Gaussian Maximum Likelihood Estimation For ARMA Models I: Time Series

Qiwei Yao
Department of Statistics
London School of Economics
London, WC2A 2AE, UK
Email: q.yao@lse.ac.uk

Peter J. Brockwell
Department of Statistics
Colorado State University
Fort Collins, CO 80532, USA
Email: pjbrock@stat.colostate.edu

Abstract

We provide a direct proof for consistency and asymptotic normality of Gaussian maximum likelihood estimators for causal and invertible ARMA time series models, which were initially established by Hannan (1973) via the asymptotic properties of a Whittle's estimator. This also paves the way to establish a similar results for spatial processes presented in the follow-up paper Yao and Brockwell (2001).

Key words and phrases: ARMA time series models, asymptotic normality, consistency, Gaussian maximum likelihood estimator, innovation algorithm, martingale-difference, prewhitening.

Running title: Gaussian MLE for ARMA time series

1 Introduction

Hannan (1973) established the asymptotic theory for the maximum likelihood estimator based on a Gaussian likelihood function for a general ARMA time series. The time series itself is not necessarily Gaussian. Under the condition that the process is causal and invertible with finite second moment, Hannan showed that the Gaussian maximum likelihood estimator is asymptotically normal and unbiased with the covariance depending on the autocorrelation function only. In fact the asymptotic variance of the estimated AR and MA coefficient vector can be nicely represented in terms of the two AR models; see Theorem 2 in §4 below. This is one of the most influential results in the classical time series. The result is simple and elegant. The imposed conditions are minimal. However its proof is indirect and is based on delicate frequency-domain arguments; see also §10.8 of Brockwell and Davis (1991). In fact, Hannan's proof essentially consists of two parts: a proof of asymptotic normality for Whittle's estimator, and a proof of the asymptotic equivalence of the Gaussian maximum likelihood estimator and Whittle's estimator.

Since the Gaussian maximum likelihood estimator is based on simple and intuitive likelihood argument, its asymptotic properties deserve a direct proof. This is the goal of our paper. We do not attempt to reproduce Hannan's result in the most general form. Instead we present the result in its most practically relevant form. Our proof is directly within the time-domain, expressing the likelihood function in terms of prewhitening. However this will not prevent us from using spectral density function occasionally to simplify some technical details. The added incentive to produce such a proof is to establish a similar result for spatial processes; see Yao and Brockwell (2001).

The paper is organised as follows. The model and estimators are defined in §2. §3 considers the consistency and §4 deals with asymptotic normality.

2 Model and estimators

The autoregressive and moving average (ARMA) process $\{X_t,\ t = 0, \pm 1, \pm 2, \cdots\}$ with orders (p, q) $(p, q \geq 0,\ p + q > 0)$ is a stationary process defined by

$$X_t - b_1 X_{t-1} - \cdots - b_p X_{t-p} = \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q}, \tag{2.1}$$

where $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$ is a sequence of uncorrelated random variables with mean 0 and variance $\sigma^2 \in (0, \infty)$, $\mathbf{b} = (b_1, \cdots, b_p)^\tau \in \mathcal{R}^p$ and $\mathbf{a} = (a_1, \cdots, a_q)^\tau \in \mathcal{R}^q$ are real parameters. We write

$\{X_t\} \sim \text{ARMA}(p, q)$.

It is convenient to represent model (2.1) in terms of the backshift operator $B^k X_t = X_{t-k}$ for $k = 0, \pm 1, \pm 2, \cdots$. To this end, we define

$$b(z) = 1 - b_1 z - \cdots - b_p z^p, \quad \text{and} \quad a(z) = 1 + a_1 z + \cdots + a_q z^q. \tag{2.2}$$

Then model (2.1) can be equivalently expressed as $b(B)X_t = a(B)\varepsilon_t$. We always assume that the orders p and q are genuine in the sense that polynomials $b(z)$ and $a(z)$ do not have common factors. It is well-known that (2.1) defines a unique stationary process $\{X_t\}$ if and only if $b(z) \neq 0$ for all complex z with $|z| = 1$. The process $\{X_t\}$ is *causal* if $b(z) \neq 0$ for all $|z| \leq 1$, and is *invertible* if $a(z) \neq 0$ for all $|z| \leq 1$.

It is easy to see that a stationary process $\{X_t\}$ defined by (2.1) has mean 0. If $b(z)a(z) \neq 0$ for all $|z| = 1$, there are 2^{p+q} stationary ARMA(p, q) models (with different $\mathbf{b}$ and $\mathbf{a}$) sharing the same autocorrelation function (ACF). To avoid the ambiguity, it is common practice to assume that $b(z) \neq 0$ for all $|z| \leq 1$, and $a(z) \neq 0$ for $|z| < 1$. This assumption guarantees that the parameters $\mathbf{b}$ and $\mathbf{a}$ are identifiable in terms of the ACF, which is a necessary condition in the context of Gaussian maximum likelihood estimation since the likelihood function depends on the parameter $(\mathbf{b}, \mathbf{a})$ through the ACF only.

In the sequel we also assume that in model (2.1) $(\mathbf{b}, \mathbf{a}) \in \mathcal{B}$, where

$$\mathcal{B} = \{(\mathbf{b}, \mathbf{a}) \in \mathcal{R}^p \times \mathcal{R}^q \;\; | \;\; b(z)a(z) \neq 0 \text{ for all complex } |z| \; \leq 1, \; b(\cdot) \text{ and } a(\cdot) \text{ do not have common factors}\}. \tag{2.3}$$

It is easy to see that $\mathcal{B}$ is an open subset in $\mathcal{R}^p \times \mathcal{R}^q$. Given the observations $X_1, \cdots, X_T$, the Gaussian likelihood function is of the form

$$L(\mathbf{b}, \mathbf{a}, \sigma^2) \; \propto \; \sigma^{-T} |\mathbf{\Sigma}|^{-1/2} \exp\{-\frac{1}{2\sigma^2} \mathbf{X}_T^\tau \mathbf{\Sigma}^{-1} \mathbf{X}_T\}, \tag{2.4}$$

where $\mathbf{X}_T = (X_1, \cdots, X_T)^\tau$ and

$$\mathbf{\Sigma} \equiv \mathbf{\Sigma}(\mathbf{b}, \mathbf{a}) = \frac{1}{\sigma^2} \text{Var}(\mathbf{X}_T),$$

which is independent of σ^2. Note that for $(\mathbf{b}, \mathbf{a}) \in \mathcal{B}$, the autocovariance function (ACVF) $\gamma(k) = \text{Cov}(X_t, X_{t-k}) \to 0$ at an exponential rate, and therefore the inverse of $\mathbf{\Sigma}(\mathbf{b}, \mathbf{a})$ always exists (Proposition 5.1.1 of Brockwell and Davis 1991). The estimators which maximise (2.4) can be expressed as

$$(\widehat{\mathbf{b}}, \widehat{\mathbf{a}}) = \arg\min_{(\mathbf{b},\mathbf{a})\in\mathcal{B}} \left(\log\{\mathbf{X}_T^\tau \mathbf{\Sigma}(\mathbf{b},\mathbf{a})^{-1}\mathbf{X}_T\} + T^{-1}\log|\mathbf{\Sigma}(\mathbf{b},\mathbf{a})|\right), \quad \widehat{\sigma}^2 = \mathbf{X}_T^\tau \mathbf{\Sigma}(\widehat{\mathbf{b}}, \widehat{\mathbf{a}})^{-1}\mathbf{X}_T/T.$$

First we establish the consistency of the estimators under the less restrictive assumptions about $\{\varepsilon_t\}$ in next section. The asymptotic normality will be proved in Section 4.

3 Consistency

The model (2.1) can be reparametrised in terms of reciprocals of the roots of equations $b(z) = 0$ and $a(z) = 0$, which form the parameter space Θ. Since $\mathcal{B}$ is open, Θ is an open subset of

$$\{\boldsymbol{\theta} \equiv (\beta_1, \cdots, \beta_p, \alpha_1, \cdots, \alpha_q)^\tau \in \mathcal{C}^{p+q} \quad | \quad 0 < |\beta_j| < 1,\ 0 < |\alpha_k| < 1, \text{and } \beta_j \neq \alpha_k \\ \text{for all } 1 \le j \le p \text{ and } 1 \le k \le q\},$$

where $\mathcal{C}$ denotes the complex number space. Thus the closure $\overline{\Theta}$ is compact.

Let $\boldsymbol{\theta}_0 \in \Theta$ be the true value of model (2.1); corresponding to $(\mathbf{b}_0, \mathbf{a}_0) \in \mathcal{B}$. Let $\widehat{\boldsymbol{\theta}} \in \Theta$ be the estimator which maximises (2.4) with $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$, namely

$$\widehat{\boldsymbol{\theta}} \equiv \widehat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\Theta}[\frac{1}{T}\log|\mathbf{\Sigma}(\boldsymbol{\theta})| + \log\{\frac{1}{T}\mathbf{X}_T^\tau \mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}_T\}]. \tag{3.1}$$

It is easy to see that $(\widehat{\mathbf{b}}, \widehat{\mathbf{a}}) \xrightarrow{a.s.} (\mathbf{b}_0, \mathbf{a}_0)$ if and only if $\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$.

We denote $b_0(\cdot)$ and $a_0(\cdot)$ the polynomials defined as in (2.2) with the coefficients corresponding to the true value $\boldsymbol{\theta}_0$, and $b(\cdot)$ and $a(\cdot)$ the polynomials corresponding to $\boldsymbol{\theta}$.

Theorem 1. Let $\{X_t\}$ be the stationary process defined by (2.1) in which $\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2)$ with $\sigma^2 > 0$, and the true value $(\mathbf{b}_0, \mathbf{a}_0) \in \mathcal{B}$. As $T \to \infty$, $\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$ and $\widehat{\sigma}^2 \xrightarrow{a.s.} \sigma^2$.

Proof. The condition $(\mathbf{b}_0, \mathbf{a}_0) \in \mathcal{B}$ implies that $\boldsymbol{\theta}_0 \in \Theta$. From (3.1) and Lemmas 1 and 2 below, it holds almost surely that

$$\limsup_{T\to\infty} \frac{1}{T}\mathbf{X}_T^\tau \mathbf{\Sigma}(\widehat{\boldsymbol{\theta}}_T)^{-1}\mathbf{X}_T \le \lim_{T\to\infty} \frac{1}{T}\mathbf{X}_T^\tau \mathbf{\Sigma}(\boldsymbol{\theta}_0)^{-1}\mathbf{X}_T = \sigma^2. \tag{3.2}$$

Define $B = \{\liminf_{T\to\infty} |\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| > 0\}$. For any $\omega \in B$, there exists a subsequence of $\{T\}$, which we still denote as $\{T\}$, for which $\widehat{\boldsymbol{\theta}}_T(\omega) \to \boldsymbol{\theta} \in \overline{\Theta}$ and $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Lemma 4 below and (3.2) ensure that $\boldsymbol{\theta} \in \Theta$. By Lemma 3(i) below, we have that for any $\epsilon > 0$,

$$\frac{1}{T}|\mathbf{X}_T^\tau \mathbf{\Sigma}(\widehat{\boldsymbol{\theta}}_T(\omega))^{-1}\mathbf{X}_T - \mathbf{X}_T^\tau \mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}_T| \le \epsilon\widehat{\gamma}(0)$$

for all sufficiently large T's, where $\widehat{\gamma}(0) = \sum_{j=1}^T X_j^2/T$. Thus

$$\lim_{T\to\infty} \frac{1}{T}\mathbf{X}_T^\tau \mathbf{\Sigma}(\widehat{\boldsymbol{\theta}}_T(\omega))^{-1}\mathbf{X}_T = \lim_{T\to\infty} \frac{1}{T}\mathbf{X}_T^\tau \mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}_T, \tag{3.3}$$

provided one of the above two limits exist. Now Lemma 2 and (3.2) imply that $P(B) = 0$, *i.e.* $\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$. It follows from Lemmas 3(i) and 2 that $\widehat{\sigma}^2 = \mathbf{X}_T^\tau \mathbf{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}_T/T \xrightarrow{a.s.} \sigma^2$. ■

In the sequel we always assume that the condition of Theorem 1 holds.

Lemma 1. $\log|\Sigma(\boldsymbol{\theta})| > 0$ for any $\boldsymbol{\theta} \in \overline{\Theta}$, and $T^{-1}\log|\Sigma(\boldsymbol{\theta})| \to 0$ for any $\boldsymbol{\theta} \in \Theta$.

Proof. First we assume $\boldsymbol{\theta} \in \Theta$. Let $\{Y_t\}$ be the stationary process defined by $b(B)Y_t = a(B)e_t$ with $\{e_t\} \sim \text{IID}(0,1)$. It is easy to see from (2.4) that $\text{Var}(\mathbf{Y}_T) = \mathbf{\Sigma}(\boldsymbol{\theta})$, where $\mathbf{Y}_T = (Y_1, \cdots, Y_T)^\tau$. For $k = 2, \cdots, T$, let

$$\widehat{Y}_{k+1} = \varphi_{k1}Y_k + \cdots + \varphi_{kk}Y_1 \tag{3.4}$$

be the best linear predictor for Y_{k+1} from $Y_k, \cdots, Y_1$ in the sense that

$$E(Y_{k+1} - \widehat{Y}_{k+1})^2 = \min_{\psi_j} E\{Y_{k+1} - \sum_{j=1}^k \psi_j Y_{k-j+1}\}^2. \tag{3.5}$$

Note that $\{Y_t\}$ is invertible, *i.e.* $Y_{t+1} = e_{t+1} - \{1 - a(B)^{-1}b(B)\}Y_{t+1} \equiv e_{t+1} + \sum_{j=1}^\infty \varphi_j Y_{t+1-j}$ with $\sum_{1\le j\le\infty}|\varphi_j| < \infty$. Since e_{t+1} is uncorrelated with $\{Y_{t+1-j}, j \ge 1\}$, it is easy to see that

$$\text{Var}(e_{t+1}) = E\{Y_{t+1} - \sum_{j=1}^\infty \varphi_j Y_{t+1-j}\}^2 = \min E\{Y_{t+1} - \sum_{j=1}^\infty \psi_j Y_{t+1-j}\}^2, \tag{3.6}$$

where the minimum is taken over all $\{\psi_j\}$ such that $\sum_{1\le j\le\infty}|\psi_j| < \infty$. Therefore, it holds for all $1 \le k \le T$,

$$r_k \equiv \text{Var}(Y_{k+1} - \widehat{Y}_{k+1}) \ge \text{Var}(e_{k+1}) = 1. \tag{3.7}$$

Let $\widehat{Y}_1 \equiv 0$, then (3.4) can be equivalently represented in the form $\widehat{Y}_{k+1} = \sum_{j=1}^k h_{kj}(Y_{k+1-j} - \widehat{Y}_{k+1-j})$ for $1 \le k < T$. This implies that

$$\mathbf{Y}_t = \mathbf{H}(\mathbf{Y}_T - \widehat{\mathbf{Y}}_T), \tag{3.8}$$

where $\mathbf{H}$ is a $T \times T$ lower triangular matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ h_{11} & 1 & 0 & \cdots & 0 \\ h_{22} & h_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ h_{T-1,T-1} & h_{T-1,T-2} & h_{T-1,T-3} & \cdots & 1 \end{pmatrix}. \tag{3.9}$$

Note that the least square property implies that the residual $Y_{k+1} - \widehat{Y}_{k+1}$ is uncorrelated with $\{Y_t, 1 \le t \le k\}$, and therefore is also uncorrelated with $\{Y_t - \widehat{Y}_t, 1 \le t \le k\}$. Hence,

$$|\mathbf{\Sigma}(\theta)| = |\mathrm{Var}(\mathbf{Y}_t)| = |\mathbf{H}\,\mathrm{Var}(\mathbf{Y}_T - \widehat{\mathbf{Y}}_T)\mathbf{H}^\tau| = |\mathrm{Var}(\mathbf{Y}_T - \widehat{\mathbf{Y}}_T)| = \prod_{k=1}^{T} r_{k-1} > 1.$$

The inequality in the above expression follows from (3.7) and the fact that $r_0 = \mathrm{Var}(Y_1 - \widehat{Y}_1) = \mathrm{Var}(Y_1) > \mathrm{Var}(e_1) = 1$ (as $p + q \ge 1$). Thus $\log|\mathbf{\Sigma}(\boldsymbol{\theta})| > 0$ for any $\boldsymbol{\theta} \in \Theta$. From (3.5) and (3.6) we may see that $r_k \to 1$ as $k \to \infty$. Thus

$$\frac{1}{T}\log|\mathbf{\Sigma}(\boldsymbol{\theta})| = \frac{1}{T}\sum_{k=1}^{T}\log r_{k-1} \to 0.$$

For $\boldsymbol{\theta} \in \overline{\Theta} - \Theta$, there exists a sequence $\{\boldsymbol{\theta}_j\} \subset \Theta$ and $\boldsymbol{\theta}_j \to \theta$. Thus it follows an obvious asymptotic argument that $|\mathbf{\Sigma}(\boldsymbol{\theta})|/\mathrm{Var}(Y_1) \ge 1$. Since $\mathrm{Var}(Y_1) > 1$, we have $\log|\mathbf{\Sigma}(\boldsymbol{\theta})| > 0$. ■

Lemma 2. For any $\boldsymbol{\theta} \in \Theta$,

$$\mathbf{X}_T^\tau \mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}_T/T \xrightarrow{a.s.} \mathrm{Var}\{a(B)^{-1}b(B)X_t\} \ge \mathrm{Var}(\varepsilon_t) = \sigma^2,$$

and the equality holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof. Let $\{Y_t\}$ be the same process as defined in the proof of Lemma 1, and $\{\varphi_{kj}\}$ and $\{\varphi_j\}$ be the same as in (3.4) and (3.6). Define $\widetilde{X}_1 \equiv 0$, $\widetilde{X}_{k+1} = \varphi_{k1}X_k + \cdots + \varphi_{kk}X_1$ for $k \ge 1$ and $\widetilde{\mathbf{X}}_T = (\widetilde{X}_1, \cdots, \widetilde{X}_T)^\tau$. It follows from (3.8) that $\mathbf{X}_T = \mathbf{H}(\mathbf{X}_T - \widetilde{\mathbf{X}}_T)$. Since $\mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{H}\,\mathrm{Var}(\mathbf{Y}_T - \widehat{\mathbf{Y}}_T)\mathbf{H}^\tau$, we have

$$\frac{1}{T}\mathbf{X}_T^\tau\mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}_T = \frac{1}{T}(\mathbf{X}_T - \widetilde{\mathbf{X}}_T)^\tau\{\mathrm{Var}(\mathbf{Y}_T - \widehat{\mathbf{Y}}_T)\}^{-1}(\mathbf{X}_T - \widetilde{\mathbf{X}}_T) = \frac{1}{T}\sum_{k=1}^{T}(X_k - \widetilde{X}_k)^2/r_{k-1},$$

where $r_k = \mathrm{Var}(Y_{k+1} - \widehat{Y}_{k+1})$. For any $\epsilon > 0$, choose $K \ge 1$ such that

$$E(\sum_{j=K+1}^{\infty}|\varphi_j X_{1-j}|)^2 < \epsilon. \tag{3.10}$$

Write for $k \ge K$

$$\begin{aligned} X_k - \widetilde{X}_k &= (X_k - \sum_{j=1}^{K}\varphi_j X_{k-j}) + \sum_{j=1}^{k}(\varphi_j - \varphi_{kj})X_{k-j} - \sum_{j=K+1}^{k}\varphi_j X_{k-j} \\ &\equiv \eta_{k1} + \eta_{k2} + \eta_{k3}, \quad \text{say.} \end{aligned} \tag{3.11}$$

It follows from the ergodic theorem that $T^{-1}\sum_{K \le k \le T}\eta_{k1}^2 \xrightarrow{a.s.} E(\eta_{t1}^2)$. By (3.10),

$$|E(\eta_{t1}^2) - E\{\varphi(B)X_t\}^2| < \epsilon + 2\{\epsilon\, E\eta_{t1}^2\}^{1/2}.$$

Thus

$$|\lim_{T\to\infty}\frac{1}{T}\sum_{k=K}^{T}\eta_{k1}^2 - E\{\varphi(B)X_t\}^2| < \epsilon + 2\{\epsilon\, E\eta_{t1}^2\}^{1/2} \quad \text{a.s..} \tag{3.12}$$

Again by the ergodic theorem and (3.10), we have that

$$\lim_{T\to\infty}\frac{1}{T}|\sum_{k=K}^{T}\eta_{k3}^2| \le \lim_{T\to\infty}\frac{1}{T}\sum_{k=K}^{T}\left(\sum_{j=K+1}^{\infty}|\varphi_j X_{k-j}|\right)^2 \xrightarrow{a.s.} E(\sum_{j=K+1}^{\infty}|\varphi_j X_{1-j}|)^2 < \epsilon. \tag{3.13}$$

Let $\boldsymbol{\varphi}_k = (\varphi_{k1}, \cdots, \varphi_{kk})^\tau$, $\widetilde{\boldsymbol{\varphi}}_t = (\varphi_1, \cdots, \varphi_k)^\tau$, and $\boldsymbol{\Sigma}_k = \text{Var}(\mathbf{X}_k)$. Then

$$E(\eta_{k2}^2) = (\widetilde{\boldsymbol{\varphi}}_k - \boldsymbol{\varphi}_k)^\tau \boldsymbol{\Sigma}_k (\widetilde{\boldsymbol{\varphi}}_k - \boldsymbol{\varphi}_k) \le \lambda_{\max}||\widetilde{\boldsymbol{\varphi}}_k - \boldsymbol{\varphi}_k||^2,$$

where $\lambda_{\max}$ is the maximal eigenvalue of $\boldsymbol{\Sigma}_k$ which is finite; see Proposition 4.5.3 of Brockwell and Davis (1991). By (10.8.50) of Brockwell and Davis (1991)*, $||\widetilde{\boldsymbol{\varphi}}_k - \boldsymbol{\varphi}_k|| \to 0$ as $k \to \infty$. Thus $E(\eta_{k2}^2) \to 0$. Based on the MA(∞) representation of $\{X_t\}$, it holds that $\eta_{k2} = \sum_{j=1}^{\infty} d_{kj}\varepsilon_{k-j}$ with $\sum_j |d_{kj}| < \infty$. By the Loève Theorem (Corollary 3, p.117 of Chow and Teicher 1997), η_{k2} converges almost surely as $k \to \infty$. Since $\eta_{k2} \xrightarrow{P} 0$, it holds that $\eta_{k2} \xrightarrow{a.s.} 0$. Consequently it also holds that,

$$\frac{1}{T}\sum_{k=1}^{K}\eta_{k2}^2 \xrightarrow{a.s.} 0. \tag{3.14}$$

Note that

$$|\frac{1}{T}\sum_{k=K}^{T}\eta_{ki}\eta_{kj}| \le \left(\frac{1}{T}\sum_{k=K}^{T}\eta_{ki}^2\,\frac{1}{T}\sum_{k=K}^{T}\eta_{kj}^2\right)^{1/2}.$$

It follows from (3.11) to (3.14) that

$$\begin{aligned}\frac{1}{T}\sum_{k=1}^{T}(X_k - \widetilde{X}_k)^2 &\xrightarrow{a.s.} E\{\varphi(B)X_t\} = \text{Var}\{a(B)^{-1}b(B)X_t\} \\ &= \text{Var}\{a(B)^{-1}b(B)b_0(B)^{-1}a_0(B)\varepsilon_t\} \ge \text{Var}(\varepsilon_t).\end{aligned} \tag{3.15}$$

The required result now follows from the fact that $r_k \to \text{Var}(e_{k+1}) = 1$, which is guaranteed by (3.5) and (3.6). ■

Lemma 3. Let $\boldsymbol{\theta}_k \in \Theta$ and $\boldsymbol{\theta}_k \to \boldsymbol{\theta}$ as $k \to \infty$. Let $\epsilon > 0$ be an arbitrary constant.

(i) If $\boldsymbol{\theta} \in \Theta$, it holds for all $T \ge 1$ and all sufficiently large k's that

$$|\mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1}\mathbf{x} - \mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{x}| \le \epsilon, \quad \mathbf{x} \in \Re^T \text{ and } ||\mathbf{x}|| = 1.$$

*There are typos in Brockwell and Davis' book: s^{-t} on p.395–6 should be s^t.

(ii) If $\boldsymbol{\theta} \in \overline{\Theta} - \Theta$ and $b(z) \neq 0$ for all $|z| \leq 1$, there exists a $\boldsymbol{\theta}_* \in \Theta$ and $||\boldsymbol{\theta}_* - \boldsymbol{\theta}||$ arbitrarily small for which it holds for all $T \geq 1$ and all sufficiently large k's that

$$\mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1}\mathbf{x} \geq \mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_*)^{-1}\mathbf{x} - \epsilon, \quad \mathbf{x} \in \mathcal{R}^T \text{ and } ||\mathbf{x}|| = 1.$$

Proof. (i) Let $g(\omega, \boldsymbol{\theta}) = \frac{\sigma^2}{2\pi}|\frac{a(e^{i\omega})}{b(e^{i\omega})}|^2$, the spectral density of the process defined by (2.1). Then $g(\omega, \cdot)$ is continuous and bounded away from both 0 and ∞ on any compact sets contained in Θ. Hence for any $\epsilon' > 0$, it holds for all large k's that

$$\sup_{\omega \in [-\pi, \pi)} |g(\omega, \boldsymbol{\theta}_k) - g(\omega, \boldsymbol{\theta})| < \epsilon'.$$

Therefore, it holds that for any $\mathbf{x} = (x_1, \cdots, x_T)^\tau, \mathbf{y} = (y_1, \cdots, y_T)^\tau \in \mathcal{R}^T$ and $||\mathbf{x}|| = ||\mathbf{y}|| = 1$,

$$\begin{aligned} |\mathbf{x}^\tau \{\boldsymbol{\Sigma}(\boldsymbol{\theta}_k) - \boldsymbol{\Sigma}(\boldsymbol{\theta})\}\mathbf{y}| &= |\int_{-\pi}^{\pi} \sum_{j=1}^{T} x_j e^{ij\omega} \sum_{k=1}^{T} y_k e^{-ik\omega} \{g(\omega, \boldsymbol{\theta}_k) - g(\omega, \boldsymbol{\theta})\} d\omega| \\ &\leq \frac{\epsilon'}{2} \int_{-\pi}^{\pi} (|\sum_{j=1}^{T} x_j e^{ij\omega}|^2 + |\sum_{k=1}^{T} y_k e^{ik\omega}|^2) d\omega = 2\pi\epsilon'. \end{aligned}$$

When $\boldsymbol{\theta} \in \Theta$, the minimum eigenvalue of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ (for all $T \geq 1$) is bounded from below by a constant $K^{-1} > 0$; see Proposition 4.5.3 of Brockwell and Davis (1991). Hence,

$$\begin{aligned} |\mathbf{x}^\tau \{\boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\}\mathbf{x}| &= |\mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \{\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)\} \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1}\mathbf{x}| \\ &\leq 2\pi\epsilon' \{|\mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\mathbf{x}|\, |\mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-2}\mathbf{x}|\}^{1/2} \leq 2\pi K^2 \epsilon'. \end{aligned} \quad (3.16)$$

Now (i) holds by choosing $\epsilon' = \epsilon/(2\pi K^2)$.

(ii) When $\boldsymbol{\theta} \in \overline{\Theta} - \Theta$, let $\boldsymbol{\theta}_* = \boldsymbol{\theta}_{k_0} \in \Theta$ for a fixed large k_0. Then for all T and large k,

$$|\mathbf{x}^\tau \{\boldsymbol{\Sigma}(\boldsymbol{\theta}_k) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_*)\}\mathbf{y}| \leq |\mathbf{x}^\tau \{\boldsymbol{\Sigma}(\boldsymbol{\theta}_k) - \boldsymbol{\Sigma}(\boldsymbol{\theta})\}\mathbf{y}| + |\mathbf{x}^\tau \{\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_{k_0})\}\mathbf{y}| < 2\epsilon'.$$

By using the same argument as (3.16) twice, we have for any $\eta > 0$ that

$$\begin{aligned} \mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1}\mathbf{x} &> \mathbf{x}^\tau \{\boldsymbol{\Sigma}(\boldsymbol{\theta}_k) + \eta\, \mathbf{I}_T\}^{-1}\mathbf{x} \geq \mathbf{x}^\tau \{\boldsymbol{\Sigma}(\boldsymbol{\theta}_*) + \eta\, \mathbf{I}_T\}^{-1}\mathbf{x} - 2\epsilon'/\eta^2 \\ &\geq \mathbf{x}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_*)^{-1}\mathbf{x} - 2\eta\, K_1^2 - 2\epsilon'/\eta^2, \end{aligned}$$

where $\mathbf{I}_T$ denotes the $T \times T$ identity matrix, and $K_1^{-1} > 0$ is a lower bound of the eigenvalues for $\boldsymbol{\Sigma}(\boldsymbol{\theta}_*)$. The proof of (ii) is completed by choosing $\eta = \epsilon/(4K_1^2)$ and $\epsilon' = \eta^2\epsilon/4$. ■

Lemma 4. Let $\boldsymbol{\theta}_T \in \Theta$ and $\boldsymbol{\theta}_T \to \boldsymbol{\theta}' \in \overline{\Theta} - \Theta$. Then it holds almost surely that

$$\liminf_{T \to \infty} \mathbf{X}_T^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_T)^{-1}\mathbf{X}_T > \sigma^2.$$

Proof. First we assume that $b'(z) \neq 0$ for all $|z| \leq 1$, where $b'(\cdot)$ and $a'(\cdot)$ are polynomials defined as in (2.2) corresponding to $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

It follows from Lemma 2 that for any $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$,

$$\text{Var}\{a(B)^{-1}b(B)X_t\} = \text{Var}\{a(B)^{-1}b(B)b_0(B)^{-1}a_0(B)\varepsilon_t\} > \text{Var}(\varepsilon_t) = \sigma^2.$$

Note that

$$a(B)^{-1}b(B)b_0(B)^{-1}a_0(B)\varepsilon_t = \varepsilon_t + \sum_{j=1}^{\infty} c_j \varepsilon_{t-1},$$

which is equal to ε_t if and only if $a(z)^{-1}b(z)b_0(z)^{-1}a_0(z) \equiv 1$, *i.e.* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Let $O(\boldsymbol{\theta}_0, \delta) = \{\boldsymbol{\theta} \in \Theta| \, ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| > \delta\}$. Then for any $\delta > 0$, there exists an $\epsilon > 0$ such that

$$\inf_{\boldsymbol{\theta} \in \Theta - O(\boldsymbol{\theta}_0, \delta)} \text{Var}\{a(B)^{-1}b(B)X_t\} > \sigma^2 + \epsilon.$$

(Otherwise, $\boldsymbol{\theta}_0$ must be in the closure of $O(\boldsymbol{\theta}_0, \delta)$.) By Lemmas 3(ii) and 2, there exists a $\boldsymbol{\theta}_* \in \Theta - O(\boldsymbol{\theta}_0, \delta)$ such that

$$\begin{aligned} \liminf_{T\to\infty} \frac{1}{T}\mathbf{X}_T^{\tau}\boldsymbol{\Sigma}(\boldsymbol{\theta}_T)^{-1}\mathbf{X}_T \;&\geq\; \lim_{T\to\infty}\{\frac{1}{T}\mathbf{X}_T^{\tau}\boldsymbol{\Sigma}(\boldsymbol{\theta}_*)^{-1}\mathbf{X}_T - \frac{1}{2\text{Var}(X_t)}\epsilon\,\widehat{\gamma}(0)\} \\ =\;& \text{Var}\{a_*(B)^{-1}b_*(B)X_t\} - \epsilon/2 \;\geq\; \sigma^2 + \epsilon/2 \;>\; \sigma^2, \quad \text{a.s.} \end{aligned} \tag{3.17}$$

In the above expression, $\widehat{\gamma}(0) = \mathbf{X}_T^{\tau}\mathbf{X}_t/T$.

Now consider the case when $b'(z) = 0$ has unit roots. To avoid cumbersome notation, we assume that there is only one unit root $1/\theta_1'$. (The cases with multiple unit roots can be dealt with in the same manner.) Define $U_t = X_t - \theta_1' X_{t-1}$ and $\mathbf{U}_{2,T} = (U_2, \cdots, U_T)^{\tau}$. Then

$$\mathbf{X}_T^{\tau}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}_T = \mathbf{U}_{2,T}^{\tau}\widetilde{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{U}_{2,T} + \xi_T^2, \tag{3.18}$$

where $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) = \text{Var}(\mathbf{U}_{2,T})$, $\boldsymbol{\gamma}(\boldsymbol{\theta}_0) = \text{Cov}(\mathbf{U}_{2,T}, X_1)$, $\gamma(\boldsymbol{\theta}_0) = \text{Var}(X_1)$, and

$$\xi_T^2 = (X_1 - \boldsymbol{\gamma}(\widehat{\boldsymbol{\theta}})^{\tau}\widetilde{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{U}_{2,T})^2/\{\gamma(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\gamma}(\widehat{\boldsymbol{\theta}})^{\tau}\widetilde{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{\gamma}(\widehat{\boldsymbol{\theta}})\}.$$

Let $b_u'(z) = b'(u)/(1 - \theta_u')$. Note that the limiting point $\boldsymbol{\theta}'$ defines a causal process $b_u'(B)U_t = a'(B)\varepsilon_t$ with $b_u'(z) \neq 0$ for all $|z| \leq 1$. Replacing $\{X_t\}$ in (3.17) by $\{U_t\}$, we have that

$$\liminf_{T\to\infty} \frac{1}{T}\mathbf{U}_{2,T}^{\tau}\widetilde{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{U}_{2,T} > \sigma^2, \quad \text{a.s.}.$$

It follows from (3.18) now that $\liminf_T \mathbf{X}_T^{\tau}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}_T/T > \sigma^2$ almost surely. ∎

4 Asymptotic normality

To state the asymptotic normality of the estimator $(\widehat{\mathbf{b}}, \widehat{\mathbf{a}})$, we introduce some notation. Let $\{W_t\} \sim \mathrm{WN}(0,1)$ be a white noise process with mean 0 and variance 1. Define

$$b(B)\xi_t = W_t \qquad \text{and} \qquad a(B)\zeta_t = W_t. \tag{4.1}$$

Let $\boldsymbol{\xi} = (\xi_{-1}, \cdots, \xi_{-p}, \zeta_{-1}, \cdots, \zeta_{-q})^\tau$, and

$$\mathbf{W}(\mathbf{b}, \mathbf{a}) = \{\mathrm{Var}(\boldsymbol{\xi})\}^{-1}. \tag{4.2}$$

Theorem 2. Let $\{X_t\}$ be the stationary process defined by (2.1) in which $\{\varepsilon_t\} \sim \mathrm{IID}(0, \sigma^2)$ with $\sigma^2 > 0$, and the true value $(\mathbf{b}_0, \mathbf{a}_0) \in \mathcal{B}$. Then as $T \to \infty$,

$$T^{\frac{1}{2}} \begin{pmatrix} \widehat{\mathbf{b}} - \mathbf{b}_0 \\ \widehat{\mathbf{a}} - \mathbf{a}_0 \end{pmatrix} \xrightarrow{D} N(\mathbf{0}, \mathbf{W}(\mathbf{b}_0, \mathbf{a}_0)).$$

Before we present the proof, we introduce some notation. Let $\boldsymbol{\beta} \equiv (\beta_1, \cdots, \beta_{p+q})^\tau = (\mathbf{b}^\tau, \mathbf{a}^\tau)^\tau$, $\widehat{\boldsymbol{\beta}} = (\widehat{\mathbf{b}}^\tau, \widehat{\mathbf{a}}^\tau)^\tau$ and $\boldsymbol{\beta}_0 = (\mathbf{b}_0^\tau, \mathbf{a}_0^\tau)^\tau$. Denote by $\boldsymbol{\Gamma}_k$ the $k \times k$ matrix with the (i,j)-th element $\gamma(j-i) \equiv \mathrm{Cov}(X_i, X_j)$ and $\boldsymbol{\gamma}_k$ the vector $(\gamma(1), \cdots, \gamma(k))^\tau$. We also write $\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k(\boldsymbol{\beta})$, $\boldsymbol{\gamma}_k = \boldsymbol{\gamma}_k(\boldsymbol{\beta})$ and etc. to indicate the dependence on the value of parameter $\boldsymbol{\beta}$. For $t \geq 1$ let

$$\widehat{X}_{t+1} = \varphi_{t1} X_t + \cdots + \varphi_{tt} X_1$$

be the best linear predictor for X_{t+1} based on $X_t, \cdots, X_1$ under model (2.1) and $\widehat{X}_1 \equiv 0$. Then

$$\boldsymbol{\varphi}_t \equiv (\varphi_{t1}, \cdots, \varphi_{tt})^\tau = \boldsymbol{\Gamma}_t^{-1} \boldsymbol{\gamma}_t, \tag{4.3}$$

and the variance of $X_{t+1} - \widehat{X}_{t+1}$ under model (2.1) is $\sigma^2 r_t \equiv \sigma^2 r_t(\boldsymbol{\beta}) = \gamma(0) - \boldsymbol{\varphi}_t^\tau \boldsymbol{\gamma}_t$. Let $X_{-k} \equiv 0$ and $Z_{-k} \equiv 0$ for all $k \geq 0$, and define for $t = 1, \cdots, T$

$$Z_t \equiv Z(\boldsymbol{\beta}) = X_t - b_1 X_{t-1} - \cdots - b_p X_{t-p} - a_1 Z_{t-1} - \cdots - a_q Z_{t-q}. \tag{4.4}$$

We write for $1 \leq j \leq p$ and $1 \leq i \leq q$

$$U_{tj} = -\frac{\partial Z_t}{\partial b_j} \qquad \text{and} \qquad V_{ti} = -\frac{\partial Z_t}{\partial a_i}. \tag{4.5}$$

Write $\mathcal{Y} = (X_1, \cdots, X_T)^\tau$ and $\mathcal{Z} = (Z_1(\boldsymbol{\beta}_0), \cdots, Z_T(\boldsymbol{\beta}_0))^\tau$. Let $\mathcal{X} = (\mathbf{X}, \mathbf{Z})$ and $\mathcal{U} = (\mathbf{U}, \mathbf{V})$, where $\mathbf{X}$ and $\mathbf{U}$ are the $T \times p$ matrices with X_{i-j} and $U_{ij}(\boldsymbol{\beta}_0)$ as their (i,j)-th elements respectively, and $\mathbf{Z}$ and $\mathbf{V}$ are $T \times q$ matrices with $Z_{i-j}(\boldsymbol{\beta}_0)$ and $V_{ij}(\boldsymbol{\beta}_0)$ as their (i,j)-th elements respectively. Denote by $\mathcal{R}$ the diagonal matrix $\text{diag}(r_0(\boldsymbol{\beta}_0), \cdots, r_{T-1}(\boldsymbol{\beta}_0))$.

We split the proof into several lemmas. In the sequel we always assume that the condition of Theorem 2 holds.

Lemma 5. For $k = 1, \cdots, p+q$,

$$T^{-1/2}\left\{\left|\frac{\partial}{\partial \beta_k}\sum_{t=1}^{T}\log r_{t-1}\right| + \left|\sum_{t=1}^{T}\frac{(X_t - \widehat{X}_t)^2}{r_{t-1}^2}\frac{\partial r_{t-1}}{\partial \beta_k}\right|\right\}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \xrightarrow{P} 0.$$

Proof. It follows the argument in p.394–6 of Brockwell and Davis (1991) that for $\boldsymbol{\beta} \in \mathcal{B}$ and $1 \le k \le p+q$,

$$|\partial r_t(\boldsymbol{\beta})/\partial \beta_k| \le C(\boldsymbol{\beta})\{s(\boldsymbol{\beta})\}^t, \quad t \ge 1,$$

where $C(\cdot) > 0$ and $s(\cdot) \in (0,1)$ are continuous By Theorem 1, $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0 \in \mathcal{B}$. Hence for any $\epsilon > 0$, there exists a sub-sample space A with $P(A) > 1 - \epsilon$ and $||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| < \epsilon$ on A for all large T's. Therefore, we may choose $C_1 \in (0, \infty)$ and $s_1 \in (0,1)$ for which $|\partial r_t(\boldsymbol{\beta})/\partial \beta_k|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \le C_1 s_1^t$ on A. Since $r_t(\boldsymbol{\beta}) \ge 1$ for all $\boldsymbol{\beta} \in \mathcal{B}$, it holds on the set A that

$$T^{-1/2}\left|\frac{\partial}{\partial \beta_k}\sum_{t=1}^{T}\log r_{t-1}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \le T^{-1/2}\sum_{t=1}^{T}\frac{1}{r_{t-1}^2}\left|\frac{\partial r_{t-1}}{\partial \beta_k}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \le C_1 T^{-1/2}\sum_{t=1}^{T}s_1^t \to 0.$$

Thus $T^{-1/2}\left|\frac{\partial}{\partial \beta_k}\sum_{t=1}^{T}\log r_{t-1}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \xrightarrow{P} 0.$

On the other hand,

$$T^{-1/2}E\left(\left|\sum_{t=1}^{T}\frac{(X_t - \widehat{X}_t)^2}{r_{t-1}^2}\frac{\partial r_{t-1}}{\partial \beta_k}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} I(A)\right) \le C C_1 T^{-1/2}\sum_{t=1}^{T}s_1^t \to 0,$$

where $I(\cdot)$ is the indicator function and $C > 0$ is a constant. Thus, $T^{-1/2}\left|\sum_{t=1}^{T}\frac{(X_t - \widehat{X}_t)^2}{r_{t-1}^2}\frac{\partial r_{t-1}}{\partial \beta_k}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \xrightarrow{P} 0.$ ■

Lemma 6. For $k = 1, \cdots, p+q$,

$$T^{-1/2}\sum_{t=1}^{T}\left\{\left|\frac{X_t - \widehat{X}_t + Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t + Z_t)}{\partial \beta_k}\right| + \left|\frac{X_t - \widehat{X}_t - Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t - Z_t)}{\partial \beta_k}\right|\right\}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \xrightarrow{P} 0.$$

Proof. We only prove that $T^{-1/2}\sum_{t=1}^{T}|\frac{X_t - \widehat{X}_t + Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t + Z_t)}{\partial \beta_k}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \xrightarrow{P} 0$, since the other half of the result may be proved in a simpler manner.

For any $\boldsymbol{\beta} \in \mathcal{B}$, let $\varphi(z) = a(z)^{-1}b(z) = 1 - \sum_{i=1}^{\infty} \varphi_i z^i$. Then

$$\frac{z^k}{a(z)} = \sum_{i=1}^{\infty} \frac{\partial \varphi_i}{\partial b_k} z^i, \quad \text{and} \quad \frac{z^k b(z)}{a(z)^2} = \sum_{i=1}^{\infty} \frac{\partial \varphi_i}{\partial a_k} z^i.$$

It may be proved from the above two equations that there exist constants $C > 0$ and $s \in (0,1)$ such that

$$|\partial \varphi_j / \partial \beta_k| \leq C s^j, \quad j \geq 1, \; 1 \leq k \leq p+q. \tag{4.6}$$

On the other hand, it follows from the identity $\gamma(j) = \sum_{i=1}^{\infty} \gamma(j-i)\varphi_i$ that

$$\frac{\partial \gamma(j)}{\partial \beta_k} = \sum_{i=1}^{\infty} \frac{\partial \gamma(j-i)}{\partial \beta_k} \varphi_i + \sum_{i=1}^{\infty} \frac{\partial \varphi_i}{\partial \beta_k} \gamma(j-i), \quad j \geq 1.$$

From (4.3) we have for $k = 1, \cdots, p+q$ that,

$$\frac{\partial \gamma(j)}{\partial \beta_k} = \sum_{i=1}^{t} \frac{\partial \gamma(j-i)}{\partial \beta_k} \varphi_{ti} + \sum_{i=1}^{t} \frac{\partial \varphi_{ti}}{\partial \beta_k} \gamma(j-i), \quad 1 \leq j \leq t.$$

Thus it holds for $j = 1, \cdots, t$ that,

$$\sum_{i=1}^{t} \frac{\partial (\varphi_{ti} - \varphi)}{\partial \beta_k} \gamma(j-i) = \sum_{i=1}^{t} (\varphi_i - \varphi_{ti}) \frac{\partial \gamma(j-i)}{\partial \beta_k} + \sum_{i>t} \{\varphi_i \frac{\partial \gamma(j-i)}{\partial \beta_k} + \frac{\partial \varphi_i}{\partial \beta_k} \gamma(j-i)\}.$$

Consequently,

$$\frac{\partial (\boldsymbol{\varphi}_t - \widetilde{\boldsymbol{\varphi}}_t)}{\partial \beta_k} = \boldsymbol{\Gamma}_t^{-1} \{ \frac{\partial \boldsymbol{\Gamma}_t}{\partial \beta_k} (\widetilde{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t) + \mathbf{d}_t \}, \tag{4.7}$$

where $\widetilde{\boldsymbol{\varphi}}_t = (\varphi_1, \cdots, \varphi_t)^\tau$, and $\mathbf{d}_t$ is a $t \times 1$ vector with $\sum_{i>t} \{\varphi_i \frac{\partial \gamma(j-i)}{\partial \beta_k} + \frac{\partial \varphi_i}{\partial \beta_k} \gamma(j-i)\}$ as its j-th component. It follows from (4.4) that under the assumption $Z_{-t} = X_{-t} = 0$ for all $t \geq 0$,

$$Z_{t+1} = a(B)^{-1} b(B) X_{t+1} = \varphi(B) X_{t+1} = X_{t+1} - \sum_{i=1}^{t} \varphi_i X_{t+1-i}, \quad t \geq 0. \tag{4.8}$$

Now it follows from (4.7) that

$$\begin{aligned}
E(\frac{\partial (\widehat{X}_{t+1} + Z_{t+1})}{\partial \beta_k})^2 &= E(\sum_{j=1}^{t} \frac{\partial (\varphi_{tj} - \varphi_j)}{\partial \beta_k} X_{t-j})^2 = \frac{\partial (\widetilde{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t)^\tau}{\beta_k} \boldsymbol{\Gamma}_t \frac{\partial (\widetilde{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t)}{\beta_k} \\
&= \{ \frac{\partial \boldsymbol{\Gamma}_t}{\partial \beta_k} (\widetilde{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t) + \mathbf{d}_t \}^\tau \boldsymbol{\Gamma}_t^{-1} \{ \frac{\partial \boldsymbol{\Gamma}_t}{\partial \beta_k} (\widetilde{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t) + \mathbf{d}_t \} \\
&\leq \frac{2}{\lambda_{\min}(\boldsymbol{\Gamma}_t)} \{ || \frac{\partial \boldsymbol{\Gamma}_t}{\partial \beta_k} (\widetilde{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t) ||^2 + ||\mathbf{d}_t||^2 \} \\
&\leq \frac{2 \max\{\alpha, \gamma(0)\}}{\lambda_{\min}(\boldsymbol{\Gamma}_t)} t [||(\widetilde{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t)||^2 + \{ \sum_{i>t} (|\varphi_i| + |\frac{\partial \varphi_i}{\partial \beta_k}|) \}^2] \leq C_1 s_1^t,
\end{aligned}$$

where $\lambda_{\min}(\boldsymbol{\Gamma}_t) > 0$ denotes the minimum eigenvalue of $\boldsymbol{\Gamma}_t$ (see Proposition 4.5.3 of Brockwell and Davis 1991), $\alpha \equiv | \int_{-\pi}^{\pi} \frac{\partial |a(e^{i\omega})/b(e^{i\omega})|^2}{\partial \beta_k} d\omega | \geq |\frac{\partial \gamma(j)}{\partial \beta_k}|$ ($j \geq 1$ and $1 \leq k \leq p+q$), and $C_1 > 0$ and

$s_1 \in (0,1)$ are some constants (which depend on $\boldsymbol{\beta} \in \mathcal{B}$ continuously). In the above expression, the first inequality makes use of (10.8.50) of Brockwell and Davis (1991) and (4.6).

Now using the similar argument as in the proof of Lemma 5, we may show that

$$E\left(\left.\frac{\partial(\widehat{X}_{t+1}+Z_{t+1})}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} I(A)\right)^2 \le C_2 s_2^t, \quad t\ge 1,\ 1\le k\le p+q,$$

where A is an event with the probability arbitrarily close to 1 and on which $||\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0||$ arbitrarily small for all large T's, and $C_2>0$ and $s_2\in(0,1)$ are some constants. Hence

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\left(E\left|\frac{X_t-\widehat{X}_t+Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t+Z_t)}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} I(A)\right)\\
\le\;& T^{-1/2}\sum_{t=1}^{T}\left(E[\{(X_t-\widehat{X}_t(\widehat{\boldsymbol{\beta}})+Z_t(\widehat{\boldsymbol{\beta}})\}^2 I(A)]\,E\{\left.\frac{\partial(\widehat{X}_{t+1}+Z_{t+1})}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} I(A)\}^2\right)^{1/2}\\
\le\;& CT^{-1/2}\sum_{t=1}^{T}s_2^{t/2}\to 0,
\end{aligned}$$

where $C>0$ is a constant. Thus $T^{-1/2}\sum_{t=1}^{T}\left|\frac{X_t-\widehat{X}_t+Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t+Z_t)}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \xrightarrow{P} 0.$ ■

Lemma 7. $T^{-1}\mathcal{U}^\tau\mathcal{R}^{-1}\mathcal{U}\xrightarrow{P}\sigma^2\mathbf{W}(\mathbf{b}_0,\mathbf{a}_0)^{-1}$ and $T^{-1/2}\mathcal{U}^\tau\mathcal{R}^{-1}\mathcal{Z}\xrightarrow{D}N(0,\sigma^4\mathbf{W}(\mathbf{b}_0,\mathbf{a}_0)^{-1})$.

Proof. In this proof, all U_{ti}, V_{ti}, Z_t and r_t are defined at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. If we adopt the notation that $B^kU_{tj}=U_{t-k,j}$ and $B^kV_{tj}=V_{t-k,j}$, it follows from (4.4) that

$$U_{tj}=a_0(B)^{-1}X_{t-j},\quad \text{and}\quad V_{tj}=a_0(B)^{-1}Z_{t-j}=a_0(B)^{-2}b_0(B)X_{t-j},$$

and these expressions are valid under the assumption that $X_{-t}=Z_{-t}=0$ for all $t\ge 0$. Let $1-\sum_{k\ge 1}\psi_j z^j=1/a_0(z)$ and $1-\sum_{k\ge 1}\eta_j z^j=b_0(z)/a_0(z)^2$. Then

$$U_{tj}=X_{t-j}-\sum_{k=1}^{t-j-1}\psi_j X_{t-j-k}=b_0(B)^{-1}\varepsilon_{t-j}+\sum_{k\ge t-j}\psi_k X_{t-j-k}\equiv\widetilde{U}_{tj}+u_{tj}, \tag{4.9}$$

$$V_{tj}=X_{t-j}-\sum_{k=1}^{t-j-1}\eta_k X_{t-j-k}=a_0(B)^{-1}\varepsilon_{t-j}+\sum_{k\ge t-j}\eta_k X_{t-j-k}\equiv\widetilde{V}_{tj}+v_{tj}. \tag{4.10}$$

Note that

$$Eu_{tj}^2=\sum_{k,l=1}^{\infty}\gamma(k-l)\psi_{t-j+k}\psi_{t-j+l}\le M\sum_{k=1}^{\infty}\psi_{t-j+k}^2\le Cs^{t-j},\quad t\ge 1,\ 1\le j\le p, \tag{4.11}$$

where $C, M > 0$ and $s \in (0,1)$ are some constants. The first inequality in the above expression follows from Proposition 4.5.3 of Brockwell and Davis (1991) and an obvious asymptotic argument. In the same vein, we have

$$Ev_{tj}^2 \le Cs^{t-j}, \quad t \ge 1,\ 1 \le j \le q. \tag{4.12}$$

The (i,j)-th element of $\mathcal{U}^\tau \mathcal{R}^{-1} \mathcal{U}/T$ is

$$\frac{1}{T}\sum_{t=1}^{T} U_{ti}U_{tj}/r_{t-1} = \frac{1}{T}\sum_{t=1}^{T}\{\widetilde{U}_{ti}\widetilde{U}_{tj} + \widetilde{U}_{ti}u_{tj} + u_{ti}\widetilde{U}_{tj} + u_{ti}u_{tj}\}/r_{t-1}, \tag{4.13}$$

see (4.9) and (4.10). By the ergodic theorem, $\sum_{t=1}^{T} \widetilde{U}_{ti}\widetilde{U}_{tj}/T \xrightarrow{a.s.} \sigma^2 \mathrm{Cov}(\xi_{t-i}, \xi_{t-j})$, where $\{\xi_t\}$ is the stationary process defined as in (4.1). Since $r_t \to 1$, we have

$$\frac{1}{T}\sum_{t=1}^{T} \widetilde{U}_{ti}\widetilde{U}_{tj}/r_{t-1} \xrightarrow{a.s.} \sigma^2 \mathrm{Cov}(\xi_{t-i}, \xi_{t-j}).$$

By the Cauchy-Schwarz inequality and (4.11),

$$\frac{1}{T}\sum_{t=1}^{T} E|u_{ti}u_{tj}|/r_{t-1} \le \frac{1}{T}\sum_{t=1}^{T} E|u_{ti}u_{tj}| \le \frac{1}{T}\sum_{t=1}^{T}\{Eu_{ti}^2 Eu_{tj}^2\}^{1/2} \le \frac{C}{T}\sum_{t=1}^{\infty} s^{2t-(i+j)} \to 0, \tag{4.14}$$

which implies $T^{-1}\sum_{t=1}^{T} u_{ti}u_{tj}/r_{t-1} \xrightarrow{P} 0$. In the same vein, the two other terms on the RHS of (4.13) also converge to 0 in probability. Hence we have shown that

$$\frac{1}{T}\sum_{t=1}^{T} U_{ti}U_{tj}/r_{t-1} \xrightarrow{P} \sigma^2 \mathrm{Cov}(\xi_{t-i}, \xi_{t-j}), \quad 1 \le i,j \le p. \tag{4.15}$$

We may prove in a similar manner that

$$\frac{1}{T}\sum_{t=1}^{T} U_{ti}V_{tj}/r_{t-1} \xrightarrow{P} \sigma^2 \mathrm{Cov}(\xi_{t-i}, \zeta_{t-j}), \quad 1 \le i \le p,\ 1 \le j \le q,$$

$$\frac{1}{T}\sum_{t=1}^{T} V_{ti}V_{tj}/r_{t-1} \xrightarrow{P} \sigma^2 \mathrm{Cov}(\zeta_{t-i}, \zeta_{t-j}), \quad 1 \le i,j \le q,$$

where $\{\zeta_t\}$ is defined as in (4.1). Combining the above three expressions together, we have $\mathcal{U}^\tau \mathcal{R}^{-1} \mathcal{U}/T \xrightarrow{P} \sigma^2 \mathbf{W}(\boldsymbol{\beta}_0)^{-1}$.

To establish the required CLT, we define $\widetilde{\mathbf{U}}_t = (\widetilde{U}_{t1}, \cdots, \widetilde{U}_{tp}, \widetilde{V}_{t1}, \cdots, \widetilde{V}_{tq})^\tau$ and $\mathbf{u}_t = (u_{t1}, \cdots, u_{tp}, v_{t1}, \cdots, v_{tq})^\tau$, where $\widetilde{U}_{tj}, \widetilde{V}_{tj}, u_{tj}$ and v_{tj} are defined in (4.9) and (4.10). From (4.8), we may write $Z_t = \varepsilon_t + z_t$ with $z_t = \sum_{j \ge t} \varphi_{j0} X_{t-j}$, where $1 - \sum_{j\ge 1} \varphi_{j0} = a_0(z)^{-1} b_0(z)$. Now

$$\frac{1}{T^{1/2}}\mathcal{U}^\tau \mathcal{R}^{-1}\mathcal{Z} = \frac{1}{T^{1/2}}\sum_{t=1}^{T}(\widetilde{\mathbf{U}}_t + \mathbf{u}_t)\frac{\varepsilon_t + z_t}{r_{t-1}} = \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\widetilde{\mathbf{U}}_t\varepsilon_t + \widetilde{\mathbf{U}}_t z_t + \mathbf{u}_t\varepsilon_t + \mathbf{u}_t z_t}{r_{t-1}}. \tag{4.16}$$

Similar to (4.11) and (4.12), we may show that $Ez_t^2 \le Cs^t$ for all $t \ge 1$, where $C > 0$ and $s \in (0,1)$ are some constants. Consequently based on the same argument as (4.14), we may show that

$$T^{-1/2} \sum_{t=1}^{T} (\widetilde{\mathbf{U}}_t z_t + \mathbf{u}_t \varepsilon_t + \mathbf{u}_t z_t)/r_{t-1} \xrightarrow{P} 0. \tag{4.17}$$

Let $\mathcal{F}_t$ be the σ-algebra generated by $\{\varepsilon_{t-k}, k \ge 0\}$. Then $\{\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t / r_{t-1}\}$ are martingale-differences with respect to $\{\mathcal{F}_t\}$ for any $\boldsymbol{\alpha} \in \mathcal{R}^{p+q}$ in the sense that $\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t / r_{t-1}$ is $\mathcal{F}_t$-measurable and $E\{\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t / r_{t-1} | \mathcal{F}_{t-1}\} = \boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t / r_{t-1} \, E\varepsilon_t = 0$; see (4.9) and (4.10). Further for any $\epsilon > 0$,

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T} E\{(\frac{\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t}{r_{t-1}})^2 I(|\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t / r_{t-1}| > T^{1/2}\epsilon) | \mathcal{F}_{t-1}\} \\
\le \; & \frac{1}{T} \sum_{t=1}^{T} E[(\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t)^2 I(|\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t| > T^{1/2}\epsilon)\{I(|\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t| > \log T) + I(|\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t| \le \log T)\} | \mathcal{F}_{t-1}] \\
\le \; & \frac{1}{T} \sum_{t=1}^{T} [\sigma^2 (\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t)^2 I(|\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t| > \log T) + (\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t)^2 E\{\varepsilon_t^2 I(|\varepsilon_t| > T^{1/2}\epsilon / \log T)\}] \\
\sim \; & \sigma^2 E\{(\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_1)^2 I(|\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_1| > \log T)\} + E(\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_1)^2 E\{\varepsilon_1^2 I(|\varepsilon_1| > T^{1/2}\epsilon / \log T)\} \to 0.
\end{aligned}$$

The last limit follows from the fact that both ε_t and $\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_1$ have finite second moments. Note that

$$\frac{1}{T} \sum_{t=1}^{T} (\frac{\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t}{r_{t-1}})^2 \sim \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t)^2 \xrightarrow{a.s.} E(\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t)^2 = \sigma^2 E(\boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t)^2 = \sigma^4 \boldsymbol{\alpha}^\tau \mathbf{W}(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\alpha},$$

it follows from Theorem 4 on p.511 of Shiryayev (1984) that

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} \boldsymbol{\alpha}^\tau \widetilde{\mathbf{U}}_t \varepsilon_t / r_{t-1} \xrightarrow{D} N(0, \sigma^4 \boldsymbol{\alpha}^\tau \mathbf{W}(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\alpha}), \quad \text{for any } \boldsymbol{\alpha} \in \mathcal{R}^{p+q}.$$

Now the limit $T^{-1/2} \sum_{t=1}^{T} \mathcal{U}^\tau \mathcal{R}^{-1} \mathcal{Z} \xrightarrow{D} N(0, \sigma^4 \mathbf{W}(\boldsymbol{\beta}_0)^{-1})$ follows from the above, (4.16) and (4.17) immediately. ■

Proof of Theorem 2. Let $\widehat{\mathbf{X}}_T = (\widehat{X}_1, \cdots, \widehat{X}_T)^\tau$. Then $\mathbf{X}_T = \mathbf{H}(\mathbf{X}_T - \widehat{\mathbf{X}}_T)$, where $\mathbf{H}$ is given as in (3.9). Thus $\boldsymbol{\Sigma}(\boldsymbol{\beta}) = \sigma^2 \mathbf{H} \mathcal{R} \mathbf{H}^\tau$ for $\{X_t - \widehat{X}_t\}$ are uncorrelated. Consequently $\mathbf{X}_T^\tau \boldsymbol{\Sigma}(\boldsymbol{\beta})^{-1} \mathbf{X}_T = \sigma^{-2} \sum_{t=1}^{T} (X_t - \widehat{X}_t)^2 / r_{t-1}$ and $|\boldsymbol{\Sigma}(\boldsymbol{\beta})| = \sigma^{2T} \prod_{t=1}^{T} r_{t-1}$. Now it follows from (2.4) that

$$\begin{aligned}
M(\boldsymbol{\beta}) \; & \equiv \; -2\sigma^2 \log L(\boldsymbol{\beta}, \sigma^2) = T\sigma^2 \log \sigma^2 + \sigma^2 \sum_{t=1}^{T} \log r_{t-1} + \sum_{t=1}^{T} (X_t - \widehat{X}_t)^2 / r_{t-1} \\
& = \; T\sigma^2 \log \sigma^2 + \sigma^2 \sum_{t=1}^{T} \log r_{t-1} + \sum_{t=1}^{T} Z_t^2 / r_{t-1} + \sum_{t=1}^{T} \frac{(X_t - \widehat{X}_t)^2 - Z_t^2}{r_{t-1}}.
\end{aligned}$$

Note that $\widehat{\boldsymbol\beta}$ is the solution of the equation $\frac{\partial}{\partial\boldsymbol\beta}M(\boldsymbol\beta)=0$, and for $1\le k\le p$, the equality $\frac{\partial}{\partial b_k}M(\boldsymbol\beta)|_{\boldsymbol\beta=\widehat{\boldsymbol\beta}}=0$ leads to

$$\begin{aligned} 0 &= \sum_{t=1}^T Z_t(\widehat{\boldsymbol\beta})U_{tk}(\widehat{\boldsymbol\beta})/r_{t-1}(\widehat{\boldsymbol\beta})+\delta_k \\ &= \sum_{t=1}^T\{X_t-\sum_{j=1}^p\widehat{b}_jX_{t-j}-\sum_{i=1}^q\widehat{a}_iZ_{t-i}(\boldsymbol\beta_0)\}U_{tk}(\boldsymbol\beta_0)/r_{t-1}(\boldsymbol\beta_0)+\boldsymbol\eta_k^\tau(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0)+\delta_k, \quad (4.18)\end{aligned}$$

where $X_{-j}=Z_{-j}\equiv 0$ for all $j\ge 0$, and

$$\begin{aligned}\delta_k &= \left(\frac{\sigma^2}{2}\frac{\partial}{\partial b_k}\sum_{t=1}^T\log r_{t-1}-\frac12\sum_{t=1}^T\frac{(X_t-\widehat{X}_t)^2}{r_{t-1}^2}\frac{\partial r_{t-1}}{\partial b_k}\right.\\ &- \left.\frac12\sum_{t=1}^T\{\frac{X_t-\widehat{X}_t+Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t+Z_t)}{\partial b_k}+\frac{X_t-\widehat{X}_t-Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t-Z_t)}{\partial b_k}\}\right)_{\boldsymbol\beta=\widehat{\boldsymbol\beta}},\end{aligned}$$

$$\begin{aligned}\boldsymbol\eta_k &= \sum_{t=1}^T\frac{U_{tk}(\boldsymbol\beta_T)}{r_{t-1}(\boldsymbol\beta_T)}\sum_{i=1}^q\widehat{a}_i\mathbf{U}_{t-i}(\boldsymbol\beta_T)+\sum_{t=1}^T\{X_t-\sum_{j=1}^p\widehat{b}_jX_{t-j}-\sum_{i=1}^q\widehat{a}_iZ_{t-i}(\boldsymbol\beta_T)\}\frac{\partial}{\partial\boldsymbol\beta}\left(\frac{U_{tk}}{r_{t-1}}\right)_{\boldsymbol\beta=\boldsymbol\beta_T}\\ &= \sum_{t=1}^T\frac{U_{tk}(\boldsymbol\beta_0)}{r_{t-1}(\boldsymbol\beta_0)}\sum_{i=1}^q a_{i0}\mathbf{U}_{t-i}(\boldsymbol\beta_0)+\sum_{t=1}^T Z_t(\boldsymbol\beta_0)\frac{\partial}{\partial\boldsymbol\beta}\left(\frac{U_{tk}}{r_{t-1}}\right)_{\boldsymbol\beta=\boldsymbol\beta_0}+O_p(T||\widehat{\boldsymbol\beta}-\boldsymbol\beta_0||). \quad (4.19)\end{aligned}$$

In the above expression, $\mathbf{U}_t=(U_{t1},\cdots,U_{tp},V_{t1},\cdots,V_{tq})^\tau$, a_{i0} is the i-th component of $\mathbf{a}_0$, and $\boldsymbol\beta_T$ is always between $\widehat{\boldsymbol\beta}$ and $\boldsymbol\beta_0$. Similarly the equation $\frac{\partial}{\partial a_k}M(\boldsymbol\beta)|_{\boldsymbol\beta=\widehat{\boldsymbol\beta}}=0$ $(1\le k\le q)$ leads to

$$0=\sum_{t=1}^T\{X_t-\sum_{j=1}^p\widehat{b}_jX_{t-j}-\sum_{i=1}^q\widehat{a}_iZ_{t-i}(\boldsymbol\beta_0)\}V_{tk}(\boldsymbol\beta_0)/r_{t-1}(\boldsymbol\beta_0)+\boldsymbol\eta_{p+k}^\tau(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0)+\delta_{p+k}, \quad (4.20)$$

where

$$\begin{aligned}\delta_{p+k} &= \left(\frac{\sigma^2}{2}\frac{\partial}{\partial a_k}\sum_{t=1}^T\log r_{t-1}-\frac12\sum_{t=1}^T\frac{(X_t-\widehat{X}_t)^2}{r_{t-1}^2}\frac{\partial r_{t-1}}{\partial a_k}\right.\\ &- \left.\frac12\sum_{t=1}^T\{\frac{X_t-\widehat{X}_t+Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t+Z_t)}{\partial a_k}+\frac{X_t-\widehat{X}_t-Z_t}{r_{t-1}}\frac{\partial(\widehat{X}_t-Z_t)}{\partial a_k}\}\right)_{\boldsymbol\beta=\widehat{\boldsymbol\beta}},\end{aligned}$$

$$\boldsymbol\eta_{p+k}=\sum_{t=1}^T\frac{U_{tk}(\boldsymbol\beta_0)}{r_{t-1}(\boldsymbol\beta_0)}\sum_{i=1}^q a_{i0}\mathbf{U}_{t-i}(\boldsymbol\beta_0)+\sum_{t=1}^T Z_t(\boldsymbol\beta_0)\frac{\partial}{\partial\boldsymbol\beta}\left(\frac{V_{tk}}{r_{t-1}}\right)_{\boldsymbol\beta=\boldsymbol\beta_0}+O_p(T||\widehat{\boldsymbol\beta}-\boldsymbol\beta_0||). \quad (4.21)$$

It follows from (4.18) and (4.20) that

$$\mathcal{U}^\tau\mathcal{R}^{-1}\mathcal{X}\widehat{\boldsymbol\beta}=\mathcal{U}^\tau\mathcal{R}^{-1}\mathcal{Y}+\mathbf{A}^\tau(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0)+\boldsymbol\delta, \quad (4.22)$$

where $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_{p+q})^\tau$, and $\mathbf{A}$ is the $(p+q) \times (p+q)$ matrix with $\boldsymbol{\eta}_k$ as its k-th column. Note that $\mathcal{Y} - \mathcal{X}\boldsymbol{\beta}_0 = \mathcal{Z}$ and

$$\mathcal{U} = \mathcal{X} - \sum_{i=1}^{q} a_{i0} \begin{pmatrix} \mathbf{U}_{1-i}(\boldsymbol{\beta}_0)^\tau \\ \vdots \\ \mathbf{U}_{T-i}(\boldsymbol{\beta}_0)^\tau \end{pmatrix}.$$

By (4.22), (4.18) and (4.20), we have

$$\mathcal{U}\mathcal{R}^{-1}\mathcal{U}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathcal{U}\mathcal{R}^{-1}\mathcal{Z} + \mathbf{B}^\tau(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \boldsymbol{\delta},$$

where $\mathbf{B}$ is the $(p+q) \times (p+q)$ matrix with the sum of the last two terms on the RHS of (4.19) as its k-th column for $k = 1, \cdots, p$, and the sum of the last two terms on the RHS of (4.21) as its $(p+k)$-th term for $k = 1, \cdots, q$. Hence

$$T^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \{\mathcal{U}\mathcal{R}^{-1}\mathcal{U}/T - \mathbf{B}^\tau/T\}^{-1}T^{-1/2}(\mathcal{U}\mathcal{R}^{-1}\mathcal{Z} - \boldsymbol{\delta}) = \{\mathcal{U}\mathcal{R}^{-1}\mathcal{U}/T\}^{-1}T^{-1/2}\mathcal{U}\mathcal{R}^{-1}\mathcal{Z} + o_p(1).$$

The last equality follows from Lemmas 5 and 6, and the fact that $\mathbf{B}/T \xrightarrow{P} 0$, which may be shown in the similar manner as (4.14). Now the theorem follows from Lemma 7 immediately. ■

Acknowledgements

QY's research was partially supported by a BBSRC/EPSRC Grant. He also thanks Professors Rainer Dahlhaus, Peter Robinson and Howell Tong for helpful discussions.

References

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

Chow, Y.S. and Teicher, H. (1997). *Probability Theory*, 3rd ed. Springer-Verlag, New York.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Prob.* **10**, 130-145.

Shiryayev, A.N. (1984). *Probability*. Springer-Verlag, New York.

Yao, Q. and Brockwell, P.J. (2001). Gaussian maximum likelihood estimation for ARMA models II: spatial processes. *A preprint.*