Rank-Based 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

July 6, 2006

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 are useful for identifying and modeling noncausal and noninvertible autoregressive-moving average processes. We establish asymptotic normality and consistency for rank-based estimators of all-pass model parameters. The estimators are obtained by minimizing the rank-based residual dispersion function in L.A. Jaeckel [Estimating regression coefficients by minimizing the dispersion of the residuals, Ann. Math. Statist. 43 (1972) 1449–1458]. These estimators can have the same asymptotic efficiency as maximum likelihood estimators and are robust. The behavior of the estimators for finite samples is studied via simulation, and rank estimation is used in the deconvolution of a simulated water gun seismogram.

1 Introduction

Autoregressive-moving average (ARMA) models, the standard linear time series models for stationary data, are often fit to observed series using Gaussian likelihood, least-squares, or related second-order moment estimation techniques. These are effective methods for finding fitted ARMA models with second-order moment properties that resemble those of an observed series, whether or not the data are Gaussian. However, because every Gaussian ARMA process has a causal, invertible ARMA representation (all roots of the autoregressive and moving average polynomials are outside the unit circle), in the non-Gaussian case, the second-order methods are unable to identify a noncausal (at least one root of the autoregressive polynomial is inside the unit circle) or noninvertible (at least one root of the moving average polynomial is inside the unit circle) ARMA series. Fitted ARMA models obtained using second-order techniques may not, therefore, most effectively capture the higher-order moment structure of the data. Consequently, an effort to identify noncausal and noninvertible series should be part of any ARMA fitting procedure. In this paper, we discuss all-pass models which are useful tools for identifying and modeling noncausal and noninvertible ARMA processes.

All-pass models are ARMA models in which the roots of the autoregressive polynomial are reciprocals of roots of the moving average polynomial and vice versa. These models generate uncorrelated (white noise) time series that are not independent in the non-Gaussian case. As discussed in Andrews, Davis, and Breidt [2], an all-pass series can be obtained by fitting a causal, invertible ARMA model to a series generated by a causal, noninvertible ARMA model. 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. Consequently, 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. Noninvertible ARMA models have appeared, for example, in vocal tract filters (Chi and Kung [8], Chien, Yang, and Chi [9]), in the analysis of unemployment rates (Huang and Pawitan [13]), and in seismogram deconvolution (Andrews, Davis, and Breidt [2], Lii and Rosenblatt [19]).

All-pass models can be used similarly to fit noncausal ARMA models (Breidt, Davis, and Trindade [6]). See Breidt, Davis, and Trindade [6] for a list of applications for noncausal models.

Estimation methods based on second-order moment techniques cannot identify all-pass models because Gaussian all-pass series are independent. Thus, cumulant-based estimators, using cumulants of order greater than two, are often used to estimate these models (Chi and Kung [8], Chien, Yang, and Chi [9], Giannakis and Swami [11]). Breidt, Davis, and Trindade [6] consider a least absolute deviations (LAD) estimation approach which is motivated by the likelihood of an all-pass model with Laplace (two-sided exponential) noise, and Andrews, Davis, and Breidt [2] consider a maximum likelihood (ML) estimation approach. The LAD and ML estimators are consistent and asymptotically normal. However, the LAD estimation procedure is limited by the assumption that the mean and median for the noise are equivalent, and the ML procedure is limited by the assumption that the probability density function for the noise is symmetric and known to within some parameter values.

In this paper, we consider a rank-based estimation technique first proposed by Jaeckel [14] for estimating linear regression parameters. Jaeckel's estimator minimizes the sum of model residuals weighted by a function of residual rank. We study the asymptotic properties of Jaeckel's rank (R) estimator in the case of all-pass parameter estimation. This R-estimator is more robust than the LAD and ML estimators; it is consistent and asymptotically normal under less stringent conditions. In addition, when R-estimation is used in lieu of LAD or ML, efficiency need not be sacrificed. There exists a weight function for which R-estimation is asymptotically equivalent to LAD estimation and, when the noise distribution is known, the weight function can be chosen so that R-estimation is asymptotically equivalent to ML estimation. We also find that when the Wilcoxon weight function (a linear weight function) is used, R-estimation is relatively very efficient for a large class of noise distributions. Another advantage of R-estimation is that one has the flexibility to choose a weight function that tends to produce relatively smooth R-objective functions which can be minimized fairly easily.

Because the objective function for Jaeckel's R-estimation method involves not only the residual ranks but also the residual values, this is not pure R-estimation. Koul and Ossiander [16], Koul and Saleh [17],

Mukherjee and Bai [20], and Terpstra, McKean, and Naranjo [23] consider related rank-based estimation approaches for autoregressive model parameters. Also, Allal, Kaaouachi, and Paindaveine [1] examine a pure R-estimator for ARMA model parameters based on correlations of weighted residual ranks. The results for this pure R-estimator are not applicable to all-pass model parameters because the parameters in the autoregressive polynomial of an all-pass model are functions of parameters in the moving average polynomial and vice versa.

In Section 2, we consider Jaeckel's R-function in the context of all-pass parameter estimation. Asymptotic normality for R-estimators is established under mild conditions and order selection is discussed in Section 3. Proofs of the lemmas used to establish the results of Section 3 can be found in the Appendix. We study the behavior of the estimators for finite samples via simulation in Section 4.1, and use R-estimation in the deconvolution of a simulated water gun seismogram in Section 4.2.

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 $\{W_t\}$, $\phi^{-1}(B)W_t = (\sum_{j=0}^{\infty} \psi_j B^j)W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}$, a function of only the past and present $\{W_t\}$. If $\phi(B)$ is causal, the filter $B^p \phi(B^{-1})$ is *purely noncausal* and hence $B^{-p} \phi^{-1}(B^{-1})W_t = (\sum_{j=0}^{\infty} \psi_j B^{-p-j})W_t = \sum_{j=0}^{\infty} \psi_j W_{t+p+j}$, 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 $r = \max\{0 \leq j \leq p : \phi_{0j} \neq 0\}$. Then, a causal all-pass time series is the ARMA series $\{X_t\}$ which satisfies the difference equations

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

or

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

where the series $\{Z_t^*\}$ is an independent and identically distributed (iid) sequence of random variables with mean 0, variance $\sigma^2 \in (0,\infty)$, and distribution function F. The true order of the all-pass model is r $(0 \le r \le p)$. Observe 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 (2.1) is

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

which is constant for $\omega \in [-\pi,\pi]$, and thus $\{X_t\}$ is an uncorrelated sequence. In the case of Gaussian $\{Z_t^*\}$, this implies that $\{X_t\}$ is iid $N(0,\sigma^2\phi_{0r}^{-2})$, but independence does not hold in the non-Gaussian case if $r \ge 1$ (see Breidt and Davis [5]). The model (2.1) 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 (2.1) as

$$\phi_0(B)X_t = \frac{B^p\phi_0(B^{-1})}{-\phi_{0r}}Z_t, \tag{2.2}$$

where $\{Z_t\} = \{Z_{t+p-r}^*\}$ is an iid sequence of random variables with mean 0, variance σ^2, and distribution function F. Rearranging (2.2) 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:

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

where $\phi := (\phi_1,\ldots,\phi_p)'$. Let $\phi_0 = (\phi_{01},\ldots,\phi_{0p})' = (\phi_{01},\ldots,\phi_{0r},0,\ldots,0)'$ denote the true parameter vector and note that $\{z_t(\phi_0)\}_{t=1}^{n-p}$ closely approximates $\{z_t\}_{t=1}^{n-p}$; the error is due to the initialization with zeros. Although $\{z_t\}$ is iid, $\{z_t(\phi_0)\}_{t=1}^{n-p}$ is not iid if $r \ge 1$.

2.2 Jaeckel's Rank Function

Suppose we have a realization of length n, $\{X_t\}_{t=1}^n$, from (2.1). Let λ be a function from $(0,1)$ to $\mathbb{R}$ such that

A1 λ *is strictly increasing and* $\lambda(s) = -\lambda(1-s)$ *for all* $s \in (0,1)$.

If ϕ forms a causal pth order autoregressive polynomial and $\{R_t(\phi)\}_{t=1}^{n-p}$ contains the ranks of $\{z_t(\phi)\}_{t=1}^{n-p}$ from (2.3), then the R-function evaluated at ϕ with weight function λ is

$$D(\phi) := \sum_{t=1}^{n-p} \lambda\left(\frac{R_t(\phi)}{n-p+1}\right) z_t(\phi). \tag{2.4}$$

Because it tends to be near zero when the elements of $\{z_t(\phi)\}$ are similar, (2.4) is a measure of the dispersion of the residuals $\{z_t(\phi)\}$. When $\{z_{(t)}(\phi)\}_{t=1}^{n-p}$ is the series $\{z_t(\phi)\}_{t=1}^{n-p}$ ordered from smallest to largest, (2.4) can also be written as $D(\phi) = \sum_{t=1}^{n-p} \lambda(t/(n-p+1)) z_{(t)}(\phi)$. A popular choice for the weight function is $\lambda(s) = s - 1/2$. In this case, the weights $\{\lambda(t/(n-p+1))\}_{t=1}^{n-p}$ are known as the Wilcoxon scores.

We give some properties for D in the following theorem. Jaeckel [14] shows that these same properties hold for the R-function in the linear regression case.

Theorem 2.1 *Assume A1 holds. For any* $\phi \in \mathbb{R}^p$*, if*

$$\{P_1(\phi), \ldots, P_{(n-p)!}(\phi)\} = \{\{z_{1,1}(\phi), \ldots, z_{1,n-p}(\phi)\}, \ldots, \{z_{(n-p)!,1}(\phi), \ldots, z_{(n-p)!,n-p}(\phi)\}\}$$

contains the $(n-p)!$ *permutations of the sequence* $\{z_t(\phi)\}_{t=1}^{n-p}$*, then*

$$D(\phi) = \sup_{j \in \{1, \ldots, (n-p)!\}} \sum_{t=1}^{n-p} \lambda\left(\frac{t}{n-p+1}\right) z_{j,t}(\phi).$$

In addition, D is a non-negative, continuous function on $\mathbb{R}^p$, and $D(\phi) = 0$ if and only if the elements of $\{z_t(\phi)\}_{t=1}^{n-p}$ are all equal.

Proof: See the proof of Theorem 1 in Jaeckel [14]. □

3 Asymptotic Results

3.1 Parameter Estimation

In order to establish asymptotic normality for R-estimators of $\boldsymbol{\phi}_0$, we make the following additional assumptions:

A2 *F, the distribution function for the noise, is strictly increasing and differentiable on $\mathbb{R}$ with density f.*

A3 *f is uniformly continuous on $\mathbb{R}$ with $\sup_{s \in \mathbb{R}} |s| f(s) < \infty$.*

A4 *The derivative of the weight function λ exists and is uniformly continuous on $(0, 1)$.*

Also, let $\tilde{J} = \int_0^1 \lambda^2(s)\,ds$, $\tilde{K} = \int_0^1 F^{-1}(s)\lambda(s)\,ds$, and $\tilde{L} = \int_0^1 f(F^{-1}(s))\lambda'(s)\,ds$, and assume

A5 *$\sigma^2 \tilde{L} > \tilde{K}$.*

Theorem 3.1 *If A1–A5 hold, there exists a sequence of minimizers $\hat{\boldsymbol{\phi}}_R$ of $D(\cdot)$ in (2.4) such that*

$$n^{1/2}(\hat{\boldsymbol{\phi}}_R - \boldsymbol{\phi}_0) \xrightarrow{d} \mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{3.1}$$

where $\boldsymbol{\Sigma} := (\sigma^2 \tilde{J} - \tilde{K}^2)/[2(\sigma^2 \tilde{L} - \tilde{K})^2]\sigma^2 \boldsymbol{\Gamma}_p^{-1}$, $\boldsymbol{\Gamma}_p := [\gamma(j-k)]_{j,k=1}^p$, and $\gamma(\cdot)$ is the autocovariance function for the autoregressive process $\{(1/\phi_0(B))Z_t\}$.

Proof: $D(\boldsymbol{\phi}) - D(\boldsymbol{\phi}_0) = S_n(\sqrt{n}(\boldsymbol{\phi} - \boldsymbol{\phi}_0))$, where $S_n(\cdot)$ is defined in Lemma 5.5 of the Appendix. Because $\mathbf{Y} := -|\phi_{0r}|\sigma^2 \boldsymbol{\Gamma}_p^{-1}\mathbf{N}/[2(\sigma^2 \tilde{L} - \tilde{K})]$ minimizes the limit $S(\cdot)$ in Lemma 5.5, the result follows by Remark 1 in Davis, Knight, and Liu [10]. $\qquad \square$

Remark 1: R-estimators of linear regression parameters are also consistent and asymptotically normal (Jaeckel [14]). Note, however, that the conditions placed on λ in assumption A4 are slightly stronger than those placed on the weight function in Jaeckel [14], where the weight function is square integrable, not necessarily bounded or continuous. The conditions in A4 can be relaxed to some extent at the expense of stronger assumptions on f, but we do not pursue those extensions here. Since piecewise continuous and

unbounded weight functions on $(0,1)$ can be well approximated by differentiable, bounded weight functions, from a practical perspective, assumption A4 is not overly restrictive.

Remark 2: Using the Cauchy-Schwarz inequality,

$$\sigma^2 \tilde{J} - \tilde{K}^2 = \sigma^2 \mathrm{E}\left\{\lambda^2(F(Z_1))\right\} - \left(\mathrm{E}\left\{Z_1\lambda(F(Z_1))\right\}\right)^2 \geq \sigma^2 \mathrm{E}\left\{\lambda^2(F(Z_1))\right\} - \mathrm{E}\left\{Z_1^2\right\}\mathrm{E}\left\{\lambda^2(F(Z_1))\right\} = 0,$$

(3.2)

with equality in (3.2) if and only if λ is proportional to F^{-1}, which is not possible since $F^{-1}(0) = -\infty$, $F^{-1}(1) = \infty$, and λ is bounded on $(0,1)$. Hence, $\sigma^2 \tilde{J} - \tilde{K}^2 > 0$. $\tilde{K} = \int_0^1 F^{-1}(s)\lambda(s)\,ds$ is also greater than zero because F^{-1} and λ are strictly increasing functions on $(0,1)$ and λ is odd about $1/2$. Without assumption A5, $\sigma^2\tilde{L} - \tilde{K}$ is not necessarily greater than zero, however. If the density function f is differentiable, using integration by parts, it can be shown that

$$\tilde{L} = \mathrm{E}\left\{f(Z_1)\lambda'(F(Z_1))\right\} = -\int_{-\infty}^{\infty} f'(s)\lambda(F(s))\,ds = -\int_0^1 \frac{f'(F^{-1}(s))}{f(F^{-1}(s))}\lambda(s)\,ds.$$

Therefore, if $Z_1 \sim \mathrm{N}(0,\sigma^2)$,

$$\sigma^2\tilde{L} = -\sigma^2 \int_0^1 \frac{f'(F^{-1}(s))}{f(F^{-1}(s))}\lambda(s)\,ds = \int_0^1 F^{-1}(s)\lambda(s)\,ds = \tilde{K},$$

and so A5 does not hold if Z_1 is Gaussian.

Remark 3: The asymptotic covariance matrix for $\hat{\phi}_R$ is a scalar multiple of $n^{-1}\sigma^2\mathbf{\Gamma}_p^{-1}$, the asymptotic covariance matrix for Gaussian likelihood estimators of the parameters of the corresponding pth order autoregressive process. The same property holds for LAD and ML estimators of all-pass model parameters, as shown in Breidt, Davis, and Trindade [6] and Andrews, Davis, and Breidt [2] respectively. The LAD estimators are quasi-maximum likelihood estimators which can be obtained by maximizing the log-likelihood of an all-pass model with Laplace noise $(f(s) = \exp(-\sqrt{2}|s|/\sigma)/(\sqrt{2}\sigma))$. The appropriate scalar multiple is

$$\frac{\mathrm{Var}|Z_1|}{2\left(2\sigma^2 f(0) - \mathrm{E}|Z_1|\right)^2}$$

(3.3)

in the LAD case (Breidt, Davis, and Trindade [6] contains an error in the calculation of the asymptotic variance; see Andrews [3] for the correction) and

$$\frac{1}{2} \left(\sigma^2 \int_{-\infty}^{\infty} \frac{(f'(s))^2}{f(s)} \, ds - 1 \right)^{-1} \tag{3.4}$$

in the ML case, while the multiple in (3.1) for R-estimation is

$$\frac{\sigma^2 \tilde{J} - \tilde{K}^2}{2(\sigma^2 \tilde{L} - \tilde{K})^2}. \tag{3.5}$$

Consequently, the asymptotic relative efficiency (ARE) for R to LAD is obtained by dividing (3.3) by (3.5), and the ARE for R to ML is obtained by dividing (3.4) by (3.5).

Remark 4: Consider the sequence of weight functions $\{\lambda_m\}$ such that $\lambda_m(s) = 2\pi^{-1} \arctan(m(s-1/2))$. It is straightforward to show that λ_m satisfies assumptions A1 and A4 for all $m > 0$. If $I\{\cdot\}$ denotes the indicator function and $\tilde{\mu} := \text{median}\{Z_1\}$, A5 is satisfied for large m when $2\sigma^2 f(\tilde{\mu}) > \text{E}\{-Z_1 I\{Z_1 < \tilde{\mu}\} + Z_1 I\{Z_1 > \tilde{\mu}\}\}$; this holds for many distributions, including the Laplace, logistic, and Students' t (with degrees of freedom greater than two) distributions, and various asymmetric distributions $(0.4N(-1,1)+0.6N(2/3,3^2)$ is one example). Because $\lambda_m(s)$ converges pointwise to $-I\{s < 1/2\} + I\{s > 1/2\}$ on $(0,1)$ as $m \to \infty$, $\tilde{J}_m = \int_0^1 \lambda_m^2(s) \, ds \to 1$, and, if Z_1 has median zero, $\tilde{K}_m = \text{E}\{Z_1 \lambda_m(F(Z_1))\} \to \text{E}|Z_1|$ and $\tilde{L}_m = \text{E}\{f(Z_1)\lambda_m'(F(Z_1))\} \to 2f(0)$. Hence, if Z_1 has median zero,

$$\frac{\sigma^2 \tilde{J}_m - \tilde{K}_m^2}{2(\sigma^2 \tilde{L}_m - \tilde{K}_m)^2} \to \frac{\sigma^2 - \text{E}^2|Z_1|}{2\left(2\sigma^2 f(0) - \text{E}|Z_1|\right)^2} = \frac{\text{Var}|Z_1|}{2\left(2\sigma^2 f(0) - \text{E}|Z_1|\right)^2},$$

and so R-estimation has virtually the same asymptotic efficiency as LAD estimation when the weight function λ_m is used with m large. If Z_1 has a Laplace distribution, LAD estimation corresponds to ML estimation. In the case of Laplace noise, therefore, R-estimation with weight function λ_m and m large also has essentially the same asymptotic efficiency as ML estimation.

Remark 5: Under general conditions, it can be shown that (3.5) equals (3.4) when the weight function is proportional to $-f'(F^{-1}(s))/f(F^{-1}(s))$. Thus, R-estimation has the same asymptotic efficiency as ML estimation when an optimal weight function $\lambda_f(s) \propto -f'(F^{-1}(s))/f(F^{-1}(s))$ is used. λ_f is also an optimal

Noise Distribution	ARE (R to LAD)	ARE (R to ML)	ARE (R to R)
Laplace	0.600	0.600	1.026
logistic	1.976	1.000	1.049
$t(3)$	1.411	0.962	1.208
$t(6)$	2.068	0.997	1.083
$t(9)$	2.354	0.980	1.023
$t(12)$	2.510	0.964	0.990
$t(15)$	2.607	0.952	0.971
$t(20)$	2.707	0.937	0.953
$t(30)$	2.810	0.921	0.938

Table 3.1: *AREs for R (with Wilcoxon weights) to LAD, R (with Wilcoxon weights) to ML, and R (with Wilcoxon weights) to R (with van der Waerden weights) for the Laplace distribution, the logistic distribution, and the Students' t-distribution with several different degrees of freedom.*

weight function in the case of R-estimation for linear regression parameters (see, for example, Jurečková and Sen [15]). Note that, if Z_1 has a Laplace distribution, then $\lambda_f(s) \propto -I\{s < 1/2\} + I\{s > 1/2\}$ for $s \in (0, 1/2) \cup (1/2, 1)$ (λ_f does not exist at $s = 1/2$).

If Z_1 has a logistic distribution, then $f(s) = \pi/(\sqrt{3}\sigma) \exp(-s\pi/(\sqrt{3}\sigma))/[1 + \exp(-s\pi/(\sqrt{3}\sigma))]^2$, and so an optimal weight function λ_f is given by the Wilcoxon weight function $\lambda(s) = s - 1/2$. For the Wilcoxon weights, assumption A5 is satisfied when $\sigma^2 E\{f(Z_1)\} > E\{Z_1 F(Z_1)\}$, which holds for the Laplace, logistic, Students' t, and 0.4N(-1,1)+0.6N(2/3,3^2) distributions, as well as many others. Columns 2 and 3 of Table 3.1 give values of ARE for R (with Wilcoxon weights) to LAD and R (with Wilcoxon weights) to ML for a number of distributions. For the logistic and Students' t-distributions, R-estimation is asymptotically much more efficient than LAD and essentially as efficient as ML. Also, even though ML estimation is asymptotically 40% more efficient than R-estimation (with Wilcoxon weights) when the noise distribution is Laplace, R-estimation can still be useful in this case because $D(\cdot)$ tends to be smoother than $\sum_{t=1}^{n-p} |z_t(\cdot)|$ and hence easier to minimize. Figure 3.1 shows ML and R objective functions for a realization of length $n = 50$ from an all-pass model with $p = 1$, $\phi_{01} = 0.5$, and Laplace noise with variance one. Observe that the ML objective function has many local minima and thus could be difficult to minimize using numerical optimization techniques.
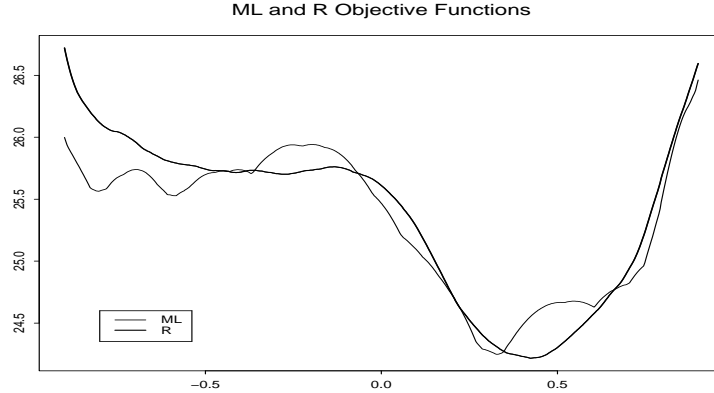
Figure 3.1: *ML and R (with Wilcoxon weights) objective functions for a realization of length $n = 50$ from an all-pass model with $p = 1$, $\phi_{01} = 0.5$, and Laplace noise with variance one.*

Remark 6: Another weight function commonly used for R-estimation is the van der Waerden weight function, $\lambda(s) = \Phi^{-1}(s)$, where Φ is the standard normal distribution function. Using results in Hallin [12], it can be shown that, if f is absolutely continuous and almost everywhere differentiable with $0 < \int_{-\infty}^{\infty} (f'(s))^2 / f(s) \, ds < \infty$, then A5 holds for the van der Waerden weights if and only if Z_1 is non-Gaussian. So, although A4 does not hold because Φ^{-1} is unbounded on $(0, 1)$, a bounded weight function approximating Φ^{-1} which does satisfy the assumptions can be found for a large class of non-Gaussian noise distributions. However, since Φ^{-1} is optimal when $Z_1 \sim N(0, \sigma^2)$ and the parameters of a Gaussian all-pass series are not identifiable, the van der Waerden weights are not particularly useful for all-pass parameter estimation. Column 4 of Table 3.1 gives the AREs for R (with Wilcoxon weights) to R (with van der Waerden weights) for various noise distributions. The van der Waerden weights are asymptotically superior to the Wilcoxon weights only when the distribution is close to Gaussian.

3.2 Order Selection

In practice, the true order r of an all-pass model is usually unknown and must be estimated. In this section, we give an order selection procedure that is analogous to using the partial autocorrelation function to identify

the order of an autoregressive model. First note that

$$\frac{\sigma^2 \tilde{J} - \tilde{K}^2}{2(\sigma^2 \tilde{L} - \tilde{K})^2} = \frac{\tilde{J} - \left(\frac{|\phi_{0r}|}{\sigma} \tilde{K}_z\right)^2}{2 \left(\frac{\sigma}{|\phi_{0r}|} \tilde{L}_z - \frac{|\phi_{0r}|}{\sigma} \tilde{K}_z\right)^2},$$

where $\tilde{K}_z := \int_0^1 F_z^{-1}(s)\lambda(s)\,ds$, $\tilde{L}_z := \int_0^1 f_z\left(F_z^{-1}(s)\right)\lambda'(s)\,ds$, and f_z and F_z are the density and distribution

functions respectively for $z_1 = \phi_{0r}^{-1} Z_1$. Because $\hat{\phi}_R \xrightarrow{P} \phi_0$,

$$\hat{s} := \left(\frac{1}{n} \sum_{t=1}^{n-p} z_t^2(\hat{\phi}_R)\right)^{1/2} \xrightarrow{P} \left(\mathrm{E}\{z_1^2\}\right)^{1/2} = \frac{\sigma}{|\phi_{0r}|} \tag{3.6}$$

and $\hat{K}_z := n^{-1} D(\hat{\phi}_R) \xrightarrow{P} \tilde{K}_z$ by Lemma 5.6 in the Appendix. Corollary 3.1 provides a consistent estimator

of $\tilde{L}_z$.

Corollary 3.1 *Consider the kernel density estimator of f_z*

$$\hat{f}_n(s) := \frac{1}{b_n n} \sum_{t=1}^{n-p} \kappa\left(\frac{s - z_t(\hat{\phi}_R)}{b_n}\right), \tag{3.7}$$

where κ is a uniformly continuous, differentiable kernel density function on $\mathbb{R}$ such that $\int |s \ln |s||^{1/2} |\kappa'(s)|\,ds <$

∞ and κ' is uniformly continuous on $\mathbb{R}$, and the bandwidth sequence $\{b_n\}$ is chosen so that $b_n \xrightarrow{P} 0$ and

$b_n^2 \sqrt{n} \xrightarrow{P} \infty$ as $n \to \infty$. If A1–A5 hold, then $\hat{L}_z := n^{-1} \sum_{t=1}^{n-p} \lambda'(t/(n-p)) \hat{f}_n(z_{(t)}(\hat{\phi}_R)) \xrightarrow{P} \tilde{L}_z$.

Proof: If $\hat{F}_n(s) := (n-p)^{-1} \sum_{t=1}^{n-p} I\{z_t(\hat{\phi}_R) \le s\}$, $\hat{F}_n^{-1}(s) := \inf\{x : \hat{F}_n(x) \ge s\}$, and

$$\lambda_n'(s) := \lambda'\left(\frac{t}{n-p}\right) \text{ for } s \in \left(\frac{t-1}{n-p}, \frac{t}{n-p}\right], \ t = 1, \ldots, n-p,$$

then $n\hat{L}_z/(n-p) = \int_0^1 \hat{f}_n(\hat{F}_n^{-1}(s))\lambda_n'(s)\,ds$. By the uniform continuity of λ', $\sup_{s \in (0,1)} |\lambda_n'(s) - \lambda'(s)| \to 0$.

Consequently, since $\sup_{s \in (0,1)} f_z(F_z^{-1}(s)) < \infty$ and $\sup_{s \in (0,1)} |\lambda'(s)| < \infty$, the proof is complete if

$$\sup_{s \in (0,1)} \left|\hat{f}_n\left(\hat{F}_n^{-1}(s)\right) - f_z\left(F_z^{-1}(s)\right)\right|$$
$$\le \sup_{s \in (0,1)} \left|\hat{f}_n\left(\hat{F}_n^{-1}(s)\right) - f_z\left(\hat{F}_n^{-1}(s)\right)\right| + \sup_{s \in (0,1)} \left|f_z\left(\hat{F}_n^{-1}(s)\right) - f_z\left(F_z^{-1}(s)\right)\right| \tag{3.8}$$

is $o_p(1)$. Because $\sup_{s \in \mathbb{R}} |\hat{f}_n(s) - f_z(s)| \xrightarrow{P} 0$ (for proof of this result, see Lemma 16 on page 88 of Andrews [3];

a similar result is given in Theorem 3 of Robinson [21]), the first term in (3.8) is $o_p(1)$. We now consider

the second term and use an argument similar to one found in the proof of Lemma 4 in Koul, Sievers, and

McKean [18]. Note that $\sup_{s\in(0,1)} |F_z(\hat{F}_n^{-1}(s)) - s| = \sup_{s\in\mathbb{R}} |\hat{F}_n(s) - F_z(s)|$, and, using the Glivenko-Cantelli theorem, it can be shown that $\sup_{s\in\mathbb{R}} |\hat{F}_n(s) - F_z(s)| \xrightarrow{P} 0$. Therefore, because $f_z\left(F_z^{-1}(\cdot)\right)$ is uniformly continuous on $(0,1)$ and $F_z^{-1}(F_z(s)) = s$ for all $s \in \mathbb{R}$ since F_z is strictly increasing on $\mathbb{R}$,

$$\sup_{s\in(0,1)} \left| f_z\left(\hat{F}_n^{-1}(s)\right) - f_z\left(F_z^{-1}(s)\right) \right| = \sup_{s\in(0,1)} \left| f_z\left(F_z^{-1}\left[F_z\left\{\hat{F}_n^{-1}(s)\right\}\right]\right) - f_z\left(F_z^{-1}(s)\right) \right| \xrightarrow{P} 0.$$

□

It follows that

$$\frac{\tilde{J} - (\hat{s}^{-1}\hat{K}_z)^2}{2(\hat{s}\hat{L}_z - \hat{s}^{-1}\hat{K}_z)^2} \xrightarrow{P} \frac{\sigma^2\tilde{J} - \tilde{K}^2}{2(\sigma^2\tilde{L} - \tilde{K})^2}. \tag{3.9}$$

Note that the Gaussian and the Students' t densities satisfy the conditions for the kernel density function κ in Corollary 3.1.

We now give the following corollary for use in order selection.

Corollary 3.2 *Assume A1–A5 hold. If the true order of the all-pass model is r and the order of the fitted model is $p > r$, then $n^{1/2}\hat{\phi}_{p,R} \xrightarrow{d} N(0, (\sigma^2\tilde{J} - \tilde{K}^2)/[2(\sigma^2\tilde{L} - \tilde{K})^2])$.*

Proof: By Problem 8.15 in Brockwell and Davis [7], the pth diagonal element of $\mathbf{\Gamma}_p^{-1}$ is σ^{-2} if $p > r$, and so the result follows from (3.1). □

A practical approach to order determination using a large sample follows:

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

2. Let the model order r be the smallest order beyond which the estimated coefficients are statistically insignificant; that is, $r = \min\{0 \le p \le P : |\hat{\phi}_{j,R}| < 1.96\,\hat{\tau}\,n^{-1/2}$ for $j > p\}$, where $\hat{\tau} := ([\tilde{J} - (\hat{s}^{-1}\hat{K}_z)^2]/[2(\hat{s}\hat{L}_z - \hat{s}^{-1}\hat{K}_z)^2])^{1/2}$ and the estimates $\hat{s}$, $\hat{K}_z$, and $\hat{L}_z$ are from the fitted Pth order model.

		Asymptotic	Empirical		
		std. dev.	mean	std. dev.	% coverage
n	mean	(LAD/Wilcoxon)	(LAD/Wilcoxon)	(LAD/Wilcoxon)	(LAD/Wilcoxon)
500	$\phi_1 = 0.5$	0.0275/0.0354	0.499/0.497	0.0332/0.0593	97.7/96.2
5000	$\phi_1 = 0.5$	0.0087/0.0112	0.500/0.499	0.0093/0.0112	97.9/96.0
500	$\phi_1 = 0.3$	0.0291/0.0374	0.299/0.299	0.0413/0.0444	96.5/94.9
	$\phi_2 = 0.4$	0.0291/0.0374	0.397/0.392	0.0479/0.0599	97.6/95.4
5000	$\phi_1 = 0.3$	0.0092/0.0118	0.300/0.300	0.0101/0.0122	97.6/95.2
	$\phi_2 = 0.4$	0.0092/0.0118	0.399/0.399	0.0099/0.0119	97.5/96.7

Table 4.1: *Empirical means, standard deviations, and percent coverages of nominal* 95% *confidence intervals for R-estimates of all-pass model parameters. The LAD-like score function* $\lambda(s) = 2\pi^{-1}\arctan(500(s-1/2))$ *and the Wilcoxon score function* $\lambda(s) = s - 1/2$ *were used. The noise distribution is Laplace with variance one.*

4   Numerical Results

4.1   Simulation Study

In this section, we give the results of a simulation study to assess the quality of the asymptotic approximations for finite samples. First, for each of 1000 replicates, we simulated all-pass data and found $\hat{\boldsymbol{\phi}}_R$ by minimizing D in (2.4). To reduce the possibility of the optimizer getting trapped at local minima, we chose 1000 random starting values for each replicate. We evaluated D at each of the 1000 candidate values and then reduced the collection of initial values to the twelve with the smallest values of D. Optimized values were found using these twelve initial values as starting points. The optimized value for which D was smallest was chosen to be $\hat{\boldsymbol{\phi}}_R$. Confidence intervals for the elements of $\boldsymbol{\phi}_0$ were constructed using (3.1) and the estimator in (3.9). For the kernel density estimator (3.7), we used the standard Gaussian kernel density function and, because of its recommendation in Silverman [22] (page 48), we used bandwidth $b_n = 0.9n^{-1/5}\min\{\hat{s}, IQR/1.34\}$, where $\hat{s}$, defined in (3.6), is the sample standard deviation for $\{z_t(\hat{\boldsymbol{\phi}}_R)\}$ and IQR is the interquartile range for $\{z_t(\hat{\boldsymbol{\phi}}_R)\}$.

Results of these simulations appear in Tables 4.1 and 4.2. We show the empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for the R-estimates of all-pass model

		Asymptotic	Empirical		
		std. dev.	mean	std. dev.	% coverage
n	mean	(LAD/Wilcoxon)	(LAD/Wilcoxon)	(LAD/Wilcoxon)	(LAD/Wilcoxon)
500	$\phi_1 = 0.5$	0.0327/0.0279	0.499/0.498	0.0405/0.0331	95.8/96.2
5000	$\phi_1 = 0.5$	0.0103/0.0088	0.500/0.500	0.0110/0.0090	95.2/95.6
500	$\phi_1 = 0.3$	0.0346/0.0296	0.301/0.299	0.0403/0.0366	95.1/94.7
	$\phi_2 = 0.4$	0.0346/0.0296	0.396/0.396	0.0418/0.0366	95.2/94.9
5000	$\phi_1 = 0.3$	0.0109/0.0093	0.300/0.300	0.0118/0.0095	94.0/95.4
	$\phi_2 = 0.4$	0.0109/0.0093	0.400/0.400	0.0115/0.0097	94.6/95.1

Table 4.2: *Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for R-estimates of all-pass model parameters. The LAD-like score function $\lambda(s) = 2\pi^{-1}\arctan(500(s-1/2))$ and the Wilcoxon score function $\lambda(s) = s - 1/2$ were used. The noise distribution is Students' t with three degrees of freedom.*

parameters. The LAD-like score function $\lambda(s) = 2\pi^{-1}\arctan(500(s-1/2))$ and the Wilcoxon score function $\lambda(s) = s-1/2$ were used. Asymptotic means and standard deviations were obtained using Theorem 3.1. Note that the R-estimates appear nearly unbiased and the confidence interval coverages are close to the nominal 95% level. The asymptotic standard deviations tend to understate the true variability of the estimates when $n = 500$, but are fairly accurate when $n = 5000$. Normal probability plots show that the R-estimates are approximately normal, particularly when $n = 5000$. The quality of the asymptotic approximations for finite samples is similar for LAD and ML estimates (see Breidt, Davis, and Trindade [6] and Andrews, Davis, and Breidt [2]).

We also ran simulations to assess the order selection procedure described in Section 3.2. For each of 100 replicates, we simulated all-pass data and estimated the model order r using the procedure in Section 3.2 with $P = 5$ and Wilcoxon scores. Table 4.3 gives the frequencies for each estimate of r. In all cases, the procedure appears to be fairly successful at identifying the true value of r. Model orders less than r were never selected, so underestimating r is clearly not a concern.

n	Model Parameters	Laplace Noise						t Noise					
		0	1	2	3	4	5	0	1	2	3	4	5
500	$\phi_0 = 0.5$	0	58	7	10	8	17	0	52	8	9	13	18
5000	$(r = 1)$	0	67	2	3	7	21	0	56	0	3	7	34
500	$\phi_0 = (0.3, 0.4)'$	0	0	69	1	16	14	0	0	57	13	16	14
5000	$(r = 2)$	0	0	82	5	5	8	0	0	81	5	6	8

Table 4.3: *The frequencies for each estimate of model order r when $P = 5$ and the Wilcoxon scores were used.*

4.2   Deconvolution

Applications for all-pass models are not limited to uncorrelated time series. As discussed in the Introduction and in Andrews, Davis, and Breidt [2], all-pass models can also be used to identify and model noncausal and noninvertible ARMA series. If, for example, a causal, invertible ARMA model is fit to a causal, noninvertible series, the residuals follow a causal all-pass model of order r, where r is the number of roots of the true moving average polynomial inside the unit circle. Therefore, the order of noninvertibility of the ARMA, r, can be determined by identifying the all-pass order of the residuals.

Consider the simulated water gun seismogram $\{X_t\}_{t=1}^{1000}$ shown in Figure 4.1(a), where $X_t = \sum_k \beta_k Z_{t-k}$, $\{\beta_k\}$ is the water gun wavelet sequence in Figure 8(2) of Lii and Rosenblatt [19], and $\{Z_t\}$ is a reflectivity sequence which was simulated as iid noise from the Students' t-distribution with five degrees of freedom. Andrews, Davis, and Breidt [2] modeled $\{X_t\}$ as a possibly noninvertible ARMA, using ML estimation for all-pass models to identify an appropriate order of noninvertibility. The wavelet and reflectivity sequences were then reconstructed from $\{X_t\}$ using the fitted ARMA model. This deconvolution procedure is of interest because, for an observed water gun seismogram, the reflectivity sequence is unknown and corresponds to reflection coefficients for layers of the earth. In this section, we identify an appropriate order of noninvertibility for $\{X_t\}$ using R-estimation for all-pass models, and we compare the R-estimation results to ML results in Andrews, Davis, and Breidt [2].

Andrews, Davis, and Breidt [2] first fit a causal, invertible ARMA(12,13) model $\phi(B)X_t = \theta(B)W_t$ to the simulated seismogram $\{X_t\}$ using Gaussian ML. The residuals from this fitted ARMA model are denoted
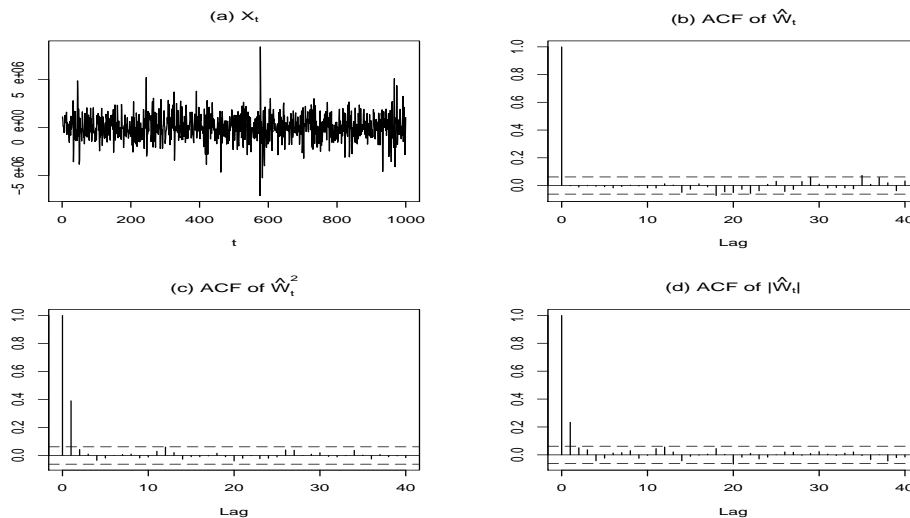
Figure 4.1: *(a) The simulated seismogram of length 1000, $\{X_t\}$, and the sample autocorrelation functions with bounds $\pm 1.96/\sqrt{1000}$ for (b) $\{\hat{W}_t\}$, (c) $\{\hat{W}_t^2\}$, and (d) $\{|\hat{W}_t|\}$.*

$\{\hat{W}_t\}$. From the sample autocorrelation functions for $\{\hat{W}_t\}$, $\{\hat{W}_t^2\}$, and $\{|\hat{W}_t|\}$ in Figure 4.1(b)–(d), it appears these ARMA residuals are uncorrelated but dependent, suggesting that a causal, invertible model is inappropriate for $\{X_t\}$. Using ML estimation and the Students' t-density, a causal all-pass model of order two was determined to be most suitable for $\{\hat{W}_t\}$ (Andrews, Davis, and Breidt [2]). The ML estimates of the all-pass model parameters are $\hat{\phi}_{ML} = (1.5286, -0.5908)'$, both with standard error 0.0338. Since the all-pass residuals appear independent, Andrews, Davis, and Breidt [2] concluded that a causal, noninvertible ARMA(12,13) with two roots of the moving average polynomial inside the unit circle is an appropriate model for $\{X_t\}$.

When the Wilcoxon weight function, the standard Gaussian kernel density function, and bandwidth $b_n = 0.9n^{-1/5}\min\{\hat{s}, IQR/1.34\}$ are used, the order selection procedure described in Section 3.2 also indicates that an all-pass model of order two is appropriate for $\{\hat{W}_t\}$. The R-estimates of the all-pass model parameters are $\hat{\phi}_R = (1.5052, -0.5700)'$, both with standard error 0.0343. In Figure 4.2, we show the sample autocorrelation functions for the squares and absolute values of $\{\hat{Z}_t\}$, the residuals from the all-pass model fit to $\{\hat{W}_t\}$ using R-estimation; these all-pass residuals appear independent. Therefore, in this example, the ML and R all-

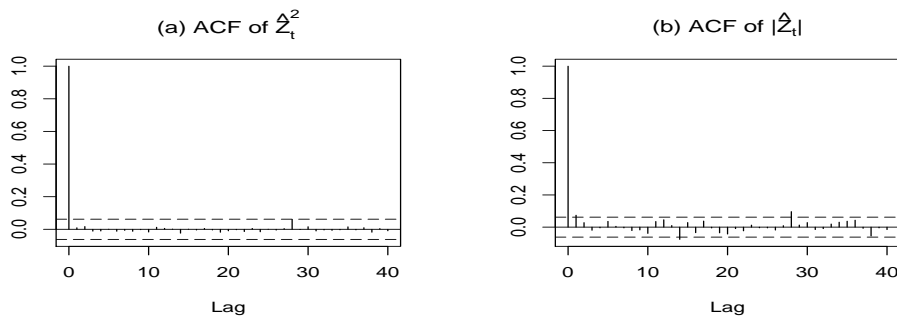(a) ACF of $\hat{Z}_t^2$ (b) ACF of $|\hat{Z}_t|$

Figure 4.2: *Diagnostics for the all-pass model of order two fit to the causal, invertible ARMA residuals using R-estimation. The sample autocorrelation functions with bounds $\pm 1.96/\sqrt{1000}$ for (a) $\{\hat{Z}_t^2\}$ and (b) $\{|\hat{Z}_t|\}$.*

pass estimation results are nearly identical, even though no specific distributional information was used for R-estimation.

Appendix

This section contains proofs of the lemmas used to establish the results of Section 3. Assume A1–A5 hold throughout. First, note that, for $j \in \{1, \ldots, p\}$ and $t \in \{1, \ldots, n - p\}$,

$$\frac{\partial z_t(\boldsymbol{\phi})}{\partial \phi_j} = \frac{1}{\phi(B^{-1})} \{X_{t+p-j} + z_{t+j}(\boldsymbol{\phi})\} \tag{5.1}$$

(see Andrews, Davis, and Breidt [2] for details). Evaluating (5.1) at the true value of $\boldsymbol{\phi}$ and ignoring the effect of recursion initialization, we have

$$\begin{aligned} \frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \phi_j} &= \frac{1}{\phi_0(B^{-1})} \left\{ \frac{-\phi_0(B^{-1}) B^p z_{t+p-j}}{\phi_0(B)} + z_{t+j}(\boldsymbol{\phi}_0) \right\} \\ &\simeq \frac{-z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}}{\phi_0(B^{-1})}, \end{aligned} \tag{5.2}$$

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 (5.2) is independent of $z_t = \phi_{0r}^{-1} Z_t$. Thus, if F_z is the distribution function of z_1 and $g_t(\boldsymbol{\phi}) := \lambda(F_z(z_t)) z_t(\boldsymbol{\phi})$, then,

for $j \in \{1, \ldots, p\}$,

$$\frac{\partial g_t(\boldsymbol{\phi}_0)}{\partial \phi_j} = \lambda(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \phi_j} \simeq \lambda(F_z(z_t))\left\{\frac{-z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}}{\phi_0(B^{-1})}\right\} =: \frac{\partial g_t^*(\boldsymbol{\phi}_0)}{\partial \phi_j}.$$

The expected value of $\partial g_t^*(\boldsymbol{\phi}_0)/\partial \phi_j$ is zero by the independence of its two terms.

We now compute the autocovariance function $\gamma^\dagger(h)$ of the zero-mean, stationary process $\{\mathbf{u}'\partial g_t^*(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi}\}$

for $\mathbf{u} \in \mathbb{R}^p$:

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

where

$$\nu_{jk}(h) := \begin{cases} 2\phi_{0r}^{-2}\tilde{J}\gamma(j-k), & \text{if } h = 0, \\[2mm] -\psi_{|h|-j}\psi_{|h|-k}\phi_{0r}^{-2}\tilde{K}^2, & \text{if } h \neq 0, \end{cases}$$

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

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

The preceding calculations lead directly to the following lemma.

Lemma 5.1 *As $n \to \infty$, $n^{-1/2}\sum_{t=1}^{n-p}\partial g_t(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi} \xrightarrow{d} \mathbf{N} \sim N(\mathbf{0}, 2\phi_{0r}^{-2}\{\sigma^2\tilde{J} - \tilde{K}^2\}\sigma^{-2}\boldsymbol{\Gamma}_p)$.*

Proof: Note that, for $t \in \{0, \ldots, n-p-1\}$,

$$z_{n-p-t} = \sum_{l=0}^\infty \psi_l\left(\phi_0(B^{-1})z_{n-p-t+l}\right) \quad \text{and} \quad z_{n-p-t}(\boldsymbol{\phi}_0) = \sum_{l=0}^t \psi_l\left(\phi_0(B^{-1})z_{n-p-t+l}\right). \tag{5.3}$$

Because there exist constants $c > 0$ and $0 < d < 1$ such that $|\psi_l| < cd^l$ for all $l \in \{0, 1, \ldots\}$ (see Brockwell

and Davis [7], Section 3.3),

$$\sum_{t=1}^{n-p}\mathrm{E}\left|\frac{\partial g_t(\boldsymbol{\phi}_0)}{\partial\phi_j} - \frac{\partial g_t^*(\boldsymbol{\phi}_0)}{\partial\phi_j}\right| = \sum_{t=1}^{n-p}\mathrm{E}\left|\lambda(F_z(z_t))\left\{\frac{z_{t+j}(\boldsymbol{\phi}_0)}{\phi_0(B^{-1})} - \frac{z_{t+j}}{\phi_0(B^{-1})}\right\}\right| = O(1)$$

for $j \in \{1, \ldots, p\}$. Consequently, $n^{-1/2}\sum_{t=1}^{n-p}[\partial g_t(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi} - \partial g_t^*(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi}] \to \mathbf{0}$ in L_1 and hence in probability.

Let $\mathbf{u} \in \mathbb{R}^p$. By the Cramér-Wold device, it suffices to show $n^{-1/2}\sum_{t=1}^{n-p}\mathbf{u}'\partial g_t^*(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi} \xrightarrow{d} \mathbf{u}'\mathbf{N} \sim$

$N(0, 2\phi_{0r}^{-2}\{\sigma^2\tilde{J} - \tilde{K}^2\}\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u})$. Elements of the infinite order moving average stationary sequence

$\{\mathbf{u}'\partial g_t^*(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi}\}$ 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 $g_t(\boldsymbol{\phi})$. For $j, k \in \{1, \ldots, p\}$,

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

and so

$$
\begin{aligned}
\frac{\partial^2 g_t(\boldsymbol{\phi}_0)}{\partial\phi_j\partial\phi_k} &= \lambda(F_z(z_t))\frac{\partial^2 z_t(\boldsymbol{\phi}_0)}{\partial\phi_j\partial\phi_k} \\
&\simeq \lambda(F_z(z_t))\left\{-\sum_{m=0}^{\infty}\sum_{\ell=0}^{\infty}\psi_m\psi_\ell(z_{t+j-k-\ell+m} + z_{t+k-j-\ell+m}) + \frac{2z_{t+j+k}}{\phi_0^2(B^{-1})}\right\} =: \frac{\partial^2 g_t^*(\boldsymbol{\phi}_0)}{\partial\phi_j\partial\phi_k}. \quad (5.4)
\end{aligned}
$$

(5.4) has expectation $-2\sigma^{-2}\gamma(j-k)\int_0^1 F_z^{-1}(s)\lambda(s)\,ds = -2|\phi_{0r}|^{-1}\tilde{K}\sigma^{-2}\gamma(j-k)$.

Lemma 5.2 *As* $n \to \infty$, $n^{-1}\sum_{t=1}^{n-p}\partial^2 g_t(\boldsymbol{\phi}_0)/(\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}') \xrightarrow{P} -2|\phi_{0r}|^{-1}\tilde{K}\sigma^{-2}\boldsymbol{\Gamma}_p$.

Proof: It can be shown that $n^{-1}\sum_{t=1}^{n-p}[\partial^2 g_t(\boldsymbol{\phi}_0)/(\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}') - \partial^2 g_t^*(\boldsymbol{\phi}_0)/(\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}')] \to \mathbf{0}$ in L_1 and probability. Because (5.4) has expectation $-2|\phi_{0r}|^{-1}\tilde{K}\sigma^{-2}\gamma(j-k)$, $n^{-1}\sum_{t=1}^{n-p}\partial^2 g_t^*(\boldsymbol{\phi}_0)/(\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}') \xrightarrow{P} -2|\phi_{0r}|^{-1}\tilde{K}\sigma^{-2}\boldsymbol{\Gamma}_p$ by the ergodic theorem. □

Lemma 5.3 *For any* $T \in (0, \infty)$, *as* $n \to \infty$,

$$
\sup_{\|\mathbf{u}\|\le T}\left|n^{-1/2}\sum_{t=1}^{n-p}\mathbf{u}'\left[\lambda\left(\frac{R_t\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1}\right) - \lambda(F_z(z_t))\right]\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial\boldsymbol{\phi}} - 2|\phi_{0r}|^{-1}\tilde{L}\mathbf{u}'\boldsymbol{\Gamma}_p\mathbf{u}\right| \xrightarrow{P} 0. \quad (5.5)
$$

Proof: Observe that the left-hand side of (5.5) is bounded above by

$$
\sup_{\|\mathbf{u}\|\le T}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n-p}\mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial\boldsymbol{\phi}}\left[\frac{R_t\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1} - F_z(z_t)\right] - 2|\phi_{0r}|^{-1}\tilde{L}\mathbf{u}'\boldsymbol{\Gamma}_p\mathbf{u}\right| \quad (5.6)
$$

$$
+ \sup_{\|\mathbf{u}\|\le T}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n-p}\mathbf{u}'\left[\lambda'(F_{t,n}^*(\mathbf{u})) - \lambda'(F_z(z_t))\right]\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial\boldsymbol{\phi}}\left[\frac{R_t\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1} - F_z(z_t)\right]\right|, \quad (5.7)
$$

where $F^*_{t,n}(\mathbf{u})$ is between $F_z(z_t)$ and $R_t(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u})/(n-p+1)$. If $F_n(x) := n^{-1}\sum_{t=1}^{n-p} I\{z_t \leq x\}$, an upper

bound for (5.6) is

$$\sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \left[\frac{R_t\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1} - F_n\left(z_t\left(\boldsymbol{\phi}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right)\right) \right] \right| \tag{5.8}$$

$$+ \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \left[F_n\left(z_t\left(\boldsymbol{\phi}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right)\right) - F_z\left(z_t\left(\boldsymbol{\phi}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right)\right) \right] \right| \tag{5.9}$$

$$+ \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \left[F_z\left(z_t\left(\boldsymbol{\phi}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right)\right) - F_z(z_t) \right] - 2|\phi_{0r}|^{-1}\tilde{L}\mathbf{u}'\boldsymbol{\Gamma}_p\mathbf{u} \right|. \tag{5.10}$$

Because

$$\sup_{\|\mathbf{u}\| \leq T} \frac{1}{n} \sum_{t=1}^{n-p} \left| \mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \right| = O_p(1)$$

and, by Lemma 3 on page 55 of Andrews [3],

$$\sup_{\|\mathbf{u}\| \leq T, t \in \{1,\ldots,n-p\}} \sqrt{n} \left| \frac{R_t\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}\right)}{n-p+1} - F_n\left(z_t\left(\boldsymbol{\phi}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right)\right) \right| \xrightarrow{P} 0, \tag{5.11}$$

(5.8) is $o_p(1)$. Lemma 10 on page 76 of Andrews [3] establishes that (5.9) is $o_p(1)$. Finally,

$$\sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \left[F_z\left(z_t\left(\boldsymbol{\phi}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right)\right) - F_z(z_t) \right] - 2|\phi_{0r}|^{-1}\tilde{L}\mathbf{u}'\boldsymbol{\Gamma}_p\mathbf{u} \right|$$

$$= \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} f_z(z^*_{t,n}(\mathbf{u})) \left(z_t\left(\boldsymbol{\phi}_0 + \frac{\mathbf{u}}{\sqrt{n}}\right) - z_t \right) - 2|\phi_{0r}|^{-1}\tilde{L}\mathbf{u}'\boldsymbol{\Gamma}_p\mathbf{u} \right|$$

$$\leq \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}'\lambda'(F_z(z_t))\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} f_z(z^*_{t,n}(\mathbf{u}))(z_t(\boldsymbol{\phi}_0) - z_t) \right|$$

$$+ \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{n} \sum_{t=1}^{n-p} \lambda'(F_z(z_t))f_z(z^*_{t,n}(\mathbf{u})) \left(\mathbf{u}'\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \right)^2 - 2|\phi_{0r}|^{-1}\tilde{L}\mathbf{u}'\boldsymbol{\Gamma}_p\mathbf{u} \right|$$

$$+ \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{2n\sqrt{n}} \sum_{t=1}^{n-p} \lambda'(F_z(z_t))f_z(z^*_{t,n}(\mathbf{u}))\mathbf{u}'\frac{\partial z_t(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}}\mathbf{u}'\frac{\partial^2 z_t(\boldsymbol{\phi}^*_{t,n}(\mathbf{u}))}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'}\mathbf{u} \right|,$$

where f_z is the density function for z_1, $z^*_{t,n}(\mathbf{u})$ is between z_t and $z_t(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u})$, and $\boldsymbol{\phi}^*_{t,n}(\mathbf{u})$ is between

$\boldsymbol{\phi}_0$ and $\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}$. From (5.3), the first term on the right-hand side is $o_p(1)$, and, since there exists a

geometrically decaying, non-negative, real-valued sequence $\{\ddot{\pi}_k\}_{k=-\infty}^{\infty}$ such that

$$\sup_{\|\mathbf{u}\| \leq T} \left| \mathbf{u}'\frac{\partial^2 z_t(\boldsymbol{\phi}^*_{t,n}(\mathbf{u}))}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'}\mathbf{u} \right| \leq \sum_{k=-\infty}^{\infty} \ddot{\pi}_k|z_{t-k}| \quad \forall t \in \{1,\ldots,n-p\}$$

for all n sufficiently large (Brockwell and Davis [7], Section 3.3), the third term is also $o_p(1)$. Using the uniform continuity of f_z, the second term equals

$$\sup_{\|\mathbf{u}\|\leq T}\left|\frac{1}{n}\sum_{t=1}^{n-p}\lambda'(F_z(z_t))f_z(z_t)\left(\mathbf{u}'\frac{\partial z_t(\phi_0)}{\partial\phi}\right)^2 - 2\phi_{0r}^{-2}\left(\int_0^1 f_z(F_z^{-1}(s))\lambda'(s)\,ds\right)\mathbf{u}'\mathbf{\Gamma}_p\mathbf{u}\right| + o_p(1),$$

which is $o_p(1)$ by the ergodic theorem. Therefore, (5.10) and consequently (5.6) are $o_p(1)$. Similarly, using the uniform continuity of λ', it can be shown that (5.7) is $o_p(1)$. $\qquad\square$

Lemma 5.4 *For any $T \in (0,\infty)$, as $n \to \infty$,*

$$\sup_{\|\mathbf{u}\|,\|\mathbf{v}\|\leq T}\left|n^{-1}\sum_{t=1}^{n-p}\mathbf{u}'\left[\lambda\left(\frac{R_t\left(\phi_0+n^{-1/2}\mathbf{v}\right)}{n-p+1}\right)-\lambda(F_z(z_t))\right]\frac{\partial^2 z_t(\phi_0)}{\partial\phi\partial\phi'}\mathbf{u}\right|\xrightarrow{P} 0.$$

Proof: Note that, because F_z is a continuous distribution function, it is also uniformly continuous on $\mathbb{R}$. Using (5.11), the Glivenko-Cantelli theorem, and the uniform continuity of F_z, for any $\epsilon,\eta > 0$, it can be shown that there exists an integer m such that

$$\mathrm{P}\left(\sup_{\|\mathbf{v}\|\leq T,t\in\{1,\ldots,n-p-m\}}\left|\frac{R_t\left(\phi_0+n^{-1/2}\mathbf{v}\right)}{n-p+1}-F_z(z_t)\right|>\eta\right)$$

$$\leq\ \mathrm{P}\left(\sup_{\|\mathbf{v}\|\leq T,t\in\{1,\ldots,n-p\}}\left|\frac{R_t\left(\phi_0+n^{-1/2}\mathbf{v}\right)}{n-p+1}-F_n\left(z_t\left(\phi_0+\frac{\mathbf{v}}{\sqrt{n}}\right)\right)\right|>\frac{\eta}{3}\right)$$

$$+\mathrm{P}\left(\sup_{x\in\mathbb{R}}|F_n(x)-F_z(x)|>\frac{\eta}{3}\right)+\mathrm{P}\left(\sup_{\|\mathbf{v}\|\leq T,t\in\{1,\ldots,n-p-m\}}\left|F_z\left(z_t\left(\phi_0+\frac{\mathbf{v}}{\sqrt{n}}\right)\right)-F_z(z_t)\right|>\frac{\eta}{3}\right)$$

is less than ϵ for all n sufficiently large. Hence,

$$\sup_{\|\mathbf{u}\|,\|\mathbf{v}\|\leq T}\left|\frac{1}{n}\sum_{t=1}^{n-p}\mathbf{u}'\left[\lambda\left(\frac{R_t\left(\phi_0+n^{-1/2}\mathbf{v}\right)}{n-p+1}\right)-\lambda(F_z(z_t))\right]\frac{\partial^2 z_t(\phi_0)}{\partial\phi\partial\phi'}\mathbf{u}\right|$$

$$=\sup_{\|\mathbf{u}\|,\|\mathbf{v}\|\leq T}\left|\frac{1}{n}\sum_{t=1}^{n-p}\mathbf{u}'\lambda'(F_{t,n}^*(\mathbf{v}))\frac{\partial^2 z_t(\phi_0)}{\partial\phi\partial\phi'}\left[\frac{R_t\left(\phi_0+n^{-1/2}\mathbf{v}\right)}{n-p+1}-F_z(z_t)\right]\mathbf{u}\right|$$

$$\xrightarrow{P}\ 0,$$

where $F_{t,n}^*(\mathbf{v})$ is between $F_z(z_t)$ and $R_t(\phi_0+n^{-1/2}\mathbf{v})/(n-p+1)$, since

$$\sup_{\|\mathbf{u}\|,\|\mathbf{v}\|\leq T}\frac{1}{n}\sum_{t=1}^{n-p}\left|\mathbf{u}'\lambda'(F_{t,n}^*(\mathbf{v}))\frac{\partial^2 z_t(\phi_0)}{\partial\phi\partial\phi'}\mathbf{u}\right|=O_p(1).$$

□

For $\mathbf{u} \in \mathbb{R}^p$ and $\delta_1, \delta_2 \in [0, 1]$, let

$$U_n(\mathbf{u}, \delta_1, \delta_2) = \sum_{t=1}^{n-p} \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_1 \mathbf{u} \right)}{n - p + 1} \right) \left[z_t \left(\phi_0 + \frac{\delta_2 \mathbf{u}}{\sqrt{n}} \right) - z_t \left(\phi_0 + \frac{\delta_1 \mathbf{u}}{\sqrt{n}} \right) \right]$$

and

$$V_n(\mathbf{u}, \delta_1, \delta_2) = \sum_{t=1}^{n-p} \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_2 \mathbf{u} \right)}{n - p + 1} \right) \left[z_t \left(\phi_0 + \frac{\delta_2 \mathbf{u}}{\sqrt{n}} \right) - z_t \left(\phi_0 + \frac{\delta_1 \mathbf{u}}{\sqrt{n}} \right) \right].$$

Using Taylor series expansions,

$$\begin{aligned}
U_n(\mathbf{u}, \delta_1, \delta_2) &= \sum_{t=1}^{n-p} \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_1 \mathbf{u} \right)}{n - p + 1} \right) \left\{ \left[z_t \left(\phi_0 + \frac{\delta_2 \mathbf{u}}{\sqrt{n}} \right) - z_t(\phi_0) \right] - \left[z_t \left(\phi_0 + \frac{\delta_1 \mathbf{u}}{\sqrt{n}} \right) - z_t(\phi_0) \right] \right\} \\
&= \frac{\delta_2 - \delta_1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_1 \mathbf{u} \right)}{n - p + 1} \right) \frac{\partial z_t(\phi_0)}{\partial \phi} \\
&\quad + \frac{1}{2} \frac{\delta_2^2 - \delta_1^2}{n} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_1 \mathbf{u} \right)}{n - p + 1} \right) \frac{\partial^2 z_t(\phi_0)}{\partial \phi \partial \phi'} \mathbf{u} \\
&\quad + \frac{1}{2} \frac{\delta_2^2}{n} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_1 \mathbf{u} \right)}{n - p + 1} \right) \left[\frac{\partial^2 z_t(\phi_n^*(\mathbf{u}, \delta_1, \delta_2))}{\partial \phi \partial \phi'} - \frac{\partial^2 z_t(\phi_0)}{\partial \phi \partial \phi'} \right] \mathbf{u} \\
&\quad - \frac{1}{2} \frac{\delta_1^2}{n} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_1 \mathbf{u} \right)}{n - p + 1} \right) \left[\frac{\partial^2 z_t(\phi_n^*(\mathbf{u}, \delta_1, \delta_1))}{\partial \phi \partial \phi'} - \frac{\partial^2 z_t(\phi_0)}{\partial \phi \partial \phi'} \right] \mathbf{u} \quad (5.12)
\end{aligned}$$

and, similarly,

$$\begin{aligned}
V_n(\mathbf{u}, \delta_1, \delta_2) &= \frac{\delta_2 - \delta_1}{\sqrt{n}} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_2 \mathbf{u} \right)}{n - p + 1} \right) \frac{\partial z_t(\phi_0)}{\partial \phi} \\
&\quad + \frac{1}{2} \frac{\delta_2^2 - \delta_1^2}{n} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_2 \mathbf{u} \right)}{n - p + 1} \right) \frac{\partial^2 z_t(\phi_0)}{\partial \phi \partial \phi'} \mathbf{u} \\
&\quad + \frac{1}{2} \frac{\delta_2^2}{n} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_2 \mathbf{u} \right)}{n - p + 1} \right) \left[\frac{\partial^2 z_t(\phi_n^*(\mathbf{u}, \delta_2, \delta_2))}{\partial \phi \partial \phi'} - \frac{\partial^2 z_t(\phi_0)}{\partial \phi \partial \phi'} \right] \mathbf{u} \\
&\quad - \frac{1}{2} \frac{\delta_1^2}{n} \sum_{t=1}^{n-p} \mathbf{u}' \lambda \left(\frac{R_t \left(\phi_0 + n^{-1/2} \delta_2 \mathbf{u} \right)}{n - p + 1} \right) \left[\frac{\partial^2 z_t(\phi_n^*(\mathbf{u}, \delta_2, \delta_1))}{\partial \phi \partial \phi'} - \frac{\partial^2 z_t(\phi_0)}{\partial \phi \partial \phi'} \right] \mathbf{u}, \quad (5.13)
\end{aligned}$$

where the values of $\phi_n^*(\mathbf{u}, \cdot, \cdot)$ lie between ϕ_0 and $\phi_0 + n^{-1/2} \mathbf{u}$.

Lemma 5.5 *For* $\mathbf{u} \in \mathbb{R}^p$, *let* $S_n(\mathbf{u}) = D(\phi_0 + n^{-1/2} \mathbf{u}) - D(\phi_0)$ *and* $S(\mathbf{u}) = \mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}(\sigma^2 \tilde{L} - \tilde{K}) \mathbf{u}' \sigma^{-2} \mathbf{\Gamma}_p \mathbf{u}$, *where* $\mathbf{N} \sim N(\mathbf{0}, 2\phi_{0r}^{-2}\{\sigma^2 \tilde{J} - \tilde{K}^2\} \sigma^{-2} \mathbf{\Gamma}_p)$. *Then* $S_n(\cdot) \xrightarrow{d} S(\cdot)$ *on* $C(\mathbb{R}^p)$, *the space of continuous functions on* $\mathbb{R}^p$ *where convergence is equivalent to uniform convergence on every compact set.*

Proof: Let $\mathbf{u} \in \mathbb{R}^p$ and suppose m is any positive integer. Because

$$D\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}\right) - D(\boldsymbol{\phi}_0) = \sum_{k=1}^{m}\left[D\left(\boldsymbol{\phi}_0 + \frac{k\mathbf{u}}{m\sqrt{n}}\right) - D\left(\boldsymbol{\phi}_0 + \frac{(k-1)\mathbf{u}}{m\sqrt{n}}\right)\right],$$

we have

$$\sum_{k=1}^{m} U_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right) \leq D\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}\right) - D(\boldsymbol{\phi}_0) \leq \sum_{k=1}^{m} V_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right) \tag{5.14}$$

by Theorem 2.1. Using (5.12), (5.13), and Lemmas 5.1, 5.2, 5.3, and 5.4,

$$
\begin{bmatrix}
U_n\left(\mathbf{u}, 0, \frac{1}{m}\right) \\[4pt]
U_n\left(\mathbf{u}, \frac{1}{m}, \frac{2}{m}\right) \\[4pt]
\vdots \\[4pt]
U_n\left(\mathbf{u}, \frac{m-1}{m}, 1\right) \\[4pt]
V_n\left(\mathbf{u}, 0, \frac{1}{m}\right) \\[4pt]
V_n\left(\mathbf{u}, \frac{1}{m}, \frac{2}{m}\right) \\[4pt]
\vdots \\[4pt]
V_n\left(\mathbf{u}, \frac{m-1}{m}, 1\right)
\end{bmatrix}
\xrightarrow{d}
\begin{bmatrix}
\frac{1}{m}\mathbf{u}'\mathbf{N} - |\phi_{0r}|^{-1}\left[\left(\frac{1}{m}\right)^2 - \left(\frac{0}{m}\right)^2\right]\tilde{K}\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u} \\[4pt]
\frac{1}{m}\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(2\frac{1}{m^2}\sigma^2\tilde{L} - \left[\left(\frac{2}{m}\right)^2 - \left(\frac{1}{m}\right)^2\right]\tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u} \\[4pt]
\vdots \\[4pt]
\frac{1}{m}\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(2\frac{m-1}{m^2}\sigma^2\tilde{L} - \left[\left(\frac{m}{m}\right)^2 - \left(\frac{m-1}{m}\right)^2\right]\tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u} \\[4pt]
\frac{1}{m}\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(2\frac{1}{m^2}\sigma^2\tilde{L} - \left[\left(\frac{1}{m}\right)^2 - \left(\frac{0}{m}\right)^2\right]\tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u} \\[4pt]
\frac{1}{m}\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(2\frac{2}{m^2}\sigma^2\tilde{L} - \left[\left(\frac{2}{m}\right)^2 - \left(\frac{1}{m}\right)^2\right]\tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u} \\[4pt]
\vdots \\[4pt]
\frac{1}{m}\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(2\frac{m}{m^2}\sigma^2\tilde{L} - \left[\left(\frac{m}{m}\right)^2 - \left(\frac{m-1}{m}\right)^2\right]\tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u}
\end{bmatrix}
$$

on $\mathbb{R}^{2m}$ since

$$\sup_{\|\mathbf{u}\|, \|\mathbf{v}\| \leq T} \frac{1}{n}\sum_{t=1}^{n-p}\left|\mathbf{u}'\left(\frac{\partial^2 z_t\left(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{v}\right)}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'} - \frac{\partial^2 z_t\left(\boldsymbol{\phi}_0\right)}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'}\right)\mathbf{u}\right| \xrightarrow{P} 0$$

for any $T > 0$ and $\sup_{s \in (0,1)} |\lambda(s)| < \infty$ by the uniform continuity of λ'. Hence,

$$
\begin{bmatrix}
\sum_{k=1}^{m} U_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right) \\[4pt]
\sum_{k=1}^{m} V_n\left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}\right)
\end{bmatrix}
\xrightarrow{d}
\begin{bmatrix}
\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(\frac{m-1}{m}\sigma^2\tilde{L} - \tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u} \\[4pt]
\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(\frac{m+1}{m}\sigma^2\tilde{L} - \tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u}
\end{bmatrix}
$$

on $\mathbb{R}^2$. For any $\epsilon > 0$, there exists an integer m large enough so that

$$\mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(\frac{m-1}{m}\sigma^2\tilde{L} - \tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u} \qquad \text{and} \qquad \mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}\left(\frac{m+1}{m}\sigma^2\tilde{L} - \tilde{K}\right)\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u}$$

are both in an ϵ-neighborhood of $S(\mathbf{u}) = \mathbf{u}'\mathbf{N} + |\phi_{0r}|^{-1}(\sigma^2\tilde{L} - \tilde{K})\mathbf{u}'\sigma^{-2}\boldsymbol{\Gamma}_p\mathbf{u}$. Thus, for any $\mathbf{u} \in \mathbb{R}^p$, $S_n(\mathbf{u}) \xrightarrow{d} S(\mathbf{u})$. It can be shown similarly that all finite-dimensional distributions of $S_n(\cdot)$ converge to those of $S(\cdot)$.

Also using (5.14), it can be shown that $\lim_{\delta \to 0^+} \limsup_{n \to \infty} \mathrm{P}(\sup_{\mathbf{u}, \mathbf{v} \in K, \|\mathbf{u}-\mathbf{v}\| \leq \delta} |S_n(\mathbf{u}) - S_n(\mathbf{v})| > \eta) = 0$

for any $\eta > 0$ and any compact subset $K \subset \mathbb{R}^p$ (see Andrews [3], pages 84-86). It follows that $S_n(\cdot)$ must be

tight on $C(K)$ and, therefore, because compact $K \subset \mathbb{R}^p$ is arbitrary, $S_n(\cdot) \overset{d}{\to} S(\cdot)$ on $C(\mathbb{R}^p)$ by Theorem 7.1

in Billingsley [4]. $\qquad \square$

Lemma 5.6 *If $\epsilon > 0$ is sufficiently small so that ϕ forms a causal polynomial for all $\phi \in \Phi := \{\phi \in \mathbb{R}^p :$*

$\|\phi - \phi_0\| \leq \epsilon\}$, then $n^{-1} \sum_{t=1}^{n-p} z_t^2(\phi) \overset{a.s.}{\to} E\{\tilde{z}_1^2(\phi)\}$ and $n^{-1}D(\phi) \overset{a.s.}{\to} \int_0^1 F_{\tilde{z}(\phi)}^{-1}(s)\lambda(s)\,ds$ uniformly on Φ,

where $\tilde{z}_t(\phi) := -\phi^{-1}(B^{-1})\phi(B)X_{t+p}$ and $F_{\tilde{z}(\phi)}(\cdot)$ is the distribution function for $\tilde{z}_1(\phi)$.

Proof: For any $\phi \in \Phi$, $n^{-1} \sum_{t=1}^{n-p} z_t^2(\phi) \overset{a.s.}{\to} E\{\tilde{z}_1^2(\phi)\}$ and $n^{-1}D(\phi) \overset{a.s.}{\to} E\{\lambda(F_{\tilde{z}(\phi)}(\tilde{z}_1(\phi)))\tilde{z}_1(\phi)\} =$

$\int_0^1 F_{\tilde{z}(\phi)}^{-1}(s)\lambda(s)\,ds$ by the ergodic theorem. Therefore, since $n^{-1} \sum_{t=1}^{n-p} z_t^2(\cdot)$ and $n^{-1}D(\cdot)$ are equicontin-

uous and uniformly bounded on Φ almost surely (see Lemma 15 on page 86 of Andrews [3]; similar results

are obtained in the proof of Proposition 1 in Breidt, Davis, and Trindade [6]), the lemma follows by the

Arzelà-Ascoli theorem. $\qquad \square$

References

[1] Allal, J., Kaaouachi, A., and Paindaveine, D. (2001). R-estimation for ARMA models. *J. Nonparametr. Stat.* **13**

815–831.

[2] Andrews, B., Davis, R.A., and Breidt, F.J. (2006). Maximum likelihood estimation for all-pass time series models.

J. Multivariate Anal. **97** 1638–1659.

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

[4] Billingsley, P. (1999). *Convergence of Probability Measures,* 2nd ed. Wiley, New York.

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

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

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

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

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

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

[11] Giannakis, G.B. and Swami, A. (1990). On estimating noncausal nonminimum phase ARMA models of non-Gaussian processes. *IEEE Trans. Acoust. Speech Signal Process.* **38** 478–495.

[12] Hallin, M. (1994). On the Pitman non-admissibility of correlogram-based methods. *J. Time Ser. Anal.* **15** 607–611.

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

[14] Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449–1458.

[15] Jurečková, J. and Sen, P.K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* Wiley, New York.

[16] Koul, H.L. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.* **22** 540–562.

[17] Koul, H.L. and Saleh, A.K.Md.E. (1993). R-estimation of the parameters of autoregressive [Ar(p)] models. *Ann. Statist.* **21** 534–551.

[18] Koul, H.L., Sievers, G.L., and McKean, J.W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Statist.* **14** 131–141.

[19] Lii, K.-S. and Rosenblatt, M. (1988). Nonminimum phase non-Gaussian deconvolution. *J. Multivariate Anal.* **27** 359–374.

[20] Mukherjee, K. and Bai, Z.D. (2002). R-estimation in autoregression with square-integrable score function. *J. Multivariate Anal.* **81** 167–186.

[21] Robinson, P.M. (1987). Time series residuals with application to probability density estimation. *J. Time Ser. Anal.* **8** 329–344.

[22] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

[23] Terpstra, J.T., McKean, J.W., and Naranjo, J.D. (2001). Weighted Wilcoxon estimates for autoregression. *Aust. N. Z. J. Stat.* **43** 399–419.