

Rank-Based Estimation for Autoregressive Moving Average Time Series Models

Beth Andrews
Northwestern University

February 6, 2007

Abstract

We establish asymptotic normality and consistency for rank-based estimators of autoregressive-moving average model parameters. The estimators are obtained by minimizing a rank-based residual dispersion function similar to the one given 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 quality of the asymptotic approximations for finite samples is studied via simulation.

AMS 2000 subject classifications. Primary 62M10; secondary 62E20, 62F10.

Key words and phrases. Autoregressive moving average models, rank estimation.

1 Introduction

In this paper, we use a rank-based technique to estimate the parameters of autoregressive-moving average (ARMA) models, the standard linear time series models for stationary data. The rank (R) estimators we consider minimize the sum of mean-corrected model residuals weighted by a function of residual rank; they are similar to the R-estimators proposed by Jaeckel (1972) for estimating linear regression parameters. As discussed in Jurečková and Sen (1996), R-estimators are, in general, robust and relatively efficient. We show this is true in the case of ARMA parameter estimation. The estimation technique is robust because the R-estimators of ARMA model parameters are consistent and asymptotically normal under very mild conditions on the noise distribution. Since the weight function can be chosen so that R-estimation has the same asymptotic efficiency as maximum likelihood (ML) estimation, the R-estimators are also relatively efficient. The relative efficiency of the estimators extends to the unknown noise distribution case since R-estimation with the Wilcoxon weight function (a linear weight function) is nearly as efficient as ML for a large class of noise distributions. In addition, R-estimation compares favorably with classical Gaussian quasi-ML estimation and least absolute deviations (LAD) estimation. We show that R-estimation with the van der Waerden weight function (the inverse of the standard normal distribution function) uniformly dominates Gaussian quasi-ML with respect to asymptotic efficiency, and we give a weight function for which R-estimation is asymptotically equivalent to LAD estimation.

Because the R-objective function we use involves not only the residual ranks but also the residual values, this is not pure R-estimation. Koul and Saleh (1993), Koul and Ossiander (1994), and Mukherjee and Bai (2002) consider related rank-based estimation approaches for autoregressive (AR) model parameters. In this paper, we consider R-estimation for ARMA models, which can have both an AR component and a moving average (MA) component. Extending the AR results to ARMA models is not straightforward, particularly since the objective function is convex only in the AR case. Rank-based estimation for ARMA models is examined in Ferretti, Kelmansky, and Yohai (1991), but the authors are only able to establish asymptotic normality for the estimators when the model is AR. Also, Allal, Kaaouachi, and Paindaveine (2001) examine

pure R-estimation for ARMA model parameters based on correlations of weighted residual ranks. These R-estimators of ARMA parameters are consistent and asymptotically normal, but under more stringent conditions on the noise distribution than those given in this paper. For instance, Allal et al. require the density function for the noise to be strongly unimodal, which does not hold for the Student's t or multimodal distributions.

The R-objective function we use is given in Section 2. In Section 3, asymptotic normality for R-estimators of ARMA model parameters is established under mild conditions, and the asymptotic efficiency for R-estimation is compared to the efficiencies for other estimation techniques. We also give a consistent estimator for the covariance matrix of the limiting normal distribution. Proofs of the lemmas used in Section 3 are relegated to the Appendix. The behavior of the R-estimators for finite samples is studied via simulation in Section 4.

2 Preliminaries

Let B denote the backshift operator ($B^k X_t = X_{t-k}$, $k \in \{0, \pm 1, \pm 2, \dots\}$) and suppose $\{X_t\}$ is the ARMA process which satisfies the difference equations

$$\phi_0(B)X_t = \theta_0(B)Z_t, \quad (2.1)$$

where the AR polynomial $\phi_0(z) := 1 - \phi_{01}z - \dots - \phi_{0p}z^p \neq 0$ for $|z| \leq 1$, the MA polynomial $\theta_0(z) := 1 + \theta_{01}z + \dots + \theta_{0q}z^q \neq 0$ for $|z| \leq 1$, $\phi_0(z)$ and $\theta_0(z)$ have no common roots, $\phi_0(z) \neq 1$ or $\theta_0(z) \neq 1$, and $\{Z_t\}$ is an independent and identically distributed (iid) sequence of random variables. We assume throughout that Z_t has mean zero, variance $\sigma^2 \in (0, \infty)$, and distribution function F . Because $\phi_0(z) \neq 0$ for $|z| \leq 1$,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (2.2)$$

for all t , where the coefficients $\{\psi_j\}_{j=0}^{\infty}$ can be obtained from the Laurent series expansion of $\theta_0(z)/\phi_0(z)$. The AR filter $\phi_0(B)$ is said to be *causal* since (2.2) is a function of only the past and present $\{Z_t\}$. Because $\theta_0(z) \neq 0$ for $|z| \leq 1$, coefficients $\{\pi_j\}_{j=0}^{\infty}$ can be obtained from the Laurent series expansion of $\phi_0(z)/\theta_0(z)$

for which

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad (2.3)$$

for all t . The MA filter $\theta_0(B)$ is said to be *invertible* since (2.3) is a function of only the past and present $\{X_t\}$. Causality and invertibility are standard assumptions in time series analysis (see, for example, Brockwell and Davis, 1991, Chapter 3).

From (2.1), $Z_t = X_t - \phi_{01}X_{t-1} - \dots - \phi_{0p}X_{t-p} - \theta_{01}Z_{t-1} - \dots - \theta_{0q}Z_{t-q}$. Therefore, for arbitrary causal AR(p) and invertible MA(q) polynomials $\phi(z) = 1 - \phi_1z - \dots - \phi_pz^p$ and $\theta(z) = 1 + \theta_1z + \dots + \theta_qz^q$, we define

$$Z_t(\boldsymbol{\alpha}) = \begin{cases} 0, & t \leq p, \\ X_t - \phi_1X_{t-1} - \dots - \phi_pX_{t-p} - \theta_1Z_{t-1}(\boldsymbol{\alpha}) - \dots - \theta_qZ_{t-q}(\boldsymbol{\alpha}), & t = p+1, \dots, n, \end{cases}$$

where $\boldsymbol{\alpha} := (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$. If $\boldsymbol{\alpha}_0 := (\phi_{01}, \dots, \phi_{0p}, \theta_{01}, \dots, \theta_{0q})'$, note that $\{Z_t(\boldsymbol{\alpha}_0)\}_{t=p+1}^n$ closely approximates $\{Z_t\}_{t=p+1}^n$; the error is due to the initialization with zeros.

Suppose

A1 λ is a nonconstant and nondecreasing function from $(0, 1)$ to \mathbb{R} .

For $\boldsymbol{\alpha} \in \mathbb{R}^{p+q}$, we introduce the R-function

$$D_n(\boldsymbol{\alpha}) = \sum_{t=p+1}^n \lambda\left(\frac{R_t(\boldsymbol{\alpha})}{n-p+1}\right) [Z_t(\boldsymbol{\alpha}) - \overline{Z(\boldsymbol{\alpha})}], \quad (2.4)$$

where $\{R_t(\boldsymbol{\alpha})\}_{t=p+1}^n$ contains the ranks of $\{Z_t(\boldsymbol{\alpha})\}_{t=p+1}^n$ and $\overline{Z(\boldsymbol{\alpha})} := (n-p)^{-1} \sum_{t=p+1}^n Z_t(\boldsymbol{\alpha})$. D_n is similar to the R-function introduced in Jaeckel (1972) for estimating linear regression parameters. We, however, consider a weighted sum of the mean-corrected residuals instead of a weighted sum of the residuals (as in Jaeckel, 1972) to avoid assuming $\sum_{t=p+1}^n \lambda((t-p)/(n-p+1)) = 0$, which is required in Jaeckel (1972). Note that, if $\{Z_{(t)}(\boldsymbol{\alpha})\}_{t=p+1}^n$ is the series $\{Z_t(\boldsymbol{\alpha})\}_{t=p+1}^n$ ordered from smallest to largest, (2.4) can also be written as $D_n(\boldsymbol{\alpha}) = \sum_{t=p+1}^n \lambda((t-p)/(n-p+1)) [Z_{(t)}(\boldsymbol{\alpha}) - \overline{Z(\boldsymbol{\alpha})}]$ and, if $\bar{\lambda} := (n-p)^{-1} \sum_{t=p+1}^n \lambda((t-p)/(n-p+1))$, $D_n(\boldsymbol{\alpha}) = \sum_{t=p+1}^n [\lambda((t-p)/(n-p+1)) - \bar{\lambda}] [Z_{(t)}(\boldsymbol{\alpha}) - \overline{Z(\boldsymbol{\alpha})}]$. Because it tends to be near zero when the elements of $\{Z_{(t)}(\boldsymbol{\alpha})\}_{t=p+1}^n$ are similar and gets larger as the values of $\{|Z_{(t)}(\boldsymbol{\alpha}) - \overline{Z(\boldsymbol{\alpha})}|\}_{t=p+1}^n$ increase, D_n

is a measure of the dispersion of the residuals $\{Z_t(\boldsymbol{\alpha})\}_{t=p+1}^n$. Given a realization of length n from (2.1), $\{X_t\}_{t=1}^n$, we estimate $\boldsymbol{\alpha}_0$ by minimizing D_n . Choices for the weight function λ are discussed in Section 3.

We give some properties for D_n in the following theorem. From Jaeckel (1972), these same properties hold in the linear regression case. It is particularly interesting that D_n is continuous on \mathbb{R}^{p+q} , since λ is not necessarily continuous on $(0, 1)$ and the residual ranks $\{R_t(\cdot)\}_{t=p+1}^n$ are only piecewise continuous on \mathbb{R}^{p+q} .

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

$$\{P_1(\boldsymbol{\alpha}), \dots, P_{(n-p)!}(\boldsymbol{\alpha})\} = \{\{Z_{1,p+1}(\boldsymbol{\alpha}), \dots, Z_{1,n}(\boldsymbol{\alpha})\}, \dots, \{Z_{(n-p)!,p+1}(\boldsymbol{\alpha}), \dots, Z_{(n-p)!,n}(\boldsymbol{\alpha})\}\}$$

contains the $(n-p)!$ permutations of the sequence $\{Z_t(\boldsymbol{\alpha})\}_{t=p+1}^n$, then

$$D_n(\boldsymbol{\alpha}) = \sup_{j \in \{1, \dots, (n-p)!\}} \sum_{t=p+1}^n \lambda\left(\frac{t-p}{n-p+1}\right) [Z_{j,t}(\boldsymbol{\alpha}) - \overline{Z}(\boldsymbol{\alpha})].$$

In addition, D_n is a nonnegative, continuous function on \mathbb{R}^{p+q} . If $q = 0$, then D_n is also convex.

Proof: Recall that $D_n(\boldsymbol{\alpha}) = \sum_{t=p+1}^n [\lambda((t-p)/(n-p+1)) - \bar{\lambda}] [Z_{(t)}(\boldsymbol{\alpha}) - \overline{Z}(\boldsymbol{\alpha})]$, where $\bar{\lambda} = (n-p)^{-1} \sum_{t=p+1}^n \lambda((t-p)/(n-p+1))$. Let $a_n(t) = \lambda((t-p)/(n-p+1)) - \bar{\lambda}$ and $z_t(\boldsymbol{\alpha}) = Z_t(\boldsymbol{\alpha}) - \overline{Z}(\boldsymbol{\alpha})$.

The results of this theorem follow from the proof of Theorem 1 in Jaeckel (1972), where properties are given for $\sum_{t=p+1}^n a_n(t) z_t(\boldsymbol{\alpha})$. □

3 Asymptotic Results

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

A2 *The weight function λ is left-continuous and bounded on $(0, 1)$.*

A3 *F , the distribution function for the noise, is strictly increasing and differentiable on \mathbb{R} with uniformly continuous density f .*

Also, for causal AR(p) polynomial $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ and invertible MA(q) polynomial $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$, let $\{U_t\}$ and $\{V_t\}$ be the autoregressive processes $\phi(B)U_t = W_t$ and $\theta(B)V_t = W_t$, where

$\{W_t\}$ is iid with mean zero and variance one. If $\mathbf{U} := (U_{-1}, \dots, U_{-p})'$ and $\mathbf{V} := (V_{-1}, \dots, V_{-q})'$, then let

$$\mathbf{\Gamma}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbb{E}\{\mathbf{U}\mathbf{U}'\} & \mathbb{E}\{\mathbf{U}\mathbf{V}'\} \\ \mathbb{E}\{\mathbf{V}\mathbf{U}'\} & \mathbb{E}\{\mathbf{V}\mathbf{V}'\} \end{bmatrix}.$$

Theorem 3.1 *If A1–A3 hold, then there exists a sequence of minimizers $\hat{\boldsymbol{\alpha}}_n^R$ of $D_n(\cdot)$ in (2.4) such that*

$$n^{1/2} \left(\hat{\boldsymbol{\alpha}}_n^R - \boldsymbol{\alpha}_0 \right) \xrightarrow{d} \mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad (3.1)$$

where $\boldsymbol{\Sigma} := \tilde{J}\tilde{K}^{-2}\sigma^{-2}\mathbf{\Gamma}^{-1}(\boldsymbol{\alpha}_0)$, $\tilde{J} := \text{Var}\{\lambda(F(Z_1))\}$, and $\tilde{K} := \int_{-\infty}^{\infty} f(x) d\lambda(F(x))$.

Proof: $D_n(\boldsymbol{\alpha}) - D_n(\boldsymbol{\alpha}_0) = S_n(\sqrt{n}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0))$, where $S_n(\cdot)$ is defined in Lemma 5.7 of the Appendix. Because $\mathbf{Y} := -\tilde{K}^{-1}\sigma^{-2}\mathbf{\Gamma}^{-1}(\boldsymbol{\alpha}_0)\mathbf{N}$ minimizes the limit $S(\cdot)$ in Lemma 5.7, the result follows by Remark 1 in Davis, Knight, and Liu (1992). \square

Remark 1: If the density function f is almost everywhere differentiable on \mathbb{R} , using integration by parts, it can be shown that $\tilde{K} = -\int_{-\infty}^{\infty} f'(x)\lambda(F(x)) dx = -\int_0^1 [f'(F^{-1}(x))/f(F^{-1}(x))]\lambda(x) dx$.

Remark 2: When the ARMA model is purely AR ($q = 0$), the limiting result in (3.1) is the same as limiting results given in Koul and Saleh (1993), Koul and Ossiander (1994), and Mukherjee and Bai (2002) for related rank-based estimators of AR model parameters.

Remark 3: From Li and McLeod (1988), ML estimators of ARMA model parameters are asymptotically normal with mean $\boldsymbol{\alpha}_0$ and covariance matrix $n^{-1}[\int_{-\infty}^{\infty} (f'(x))^2/f(x) dx]^{-1}\sigma^{-2}\mathbf{\Gamma}^{-1}(\boldsymbol{\alpha}_0)$. The asymptotic relative efficiency (ARE) of R-estimation with respect to ML estimation is therefore $\tilde{J}^{-1}\tilde{K}^2[\int_{-\infty}^{\infty} (f'(x))^2/f(x) dx]^{-1}$; this same property holds for R-estimators of linear regression parameters (Jurečková and Sen, 1996, Section 3.4). Because $\tilde{J}^{-1}\tilde{K}^2 = \int_{-\infty}^{\infty} (f'(x))^2/f(x) dx$ when the weight function is proportional to $-f'(F^{-1}(x))/f(F^{-1}(x))$ (Jurečková and Sen, 1996, Section 3.4), R-estimation has the same asymptotic efficiency as ML estimation when an optimal weight function $\lambda_f(x) \propto -f'(F^{-1}(x))/f(F^{-1}(x))$ is used.

Remark 4: When $Z_t \sim N(0, \sigma^2)$, an optimal weight function λ_f is given by the van der Waerden weight function $\lambda(x) = \Phi^{-1}(x)$, where Φ is the standard normal distribution function. Hence, in the case of Gaussian noise, R-estimation with the van der Waerden weight function is asymptotically equivalent to ML estimation.

From Brockwell and Davis (1991, Section 8.8), Gaussian ML estimators of ARMA model parameters are consistent even when the noise distribution is non-Gaussian. These estimators are asymptotically normal with mean α_0 and covariance matrix $n^{-1}\mathbf{\Gamma}^{-1}(\alpha_0)$. When $\lambda(x) = \Phi^{-1}(x)$, $\tilde{J}\tilde{K}^{-2}\sigma^{-2} \leq 1$, with equality if and only if Z_t is Gaussian (Chernoff and Savage, 1958; Gastwirth and Wolff, 1968). Consequently, R-estimation with the van der Waerden weight function is at least as asymptotically efficient as Gaussian ML estimation, and strictly more efficient when the noise distribution is non-Gaussian.

Note that Φ^{-1} does not satisfy assumption A2 since it is unbounded on $(0, 1)$, but a bounded weight function closely approximating Φ^{-1} which does satisfy the assumptions can be found. For example, let

$$\lambda_m(x) = \Phi^{-1}(x)I\{1/m \leq x \leq 1 - 1/m\} + \Phi^{-1}(1/m)I\{x < 1/m\} + \Phi^{-1}(1 - 1/m)I\{x > 1 - 1/m\}, \quad (3.2)$$

where $m > 2$ and $I\{\cdot\}$ represents the indicator function. This weight function λ_m satisfies assumptions A1 and A2. In addition, as $m \rightarrow \infty$, λ_m converges pointwise to Φ^{-1} on $(0, 1)$ and, if $\tilde{J}_m := \text{Var}\{\lambda_m(F(Z_1))\}$ and $\tilde{K}_m := \int_{-\infty}^{\infty} f(x) d\lambda_m(F(x))$, $\tilde{J}_m \rightarrow \text{Var}\{\Phi^{-1}(F(Z_1))\}$ and $\tilde{K}_m \rightarrow \int_{-\infty}^{\infty} f(x) d\Phi^{-1}(F(x))$. Therefore, $m > 2$ can be chosen so that the asymptotic efficiency of R-estimation with weight function λ_m is arbitrarily close to the asymptotic efficiency of R-estimation with the van der Waerden weight function.

Remark 5: If $\tilde{\mu} := \text{median}\{Z_t\}$ is zero, LAD estimators of ARMA model parameters are asymptotically normal with mean α_0 and covariance matrix $n^{-1}[4f^2(0)]^{-1}\sigma^{-2}\mathbf{\Gamma}^{-1}(\alpha_0) = n^{-1}[4f^2(\tilde{\mu})]^{-1}\sigma^{-2}\mathbf{\Gamma}^{-1}(\alpha_0)$ (Davis and Dunsmuir, 1997). If Z_t is Laplace ($f(x) = \exp(-\sqrt{2}|x|/\sigma)/(\sqrt{2}\sigma)$), LAD estimation corresponds to ML estimation. Also for Laplace noise, $\lambda_f(x) \propto -I\{x \leq 1/2\} + I\{x > 1/2\}$ for $x \in (0, 1/2) \cup (1/2, 1)$ (λ_f does not exist at $x = 1/2$). Therefore, when the noise distribution is Laplace, R-estimation with weight function $\lambda(x) = -I\{x \leq 1/2\} + I\{x > 1/2\}$ is asymptotically equivalent to LAD/ML estimation. R-estimation with this weight function is also asymptotically equivalent to LAD estimation in general since, for weight function $\lambda(x) = -I\{x \leq 1/2\} + I\{x > 1/2\}$ and any noise distribution, $\tilde{J}\tilde{K}^{-2} = [4f^2(\tilde{\mu})]^{-1}$.

Remark 6: Another weight function commonly used for R-estimation is the Wilcoxon weight function $\lambda(x) = x - 1/2$, which is optimal when Z_t is logistic ($f(x) = \pi/(\sqrt{3}\sigma) \exp(-x\pi/(\sqrt{3}\sigma))/[1 + \exp(-x\pi/(\sqrt{3}\sigma))]^2$). The Wilcoxon weight function is also nearly optimal for Student's t noise; optimal weight functions corre-

Noise Distribution	ARE (R to ML)		
	van der Waerden	LAD-like	Wilcoxon
Laplace	0.637	1.000	0.750
logistic	0.955	0.750	1.000
$N(0, \sigma^2)$	1.000	0.637	0.955
$t(3)$	0.820	0.811	0.950
$t(5)$	0.915	0.769	0.993
$t(7)$	0.951	0.741	0.999
$t(9)$	0.968	0.723	0.998
$t(12)$	0.981	0.705	0.993
$t(15)$	0.987	0.693	0.988
$t(20)$	0.993	0.680	0.983
$t(30)$	0.997	0.666	0.975
$0.6(t(3) - 2/3) + 0.4(t(5) + 1)$	0.899	0.594	0.937

Table 3.1: AREs for R (with van der Waerden, LAD-like, and Wilcoxon weights) to ML for the Laplace, logistic, $N(0, \sigma^2)$ distributions, the Student's t -distribution with several different degrees of freedom, and the asymmetric distribution $0.6(t(3) - 2/3) + 0.4(t(5) + 1)$.

sponding to the t -distribution do not satisfy the assumptions, however, since $-f'(F^{-1}(x))/f(F^{-1}(x))$ is both increasing and decreasing on $(0, 1)$. The fourth column of Table 3.1 contains the AREs for R-estimation with Wilcoxon weights with respect to ML estimation for the Student's t -distribution with several different degrees of freedom. Because the AREs for t noise are all very close to one, R-estimation is essentially as efficient as ML.

AREs for R-estimation with Wilcoxon weights with respect to Gaussian ML estimation are ≥ 0.864 (Hodges and Lehmann, 1956), and sequences of distributions exist for which the AREs go to infinity. For the Student's t -distribution, $\text{ARE} \rightarrow \infty$ as degrees of freedom $\downarrow 2$, for example. Therefore, R-estimation with the Wilcoxon weight function is overall superior to Gaussian ML estimation with regard to asymptotic efficiency. R-estimation with the Wilcoxon weight function cannot be so easily compared to R-estimation with the van der Waerden weight function $\lambda(x) = \Phi^{-1}(x)$ or the LAD-like weight function $\lambda(x) = -I\{x \leq 1/2\} + I\{x > 1/2\}$, since distributions exist for which corresponding AREs are much smaller than 0.864. However, as can be seen in Table 3.1, the Wilcoxon weight function performs nearly as well as ML for both Gaussian and heavy-tailed distributions, and so the Wilcoxon function is often the recommended weight

function for R-estimation when the noise distribution is unknown (see, for example, Hettmansperger and McKean, 1998). For comparison, AREs corresponding to the van der Waerden and LAD-like weight functions are also included in Table 3.1.

Remark 7: AREs for the noise distribution $0.6(t(3) - 2/3) + 0.4(t(5) + 1)$, which is moderately skewed to the right, are given in Table 3.1 to demonstrate that the Wilcoxon function can perform relatively well for asymmetric distributions. The Wilcoxon function is not particularly efficient for significantly skewed distributions, however (see Choi and Öztürk, 2002). Weight functions that are relatively efficient when the noise distribution is significantly skewed to the right or left are given in McKean and Sievers (1989) and Choi and Öztürk (2002). McKean and Sievers (1989) give the optimal score function corresponding to the generalized F -distribution, which can be skewed, and Choi and Öztürk (2002) consider a weight function which they show is nearly optimal for generalized F noise and also relatively efficient for other skewed noise distributions.

We now estimate the covariance matrix $\Sigma = \tilde{J}\tilde{K}^{-2}\sigma^{-2}\Gamma^{-1}(\alpha_0)$ in (3.1) so that confidence intervals can be computed for the elements of α_0 . First, note that $\tilde{J} = \int_0^1 \lambda^2(x) dx - (\int_0^1 \lambda(x) dx)^2$ is known. Because $\hat{\alpha}_n^R \xrightarrow{P} \alpha_0$, it can be shown that

$$\hat{\sigma}_n := \left(\frac{1}{n} \sum_{t=p+1}^n Z_t^2(\hat{\alpha}_n^R) \right)^{1/2} \xrightarrow{P} (\mathbb{E} \{Z_1^2\})^{1/2} = \sigma;$$

similar results are established in Section 3.3.2 of Andrews (2003) and Section 3.2 of Andrews, Davis, and Breidt (2007), so we omit the details here. Also, since $\Gamma^{-1}(\cdot)$ is continuous at α_0 (see Brockwell and Davis, 1991, Section 8.8), $\Gamma^{-1}(\hat{\alpha}_n^R)$ is a consistent estimator of $\Gamma^{-1}(\alpha_0)$. A consistent estimator for \tilde{K} is given in the following corollary. For pure AR models ($q = 0$), other consistent estimators for \tilde{K} can be found in Section 7.3.3 of Koul (2002).

Corollary 3.1 *Consider the empirical distribution function $\hat{F}_n(x) := (n-p)^{-1} \sum_{t=p+1}^n I\{Z_t(\hat{\alpha}_n^R) \leq x\}$ and the kernel density estimator of f*

$$\hat{f}_n(x) := \frac{1}{h_n n} \sum_{t=p+1}^n \kappa \left(\frac{Z_t(\hat{\alpha}_n^R) - x}{h_n} \right), \quad (3.3)$$

where κ is a kernel density function on \mathbb{R} that satisfies condition A10 in Robinson (1987) and the bandwidth sequence $\{h_n\}$ is chosen so that $h_n \xrightarrow{P} 0$ and $h_n^3 n \xrightarrow{P} \infty$ as $n \rightarrow \infty$. If A1–A3 hold, then

$$\begin{aligned} \hat{K}_n &:= \int_{-\infty}^{\infty} \hat{f}_n(x) d\lambda(\hat{F}_n(x)) = \sum_{t=p+1}^n \hat{f}_n(Z_{(t)}(\hat{\alpha}_n^R)) \left[\lambda\left(\frac{t-p}{n-p}\right) - \lambda\left(\frac{t-p-1}{n-p}\right) \right] \\ &\xrightarrow{P} \tilde{K}. \end{aligned}$$

Proof: If $\tilde{F}_n(x) := (n-p)^{-1} \sum_{t=p+1}^n I\{Z_t \leq x\}$, then $\sup_{x \in \mathbb{R}} n^{1/2} |\hat{F}_n(x) - \tilde{F}_n(x)| \xrightarrow{P} 0$ by Theorem 1 of Bai (1994). By Theorem 3 of Robinson (1987), $\sup_{x \in \mathbb{R}} |\hat{f}_n(x) - f(x)| \xrightarrow{P} 0$. As a result, $\hat{K}_n \xrightarrow{P} \tilde{K}$ follows from the proof of Theorem 7.3.3 in Koul (2002). \square

Therefore, $\tilde{J}\hat{K}_n^{-2}\hat{\sigma}_n^{-2}\mathbf{\Gamma}^{-1}(\hat{\alpha}_n^R)$ is a consistent estimator for $\mathbf{\Sigma}$. Note that Gaussian densities can be used for the kernel density function κ since they satisfy condition A10 in Robinson (1987).

4 Simulation Study

In this section, we give the results of a simulation study to assess the quality of the asymptotic approximations for finite samples. For each of 1000 replicates, we simulated an ARMA series and found $\hat{\alpha}_n^R$ by minimizing D_n in (2.4). To reduce the possibility of the optimizer getting trapped at local minima, we used 200 random starting values for each replicate. We evaluated D_n at each of the 200 candidate values and then reduced the collection of initial values to the ten with the smallest values of D_n . Using these ten initial values as starting points, we found optimized values by implementing the Nelder-Mead algorithm (Lagarias, Reeds, Wright, and Wright, 1998). The optimized value for which D_n was smallest was chosen to be $\hat{\alpha}_n^R$. Confidence intervals for the elements of α_0 were constructed using (3.1) and $\tilde{J}\hat{K}_n^{-2}\hat{\sigma}_n^{-2}\mathbf{\Gamma}^{-1}(\hat{\alpha}_n^R)$, the consistent estimator of $\mathbf{\Sigma}$. For the kernel density estimator (3.3), we used the standard Gaussian kernel density function and, because of its recommendation in Silverman (1986, page 48), we used bandwidth $h_n = 0.9n^{-1/5} \min\{\hat{\sigma}_n, IQR/1.34\}$, where IQR is the interquartile range for $\{Z_t(\hat{\alpha}_n^R)\}_{t=p+1}^n$.

Results of simulations for the $N(0,1)$ and $t(3)$ noise distributions and van der Waerden (to be specific,

n	Asymptotic		Empirical		
	mean	std. dev. ($N(0,1)/t(3)$)	mean ($N(0,1)/t(3)$)	std. dev. ($N(0,1)/t(3)$)	% coverage ($N(0,1)/t(3)$)
100	$\phi_1 = 0.3$	0.0954/0.0745	0.280/0.285	0.0966/0.0836	91.7/92.3
500	$\phi_1 = 0.3$	0.0427/0.0333	0.298/0.296	0.0405/0.0354	95.5/95.0
1500	$\phi_1 = 0.3$	0.0246/0.0192	0.299/0.300	0.0243/0.0196	94.8/94.6
100	$\theta_1 = 0.4$	0.0917/0.0716	0.394/0.398	0.0972/0.0829	89.9/90.5
500	$\theta_1 = 0.4$	0.0410/0.0320	0.398/0.399	0.0406/0.0330	95.5/94.7
1500	$\theta_1 = 0.4$	0.0237/0.0185	0.400/0.400	0.0232/0.0192	95.1/94.5
100	$\phi_1 = 0.3$	0.1526/0.1192	0.268/0.283	0.1538/0.1308	91.3/91.8
	$\theta_1 = 0.4$	0.1466/0.1145	0.421/0.412	0.1587/0.1339	87.5/88.2
500	$\phi_1 = 0.3$	0.0683/0.0533	0.300/0.296	0.0694/0.0560	93.5/94.3
	$\theta_1 = 0.4$	0.0656/0.0512	0.400/0.401	0.0661/0.0532	94.1/94.3
1500	$\phi_1 = 0.3$	0.0394/0.0308	0.299/0.299	0.0385/0.0326	96.3/94.2
	$\theta_1 = 0.4$	0.0379/0.0296	0.400/0.401	0.0381/0.0307	94.7/94.2

Table 4.1: Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for R-estimates of ARMA model parameters. The $N(0,1)$ and $t(3)$ noise distributions and van der Waerden weights were used.

weight function (3.2) was used, with $m = 10,000$), LAD-like, and Wilcoxon weights are given in Tables 4.1–4.3. We show the empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for R-estimates of ARMA model parameters. Asymptotic means and standard deviations were obtained using Theorem 3.1. Particularly when $n = 1500$, the R-estimates appear nearly unbiased, the asymptotic standard deviations fairly accurately reflect the true variability of the estimates, and the confidence interval coverages are close to the nominal 95% level. Normal probability plots show that the R-estimates are approximately normal in all cases. Simulation results in Li and McLeod (1988) indicate that asymptotic approximations for ML estimates of ARMA model parameters are also fairly accurate for large sample sizes. Therefore, it appears that the AREs in Table 3.1 for R-estimation with respect to ML are indicative of finite, large sample behavior.

n	Asymptotic		Empirical		
	mean	std. dev. (N(0,1)/t(3))	mean (N(0,1)/t(3))	std. dev. (N(0,1)/t(3))	% coverage (N(0,1)/t(3))
100	$\phi_1 = 0.3$	0.1196/0.0749	0.275/0.288	0.1210/0.0882	95.2/94.7
500	$\phi_1 = 0.3$	0.0535/0.0335	0.296/0.300	0.0538/0.0355	95.6/96.0
1500	$\phi_1 = 0.3$	0.0309/0.0193	0.300/0.299	0.0311/0.0195	95.5/95.8
100	$\theta_1 = 0.4$	0.1149/0.0720	0.392/0.403	0.1210/0.0847	92.9/94.1
500	$\theta_1 = 0.4$	0.0514/0.0322	0.400/0.399	0.0522/0.0345	94.7/95.9
1500	$\theta_1 = 0.4$	0.0297/0.0186	0.398/0.400	0.0294/0.0197	95.8/95.0
100	$\phi_1 = 0.3$	0.1913/0.1199	0.273/0.285	0.2000/0.1435	90.6/93.2
	$\theta_1 = 0.4$	0.1838/0.1152	0.407/0.407	0.1984/0.1433	90.6/92.2
500	$\phi_1 = 0.3$	0.0855/0.0536	0.301/0.299	0.0870/0.0585	94.8/94.1
	$\theta_1 = 0.4$	0.0822/0.0515	0.397/0.400	0.0838/0.0583	95.0/92.8
1500	$\phi_1 = 0.3$	0.0494/0.0310	0.300/0.300	0.0488/0.0305	95.6/96.3
	$\theta_1 = 0.4$	0.0475/0.0297	0.399/0.400	0.0467/0.0301	95.5/95.9

Table 4.2: Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for R -estimates of ARMA model parameters. The $N(0,1)$ and $t(3)$ noise distributions and LAD-like weights were used.

n	Asymptotic		Empirical		
	mean	std. dev. (N(0,1)/t(3))	mean (N(0,1)/t(3))	std. dev. (N(0,1)/t(3))	% coverage (N(0,1)/t(3))
100	$\phi_1 = 0.3$	0.0976/0.0692	0.282/0.288	0.0982/0.0764	94.8/94.4
500	$\phi_1 = 0.3$	0.0437/0.0310	0.296/0.298	0.0433/0.0330	96.0/93.7
1500	$\phi_1 = 0.3$	0.0252/0.0179	0.300/0.299	0.0252/0.0176	94.9/96.0
100	$\theta_1 = 0.4$	0.0938/0.0665	0.400/0.406	0.0988/0.0751	91.8/94.0
500	$\theta_1 = 0.4$	0.0419/0.0297	0.399/0.399	0.0413/0.0319	95.3/94.4
1500	$\theta_1 = 0.4$	0.0242/0.0172	0.400/0.399	0.0241/0.0178	95.1/94.7
100	$\phi_1 = 0.3$	0.1562/0.1107	0.283/0.285	0.1699/0.1308	91.3/92.8
	$\theta_1 = 0.4$	0.1501/0.1064	0.406/0.406	0.1705/0.1236	89.9/93.7
500	$\phi_1 = 0.3$	0.0699/0.0495	0.296/0.296	0.0688/0.0533	94.9/92.8
	$\theta_1 = 0.4$	0.0671/0.0476	0.403/0.403	0.0646/0.0509	94.8/94.2
1500	$\phi_1 = 0.3$	0.0403/0.0286	0.297/0.299	0.0401/0.0294	95.2/95.6
	$\theta_1 = 0.4$	0.0387/0.0275	0.402/0.400	0.0387/0.0283	95.2/95.0

Table 4.3: Empirical means, standard deviations, and percent coverages of nominal 95% confidence intervals for R -estimates of ARMA model parameters. The $N(0,1)$ and $t(3)$ noise distributions and Wilcoxon weights were used.

Appendix

This section contains proofs of the lemmas used to establish Theorem 3.1. Assume A1–A3 hold throughout.

First, from Section 8.11 of Brockwell and Davis (1991), for $t \in \{p+1, \dots, n\}$ and $\boldsymbol{\alpha} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$, we have

$$\frac{\partial Z_t(\boldsymbol{\alpha})}{\partial \alpha_j} = \frac{-X_{t-j}}{\theta(B)} = \frac{-\theta_0(B)Z_{t-j}}{\theta(B)\phi_0(B)}, \quad j \in \{1, \dots, p\}, \quad \text{and} \quad \frac{\partial Z_t(\boldsymbol{\alpha})}{\partial \alpha_j} = \frac{-Z_{t+p-j}(\boldsymbol{\alpha})}{\theta(B)}, \quad j \in \{p+1, \dots, p+q\}.$$

Evaluating the partial derivatives at $\boldsymbol{\alpha}_0$ and ignoring the effect of recursion initialization, we have

$$\begin{aligned} \frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \alpha_j} &= \begin{cases} -Z_{t-j}/\phi_0(B), & j \in \{1, \dots, p\} \\ -Z_{t+p-j}(\boldsymbol{\alpha}_0)/\theta_0(B), & j \in \{p+1, \dots, p+q\} \end{cases} \\ &\simeq \begin{cases} -Z_{t-j}/\phi_0(B), & j \in \{1, \dots, p\} \\ -Z_{t+p-j}/\theta_0(B), & j \in \{p+1, \dots, p+q\} \end{cases} \\ &=: \frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j}. \end{aligned}$$

The expected value of $\partial Z_t^*(\boldsymbol{\alpha}_0)/\partial \boldsymbol{\alpha}$ is zero, the covariance matrix is $\sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0)$, and, because $\phi_0(B)$ is a causal AR operator and $\theta_0(B)$ is an invertible MA operator, $\partial Z_t^*(\boldsymbol{\alpha}_0)/\partial \alpha_j \in \sigma(Z_{t-1}, Z_{t-2}, \dots)$ for all $j \in \{1, \dots, p+q\}$.

Now, let

$$\overline{\frac{\partial Z(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}} = \frac{1}{n-p} \sum_{t=p+1}^n \frac{\partial Z_t(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \quad \text{and} \quad \overline{\frac{\partial Z^*(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}} = \frac{1}{n-p} \sum_{t=p+1}^n \frac{\partial Z_t^*(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}.$$

Lemma 5.1 *As $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{t=p+1}^n \lambda(F(Z_t)) \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] \xrightarrow{d} \mathbf{N} \sim N(\mathbf{0}, \tilde{J} \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0)).$$

Proof: If the Laurent series expansion for $1/\theta_0(z)$ is given by $\sum_{j=0}^{\infty} \beta_j z^j$, then, for $t \in \{0, \dots, n-p-1\}$,

$$Z_{p+1+t} = \sum_{j=0}^{\infty} \beta_j (\phi_0(B)X_{p+1+t-j}) \quad \text{and} \quad Z_{p+1+t}(\boldsymbol{\alpha}_0) = \sum_{j=0}^t \beta_j (\phi_0(B)X_{p+1+t-j}). \quad (5.1)$$

Because there exist constants $a > 0$ and $0 < b < 1$ such that $|\beta_j| < ab^j$ for all $j \in \{0, 1, \dots\}$ (see Brockwell and Davis, 1991, Section 3.3), it follows that $\sum_{t=p+1}^n \mathbb{E}|\partial Z_t(\boldsymbol{\alpha}_0)/\partial \alpha_j - \partial Z_t^*(\boldsymbol{\alpha}_0)/\partial \alpha_j| = O(1)$ for all $j \in$

$\{1, \dots, p+q\}$. Therefore,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \lambda(F(Z_t)) \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] &= \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \lambda(F(Z_t)) \left[\frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z^*(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=p+1}^n [\lambda(F(Z_t)) - \mathbb{E}\{\lambda(F(Z_1))\}] \frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \\ &\quad - \left(\frac{1}{\sqrt{n}} \sum_{t=p+1}^n [\lambda(F(Z_t)) - \mathbb{E}\{\lambda(F(Z_1))\}] \right) \overline{\frac{\partial Z^*(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} + o_p(1). \end{aligned} \quad (5.2)$$

By the central limit theorem, $n^{-1/2} \sum_{t=p+1}^n [\lambda(F(Z_t)) - \mathbb{E}\{\lambda(F(Z_1))\}] \xrightarrow{d} N(0, \tilde{J})$ and, by the ergodic theorem, $\overline{\frac{\partial Z^*(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} = (n-p)^{-1} \sum_{t=p+1}^n \partial Z_t^*(\boldsymbol{\alpha}_0) / \partial \boldsymbol{\alpha} \xrightarrow{P} \mathbf{0}$. Thus, (5.2) equals

$$\frac{1}{\sqrt{n}} \sum_{t=p+1}^n [\lambda(F(Z_t)) - \mathbb{E}\{\lambda(F(Z_1))\}] \frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} + o_p(1).$$

Let $\mathbf{u} \in \mathbb{R}^{p+q}$. By the Cramér-Wold device, it suffices to show $n^{-1/2} \sum_{t=p+1}^n Y_t \xrightarrow{d} \mathbf{u}' \mathbf{N} \sim N(0, \tilde{J} \mathbf{u}' \sigma^2 \mathbf{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u})$, where $Y_t := [\lambda(F(Z_t)) - \mathbb{E}\{\lambda(F(Z_1))\}] \mathbf{u}' \partial Z_t^*(\boldsymbol{\alpha}_0) / \partial \boldsymbol{\alpha}$. Elements of the infinite order moving average stationary sequence $\{Y_t\}$ can be truncated to create a finite order moving average stationary sequence. By applying a central limit theorem (Brockwell and Davis, 1991, Theorem 6.4.2) to each truncation level, asymptotic normality can be deduced. The details are omitted. \square

For $\mathbf{u} \in \mathbb{R}^{p+q}$, define $Z_{t,n}(\mathbf{u}) = Z_t - Z_t(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{u})$.

Lemma 5.2 For any $T \in (0, \infty)$, as $n \rightarrow \infty$,

$$\sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' [I\{Z_t \leq x + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq x\} - F(x + Z_{t,n}(\mathbf{u})) + F(x)] \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] \right| \quad (5.3)$$

is $o_p(1)$.

Proof: Observe that (5.3) is bounded above by

$$\sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' [I\{Z_t \leq x + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq x\} - F(x + Z_{t,n}(\mathbf{u})) + F(x)] \frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \right| \quad (5.4)$$

$$+ \sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n [I\{Z_t \leq x + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq x\} - F(x + Z_{t,n}(\mathbf{u})) + F(x)] \right| \quad (5.5)$$

$$\times \sup_{\|\mathbf{u}\| \leq T} \left| \mathbf{u}' \frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \right|. \quad (5.6)$$

Andrews (2003, see pages 56–57) shows that (5.5) is $o_p(1)$ when the observed time series is an all-pass process.

A similar argument can be used to show that (5.5) is $o_p(1)$ in the case of an ARMA process. By the ergodic theorem, (5.6) is also $o_p(1)$. Hence, following (5.1), the proof is complete if, for any $j \in \{1, \dots, p+q\}$,

$$\sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n [I\{Z_t \leq x + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq x\} - F(x + Z_{t,n}(\mathbf{u})) + F(x)] \frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right| \quad (5.7)$$

is $o_p(1)$.

Fix a nondecreasing sequence of positive integers $\{m_n\}_{n=1}^\infty$ such that $m_n \sim n^{1/10}$ ($n^{1/10}/m_n \rightarrow 1$ as $n \rightarrow \infty$). (5.7) is bounded above by

$$\begin{aligned} & \sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^{m_n} [I\{Z_t \leq x + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq x\} - F(x + Z_{t,n}(\mathbf{u})) + F(x)] \frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right| \quad (5.8) \\ & + \sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=m_n+1}^n [I\{Z_t \leq x + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq x\} - F(x + Z_{t,n}(\mathbf{u})) + F(x)] \frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right|, \quad (5.9) \end{aligned}$$

and, since $m_n/n^{1/2} \rightarrow 0$, (5.8) is $o_p(1)$. Now, fix $\|\mathbf{u}\| \leq T$ and note that

$$Z_{t,n}(\mathbf{u}) = Z_t - Z_t \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) = Z_t - Z_t(\boldsymbol{\alpha}_0) - \frac{\mathbf{u}'}{\sqrt{n}} \frac{\partial Z_t(\boldsymbol{\alpha}_{t,n}^*(\mathbf{u}))}{\partial \boldsymbol{\alpha}}, \quad (5.10)$$

where $\boldsymbol{\alpha}_{t,n}^*(\mathbf{u})$ is between $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_0 + n^{-1/2}\mathbf{u}$. Following Section 3.3 of Brockwell and Davis (1991), there exists a geometrically decaying, non-negative, real-valued sequence $\{\ddot{\psi}_k\}_{k=1}^\infty$ such that

$$\left| \mathbf{u}' \frac{\partial Z_t(\boldsymbol{\alpha}_{t,n}^*(\mathbf{u}))}{\partial \boldsymbol{\alpha}} \right| \leq \sum_{k=1}^{\infty} \ddot{\psi}_k |Z_{t-k}| \quad \forall t \in \{p+1, \dots, n\}$$

for all n sufficiently large, and so $\max_{m_n+1 \leq t \leq n} |Z_{t,n}(\mathbf{u})| = o_p(1)$. Because $\max_{m_n+1 \leq t \leq n} |Z_{t,n}(\mathbf{u})| = o_p(1)$,

$$\left[\frac{1}{n - m_n} \sum_{t=m_n+1}^n \left(\frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right)^2 \right]^{1/2} \xrightarrow{P} \left(\mathbb{E} \left\{ \frac{\partial Z_1^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right\}^2 \right)^{1/2}, \quad \text{and} \quad \frac{1}{\sqrt{n - m_n}} \max_{m_n+1 \leq t \leq n} \left| \frac{Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right| \xrightarrow{P} 0,$$

by Theorem 1.1 in Koul and Ossiander (1994),

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=m_n+1}^n [I\{Z_t \leq x + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq x\} - F(x + Z_{t,n}(\mathbf{u})) + F(x)] \frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right| \xrightarrow{P} 0$$

for fixed $\mathbf{u} \in \mathbb{R}^{p+q}$. As discussed in the proof of Lemma 1.1 in Koul and Ossiander (1994), results in the proof of Theorem 1.2 in Koul (1991) can be used to obtain uniformity with respect to $\|\mathbf{u}\| \leq T$, and so (5.9)

is $o_p(1)$. \square

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

$$\sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' [F(x + Z_{t,n}(\mathbf{u})) - F(x)] \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] + f(x) \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right| \quad (5.11)$$

is $o_p(1)$.

Proof: (5.11) is bounded above by

$$\sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' Z_{t,n}(\mathbf{u}) f(x) \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] + f(x) \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right| \quad (5.12)$$

$$+ \sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' Z_{t,n}(\mathbf{u}) [f(x_{t,n}^*(\mathbf{u})) - f(x)] \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] \right|, \quad (5.13)$$

where $x_{t,n}^*(\mathbf{u})$ lies between x and $x + Z_{t,n}(\mathbf{u})$. Following (5.10), (5.12) is bounded above by

$$\left(\sup_{x \in \mathbb{R}} f(x) \right) \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' (Z_t - Z_t(\boldsymbol{\alpha}_0)) \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] \right| \quad (5.14)$$

$$+ \left(\sup_{x \in \mathbb{R}} f(x) \right) \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{n} \sum_{t=p+1}^n \mathbf{u}' \frac{\partial Z_t(\boldsymbol{\alpha}_{t,n}^*(\mathbf{u}))}{\partial \boldsymbol{\alpha}} \mathbf{u}' \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] - \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right|. \quad (5.15)$$

Since f is uniformly continuous on \mathbb{R} , $\sup_{x \in \mathbb{R}} f(x) < \infty$, and so, using (5.1), it can be shown that (5.14) is $o_p(1)$. Because $\sup_{p+1 \leq t \leq n, \|\mathbf{u}\| \leq T} \|\boldsymbol{\alpha}_{t,n}^*(\mathbf{u}) - \boldsymbol{\alpha}_0\| = o(1)$ and $\overline{\partial Z(\boldsymbol{\alpha}_0)/\partial \boldsymbol{\alpha}} = (n-p)^{-1} \sum_{t=p+1}^n \partial Z_t(\boldsymbol{\alpha}_0)/\partial \boldsymbol{\alpha} = o_p(1)$, (5.15) equals

$$\left(\sup_{x \in \mathbb{R}} f(x) \right) \sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{n} \sum_{t=p+1}^n \left[\mathbf{u}' \frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \right]^2 - \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right| + o_p(1) \xrightarrow{P} 0.$$

Thus, (5.12) is $o_p(1)$. Using the uniform continuity of f , it can be shown similarly that (5.13) is also $o_p(1)$.

□

For $\mathbf{u} \in \mathbb{R}^{p+q}$ and $x \in \mathbb{R}$, let $F_n(x, \mathbf{u}) = (n-p+1)^{-1} \sum_{t=p+1}^n I\{Z_t(\boldsymbol{\alpha}_0 + n^{-1/2}\mathbf{u}) \leq x\}$ and, for $y \in [0, 1]$,

let

$$F_n^{-1}(y, \mathbf{u}) = \begin{cases} \inf \{x : F_n(x, \mathbf{u}) \geq y\}, & 0 \leq y \leq \frac{n-p}{n-p+1} \\ +\infty, & \frac{n-p}{n-p+1} < y \leq 1. \end{cases} \quad (5.16)$$

Combining Lemmas 5.2 and 5.3, we get

$$\sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' [I\{Z_t \leq F_n^{-1}(y, \mathbf{u}) + Z_{t,n}(\mathbf{u})\} - I\{Z_t \leq F_n^{-1}(y, \mathbf{u})\}] \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] + f(F_n^{-1}(y, \mathbf{u})) \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right| \xrightarrow{P} 0 \quad (5.17)$$

for any $T \in (0, \infty)$.

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

$$\sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' [I\{Z_t \leq F_n^{-1}(y, \mathbf{u})\} - I\{Z_t \leq F^{-1}(y)\}] \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \overline{\frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}} \right] \right| \xrightarrow{P} 0. \quad (5.18)$$

Proof: Choose arbitrary $j \in \{1, \dots, p+q\}$. To prove this lemma, we show that

$$\sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n [I\{Z_t \leq F^{-1}(F(F_n^{-1}(y, \mathbf{u})))\} - I\{Z_t \leq F^{-1}(y)\}] \left[\frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} - \overline{\frac{\partial Z^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j}} \right] \right| \quad (5.19)$$

is $o_p(1)$, where $\overline{\partial Z^*(\boldsymbol{\alpha})/\partial \alpha_j} = (n-p)^{-1} \sum_{t=p+1}^n \partial Z_t^*(\boldsymbol{\alpha})/\partial \alpha_j$. First, for $y \in [0, 1]$, let

$$T_n^*(y) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^n [I\{Z_t \leq F^{-1}(y)\} - y] \left[\frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} - \overline{\frac{\partial Z^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j}} \right].$$

Because (5.19) is bounded above by

$$\sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| T_n^*(F(F_n^{-1}(y, \mathbf{u}))) - T_n^*(y) \right| \quad (5.20)$$

$$+ \sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| F(F_n^{-1}(y, \mathbf{u})) - y \right| \quad (5.21)$$

$$\times \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \left[\frac{\partial Z_t^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} - \overline{\frac{\partial Z^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j}} \right] \right| \quad (5.22)$$

and (5.22) equals zero, the proof is complete if (5.21) is $o_p(1)$ and, for any $\eta > 0$,

$$\lim_{\delta \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{x, y \in [0,1], |x-y| \leq \delta} |T_n^*(x) - T_n^*(y)| > \eta \right) = 0.$$

Note that (5.21) equals $\sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} |F_n(x, \mathbf{u}) - F(x)|$, which is bounded above by

$$\sup_{\|\mathbf{u}\| \leq T, x \in \mathbb{R}} \left| \frac{1}{n-p+1} \sum_{t=p+1}^n \left[I\left\{ Z_t \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) \leq x \right\} - I\{Z_t \leq x\} \right] \right| \quad (5.23)$$

$$+ \sup_{x \in \mathbb{R}} \left| \frac{1}{n-p+1} \sum_{t=p+1}^n I\{Z_t \leq x\} - F(x) \right|. \quad (5.24)$$

Following the proof of Lemma 3 on page 55 of Andrews (2003), (5.23) is $o_p(1)$ and, by the Glivenko-Cantelli theorem, (5.24) is also $o_p(1)$.

Now, for $y \in [0, 1]$, let $U_n^*(y) = n^{-1/2} \sum_{t=p+1}^n [I\{F(Z_t) \leq y\} - y] \partial Z_t^*(\boldsymbol{\alpha}_0) / \partial \alpha_j$, and consider

$$\begin{aligned} & \mathbb{P} \left(\sup_{x, y \in [0, 1], |x-y| \leq \delta} |T_n^*(x) - T_n^*(y)| > \eta \right) \\ & \leq \mathbb{P} \left(\sup_{x, y \in [0, 1], |x-y| \leq \delta} |U_n^*(x) - U_n^*(y)| > \frac{\eta}{2} \right) \end{aligned} \quad (5.25)$$

$$+ \mathbb{P} \left(\sup_{x, y \in [0, 1], |x-y| \leq \delta} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n [I\{F(Z_t) \leq x\} - I\{F(Z_t) \leq y\} - x + y] \left\| \frac{\partial Z^*(\boldsymbol{\alpha}_0)}{\partial \alpha_j} \right\| > \frac{\eta}{2} \right). \quad (5.26)$$

By Theorem 1.1 in Koul and Ossiander (1994), $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty}$ of (5.25) is zero. If $W^*(\cdot)$ denotes a Brownian bridge on $[0, 1]$, then, following Theorem 14.3 in Billingsley (1999),

$$\sup_{x, y \in [0, 1], |x-y| \leq \delta} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n [I\{F(Z_t) \leq x\} - I\{F(Z_t) \leq y\} - x + y] \right| \xrightarrow{d} \sup_{x, y \in [0, 1], |x-y| \leq \delta} |W^*(x) - W^*(y)|$$

for any $\delta > 0$. Therefore, since $\overline{\partial Z^*(\boldsymbol{\alpha}_0) / \partial \alpha_j} = o_p(1)$, $\limsup_{n \rightarrow \infty}$ of (5.26) equals zero for any fixed $\delta > 0$.

□

Lemma 5.5 For any $T \in (0, \infty)$, as $n \rightarrow \infty$,

$$\sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \left[\lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{u})}{n-p+1} \right) - \lambda(F(Z_t)) \right] \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \frac{\partial Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \right] - \tilde{K} \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right| \quad (5.27)$$

is $o_p(1)$.

Proof: If $R_{t,n}(\mathbf{u}) := R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{u})$ and $\mathbf{V}_{t,n}^* := \partial Z_t(\boldsymbol{\alpha}_0) / \partial \boldsymbol{\alpha} - \overline{\partial Z(\boldsymbol{\alpha}_0) / \partial \boldsymbol{\alpha}}$, then (5.27) can be expressed

as

$$\sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \left[\lambda \left(\frac{R_{t,n}(\mathbf{u})}{n-p+1} \right) - \lambda(F(Z_t)) \right] \mathbf{V}_{t,n}^* - \tilde{K} \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right|.$$

Because the weight function λ is left-continuous, (5.27) equals

$$\sup_{\|\mathbf{u}\| \leq T} \left| \int_0^1 \left(\frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \left[I \left\{ \frac{R_{t,n}(\mathbf{u})}{n-p+1} \leq y \right\} - I\{F(Z_t) \leq y\} \right] \mathbf{V}_{t,n}^* + f(F^{-1}(y)) \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right) d\lambda(y) \right|,$$

which is bounded above by

$$\sup_{\|\mathbf{u}\| \leq T, y \in (0, 1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \left[I \left\{ \frac{R_{t,n}(\mathbf{u})}{n-p+1} \leq y \right\} - I\{F(Z_t) \leq y\} \right] \mathbf{V}_{t,n}^* + f(F^{-1}(y)) \mathbf{u}' \sigma^2 \boldsymbol{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right| \quad (5.28)$$

$$\times [\lambda(1) - \lambda(0)]. \quad (5.29)$$

Since λ is bounded, the proof is complete if (5.28) is $o_p(1)$.

Because $f(F^{-1}(\cdot))$ is uniformly continuous on $[0, 1]$ and, from the proof of Lemma 5.4, (5.21) is $o_p(1)$,

$$\sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| f(F_n^{-1}(y, \mathbf{u})) - f(F^{-1}(y)) \right| \xrightarrow{P} 0. \quad (5.30)$$

Combining (5.17), (5.18), and (5.30),

$$\sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' [I\{Z_t \leq F_n^{-1}(y, \mathbf{u}) + Z_{t,n}(\mathbf{u})\} - I\{F(Z_t) \leq y\}] \mathbf{V}_{t,n}^* + f(F^{-1}(y)) \mathbf{u}' \sigma^2 \mathbf{\Gamma}(\boldsymbol{\alpha}_0) \mathbf{u} \right|$$

is $o_p(1)$. Therefore, all that remains is to show that

$$\begin{aligned} & \sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \left[I\left\{ \frac{R_{t,n}(\mathbf{u})}{n-p+1} \leq y \right\} - I\{Z_t \leq F_n^{-1}(y, \mathbf{u}) + Z_{t,n}(\mathbf{u})\} \right] \mathbf{V}_{t,n}^* \right| \\ &= \sup_{\|\mathbf{u}\| \leq T, y \in (0,1)} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \left[I\left\{ \frac{R_{t,n}(\mathbf{u})}{n-p+1} \leq y \right\} - I\left\{ Z_t \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) \leq F_n^{-1}(y, \mathbf{u}) \right\} \right] \mathbf{V}_{t,n}^* \right| \end{aligned} \quad (5.31)$$

is $o_p(1)$. Observe that

$$\begin{aligned} & \frac{R_{t,n}(\mathbf{u})}{n-p+1} \\ &= \frac{1}{n-p+1} \left[\sum_{j=p+1}^n I\left\{ Z_j \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) \leq Z_t \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) \right\} - \sum_{j=t+1}^n I\left\{ Z_j \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) = Z_t \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) \right\} \right] \end{aligned}$$

and, since Z_t has a continuous distribution,

$$\mathbb{P} \left(\sup_{\|\mathbf{u}\| \leq T} \sum_{t=p+1}^n \sum_{j=t+1}^n I\left\{ Z_j \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) = Z_t \left(\boldsymbol{\alpha}_0 + \frac{\mathbf{u}}{\sqrt{n}} \right) \right\} > 1 \right) = 0$$

for all n . Hence, following the definition of $F_n^{-1}(\cdot, \cdot)$ in (5.16), (5.31) is almost surely bounded above by

$\sup_{p+1 \leq t \leq n, \|\mathbf{u}\| \leq T} 2n^{-1/2} |\mathbf{u}' \mathbf{V}_{t,n}^*|$, which is $o_p(1)$. \square

Now consider the mixed partial derivatives of $Z_t(\cdot)$:

$$\frac{\partial^2 Z_t(\boldsymbol{\alpha})}{\partial \alpha_j \partial \alpha_k} = \begin{cases} 0, & j, k \in \{1, \dots, p\}, \\ X_{t+p-j-k} / \theta^2(B), & j \in \{1, \dots, p\}, k \in \{p+1, \dots, p+q\}, \\ 2Z_{t+2p-j-k}(\boldsymbol{\alpha}) / \theta^2(B), & j, k \in \{p+1, \dots, p+q\}, \end{cases}$$

and so

$$\frac{\partial^2 Z_t(\boldsymbol{\alpha}_0)}{\partial \alpha_j \partial \alpha_k} = \begin{cases} 0, & j, k \in \{1, \dots, p\}, \\ Z_{t+p-j-k} / [\phi_0(B) \theta_0(B)], & j \in \{1, \dots, p\}, k \in \{p+1, \dots, p+q\}, \\ 2Z_{t+2p-j-k}(\boldsymbol{\alpha}_0) / \theta_0^2(B), & j, k \in \{p+1, \dots, p+q\}. \end{cases}$$

Also, define

$$\overline{\frac{\partial^2 Z(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}} = \frac{1}{n-p} \sum_{t=p+1}^n \frac{\partial^2 Z_t(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}$$

Lemma 5.6 For any $T \in (0, \infty)$, as $n \rightarrow \infty$,

$$\sup_{\|\mathbf{u}\|, \|\mathbf{v}\| \leq T} \left| \frac{1}{n} \sum_{t=p+1}^n \mathbf{u}' \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{u})}{n-p+1} \right) \left[\frac{\partial^2 Z_t(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{v})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \overline{\frac{\partial^2 Z(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{v})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}} \right] \mathbf{u} \right| \xrightarrow{P} 0.$$

Proof: Following Lemmas 5.1 and 5.5, it can be shown that

$$\sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{n} \sum_{t=p+1}^n \mathbf{u}' \lambda(F(Z_t)) \left[\frac{\partial^2 Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \overline{\frac{\partial^2 Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}} \right] \mathbf{u} \right| \xrightarrow{P} 0$$

and

$$\sup_{\|\mathbf{u}\| \leq T} \left| \frac{1}{n} \sum_{t=p+1}^n \mathbf{u}' \left[\lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{u})}{n-p+1} \right) - \lambda(F(Z_t)) \right] \left[\frac{\partial^2 Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \overline{\frac{\partial^2 Z(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'}} \right] \mathbf{u} \right| \xrightarrow{P} 0.$$

Because it can also be shown that

$$\sup_{\|\mathbf{v}\| \leq T} \frac{1}{n} \sum_{t=p+1}^n \left| \frac{\partial^2 Z_t(\boldsymbol{\alpha}_0 + n^{-1/2} \mathbf{v})}{\partial \alpha_j \partial \alpha_k} - \frac{\partial^2 Z_t(\boldsymbol{\alpha}_0)}{\partial \alpha_j \partial \alpha_k} \right| \xrightarrow{P} 0 \quad \forall j, k \in \{1, \dots, p+q\}$$

(see Brockwell and Davis, 1991, Section 3.3), the proof is complete. \square

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

$$\begin{aligned} \tilde{U}_n(\mathbf{u}, \delta_1, \delta_2) &= \sum_{t=p+1}^n \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \delta_1 \mathbf{u})}{n-p+1} \right) \left[Z_t \left(\boldsymbol{\alpha}_0 + \frac{\delta_2 \mathbf{u}}{\sqrt{n}} \right) - \overline{Z \left(\boldsymbol{\alpha}_0 + \frac{\delta_2 \mathbf{u}}{\sqrt{n}} \right)} \right] \\ &\quad - \sum_{t=p+1}^n \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \delta_1 \mathbf{u})}{n-p+1} \right) \left[Z_t \left(\boldsymbol{\alpha}_0 + \frac{\delta_1 \mathbf{u}}{\sqrt{n}} \right) - \overline{Z \left(\boldsymbol{\alpha}_0 + \frac{\delta_1 \mathbf{u}}{\sqrt{n}} \right)} \right] \end{aligned}$$

and

$$\begin{aligned} \tilde{V}_n(\mathbf{u}, \delta_1, \delta_2) &= \sum_{t=p+1}^n \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \delta_2 \mathbf{u})}{n-p+1} \right) \left[Z_t \left(\boldsymbol{\alpha}_0 + \frac{\delta_2 \mathbf{u}}{\sqrt{n}} \right) - \overline{Z \left(\boldsymbol{\alpha}_0 + \frac{\delta_2 \mathbf{u}}{\sqrt{n}} \right)} \right] \\ &\quad - \sum_{t=p+1}^n \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2} \delta_2 \mathbf{u})}{n-p+1} \right) \left[Z_t \left(\boldsymbol{\alpha}_0 + \frac{\delta_1 \mathbf{u}}{\sqrt{n}} \right) - \overline{Z \left(\boldsymbol{\alpha}_0 + \frac{\delta_1 \mathbf{u}}{\sqrt{n}} \right)} \right]. \end{aligned}$$

(Recall that $\overline{Z(\boldsymbol{\alpha})} = (n-p)^{-1} \sum_{t=p+1}^n Z_t(\boldsymbol{\alpha})$.) Using Taylor series expansions,

$$\tilde{U}_n(\mathbf{u}, \delta_1, \delta_2)$$

$$\begin{aligned}
&= \sum_{t=p+1}^n \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_1\mathbf{u})}{n-p+1} \right) \left[Z_t \left(\boldsymbol{\alpha}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}} \right) - \overline{Z \left(\boldsymbol{\alpha}_0 + \frac{\delta_2\mathbf{u}}{\sqrt{n}} \right)} - Z_t(\boldsymbol{\alpha}_0) + \overline{Z(\boldsymbol{\alpha}_0)} \right] \\
&\quad - \sum_{t=p+1}^n \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_1\mathbf{u})}{n-p+1} \right) \left[Z_t \left(\boldsymbol{\alpha}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}} \right) - \overline{Z \left(\boldsymbol{\alpha}_0 + \frac{\delta_1\mathbf{u}}{\sqrt{n}} \right)} - Z_t(\boldsymbol{\alpha}_0) + \overline{Z(\boldsymbol{\alpha}_0)} \right] \\
&= \frac{\delta_2 - \delta_1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_1\mathbf{u})}{n-p+1} \right) \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \frac{\partial \overline{Z(\boldsymbol{\alpha}_0)}}{\partial \boldsymbol{\alpha}} \right] \\
&\quad + \frac{\delta_2^2}{2n} \sum_{t=p+1}^n \mathbf{u}' \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_1\mathbf{u})}{n-p+1} \right) \left[\frac{\partial^2 Z_t(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_1, \delta_2))}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \frac{\partial^2 \overline{Z(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_1, \delta_2))}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \right] \mathbf{u} \\
&\quad - \frac{\delta_1^2}{2n} \sum_{t=p+1}^n \mathbf{u}' \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_1\mathbf{u})}{n-p+1} \right) \left[\frac{\partial^2 Z_t(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_1, \delta_1))}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \frac{\partial^2 \overline{Z(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_1, \delta_1))}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \right] \mathbf{u} \quad (5.32)
\end{aligned}$$

and, similarly,

$$\begin{aligned}
&\tilde{V}_n(\mathbf{u}, \delta_1, \delta_2) \\
&= \frac{\delta_2 - \delta_1}{\sqrt{n}} \sum_{t=p+1}^n \mathbf{u}' \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_2\mathbf{u})}{n-p+1} \right) \left[\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} - \frac{\partial \overline{Z(\boldsymbol{\alpha}_0)}}{\partial \boldsymbol{\alpha}} \right] \\
&\quad + \frac{\delta_2^2}{2n} \sum_{t=p+1}^n \mathbf{u}' \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_2\mathbf{u})}{n-p+1} \right) \left[\frac{\partial^2 Z_t(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_2, \delta_2))}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \frac{\partial^2 \overline{Z(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_2, \delta_2))}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \right] \mathbf{u} \\
&\quad - \frac{\delta_1^2}{2n} \sum_{t=p+1}^n \mathbf{u}' \lambda \left(\frac{R_t(\boldsymbol{\alpha}_0 + n^{-1/2}\delta_2\mathbf{u})}{n-p+1} \right) \left[\frac{\partial^2 Z_t(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_2, \delta_1))}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} - \frac{\partial^2 \overline{Z(\boldsymbol{\alpha}_n^*(\mathbf{u}, \delta_2, \delta_1))}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \right] \mathbf{u}, \quad (5.33)
\end{aligned}$$

where the values of $\boldsymbol{\alpha}_n^*(\mathbf{u}, \cdot, \cdot)$ lie between $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_0 + n^{-1/2}\mathbf{u}$.

Lemma 5.7 For $\mathbf{u} \in \mathbb{R}^{p+q}$, let $S_n(\mathbf{u}) = D_n(\boldsymbol{\alpha}_0 + n^{-1/2}\mathbf{u}) - D_n(\boldsymbol{\alpha}_0)$ and $S(\mathbf{u}) = \mathbf{u}'\mathbf{N} + 2^{-1}\tilde{K}\mathbf{u}'\sigma^2\boldsymbol{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u}$, where $\mathbf{N} \sim N(\mathbf{0}, \tilde{J}\sigma^2\boldsymbol{\Gamma}(\boldsymbol{\alpha}_0))$. Then $S_n(\cdot) \xrightarrow{d} S(\cdot)$ on $C(\mathbb{R}^{p+q})$, the space of continuous functions on \mathbb{R}^{p+q} where convergence is equivalent to uniform convergence on every compact set.

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

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

we have

$$\sum_{k=1}^m \tilde{U}_n \left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m} \right) \leq S_n(\mathbf{u}) \leq \sum_{k=1}^m \tilde{V}_n \left(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m} \right) \quad (5.34)$$

by Theorem 2.1. Using (5.32), (5.33), and Lemmas 5.1, 5.5, and 5.6,

$$\begin{bmatrix} \tilde{U}_n(\mathbf{u}, 0, \frac{1}{m}) \\ \vdots \\ \tilde{U}_n(\mathbf{u}, \frac{m-1}{m}, 1) \\ \tilde{V}_n(\mathbf{u}, 0, \frac{1}{m}) \\ \vdots \\ \tilde{V}_n(\mathbf{u}, \frac{m-1}{m}, 1) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{0}{m^2}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u} \\ \vdots \\ \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{m-1}{m^2}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u} \\ \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{1}{m^2}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u} \\ \vdots \\ \frac{1}{m}\mathbf{u}'\mathbf{N} + \frac{m}{m^2}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u} \end{bmatrix}$$

on \mathbb{R}^{2m} . Hence,

$$\begin{bmatrix} \sum_{k=1}^m \tilde{U}_n(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}) \\ \sum_{k=1}^m \tilde{V}_n(\mathbf{u}, \frac{k-1}{m}, \frac{k}{m}) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathbf{u}'\mathbf{N} + \frac{m-1}{2m}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u} \\ \mathbf{u}'\mathbf{N} + \frac{m+1}{2m}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u} \end{bmatrix}$$

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

$$\mathbf{u}'\mathbf{N} + \frac{m-1}{2m}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u} \quad \text{and} \quad \mathbf{u}'\mathbf{N} + \frac{m+1}{2m}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u}$$

are both in an ϵ -neighborhood of $S(\mathbf{u}) = \mathbf{u}'\mathbf{N} + 2^{-1}\tilde{K}\mathbf{u}'\sigma^2\mathbf{\Gamma}(\boldsymbol{\alpha}_0)\mathbf{u}$. Thus, for any $\mathbf{u} \in \mathbb{R}^{p+q}$, $S_n(\mathbf{u}) \xrightarrow{d} S(\mathbf{u})$.

It can be shown similarly that all finite-dimensional distributions of $S_n(\cdot)$ converge to those of $S(\cdot)$.

Using (5.34) and an argument in Andrews (2003, pages 84–86), it can be shown that $\lim_{\delta \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}(\sup_{\mathbf{u}, \mathbf{v} \in K, \|\mathbf{u}-\mathbf{v}\| \leq \delta} |S_n(\mathbf{u}) - S_n(\mathbf{v})| > \eta) = 0$ for any $\eta > 0$ and any compact subset $K \subset \mathbb{R}^{p+q}$. Therefore, $S_n(\cdot)$ is tight on $C(K)$ for any compact set $K \subset \mathbb{R}^{p+q}$. It follows that $S_n(\cdot) \xrightarrow{d} S(\cdot)$ on $C(\mathbb{R}^{p+q})$ by Theorem 7.1 in Billingsley (1999). \square

Acknowledgment We thank an anonymous reviewer for his helpful comments.

References

ALLAL, J., KAAOUACHI, A., and PAINDAVEINE, D. (2001). R-estimation for ARMA models. *Journal of Nonparametric Statistics* 13, 815–831.

ANDREWS, B., DAVIS, R.A., and BREIDT, F.J. (2007). Rank-based estimation for all-pass time series models. *Annals of Statistics* 35, to appear.

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

BAI, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *Annals of Statistics* 22, 2051–2061.

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

BROCKWELL, P.J. and DAVIS, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

CHERNOFF, H. and SAVAGE, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Annals of Mathematical Statistics* 29, 972–994.

CHOI, Y.H. and ÖZTÜRK, Ö. (2002). A new class of score generating functions for regression models. *Statistics & Probability Letters* 57, 205–214.

DAVIS, R.A. and DUNSMUIR, W.T.M. (1997). Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability* 10, 481–497.

DAVIS, R.A., KNIGHT, K., and LIU, J. (1992). M -estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications* 40, 145–180.

FERRETTI, N.E., KELMANSKY, D.M., and YOHAI, V.J. (1991). Estimators based on ranks for ARMA models. *Communications in Statistics: Theory and Methods* 20, 3879–3907.

GASTWIRTH, J.L. and WOLFF, S.S. (1968). An elementary method for obtaining lower bounds on the asymptotic power of rank tests. *Annals of Mathematical Statistics* 39, 2128–2130.

HETTMANSPERGER, T.P. and MCKEAN, J.W. (1998). *Robust Nonparametric Statistical Methods*. Arnold, London.

HODGES, J.L. and LEHMANN, E.L. (1956). The efficiency of some nonparametric competitors of the t -test. *Annals of Mathematical Statistics* 27, 324–335.

JAECKEL, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics* 43, 1449–1458.

JUREČKOVÁ, J. and SEN, P.K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*. Wiley, New York.

KOUL, H.L. (1991). A weak convergence result useful in robust autoregression. *Journal of Statistical Planning and Inference* 29, 291–308.

KOUL, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed. Springer, New York.

KOUL, H.L. and OSSIANDER, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* 22, 540–562.

KOUL, H.L. and SALEH, A.K.MD.E. (1993). R-estimation of the parameters of autoregressive $[Ar(p)]$ models. *Annals of Statistics* 21, 534–551.

LAGARIAS, J.C., REEDS, J.A., WRIGHT, M.H., and WRIGHT, P.E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization* 9, 112–147.

LI, W.K. and MCLEOD, A.I. (1988). ARMA modelling with non-Gaussian innovations. *Journal of Time Series Analysis* 9, 155–168.

MCKEAN, J.W. and SIEVERS, G.L. (1989). Rank scores suitable for analyses of linear models under asymmetric error distributions. *Technometrics* 31, 207–218.

MUKHERJEE, K. and BAI, Z.D. (2002). R-estimation in autoregression with square-integrable score function. *Journal of Multivariate Analysis* 81, 167–186.

ROBINSON, P.M. (1987). Time series residuals with application to probability density estimation. *Journal of Time Series Analysis* 8, 329–344.

SILVERMAN, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.