: By Problem 8.15 in Brockwell and Davis [7], the pth diagonal element of $\mathbf{\Gamma}_p^{-1}$ is σ_0^{-2} if $p > r$, and so the result follows from (5). □

A practical approach to order determination using a large sample follows. This procedure is analogous to using the partial autocorrelation function to identify the order of an autoregressive model.

1. For some large P, fit all-pass models of order p, $p = 1, 2, \ldots, P$, via ML and obtain the pth coefficient, $\hat{\phi}_{p,ML}$, for each.

2. Let the model order r be the smallest order beyond which the estimated coefficients are statistically insignificant; that is,

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

   where $\hat{\lambda} := \left(2\int (f'(s;\hat{\boldsymbol{\theta}}_{ML}))^2/f(s;\hat{\boldsymbol{\theta}}_{ML})\,ds - 2\right)^{-1/2}$ and $\hat{\boldsymbol{\theta}}_{ML}$ is the MLE from the fitted Pth-order model.

4 Numerical Results

4.1 Simulation Study

In this section, we describe a simulation experiment to assess the quality of the asymptotic approximations for finite samples. We used both the rescaled Students' t-density (10) and the Gaussian scale mixture density (12). For the rescaled Students' t-density, we let $\sigma_0 = (\theta_0/(\theta_0 - 2))^{1/2}$, so Z_1 followed the Students' t-distribution with θ_0 degrees of freedom.

To reduce the possibility of the optimizer being trapped at local maxima, we used 250 starting values for each of the 1000 replicates. The initial values for $\phi_1, \ldots, \phi_p$ were uniformly distributed in the space of partial autocorrelations and then mapped to the space of autoregressive coefficients using the Durbin-Levinson algorithm (Brockwell and Davis [7], Proposition 5.2.1). That is, for a model of order p, the kth 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

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

With $(\phi_{p1}^{(k)}, \ldots, \phi_{pp}^{(k)})'$ and a realization of length n, we obtained residuals using (3). To get the kth starting value $\alpha_{p+1}^{(k)}$, we divided the standard deviation of the residuals by $|\phi_{pq}^{(k)}|$, where $q := \max\{0 \leq j \leq p : \phi_{pj}^{(k)} \neq 0\}$. Finally, we randomly chose the starting values for $\boldsymbol{\theta}$.

The log-likelihood was evaluated at each of the 250 candidate values. When Z_1 follows the Students' t-distribution, the likelihood function is almost constant with respect to $(\alpha_{p+1}, \theta)'$ near $(\boldsymbol{\phi}_0', \alpha_{0,p+1}, \theta_0)'$, and so the maximum can be difficult to find. This is not the case when Z_1 is a Gaussian scale mixture. Therefore, when using (10), the collection of initial values was reduced to the nine with the highest likelihoods plus $\boldsymbol{\alpha}_0$, and, when using (12), the collection of initial values was reduced to the two with the highest likelihoods plus $\boldsymbol{\alpha}_0$. We found optimized values by implementing the Hooke and Jeeves [14] algorithm and using the ten or three values as starting points. The optimized value with the greatest likelihood was selected to be $\hat{\boldsymbol{\alpha}}_{ML}$. We constructed confidence intervals for the elements of $\boldsymbol{\alpha}_0$ using (5) and the estimators (13)–(16).

Results of the simulations appear in Tables 2 and 3. In the tables, we see that the MLEs are approximately unbiased and the confidence interval coverages are fairly close to the nominal 95% level, particularly when

$n = 5000$. For the Students' t-distribution, the asymptotic standard deviations tend to understate the true variability of the MLEs when $n = 500$, but are more accurate when $n = 5000$. The asymptotic standard deviations are close to the empirical standard deviations in the Gaussian scale mixture case when $n = 500$ and $n = 5000$. Normal probability plots of the MLEs show that approximate normality is achieved when $n = 5000$.

4.2 Noninvertible ARMA Modeling

As mentioned in the introduction, all-pass models can be used to fit causal, noninvertible ARMA models. Suppose the series $\{X_t\}$ follows the model

$$\phi(B)X_t = \theta_i(B)\theta_{ni}(B)Z_t, \tag{17}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ is a causal AR(p) polynomial (all the roots of $\phi(z)$ fall outside the unit circle), $\theta_i(z) = 1 + \theta_{i,1}z + \cdots + \theta_{i,q}z^q$ is an invertible MA(q) polynomial (all the roots of $\theta_i(z)$ fall outside the unit circle), $\theta_{ni}(z) = 1 + \theta_{ni,1}z + \cdots + \theta_{ni,r}z^r$ is a purely noninvertible MA(r) polynomial such that $\theta_{ni}(z) \neq 0$ for $|z| = 1$ (all the roots of $\theta_{ni}(z)$ fall inside the unit circle), and $\{Z_t\}$ is iid. If $\theta_{ni}^{(i)}(z)$ is the invertible rth order polynomial with roots that are the reciprocals of the roots of $\theta_{ni}(z)$ and $\{X_t\}$ is mistakenly modeled as the causal, invertible ARMA

$$\phi(B)X_t = \theta_i(B)\theta_{ni}^{(i)}(B)W_t,$$

then $\{W_t\}$ satisfies

$$\begin{aligned} W_t &= \frac{\theta_{ni}(B)}{\theta_{ni}^{(i)}(B)} Z_t \\ &= \frac{B^r \theta_{ni}^{(i)}(B^{-1})}{\theta_{ni,r}^{(i)} \theta_{ni}^{(i)}(B)} Z_t, \end{aligned}$$

where $\theta_{ni,r}^{(i)}$ is the coefficient of z^r in $\theta_{ni}^{(i)}(z)$. So, $\{W_t\}$ follows the causal all-pass model

$$\theta_{ni}^{(i)}(B)W_t = \frac{B^r \theta_{ni}^{(i)}(B^{-1})}{\theta_{ni,r}^{(i)}} Z_t.$$

	Asymptotic		Empirical		
n	mean	std.dev.	mean (c.i.)	std.dev. (c.i.)	% coverage (c.i.)
500	$\phi_1 = 0.5$	0.0274	0.4971	0.0315	93.0
			(0.4951,0.4990)	(0.0301,0.0329)	(91.4,94.6)
	$\alpha_2 = 3.4641$	0.4177	3.5533	0.6277	90.0
			(3.5144,3.5922)	(0.5995,0.6546)	(88.1,91.9)
	$\theta = 3.0$	0.4480	3.1123	0.5008	95.8
			(3.0812,3.1433)	(0.4783,0.5223)	(94.6,97.0)
5000	$\phi_1 = 0.5$	0.0087	0.4997	0.0091	93.4
			(0.4991,0.5003)	(0.0087,0.0095)	(91.9,94.9)
	$\alpha_2 = 3.4641$	0.1321	3.4787	0.1427	94.0
			(3.4699,3.4876)	(0.1363,0.1489)	(92.5,95.5)
	$\theta = 3.0$	0.1417	3.0084	0.1533	94.0
			(2.9989,3.0179)	(0.1464,0.1599)	(92.5,95.5)
500	$\phi_1 = 0.3$	0.0290	0.2993	0.0345	90.6
			(0.2971,0.3014)	(0.0330,0.0360)	(88.8,92.4)
	$\phi_2 = 0.4$	0.0290	0.3964	0.0350	90.1
			(0.3942,0.3986)	(0.0335,0.0365)	(88.2,92.0)
	$\alpha_3 = 4.3301$	0.5222	4.4842	0.9460	94.0
			(4.4256,4.5428)	(0.9036,0.9866)	(92.5,95.5)
	$\theta = 3.0$	0.4480	3.0789	0.4722	94.8
			(3.0497,3.1082)	(0.4510,0.4925)	(93.4,96.2)
5000	$\phi_1 = 0.3$	0.0092	0.2999	0.0095	94.0
			(0.2993,0.3005)	(0.0091,0.0099)	(92.5,95.5)
	$\phi_2 = 0.4$	0.0092	0.3999	0.0094	94.6
			(0.3993,0.4005)	(0.0090,0.0098)	(93.2,96.0)
	$\alpha_3 = 4.3301$	0.1651	4.3421	0.1740	94.6
			(4.3313,4.3528)	(0.1662,0.1815)	(93.2,96.0)
	$\theta = 3.0$	0.1417	3.0079	0.1458	95.2
			(2.9989,3.0169)	(0.1393,0.1521)	(93.9,96.5)

Table 2: Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for maximum likelihood estimates of all-pass model parameters when f is the rescaled Students' t-density. For each sample size n, empirical confidence intervals were computed using standard asymptotic theory for 1000 iid replicates. Asymptotic means and standard deviations were computed using Theorem 1.

	Asymptotic		Empirical		
n	mean	std.dev.	mean (c.i.)	std.dev. (c.i.)	% coverage (c.i.)
500	$\phi_1 = 0.5$	0.0218	0.4989	0.0232	93.3
			(0.4975,0.5004)	(0.0222,0.0242)	(91.8,94.8)
	$\alpha_2 = 4.0$	0.2045	3.9984	0.1928	96.2
			(3.9865,4.0104)	(0.1841,0.2010)	(95.0,97.4)
	$\theta_1 = 0.6$	0.0476	0.6001	0.0492	92.5
			(0.5970,0.6031)	(0.0470,0.0513)	(90.9,94.1)
	$\theta_2 = 0.4$	0.0370	0.3995	0.0378	94.1
			(0.3972,0.4019)	(0.0361,0.0395)	(92.6,95.6)
5000	$\phi_1 = 0.5$	0.0069	0.5000	0.0070	94.1
			(0.4995,0.5004)	(0.0067,0.0073)	(92.6,95.6)
	$\alpha_2 = 4.0$	0.0646	3.9976	0.0643	94.3
			(3.9936,4.0016)	(0.0614,0.0671)	(92.9,95.7)
	$\theta_1 = 0.6$	0.0150	0.6001	0.0141	95.7
			(0.5992,0.6010)	(0.0135,0.0147)	(94.4,97.0)
	$\theta_2 = 0.4$	0.0117	0.3998	0.0114	95.5
			(0.3991,0.4005)	(0.0109,0.0119)	(94.2,96.8)
500	$\phi_1 = 0.3$	0.0230	0.2989	0.0239	93.7
			(0.2974,0.3004)	(0.0228,0.0249)	(92.2,95.2)
	$\phi_2 = 0.4$	0.0230	0.3990	0.0233	95.2
			(0.3975,0.4004)	(0.0223,0.0243)	(93.9,96.5)
	$\alpha_3 = 5.0$	0.2555	4.9902	0.2591	93.3
			(4.9742,5.0063)	(0.2475,0.2702)	(91.8,94.8)
	$\theta_1 = 0.6$	0.0476	0.5972	0.0483	94.5
			(0.5942,0.6001)	(0.0461,0.0503)	(93.1,95.9)
	$\theta_2 = 0.4$	0.0370	0.3977	0.0367	94.7
			(0.3954,0.4000)	(0.0351,0.0383)	(93.3,96.1)
5000	$\phi_1 = 0.3$	0.0073	0.3000	0.0074	95.1
			(0.2995,0.3004)	(0.0070,0.0077)	(93.8,96.4)
	$\phi_2 = 0.4$	0.0073	0.3996	0.0072	95.5
			(0.3991,0.4000)	(0.0069,0.0075)	(94.2,96.8)
	$\alpha_3 = 5.0$	0.0806	4.9960	0.0795	94.9
			(4.9911,5.0010)	(0.0759,0.0829)	(93.5,96.3)
	$\theta_1 = 0.6$	0.0150	0.6005	0.0147	94.5
			(0.5996,0.6014)	(0.0141,0.0154)	(93.1,95.9)
	$\theta_2 = 0.4$	0.0117	0.4006	0.0117	95.3
			(0.3999,0.4013)	(0.0112,0.0122)	(94.0,96.6)

Table 3: Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for maximum likelihood estimates of all-pass model parameters when f is the Gaussian scale mixture density. For each sample size n, empirical confidence intervals were computed using standard asymptotic theory for 1000 iid replicates. Asymptotic means and standard deviations were computed using Theorem 1.

When fitting (17), it is, therefore, not necessary to look at all possible configurations of roots of the MA polynomial inside and outside the unit circle. First, fit a causal, invertible ARMA$(p, q+r)$ model to the data using a method such as Gaussian maximum likelihood, and obtain the residuals $\{\hat{W}_t\}$ and estimates of $\phi(z)$ and $\theta_i(z)\theta_{ni}^{(i)}(z)$. Appropriate values for p and $q+r$ can be found using a standard order selection procedure such as the Akaike information criterion. Then fit a causal all-pass model of order r to $\{\hat{W}_t\}$ and obtain $\hat{\theta}_{ni}^{(i)}(z)$, an estimate of $\theta_{ni}^{(i)}(z)$. An appropriate r can be determined using the all-pass order selection procedure described in Section 3.2. The rth order polynomial with roots that are reciprocals of the roots of $\hat{\theta}_{ni}^{(i)}(z)$ is an estimate of $\theta_{ni}(z)$. An estimate of $\theta_i(z)$ can be obtained by canceling the roots of the invertible MA$(q+r)$ polynomial from the Gaussian likelihood fit which correspond to roots of $\hat{\theta}_{ni}^{(i)}(z)$.

4.3 Deconvolution

In this example, we simulate a seismogram $\{X_t\}_{t=1}^{1000}$ via

$$X_t = \sum_k \beta_k Z_{t-k},$$

where $\{\beta_k\}$ is the water gun wavelet sequence shown in Figure 8(2) of Lii and Rosenblatt [17] and $\{Z_t\}$ is a reflectivity sequence simulated here as iid noise from the Students' t distribution with five degrees of freedom. It is assumed that the seismogram is observed, but the wavelet and reflectivity sequences are unknown, as would be the case in a real deconvolution problem. We model the seismogram as a possibly noninvertible ARMA using the procedure described in Section 4.2 and attempt to reconstruct the wavelet and reflectivity sequences. This problem is of interest because, for an observed water gun seismogram, the reflectivity sequence corresponds to reflection coefficients for layers of the earth.

The simulated seismogram $\{X_t\}$ is shown in Figure 1(a). The corrected Akaike information criterion indicates that an ARMA$(12, 13)$ model is appropriate for the data, and the causal, invertible ARMA fit to $\{X_t\}$ using Gaussian maximum likelihood is $X_t = \phi^{-1}(B)\theta(B)W_t$, where

$$\begin{aligned}\phi(B) &= 1 - 0.1013B + 0.1137B^2 - 0.0776B^3 + 0.0542B^4 - 0.0326B^5 + 0.0086B^6 - 0.2280B^7 \\ &\quad + 0.1135B^8 + 0.2242B^9 - 0.0263B^{10} - 0.0793B^{11} + 0.1587B^{12}\end{aligned}$$

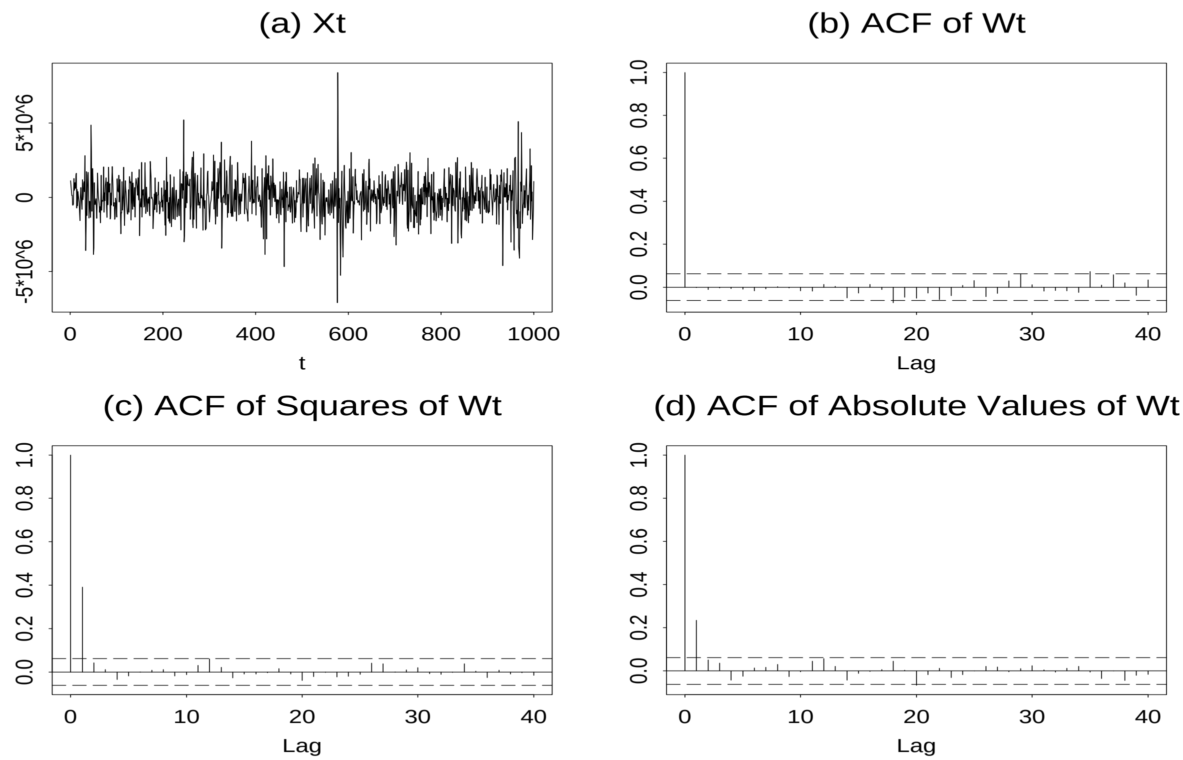

Figure 1: (a) The simulated seismogram of length 1000, $\{X_t\}$, and the sample autocorrelation functions with bounds $\pm 1.96/\sqrt{1000}$ for (b) $\{\hat{W}_t\}$, (c) $\{\hat{W}_t^2\}$, and (d) $\{|\hat{W}_t|\}$.

and

$$\begin{aligned}\theta(B) \quad = \quad & 1 - 0.0589B - 0.1843B^2 + 0.0918B^3 - 0.1068B^4 - 0.0226B^5 - 0.2400B^6 + 0.1196B^7 \\ & + 0.2206B^8 + 0.3376B^9 - 0.2004B^{10} - 0.0154B^{11} + 0.2872B^{12} + 0.2851B^{13}.\end{aligned}$$

The residuals from this fitted model are denoted $\{\hat{W}_t\}$. From the sample autocorrelation functions of $\{\hat{W}_t\}$, $\{\hat{W}_t^2\}$, and $\{|\hat{W}_t|\}$ in Figure 1(b)–(d), it appears the ARMA residuals are uncorrelated but dependent, suggesting the inappropriateness of a causal, invertible ARMA model.

The all-pass order selection procedure described in Section 3.2 indicates that an all-pass model of order two provides a good fit for $\{\hat{W}_t\}$ and, when (10) is used for f, the MLEs for this fitted all-pass model are

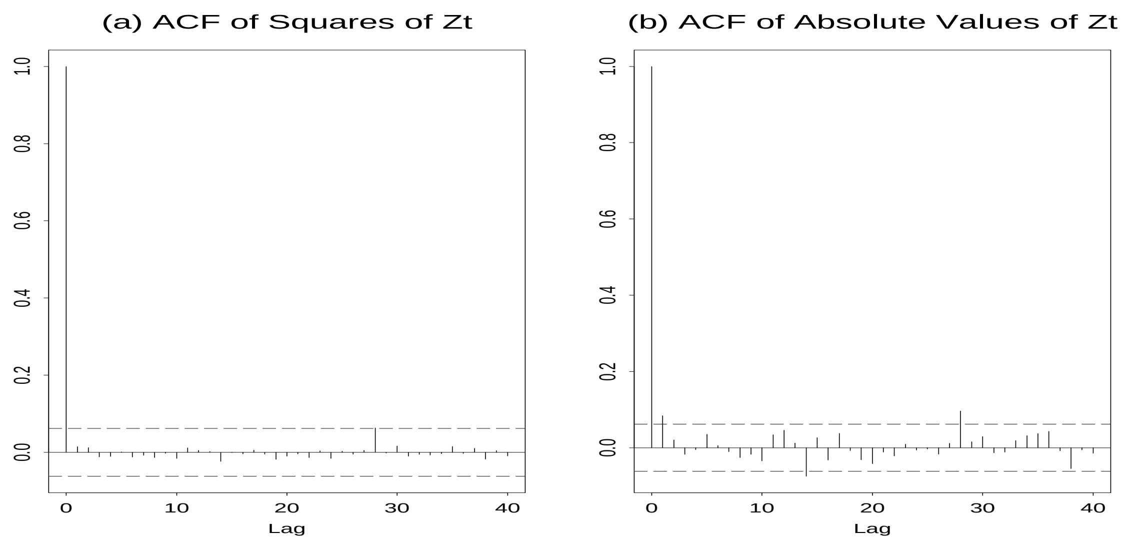

Figure 2: Diagnostics for the all-pass model of order two fit to the causal, invertible ARMA residuals: the sample autocorrelation functions with bounds $\pm 1.96/\sqrt{1000}$ for (a) $\{\hat{Z}_t^2\}$ and (b) $\{|\hat{Z}_t|\}$.

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_{ML} &= (\hat{\phi}_1, \hat{\phi}_2, \hat{\alpha}_3, \hat{\theta})' \\ &= (1.5286, -0.5908, 1097690.917, 4.7232)',\end{aligned}$$

with standard errors 0.0338, 0.0338, 44515.2174, and 0.7057 respectively. The sample autocorrelation functions for the squares and absolute values of $\{\hat{Z}_t\}$, the residuals from the fitted all-pass model, are shown in Figure 2. Because the series $\{\hat{Z}_t\}$ appears independent,

$$X_t = \frac{B^2(1 - 1.5286B^{-1} + 0.5908B^{-2})}{0.5908(1 - 1.5286B + 0.5908B^2)} \frac{\theta(B)}{\phi(B)} Z_t \tag{18}$$

seems to be a more appropriate model for $\{X_t\}$.

Since the simulated reflectivity sequence is iid with the Students' t-distribution and five degrees of freedom, in an effort to reconstruct the wavelet and reflectivity sequences, we express the right hand side of (18) as

$$-(0.5908)(1097690.917)\sqrt{3/5}\frac{B^2(1 - 1.5286B^{-1} + 0.5908B^{-2})}{0.5908(1 - 1.5286B + 0.5908B^2)} \frac{\theta(B)}{\phi(B)} \tilde{Z}_t, \tag{19}$$

where $\{\tilde{Z}_t\}$ is iid with density $\sqrt{3/5}\, f(\sqrt{3/5}\, s; 5)$, the Students' t-density with five degrees of freedom. Note that no roots of the polynomial in the denominator of (19) cancel exactly with roots of the polynomial in the numerator. So, for further model accuracy, we can directly fit a causal, noninvertible ARMA(12, 13) with two roots of the moving average polynomial inside the unit circle. Using maximum likelihood estimation with the Students' t-density and five degrees of freedom, this yields

$$X_t = -(0.5908)(1097690.917)\sqrt{3/5}\,\tilde{\phi}^{-1}(B)\tilde{\theta}(B)\tilde{Z}_t,$$

where

$$\begin{aligned}\tilde{\phi}(B) \quad &= \quad 1 - 0.0501B + 0.4967B^2 - 0.0664B^3 + 0.3255B^4 - 0.0552B^5 + 0.2254B^6 - 0.2374B^7 \\ &\quad + 0.0420B^8 - 0.0323B^9 - 0.0669B^{10} - 0.1950B^{11} - 0.0690B^{12}\end{aligned}$$

and

$$\begin{aligned}\tilde{\theta}(B) \quad &= \quad 1 - 1.1726B - 0.3091B^2 - 0.3545B^3 - 0.1261B^4 + 0.0095B^5 + 0.0673B^6 + 0.6313B^7 \\ &\quad + 0.2687B^8 + 0.7172B^9 - 0.0335B^{10} + 0.5541B^{11} + 0.5199B^{12} + 0.8270B^{13}.\end{aligned}$$

As shown in Figures 3 and 4,

$$-(0.5908)(1097690.917)\sqrt{3/5}\,\tilde{\phi}^{-1}(B)\tilde{\theta}(B)$$

and the residuals provide good estimates of the water gun wavelet and reflectivity sequences respectively.

Appendix

This section contains proofs of the lemmas used to establish the results of Section 3. First, for an arbitrary, causal autoregressive polynomial $\phi(z)$, define $\varphi(z) = \phi_1 z + \cdots + \phi_p z^p = 1 - \phi(z)$, and define $\varphi_0(z) = 1 - \phi_0(z)$. Note that, for $t = 1, \ldots, n - p$,

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

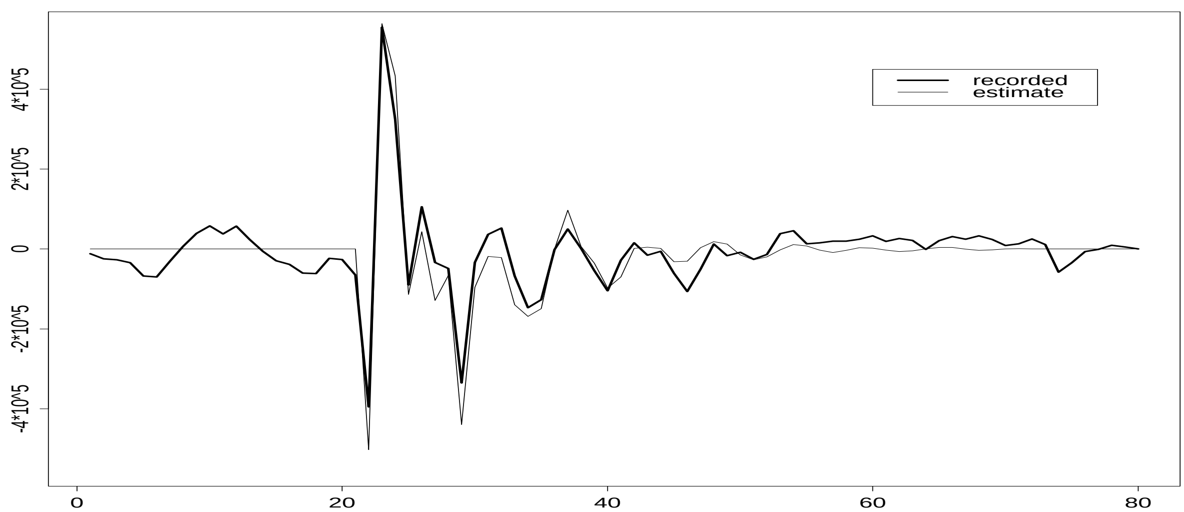

Figure 3: The recorded water gun wavelet and its estimate.

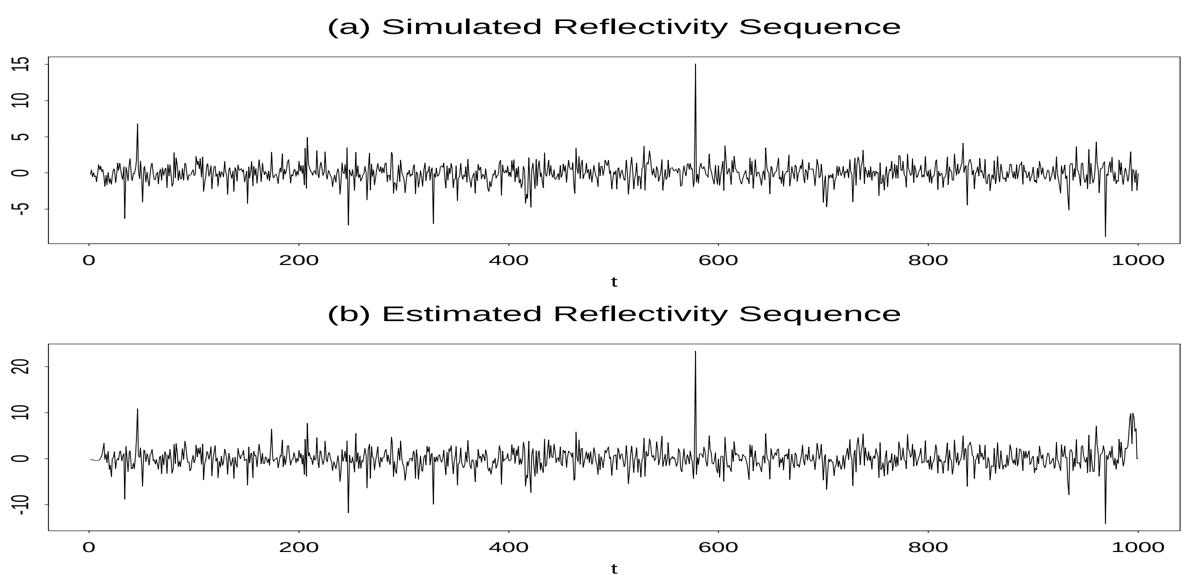

Figure 4: The simulated reflectivity sequence and its estimate.

so, if $j = 1, \ldots, p$, then

$$\frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1}) z_t(\boldsymbol{\phi}) \right\} = -X_{t+p-j} + \frac{\partial z_t(\boldsymbol{\phi})}{\partial \phi_j}. \tag{20}$$

Also, if $j = 1, \ldots, p$, then

$$\begin{aligned} \frac{\partial}{\partial \phi_j} \left\{ \varphi(B^{-1}) z_t(\boldsymbol{\phi}) \right\} &= \frac{\partial}{\partial \phi_j} \left\{ \phi_1 z_{t+1}(\boldsymbol{\phi}) + \cdots + \phi_p z_{t+p}(\boldsymbol{\phi}) \right\} \\ &= \varphi(B^{-1}) \frac{\partial z_t(\boldsymbol{\phi})}{\partial \phi_j} + z_{t+j}(\boldsymbol{\phi}). \end{aligned} \tag{21}$$

Equating (20) and (21) and solving for $\partial z_t(\boldsymbol{\phi}) / \partial \phi_j$, we obtain

$$\frac{\partial z_t(\boldsymbol{\phi})}{\partial \phi_j} = \frac{1}{\phi(B^{-1})} \left\{ X_{t+p-j} + z_{t+j}(\boldsymbol{\phi}) \right\}. \tag{22}$$

Evaluating (22) at the true value of $\boldsymbol{\phi}$ and ignoring the effect of recursion initialization, we have

$$\begin{aligned} \frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \phi_j} &= \frac{1}{\phi_0(B^{-1})} \left\{ \frac{-\phi_0(B^{-1}) B^p z_{t+p-j}}{\phi_0(B)} + z_{t+j}(\boldsymbol{\phi}_0) \right\} \\ &\simeq \frac{-z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}}{\phi_0(B^{-1})}, \end{aligned} \tag{23}$$

where the first term is an element of $\sigma(z_{t-1}, z_{t-2}, \ldots)$ and the second term is an element of $\sigma(z_{t+1}, z_{t+2}, \ldots)$ because $\phi_0(B)$ is a causal operator and $\phi_0(B^{-1})$ is a purely noncausal operator. It follows that (23) is independent of $z_t = \phi_{0r}^{-1} Z_t$.

Thus, for $j = 1, \ldots, p$,

$$\begin{aligned} \frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \alpha_j} &= \frac{f'\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1}; \boldsymbol{\theta}_0\right)}{f\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1}; \boldsymbol{\theta}_0\right)} \frac{1}{\alpha_{0,p+1}} \frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \phi_j} \\ &\simeq \frac{f'\left(z_t/\alpha_{0,p+1}; \boldsymbol{\theta}_0\right)}{f\left(z_t/\alpha_{0,p+1}; \boldsymbol{\theta}_0\right)} \frac{1}{\alpha_{0,p+1}} \left\{ \frac{-z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}}{\phi_0(B^{-1})} \right\} \\ &=: \frac{\partial g_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j}. \end{aligned} \tag{24}$$

The expected value of (24) is zero by the independence of its two terms.

We now compute the autocovariance function $\gamma^\dagger(h)$ of the zero-mean, stationary process $\left\{ \mathbf{u}_p' \left[\partial g_t^*(\boldsymbol{\alpha}_0) / \partial \alpha_j \right]_{j=1}^p \right\}$ for $\mathbf{u}_p \in \mathbb{R}^p$:

$$\begin{aligned} \gamma^\dagger(h) &= \mathrm{E} \left\{ \mathbf{u}_p' \left[\frac{\partial g_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right]_{j=1}^p \left(\left[\frac{\partial g_{t+h}^*(\boldsymbol{\alpha}_0)}{\partial \alpha_k} \right]_{k=1}^p \right)' \mathbf{u}_p \right\} \\ &= \mathbf{u}_p' \left[\nu_{jk}(h) \right]_{j,k=1}^p \mathbf{u}_p, \end{aligned}$$

where

$$\nu_{jk}(h) := \begin{cases} 2\gamma(j-k)\tilde{J}, & h = 0, \\ -\psi_{|h|-j}\psi_{|h|-k}, & h \neq 0, \end{cases}$$

and the ψ_ℓ are given by $\sum_{\ell=0}^{\infty} \psi_\ell z^\ell = 1/\phi_0(z)$ with $\psi_\ell = 0$ for $\ell < 0$. Thus,

$$\begin{aligned} \gamma^\dagger(0) + 2\sum_{h=1}^{\infty}\gamma^\dagger(h) &= \mathbf{u}_p' \left\{ [2\tilde{J}\gamma(j-k)]_{j,k=1}^p - 2\left[\sum_{h=1}^{\infty}\psi_{h-j}\psi_{h-k}\right]_{j,k=1}^p \right\} \mathbf{u}_p \\ &= 2(\sigma_0^2\tilde{J} - 1)\mathbf{u}_p'\sigma_0^{-2}\mathbf{\Gamma}_p\mathbf{u}_p. \end{aligned}$$

By A5, $\sigma_0^2\tilde{J} - 1 > 0$.

Next,

$$\begin{aligned} \frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \alpha_{p+1}} &= \frac{-f'\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}{f\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}\frac{z_t(\boldsymbol{\phi}_0)}{\alpha_{0,p+1}^2} - \frac{1}{\alpha_{0,p+1}} \\ &\simeq \frac{-f'\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}{f\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}\frac{z_t}{\alpha_{0,p+1}^2} - \frac{1}{\alpha_{0,p+1}} \\ &=: \frac{\partial g_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_{p+1}}. \end{aligned} \tag{25}$$

The expected value of (25) is zero and the variance is $\tilde{K}$. Also, the sequence (25) is iid and orthogonal to the corresponding partials for α_j, $j = 1, \ldots, p$, in (24).

For $j = p+2, \ldots, p+d+1$,

$$\begin{aligned} \frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \alpha_j} &= \frac{1}{f\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}\frac{\partial f\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}{\partial \theta_{j-p-1}} \\ &\simeq \frac{1}{f\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}\frac{\partial f\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}{\partial \theta_{j-p-1}} \\ &=: \frac{\partial g_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j}. \end{aligned} \tag{26}$$

By A7 and the dominated convergence theorem, the expected value of (26) is zero. In addition, the series $\left\{[\partial g_t^*(\boldsymbol{\alpha}_0)/\partial\alpha_j]_{j=p+2}^{p+d+1}\right\}$ is iid, has covariance matrix I, and is orthogonal to the partials for α_j, $j = 1, \ldots, p$, in (24). The expectation of $(\partial g_t^*(\boldsymbol{\alpha}_0)/\partial\alpha_{p+1})\,[\partial g_t^*(\boldsymbol{\alpha}_0)/\partial\alpha_j]_{j=p+2}^{p+d+1}$ is L.

The preceding calculations lead directly to the following lemma.

Lemma 1 *If f satisfies A1–A7, then, as $n \to \infty$,*

$$n^{-1/2}\sum_{t=1}^{n-p}\frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \xrightarrow{d} \mathbf{N} \sim N\left(\mathbf{0}, \mathbf{\Sigma}\right),$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} 2(\sigma_0^2\tilde{J}-1)\sigma_0^{-2}\mathbf{\Gamma}_p & \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times d} \\ \mathbf{0}_{1\times p} & \tilde{K} & L' \\ \mathbf{0}_{d\times p} & L & I \end{bmatrix}.$$

Proof: Note that, for $t = 0, \ldots, n-p-1$,

$$z_{n-p-t} = \sum_{l=0}^{\infty} \psi_l \left(\phi_0(B^{-1}) z_{n-p-t+l}\right) \quad \text{and} \quad z_{n-p-t}(\boldsymbol{\phi}_0) = \sum_{l=0}^{t} \psi_l \left(\phi_0(B^{-1}) z_{n-p-t+l}\right).$$

Because there exist constants $C \in (0,\infty)$ and $D \in (0,1)$ such that $|\psi_l| < CD^l$ for all $l \in \{0,1,\ldots\}$ (see Brockwell and Davis [7], Section 3.3), using A7 and the mean value theorem we can show that

$$n^{-1/2} \sum_{t=1}^{n-p} \frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - n^{-1/2} \sum_{t=1}^{n-p} \frac{\partial g_t^*(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \to \mathbf{0}$$

in L_1 and hence in probability.

Let $\mathbf{u} = (\mathbf{u}_p', u_1, \mathbf{u}_d')' \in \mathbb{R}^{p+d+1}$. By the Cramér-Wold device, it suffices to show

$$n^{-1/2} \sum_{t=1}^{n-p} V_t \xrightarrow{d} \mathrm{N}\left(0, 2(\sigma_0^2\tilde{J}-1)\mathbf{u}_p'\sigma_0^{-2}\mathbf{\Gamma}_p\mathbf{u}_p + u_1^2\tilde{K} + 2u_1\mathbf{u}_d'L + \mathbf{u}_d' I \mathbf{u}_d\right),$$

where $V_t := \mathbf{u}'\partial g_t^*(\boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}$. Elements of the infinite order moving average stationary sequence $\{V_t\}$ can be truncated to create a finite order moving average stationary sequence. By applying a central limit theorem (Brockwell and Davis [7], Theorem 6.4.2) to each truncation level, asymptotic normality can be deduced. The details are omitted. □

Now consider the mixed partials of $g_t(\boldsymbol{\alpha})$. For $j, k = 1, \ldots, p$,

$$\begin{aligned} \frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial\alpha_j\partial\alpha_k} = & \frac{f'\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}{f\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)} \frac{1}{\alpha_{0,p+1}} \frac{\partial^2 z_t(\boldsymbol{\phi}_0)}{\partial\phi_j\partial\phi_k} \\ & + \frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial\phi_j} \frac{f''\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}{\alpha_{0,p+1}^2 f\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)} \frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial\phi_k} \\ & - \frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial\phi_j} \frac{\left(f'\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)\right)^2}{\alpha_{0,p+1}^2 f^2\left(z_t(\boldsymbol{\phi}_0)/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)} \frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial\phi_k}. \end{aligned} \tag{27}$$

Because

$$\begin{aligned}\frac{\partial^2 z_t(\boldsymbol{\phi}_0)}{\partial\phi_j\partial\phi_k} &= \frac{1}{\phi_0^2(B^{-1})}\left\{X_{t+p+j-k}+X_{t+p+k-j}+2z_{t+j+k}(\boldsymbol{\phi}_0)\right\}\\ &\simeq \frac{-z_{t+j-k}-z_{t+k-j}}{\phi_0(B^{-1})\phi_0(B)}+\frac{2z_{t+j+k}}{\phi_0^2(B^{-1})}\\ &= -\sum_{m=0}^{\infty}\sum_{\ell=0}^{\infty}\psi_m\psi_\ell(z_{t+j-k-\ell+m}+z_{t+k-j-\ell+m})+\frac{2z_{t+j+k}}{\phi_0^2(B^{-1})},\end{aligned}$$

(27) is approximately

$$\begin{aligned}\frac{\partial^2 g_t^*(\boldsymbol{\alpha}_0)}{\partial\alpha_j\partial\alpha_k} := &\frac{f'\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}{f\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}\frac{1}{\alpha_{0,p+1}}\\ &\times\left\{-\sum_{m=0}^{\infty}\sum_{\ell=0}^{\infty}\psi_m\psi_\ell(z_{t+j-k-\ell+m}+z_{t+k-j-\ell+m})+\frac{2z_{t+j+k}}{\phi_0^2(B^{-1})}\right\}\\ &+\frac{f\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)f''\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)-\left(f'\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)\right)^2}{\alpha_{0,p+1}^2 f^2\left(z_t/\alpha_{0,p+1};\boldsymbol{\theta}_0\right)}\\ &\times\left\{\frac{-z_{t-j}}{\phi_0(B)}+\frac{z_{t+j}}{\phi_0(B^{-1})}\right\}\left\{\frac{-z_{t-k}}{\phi_0(B)}+\frac{z_{t+k}}{\phi_0(B^{-1})}\right\},\end{aligned}$$

which has expectation $-2\sigma_0^{-2}\gamma(j-k)(\sigma_0^2\tilde{J}-1)$. Similar arguments show that the approximations of the mixed partials evaluated at the true parameter values have expectation zero for $j=1,\ldots,p$, $k=p+1,\ldots,p+d+1$, $-\tilde{K}$ for $j=k=p+1$,

$$\frac{1}{\alpha_{0,p+1}}\int\frac{sf'\left(s;\boldsymbol{\theta}_0\right)}{f\left(s;\boldsymbol{\theta}_0\right)}\frac{\partial f\left(s;\boldsymbol{\theta}_0\right)}{\partial\theta_{k-p-1}}\,ds$$

for $j=p+1$, $k=p+2,\ldots,p+d+1$, and

$$-\int\frac{1}{f\left(s;\boldsymbol{\theta}_0\right)}\frac{\partial f\left(s;\boldsymbol{\theta}_0\right)}{\partial\theta_{j-p-1}}\frac{\partial f\left(s;\boldsymbol{\theta}_0\right)}{\partial\theta_{k-p-1}}\,ds$$

for $j,k=p+2,\ldots,p+d+1$.

Lemma 2 *If f satisfies A1–A7, then, as $n\to\infty$,*

$$n^{-1}\sum_{t=1}^{n-p}\frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'}\xrightarrow{P}\begin{bmatrix}-2(\sigma_0^2\tilde{J}-1)\sigma_0^{-2}\boldsymbol{\Gamma}_p & \mathbf{0}_{p\times1} & \mathbf{0}_{p\times d}\\ \mathbf{0}_{1\times p} & -\tilde{K} & -L'\\ \mathbf{0}_{d\times p} & -L & -I\end{bmatrix}=-\boldsymbol{\Sigma}.$$

Proof: By A7,

$$n^{-1}\sum_{t=1}^{n-p}\frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'} - n^{-1}\sum_{t=1}^{n-p}\frac{\partial^2 g_t^*(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'} \xrightarrow{P} \mathbf{0},$$

and, by the ergodic theorem and the computations preceding the lemma,

$$n^{-1}\sum_{t=1}^{n-p}\frac{\partial^2 g_t^*(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'} \xrightarrow{P} -\boldsymbol{\Sigma}.$$

□

Lemma 3 *For $\mathbf{u}\in\mathbb{R}^{p+d+1}$, define*

$$S_n^{\dagger}(\mathbf{u}) = n^{-1/2}\sum_{t=1}^{n-p}\mathbf{u}'\frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}} + \frac{1}{2}n^{-1}\sum_{t=1}^{n-p}\mathbf{u}'\frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'}\mathbf{u}$$

and

$$S_n(\mathbf{u}) = \sum_{t=1}^{n-p}\left[g_t\left(\boldsymbol{\alpha}_0 + n^{-1/2}\mathbf{u}\right) - g_t\left(\boldsymbol{\alpha}_0\right)\right].$$

If f satisfies A1–A7,

1. *$S_n^{\dagger} \xrightarrow{d} S$ on $C(\mathbb{R}^{p+d+1})$, where*

$$S(\mathbf{u}) := \mathbf{u}'\mathbf{N} - \frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}\mathbf{u},$$

   $\mathbf{N}\sim N(\mathbf{0},\boldsymbol{\Sigma})$, and $C(\mathbb{R}^{p+d+1})$ is the space of continuous functions on $\mathbb{R}^{p+d+1}$ where convergence is equivalent to uniform convergence on every compact set.

2. *$S_n \xrightarrow{d} S$ on $C(\mathbb{R}^{p+d+1})$.*

Proof:

1. The finite dimensional distributions of $S_n^{\dagger}$ converge to those of S by Lemmas 1 and 2. Since $S_n^{\dagger}$ is quadratic in $\mathbf{u}$, $\{S_n^{\dagger}\}$ is tight on $C(K)$ for any compact set $K\subset\mathbb{R}^{p+d+1}$. Therefore, $S_n^{\dagger}$ converges to S on $C(\mathbb{R}^{p+d+1})$ by Theorem 7.1 in Billingsley [2].

2. By a Taylor series expansion,

$$\begin{aligned} S_n(\mathbf{u}) &= \sum_{t=1}^{n-p}\left[g_t\left(\boldsymbol{\alpha}_0+n^{-1/2}\mathbf{u}\right)-g_t\left(\boldsymbol{\alpha}_0\right)\right] \\ &= n^{-1/2}\sum_{t=1}^{n-p}\mathbf{u}'\frac{\partial g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}} + \frac{1}{2}n^{-1}\sum_{t=1}^{n-p}\mathbf{u}'\frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'}\mathbf{u} \\ &\quad + \frac{1}{2}n^{-1}\sum_{t=1}^{n-p}\mathbf{u}'\left(\frac{\partial^2 g_t(\boldsymbol{\alpha}_n^*(\mathbf{u}))}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'}-\frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'}\right)\mathbf{u} \end{aligned}$$

for some $\boldsymbol{\alpha}_n^*(\mathbf{u})$ on the line segment connecting $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_0+n^{-1/2}\mathbf{u}$. If $\|\cdot\|$ measures Euclidean distance,

$$\sup_{\mathbf{u}\in K}\|\boldsymbol{\alpha}_n^*(\mathbf{u})-\boldsymbol{\alpha}_0\|\to 0$$

for any compact set $K\subset\mathbb{R}^{p+d+1}$, and so using A7 we can show that

$$n^{-1}\sum_{t=1}^{n-p}\mathbf{u}'\left(\frac{\partial^2 g_t(\boldsymbol{\alpha}_n^*(\mathbf{u}))}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'}-\frac{\partial^2 g_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}'}\right)\mathbf{u}\xrightarrow{P}\mathbf{0}$$

on $C(\mathbb{R}^{p+d+1})$. Thus, $\{S_n\}$ must have the same limiting distribution as $\{S_n^\dagger\}$ on $C(\mathbb{R}^{p+d+1})$.

□

Now let $\tilde{z}_t(\boldsymbol{\phi})=-(\phi(B)/\phi(B^{-1}))X_{t+p}$ and consider $\boldsymbol{\Xi}\subset\mathbb{R}^{p+d+1}$, any compact, convex parameter space such that $\boldsymbol{\phi}$ forms a causal polynomial, $\alpha_{p+1}>0$, and $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ for all $\boldsymbol{\alpha}=(\boldsymbol{\phi}',\alpha_{p+1},\boldsymbol{\theta}')'\in\boldsymbol{\Xi}$.

Lemma 4 *If f satisfies A1, $\int|\ln f(s;\boldsymbol{\theta}_0)|f(s;\boldsymbol{\theta}_0)\,ds<\infty$, and $|sf'(s;\boldsymbol{\theta})|/f(s;\boldsymbol{\theta})$ and $|\frac{\partial}{\partial\theta_j}f(s;\boldsymbol{\theta})|/f(s;\boldsymbol{\theta})$, $j=1,\ldots,d$, are dominated by $a_3+a_4|s|^{c_2}$ for all $\boldsymbol{\alpha}\in\boldsymbol{\Xi}$, where a_3,a_4, c_2 are non-negative constants such that $\int|s|^{c_2}f(s;\boldsymbol{\theta}_0)<\infty$, then, as $n\to\infty$,*

$$n^{-1}\mathcal{L}(\boldsymbol{\alpha})\xrightarrow{a.s.}E\{\ln f(\tilde{z}_1(\boldsymbol{\phi})/\alpha_{p+1};\boldsymbol{\theta})-\ln\alpha_{p+1}\}$$

uniformly on $\boldsymbol{\Xi}$.

Proof: For any $\boldsymbol{\alpha}\in\boldsymbol{\Xi}$,

$$\mathrm{E}|\ln f(\tilde{z}_1(\boldsymbol{\phi})/\alpha_{p+1};\boldsymbol{\theta})-\ln\alpha_{p+1}|<\infty,$$

and so

$$n^{-1}\mathcal{L}(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathrm{E}\{\ln f(\tilde{z}_1(\boldsymbol{\phi})/\alpha_{p+1};\boldsymbol{\theta}) - \ln \alpha_{p+1}\}$$

by the ergodic theorem. Therefore, the lemma follows by the Arzela-Ascoli theorem if $n^{-1}\mathcal{L}(\boldsymbol{\alpha})$ is equicontinuous and uniformly bounded on Ξ almost surely.

If $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \Xi$, then, for some $\boldsymbol{\alpha}_n^*$ on the line segment connecting $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$,

$$|n^{-1}\mathcal{L}(\boldsymbol{\alpha}_1) - n^{-1}\mathcal{L}(\boldsymbol{\alpha}_2)| \le \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| n^{-1} \left\| \sum_{t=1}^{n-p} \frac{\partial}{\partial \boldsymbol{\alpha}} g_t(\boldsymbol{\alpha}_n^*) \right\|.$$

Because $\boldsymbol{\phi}$ forms a causal polynomial for all $\boldsymbol{\phi}$ in Ξ, there exist coefficients $\pi_k \ge 0$, $k = 0, \pm 1, \ldots$, decaying at a geometric rate such that

$$\sup_{\boldsymbol{\alpha}\in\Xi} |z_t(\boldsymbol{\phi})| \le \sum_{k=-\infty}^{\infty} \pi_k |z_{t-k}| \qquad \text{and} \qquad \sup_{\boldsymbol{\alpha}\in\Xi} \left| \frac{\partial z_t(\boldsymbol{\phi})}{\partial \phi_j} \right| \le \sum_{k=-\infty}^{\infty} \pi_k |z_{t-k}|,\ j = 1, \ldots, p,$$

for $t = 1, \ldots, n-p$ and all n. Also, because $\alpha_{p+1} > 0$ for every value of α_{p+1} in Ξ, there exists a constant $M > 0$ such that

$$\sup_{\boldsymbol{\alpha}\in\Xi} \frac{1}{\alpha_{p+1}} \le M.$$

Consequently, for $j = 1, \ldots, p$, $t = 1, \ldots, n-p$, and all n, we have

$$\begin{aligned} \left| \frac{\partial g_t(\boldsymbol{\alpha}_n^*)}{\partial \alpha_j} \right| &= \left| \frac{f'(z_t(\boldsymbol{\phi}_n^*)/\alpha_{n,p+1}^*;\boldsymbol{\theta}_n^*)}{f(z_t(\boldsymbol{\phi}_n^*)/\alpha_{n,p+1}^*;\boldsymbol{\theta}_n^*)} \frac{1}{\alpha_{n,p+1}^*} \frac{\partial z_t(\boldsymbol{\phi}_n^*)}{\partial \phi_j} \right| \\ &\le a_5 M \sum_{k=-\infty}^{\infty} \pi_k |z_{t-k}| + a_4 \left(M \sum_{k=-\infty}^{\infty} \pi_k |z_{t-k}| \right)^{c_2 \vee 1}, \end{aligned}$$

where $a_5 = a_3 + \sup_{|s|\le 1, \boldsymbol{\alpha}\in\Xi} |f'(s;\boldsymbol{\theta})|/f(s;\boldsymbol{\theta})$. Also for $t = 1, \ldots, n-p$ and all n,

$$\begin{aligned} \left| \frac{\partial g_t(\boldsymbol{\alpha}_n^*)}{\partial \alpha_{p+1}} \right| &= \left| \frac{f'(z_t(\boldsymbol{\phi}_n^*)/\alpha_{n,p+1}^*;\boldsymbol{\theta}_n^*)}{f(z_t(\boldsymbol{\phi}_n^*)/\alpha_{n,p+1}^*;\boldsymbol{\theta}_n^*)} \frac{z_t(\boldsymbol{\phi}_n^*)}{(\alpha_{n,p+1}^*)^2} + \frac{1}{\alpha_{n,p+1}^*} \right| \\ &\le M \left[a_3 + a_4 \left(M \sum_{k=-\infty}^{\infty} \pi_k |z_{t-k}| \right)^{c_2} \right] + M \end{aligned}$$

and, for $j = p+2, \ldots, p+d+1$,

$$\begin{aligned} \left| \frac{\partial g_t(\boldsymbol{\alpha}_n^*)}{\partial \alpha_j} \right| &= \left| \frac{1}{f(z_t(\boldsymbol{\phi}_n^*)/\alpha_{n,p+1}^*;\boldsymbol{\theta}_n^*)} \frac{\partial f(z_t(\boldsymbol{\phi}_n^*)/\alpha_{n,p+1}^*;\boldsymbol{\theta}_n^*)}{\partial \theta_{j-p-1}} \right| \\ &\le a_3 + a_4 \left(M \sum_{k=-\infty}^{\infty} \pi_k |z_{t-k}| \right)^{c_2}. \end{aligned}$$

It follows that

$$n^{-1}\left\| \sum_{t=1}^{n-p} \frac{\partial}{\partial \boldsymbol{\alpha}} g_t(\boldsymbol{\alpha}_n^*) \right\| \leq constant$$

for all n sufficiently large almost surely, and thus $n^{-1}\mathcal{L}(\boldsymbol{\alpha})$ is equicontinuous on Ξ almost surely. It can be shown similarly that $n^{-1}\mathcal{L}(\boldsymbol{\alpha})$ is uniformly bounded on Ξ almost surely. □

Acknowledgments We would like to thank Professor Keh-Shin Lii for supplying the water gun wavelet used in Section 4.3. Also, the work reported here was developed in part under STAR Research Assistance Agreement CR-829095 awarded by the U.S. Environmental Protection Agency (EPA) to Colorado State University. This manuscript has not been formally reviewed by EPA. The views expressed here are solely those of the authors. EPA does not endorse any products or commercial services mentioned in this report.

References

[1] M.E. Andrews, Parameter Estimation for All-Pass Time Series Models, Ph.D. dissertation, Department of Statistics, Colorado State University, Fort Collins, CO, 2003.

[2] P. Billingsley, Convergence of Probability Measures, 2nd edition, Wiley, New York, 1999.

[3] W.E. Blass, G.W. Halsey, Deconvolution of Absorption Spectra, Academic Press, New York, 1981.

[4] F.J. Breidt, R.A. Davis, Time-reversibility, identifiability and independence of innovations for stationary time series, Journal of Time Series Analysis 13 (1991) 377–390.

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

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

[7] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, 2nd edition, Springer-Verlag, New York, 1991.

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

[9] H.-M. Chien, H.-L. Yang, C.-Y. Chi, Parametric cumulant based phase estimation of 1-d and 2-d nonminimum phase systems by allpass filtering, IEEE Transactions on Signal Processing 45 (1997) 1742–1762.

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

[11] D. Donoho, On minimum entropy deconvolution, in: D.F. Findley (Ed.), Applied Time Series Analysis II, Academic Press, New York, 1981, 565–608.

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

[13] R. Godfrey, F. Rocca, Zero memory nonlinear deconvolution, Geophysical Prospecting 29 (1981) 189–228.

[14] R. Hooke, T. Jeeves, A direct search solution of numerical and statistical problems, Journal of the Association for Computing Machinery 8 (1961) 212–229.

[15] A.-C. Hsueh, J.M. Mendel, Minimum-variance and maximum-likelihood deconvolution for noncausal channel models, IEEE Transactions on Geoscience and Remote Sensing 23 (1985) 797–808.

[16] J. Huang, Y. Pawitan, Quasi-likelihood estimation of non-invertible moving average processes, Scandinavian Journal of Statistics 27 (2000) 689–702.

[17] K.-S. Lii, M. Rosenblatt, Nonminimum phase non-Gaussian deconvolution, Journal of Multivariate Analysis 27 (1988) 359–374.

[18] K.-S. Lii, M. Rosenblatt, An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes, Journal of Multivariate Analysis 43 (1992) 272–299.

[19] K.-S. Lii, M. Rosenblatt, Maximum likelihood estimation for nonGaussian nonminimum phase ARMA sequences, Statistica Sinica 6 (1996) 1–22.

[20] M. Ooe, T.J. Ulrych, Minimum entropy deconvolution with an exponential transformation, Geophysical Prospecting 27 (1979) 458–473.

[21] R.A. Wiggins, Minimum entropy deconvolution, Geoexploration 16 (1978) 21–35.