The Annals of Statistics
2016, Vol. 44, No. 4, 1618–1660
DOI: 10.1214/15-AOS1430

ESTIMATION OF SEMIVARYING COEFFICIENT TIME SERIES MODELS WITH ARMA ERRORS

BY HUANG LEI*, YINGCUN XIA†,‡,1 AND XU QIN‡

Southwest Jiaotong University of China, National University of Singapore† and University of Electronic Science and Technology of China‡*

Serial correlation in the residuals of time series models can cause bias in both model estimation and prediction. However, models with such serially correlated residuals are difficult to estimate, especially when the regression function is nonlinear. Existing estimation methods require strong assumption for the relation between the residuals and the regressors, which excludes the commonly used autoregressive models in time series analysis. By extending the Whittle likelihood estimation, this paper investigates in details a semi-parametric autoregressive model with ARMA sequence of residuals. Asymptotic normality of the estimators is established, and a model selection procedure is proposed. Numerical examples are employed to illustrate the performance of the proposed estimation method and the necessity of incorporating the serial correlation in the residuals.

1. Introduction. Serial correlation in the residuals of nonparametric regression has been noticed for many years and its impact has been investigated intensively; see, for example, Hall and Van Keilegom (2003), Hart (1991), Opsomer, Wang and Yang (2001), Ray and Tsay (1997) and Cai (2007). For illustration, consider the simple nonparametric regression

$$Y_t = g(X_t) + \xi_t, \qquad t = 1, 2, \dots.$$

It is known from the aforementioned work that there is big difference between the estimation for the case with i.i.d. residuals ξ_t and that with serial correlated residuals. However, in those works, it is usually assumed that $\{X_t\}$ and $\{\xi_t\}$ are independent, or at least $E(\xi_t|X_t) = 0$, which is an exception for the autoregressive models where X_t is a lagged variable of Y_t, for example, $X_t = Y_{t-1}$. On the other hand, most of the existing nonparametric or semi-parametric autoregressive time series models do not allow the residuals to be serially correlated, because otherwise the consistency of estimation is not easy to obtain; see, for example, Cai, Fan and Yao (2000), Tjøstheim and Auestad (1994), Xue and Yang (2006) and Gao (2007).

Received August 2015; revised December 2015.

[1]Supported in part by Ministry of Education of Singapore: MOE2014-T2-1-072, and National Natural Science Foundation of China: 71371095.

MSC2010 subject classifications. 60K35.

Key words and phrases. ARMA process, B-spline, correlated errors, semi-varying coefficient model, spectral density function, Whittle likelihood estimation.

To accommodate such serial correlation, linear autoregressive models were commonly used in the literature. For example, Xiao et al. (2003) proposed an efficient approach to nonparametric regression with residuals being a general linear process that can be approximated by a truncated AR process. Chen, Li and Li (2015) considered a varying coefficient model that has AR residuals. More generally, Opsomer, Wang and Yang (2001) and Liu, Chen and Yao (2010) used a stationary ARMA process to model the residuals, that is,

$$\xi_t = \theta_a(B)^{-1}\theta_m(B)\varepsilon_t, \tag{1.1}$$

where $\theta_a(B) = 1 + \theta_{a1}B + \cdots + \theta_{aq_1}B^{q_1}$ and $\theta_m(B) = 1 + \theta_{m1}B + \cdots + \theta_{mq_2}B^{q_2}$ are irreducible, and B is the back-shift operator. They found that the estimation efficiency improves substantially by considering the autocorrelated errors when serial correlation exists. More complicated structure for the errors was also considered in the literature. Su and Ullah (2006) considered the residual process without assuming any explicit parametric form, which however may lead to slow convergence of estimation, and hence affects model's prediction. Again, most of the existing autoregressive models still assume

$$E(\xi_t|X_t) = 0 \qquad \text{almost surely}, \tag{1.2}$$

in order to obtain their theoretical properties. As we discussed, (1.2) is not typical for autoregressive model.

In the conventional linear ARMA model, one way to cope with the problem of (1.2) being violated is using higher order AR(m) to approximate the ARMA model. However, this is not so efficient as m must tend to infinity with sample size; see, for example, Pierce (1971). As a consequence, the higher order AR model is less efficient than the original ARMA model in the estimation. For nonlinear time series models, to use a higher order nonlinear AR model to approximate a nonlinear ARMA model is even more intractable because the resultant model might have a very complicated functional structure. Therefore, investigating the estimation of AR models with serial correlated residuals is very important in time series modeling. However, estimation of the resultant model with serial correlated residuals is difficult, even for the simple AR model with serial correlated errors (1.1). First, the least squares estimator (LSE) might not be consistent when (1.2) is not satisfied. Second, though iterative estimation between the regressive or autoregressive part and the error part (1.1) can be used in calculation, its theory can only be justified in some special case; see, for example, Liu, Chen and Yao (2010). Third, the maximum likelihood estimation (MLE) is also not easily tractable because the likelihood function contains the inverse of covariance matrix of $(\xi_1, \ldots, \xi_N)^\top$, where dimension of the matrix goes to infinity with sample size N; see Yao and Brockwell (2006). To overcome these problems, Whittle (1953) used several ingenious matrix calculus and approximated the maximum likelihood function by a summation of the ratios between the periodogram of the observations and the corresponding spectral density function (SDF), called Whittle likelihood estimation (WLE). Hannan (1973) proved the asymptotic normality of Whittle likelihood estimator and its equivalence in estimation efficiency to the maximum likelihood estimator. However, WLE essentially requires that the time series has a closed form for its spectral density function (SDF), while nonlinear time series models do not meet this requirement, and thus cannot be estimated directly by WLE.

In this paper, a new estimation method is proposed for a general varying coefficient model with ARMA errors. The method can be analogously extended to other semi-parametric models. The estimation is based on (i) an extension of the traditional Whittle likelihood estimation (WLE), and (ii) B-spline approximation of functions in the model. Compared with Liu, Chen and Yao (2010) and Ma and Yang (2011), iteration is not needed for our method, and our settings are more general and do not require (1.2). The general ARMA process of the residuals and nonparametric setting for the model considered in this paper differentiates it from Wang and Xia (2014) in both modeling and mathematical techniques. First, extending AR residuals to ARMA residuals encounters the identification problem; second, our Whittle likelihood has a different formulation of loss function, the existing techniques for splines approximation cannot be used directly, and thus new theories must be developed. On the other hand, as discussed above semiparametric dynamical models with serial correlation in the residuals have received great attention, but as Xiao et al. (2003) openly discussed, there exists difficulty in the modeling and estimation. This paper will address these problems under more realistic model assumptions, and provide an approach to a general collection of nonlinear dynamical models.

The rest of this paper is organized as follows. Section 2 proposes the model to be estimated and its identifiability. An extension of the Whittle likelihood estimation (WLE) is proposed for the model estimation, while a spline-based method is used to approximate the varying coefficients. Section 3 studies consistency of the estimators; Section 4 elaborates a model selection procedure, provides some implementation details for selection of variables, identification of varying coefficient variables and linear variables and determination of the threshold variable and placement of knots; Section 5 reports some examples to show the simulation performance of the method, and demonstrate the applicability of the proposed method for real data analyses.

2. Estimation method. We have discussed in Section 1 the prevalence of serial correlation in the residuals of regression models, and the lack of estimation methods for autoregressive models with correlated residuals. In this section, we propose an estimation method for such purpose. We illustrate the estimation details by using a popular time series model with ARMA process as its residuals, called semi-varying coefficient model with ARMA errors (SVCARMA). The model takes the following form:

$$X_t = g_0(X_{t-d}) + g_1(X_{t-d})X_{t-1} + \cdots + g_r(X_{t-d})X_{t-r} + \beta_1 X_{t-r-1} + \cdots + \beta_p X_{t-r-p} + \xi_t, \quad \xi_t = \theta_a(B)^{-1}\theta_m(B)\varepsilon_t, \tag{2.1}$$

where ξ_t is a general ARMA(q_1, q_2) process with $\theta_a(B)$ and $\theta_m(B)$ defined below (1.1). This model not only allows the random errors to be autocorrelated, but also relaxes the restriction $E(\xi_t|\mathbb{F}_{t-1}) = 0$, where $\mathbb{F}_{t-1} = \sigma\{X_{t-1}, X_{t-2}, \ldots\}$ is the σ-field containing information on and before $t-1$. Without losing generality, we assume lagged variables $\{X_{t-1}, \ldots, X_{t-r}\}$ with varying coefficients, $g_1(X_{t-d}), \ldots, g_r(X_{t-d})$, and others with constant coefficients, and that X_{t-d} is the threshold variable. We also assume $g_0(X_{t-d})$ is a varying intercept which could also be downgraded to constant intercept β_0. Determination of the two types of coefficients will be discussed later. It should be noted that the results obtained below can be easily extended to the case that part or all of $\{X_{t-j}, j = 1, \ldots, (r+p)\}$ are exogenous variables. It can be seen that (2.1) originates from the varying (or functional) coefficient models; see, for example, Cai, Fan and Yao (2000), Chen and Tsay (1993), Hastie and Tibshirani (1993), Zhang, Lee and Song (2002) and Li et al. (2002). These types of models were widely used in practice, and were studied under (1.2). Other models with ARMA residuals can be estimated using the same idea; other patterns of residuals can also be studied similarly providing that its SDF exists, such as residuals of the ARCH process. See, for example, Giraitis and Robinson (2001).

Model (2.1) may not be identifiable generally. For example, when all the varying intercept and varying coefficient functions, $g_0, g_1, \ldots, g_r$, are constant, the model will be a linear autoregressive model AR(p) with ARMA(q_1, q_2) errors,

$$\beta(B)X_t = \xi_t, \qquad \xi_t = \theta_a(B)^{-1}\theta_m(B)\varepsilon_t,$$

where $\beta(B) = 1 + \beta_1 B + \cdots + \beta_p B^p$ is the polynomial of B for the AR part. This model is not identifiable, because it could also be represented by

$$\theta_a(B)X_t = \xi_t, \qquad \xi_t = \beta(B)^{-1}\theta_m(B)\varepsilon_t.$$

However, this identification problem can be fixed by imposing the following assumptions.

(A1) The innovation errors $\varepsilon_t, t = 1, \ldots, N$ in model (2.1) are white noise, and $0 < \sigma_0^2 = \mathrm{Var}(\varepsilon_t) < \infty$.

(A2) Let

$$g(X_{t-\mathbb{S}_g}) = g_0(X_{t-d}) + g_1(X_{t-d})X_{t-1} + \cdots + g_r(X_{t-d})X_{t-r} + \beta_1 X_{t-r-1} + \cdots + \beta_p X_{t-r-p},$$

where $\mathbb{S}_g$ is the set of all lagged variables. Similarly, $g(X_{t-1-\mathbb{S}_g})$ represents the same structure at time point $t-1$. We assume $G(X_{t-\mathbb{S}_g}) = X_t - g(X_{t-\mathbb{S}_g})$ could never be factorized as

$$\left(1 + \sum_{j=1}^{\infty} \phi_j B^j\right) H(X_{t-\mathbb{S}_g})$$

for any set of nonzero $\{\phi_j, j = 1, \dots, \infty\}$, that is, $\sum_{j=1}^{\infty} \phi_j^2 \neq 0$, and any $H(X_{t-\mathbb{S}_g})$ of semi-parametric or parametric structure, where the backshift operator B to $H(X_{t-\mathbb{S}_g})$ is defined as $B^j H(X_{t-\mathbb{S}_g}) = H(X_{t-j-\mathbb{S}_g})$.

LEMMA 2.1. *Suppose assumptions* (A1) *and* (A2) *hold for model* (2.1). *If model* (2.1) *could be written as*

$$X_t = g(X_{t-\mathbb{S}_g}) + \xi_t, \qquad \text{where } \xi_t = \theta_a(B)^{-1}\theta_m(B)\varepsilon_t, \tag{2.2}$$

then both polynomial $\theta_a^{-1}(B)\theta_m(B)$ *and nonlinear function* $g(X_{t-\mathbb{S}_g})$ *are unique.*

For model (2.1), the first part to be estimated is the varying coefficients and the second part is the constant coefficients including linear variables' coefficients and those in random errors ξ_t. Determination of the corresponding varying coefficient variables and linear variables will be studied in Section 4. Denote the parameters of linear part by $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ and that of ξ_t by $\boldsymbol{\theta} = (\theta_{a1}, \dots, \theta_{aq_1}, \theta_{m1}, \dots, \theta_{mq_2})^\top$.

Write model (2.1) in matrix form as

$$\mathbf{Y} = g_0(\mathbf{X}_{t-d}) + D(\mathbf{X}_1)g_1(\mathbf{X}_{t-d}) + \cdots + D(\mathbf{X}_r)g_r(\mathbf{X}_{t-d}) + \mathbf{X}_{Np}\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{2.3}$$

where

$$\begin{aligned}
\mathbf{Y} &= (X_1, X_2, \dots, X_N)^\top, \qquad \mathbf{X}_j = (X_{1-j}, \dots, X_{N-j})^\top, \\
& j = 1, 2, \dots, r+p, \\
g_j(\mathbf{X}_{t-d}) &= \big(g_j(X_{1-d}), \dots, g_j(X_{N-d})\big)^\top, \qquad j = 0, 1, 2, \dots, r, \\
\mathbf{X}_{Np} &= (\mathbf{X_{r+1}}, \dots, \mathbf{X_{r+p}}), \qquad \mathbf{1} = (1, 1, \dots, 1)^\top, \quad \text{and} \\
\boldsymbol{\xi} &= (\xi_1, \dots, \xi_N)^\top,
\end{aligned}$$

where $D(\mathbf{X}_j)$ is a diagonal matrix of $\mathbf{X}_j$, and $\mathbf{Y}$ is a column vector of time series, $\mathbf{X}_j$ is a column vector of jth lagged variable, and $g_j(\mathbf{X}_{t-d})$ is a column vector of varying coefficient at all $\mathbf{X}_{t-d}$, $\boldsymbol{\xi}$ is a column vector of the random errors. Let $\mathbf{g}(u) = (g_1(u), \dots, g_r(u))^\top$ and $\boldsymbol{\alpha}^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)$.

We first introduce the Whittle likelihood function for errors ξ_t in model (2.1), $\xi_t = \theta_a(B)^{-1}\theta_m(B)\varepsilon_t$. Define a set of frequencies as

$$\omega_n \in \mathbf{W}_N = \left\{\omega_n : \omega_n = -\frac{2\pi}{N}\left[\frac{N-1}{2}\right], \dots, 0, \dots, \frac{2\pi}{N}\left[\frac{N}{2}\right]\right\}.$$

The periodogram of $\{\xi_t, t = 1, \dots, N\}$ at frequency ω_n, denoted by $I(\omega_n)$, has the following two explicit summation forms:

$$I(\omega_n, \boldsymbol{\xi}) = \frac{1}{2\pi N}\left|\sum_{t=1}^{N} \xi_t e^{-it\omega_n}\right|^2, \qquad \text{where } i = \sqrt{-1}, \tag{2.4}$$

and

$$I(\omega_n, \boldsymbol{\xi}) = \frac{1}{2\pi}\sum_{\kappa=-N+1}^{N-1} c(\kappa) e^{-i\kappa\omega_n}, \qquad \text{where } c(\kappa) = \frac{\sum_{t=1}^{N-|\kappa|} \xi_t \xi_{t+|\kappa|}}{N},$$

here $c(\kappa)$ is similar to the sample auto-covariance function (ACVF) for ξ_t except that there is no substraction of $\bar{\xi} = 1/N \sum_{t=1}^{N} \xi_t$. Note that the theoretical SDF $f_*(\omega, \boldsymbol{\theta})$ of $\{\xi_t\}$ takes the following parametric formula:

$$f_*(\omega_n, \boldsymbol{\theta}) = \sigma^2 f(\omega_n, \boldsymbol{\theta}) = \frac{\sigma^2}{2\pi} \frac{|1 + \sum_{j=1}^{q_1} \theta_{aj} e^{-ij\omega_n}|^2}{|1 + \sum_{j=1}^{q_2} \theta_{mj} e^{-ij\omega_n}|^2}, \tag{2.5}$$

where $f(\omega, \boldsymbol{\theta})$ is the standardized SDF when $\mathrm{Var}(\varepsilon_t) = 1$ in $\xi_t = \theta_a(B)^{-1}\theta_m(B)\varepsilon_t$:

$$Q_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{\omega_n \in \mathbf{W}_N} \frac{I(\omega_n, \boldsymbol{\xi})}{f(\omega_n, \boldsymbol{\theta})}. \tag{2.6}$$

Let $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} Q_N(\boldsymbol{\theta})$. Then σ^2 can be estimated by replacing the $\boldsymbol{\theta}$ in (2.6) with $\hat{\boldsymbol{\theta}}$,

$$\hat{\sigma}^2 = Q_N(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{\omega_n \in \mathbf{W}_N} \frac{I(\omega_n, \boldsymbol{\xi})}{f(\omega_n, \hat{\boldsymbol{\theta}})}.$$

Note that the above estimation is practicable for linear ARMA(q_1, q_2) model when each ξ_t is observable. For the prime model (2.1), however, ξ_t is not observable. A natural way is to replace it with

$$\boldsymbol{\xi}^* = \{\mathbf{Y} - (g_0(\mathbf{X}_{t-d}) + D(\mathbf{X}_{t-y1}) g_1(\mathbf{X}_{t-d}) + \cdots + D(\mathbf{X}_{t-yr}) g_r(\mathbf{X}_{t-d}) + \mathbf{X}_{Np}\boldsymbol{\beta})\},$$

which could be regarded as an estimator of $\boldsymbol{\xi}$ if varying coefficients $\mathbf{g}(u)$ and coefficients $\boldsymbol{\beta}$ were known. Therefore, we propose to estimate varying coefficients $\mathbf{g}(u)$ and constant coefficients $\boldsymbol{\alpha}^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)$ together by minimizing the following formula:

$$Q_N^*(\boldsymbol{\theta}) = \frac{1}{N} \sum_{\omega_n \in \mathbf{W}_N} \frac{I(\omega_n, \boldsymbol{\xi}^*)}{f(\omega_n, \boldsymbol{\theta})}. \tag{2.7}$$

To estimate $\mathbf{g}(u)$, we use B-spline approach. Replace those $\mathbf{g}(u)$ in $\boldsymbol{\xi}^*$ with B-spline bases multiplying their coefficients $\boldsymbol{\gamma}^\top = (\gamma_1^\top, \dots, \gamma_r^\top)$, where $\{\gamma_j, j = 1, \dots, r\}$ are the coefficients of B-spline bases for approximating each $\{g_j(u), j =$

$1,\dots,r\}$. Then, $\boldsymbol{\xi}^*$ could be written as a function of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, and thus (2.7) becomes

$$Q_N(\boldsymbol{\gamma},\boldsymbol{\alpha})=\frac{1}{N}\sum_{\omega_n\in\mathbf{W}_N}\frac{I(\omega_n,\boldsymbol{\gamma},\boldsymbol{\beta})}{f(\omega_n,\boldsymbol{\theta})}. \tag{2.8}$$

This formulation is an extension of the Whittle likelihood, which is also applicable to any other model that has theoretical SDF for its residuals.

Next, we discuss in details on using B-spline approach to estimate the varying coefficients $\mathbf{g}(u)=(g_1(u),\dots,g_r(u))^\top$. Let $C^m[a,b]$ be a collection of functions defined on $[a,b]$ such that their mth order derivative exists and is continuous. The following assumptions will be used in deriving the asymptotic properties.

(A3) Function $g_j(u)\in C^m[a,b]$, for all $j=1,\dots,r$. The number of inter knots $k_0\ge C_0N^{1/(2m+1)}$, where $m\ge 2$ and $C_0>0$ is a constant.

(A4) Let $\underline{c}=\{a=c_0<c_1<\cdots<c_{k_0}<c_{k_0+1}=b\}$ be the knots sequence for u. Assume the distance $h_j=c_j-c_{j-1}$ between adjacent knots and k_0 satisfy

$$\max_{1\le j\le k_0}|h_{j+1}-h_j|=o(k_0^{-1})\quad\text{and}\quad h\Big/\min_{1\le j\le k_0}h_j\le M_0,$$

where $h=\max_{1\le j\le k_0}h_j$, and $M_0>0$ is independent of N and $\{g_j(u),j=1,\dots,r\}$.

These two assumptions ensure that $h=O(1/k_0)=O(N^{-1/(2m+1)})$. Let $\{B_{j,m}(\cdot)\}_{j=1}^k$ be the collection of B-spline bases of order m built on knots sequence $\underline{c}$. Define the function space built on the bases as

$$\begin{aligned}S(m,\underline{c})&=\{B(u)\boldsymbol{\gamma}:\gamma\in R^k\}\\&=\{s(x)\in C^{m-2}[a,b]:s(x)\text{ is a polynomial of degree }(m-1)\\&\qquad\text{on each subinterval }[c_{j-1},c_j]\},\end{aligned}$$

where $B(u)=\{B_{1,m}(u),\dots,B_{k,m}(u)\}$. Hereafter for fixed m, $B_{j,m}(\cdot)$ is abbreviated as $B_j(\cdot)$ for convenience. We outline the three main steps in constructing estimators for $\boldsymbol{\alpha}$ and $\mathbf{g}(u)$ as follows.

Step 1. Estimate $\boldsymbol{\gamma}=(\gamma_1^\top,\dots,\gamma_r^\top)^\top$ by assuming $\boldsymbol{\alpha}$ is known. First, let

$$B_{Nk}(\mathbf{X}_{t-d})=\begin{pmatrix}B_1(u_1)&B_2(u_1)&\dots&B_k(u_1)\\B_1(u_2)&\dots&\dots&B_k(u_2)\\\vdots&\ddots&\ddots&\vdots\\B_1(u_N)&\dots&\dots&B_k(u_N)\end{pmatrix}$$

and $s_j(\mathbf{X}_{t-d})=B_{Nk}(\mathbf{X}_{t-d})\gamma_j$, $j=1,\dots,r$. Then

$$\begin{aligned}&\mathbf{X}_{Np}\boldsymbol{\beta}+D(\mathbf{X}_1)s_1(\mathbf{X}_{t-d})+\cdots+D(\mathbf{X}_r)s_r(\mathbf{X}_{t-d})=\mathbf{X}_{Np}\boldsymbol{\beta}+\mathbb{D}\boldsymbol{\gamma},\\&\qquad\text{where }\mathbb{D}=\{D(\mathbf{X}_1)B_{Nk}(\mathbf{X}_{t-d}),\dots,D(\mathbf{X}_r)B_{Nk}(\mathbf{X}_{t-d})\}_{(N\times rk)}.\end{aligned}\tag{2.9}$$

Corresponding to (2.4), the periodogram in (2.8) is

$$\begin{aligned}&I(\omega_n, \boldsymbol{\gamma}, \boldsymbol{\beta})\\ &\quad= \frac{1}{2\pi N}(\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta} - \mathbb{D}\boldsymbol{\gamma})^\top \mathbf{e_N}(\omega_n)\overline{\mathbf{e_N}}(\omega_n)^\top (\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta} - \mathbb{D}\boldsymbol{\gamma}),\end{aligned} \tag{2.10}$$

where $\mathbf{e_N}(\omega_n) = \{e^{-i\omega_n}, \dots, e^{-Ni\omega_n}\}^\top$, and $\overline{\mathbf{e_N}}(\omega_n)$ is the conjugate of $\mathbf{e_N}(\omega_n)$. Let

$$E_N = \frac{1}{2\pi N} \sum_{\omega_n \in \mathbf{W}_N} \frac{\mathbf{e_N}(\omega_n)\overline{\mathbf{e_N}}(\omega_n)^\top}{f(\omega_n, \boldsymbol{\theta})}.$$

Then (2.8) could be written as

$$Q_N(\gamma, \boldsymbol{\alpha}) = \frac{1}{N}(\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta} - \mathbb{D}\boldsymbol{\gamma})^\top E_N(\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta} - \mathbb{D}\boldsymbol{\gamma}). \tag{2.11}$$

When $\boldsymbol{\alpha}$ is fixed, the estimator of B-spline coefficients that minimize (2.11) is

$$\hat{\boldsymbol{\gamma}}(\alpha) = (\mathbb{D}^\top E_N \mathbb{D})^{-1}\mathbb{D}^\top E_N(\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta}). \tag{2.12}$$

For convenience, we also write $\hat{\boldsymbol{\gamma}}(\alpha)$ as $\hat{\boldsymbol{\gamma}}$.

Step 2. Estimate $\mathbf{g}(u)$ by the B-splines approximation using $\hat{\boldsymbol{\gamma}}(\alpha)$, but still assuming $\boldsymbol{\alpha}$ is known.

Note that $\hat{\boldsymbol{\gamma}}$ is still a function of $\boldsymbol{\alpha}$. Thus, the estimator of $\mathbf{g}(u)$ at $X_{t-d} = u$ is also a function of $\boldsymbol{\alpha}$, that is,

$$\hat{\mathbf{g}}(u, \boldsymbol{\alpha}) = D(B(u))\hat{\boldsymbol{\gamma}} = D(B(u))(\mathbb{D}^\top E_N \mathbb{D})^{-1}\mathbb{D}^\top E_N(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),$$

where $\hat{\mathbf{g}}(u, \boldsymbol{\alpha}) = (\hat{g}_1(u, \boldsymbol{\alpha}), \dots, \hat{g}_r(u, \boldsymbol{\alpha}))^\top$, $B(u) = \{B_1(u), \dots, B_k(u)\}_{1\times k}$ is the B-spline bases at $X_{t-d} = u$. Let $D(B(u)) = \operatorname{diag}(B(u), \dots, B(u))_{r\times rk}$ be the block diagonal matrix of $B(u)$,

$$D(B(u)) = \begin{pmatrix} B(u) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & B(u) & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & B(u) \end{pmatrix}.$$

Step 3. Estimate $\boldsymbol{\alpha}$ by minimizing (2.11) with $\boldsymbol{\gamma}$ being substituted by $\hat{\boldsymbol{\gamma}}(\alpha)$:

$$\begin{aligned}&\mathbb{Q}_N(\boldsymbol{\alpha}) \overset{\text{def}}{=} Q_N(\hat{\gamma}, \boldsymbol{\alpha})\\ &\quad= \frac{1}{N}(\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta} - \mathbb{D}\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}))^\top E_N(\boldsymbol{\theta})(\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta} - \mathbb{D}\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}))\end{aligned} \tag{2.13}$$

which is only related to $\boldsymbol{\alpha}^\top = \{\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top\}$, that is, the linear coefficients of (2.1) and coefficients in ξ_t. As the above procedure is similar to profile likelihood estimation of Severini and Wong (1992) and Carroll et al. (1997), we call (2.13) the Profile

Based Whittle Likelihood Estimation (PBWLE). Finally, we can estimate $\boldsymbol{\alpha}$ by minimizing (2.13),

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \mathbb{Q}_N(\boldsymbol{\alpha}), \tag{2.14}$$

and $\mathbf{g}(u)$ by

$$\hat{\mathbf{g}}(u, \hat{\boldsymbol{\alpha}}) = D\big(B(u)\big)\big(\mathbb{D}^\top E_N(\hat{\boldsymbol{\theta}})\mathbb{D}\big)^{-1}\mathbb{D}^\top E_N(\hat{\boldsymbol{\theta}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{2.15}$$

3. Asymptotic properties of estimators. We first consider the asymptotic properties of estimators for the varying coefficients by assuming constant coefficients to be known, and we shall come back to the case when they are unknown later. The following assumptions will be used to derive the theoretical results.

(A5) Let Θ be a compact subset of $\boldsymbol{\theta} = \{\theta_{a1}, \dots, \theta_{aq_1}, \theta_{m1}, \dots, \theta_{mq_2}\}$, such that ξ_t, denoted as ARMA$(q_1, q_2, \boldsymbol{\theta})$, is stationary and invertible for all $\boldsymbol{\theta} \in \Theta$. More specifically, there exists a $\delta > 0$, such that all roots of $\theta_a(z)\theta_m(z) = 0$ are outside the circle $\{z \in \mathbb{C} : |z| = 1 + \delta\}$. This leads to the existence of uniform values C_{f1} and C_{f2} such that

$$0 < C_{f1} < f(\omega, \boldsymbol{\theta}) < C_{f2} < \infty. \tag{3.1}$$

Furthermore, assume that for any $\alpha \in \mathbb{A}$, $\lim_{N\to\infty} \mathbb{Q}_N(\boldsymbol{\alpha}) \geq \lim_{N\to\infty} \mathbb{Q}_N(\boldsymbol{\alpha}_0)$, which means that the Whittle Likelihood achieves minimum value at the true parameters $\boldsymbol{\alpha}_0 \in \mathbb{A}$.

By Brockwell and Davis (1991), page 391, (A5) guarantees that $f(\omega, \boldsymbol{\theta})$ and its first derivative $\partial f(\omega, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ are Lipschitz class Λ_a with $a > 1/2$, that is, for any $\boldsymbol{\theta} \in \Theta$,

$$\sup_\omega \big| f(\omega, \boldsymbol{\theta}) - f(\omega + \Delta, \boldsymbol{\theta})\big| = O\big(\Delta^a\big),$$
$$\sup_\omega \left| \frac{\partial f(\omega, \boldsymbol{\theta})}{\partial \theta_j} - \frac{\partial f(\omega + \Delta, \boldsymbol{\theta})}{\partial \theta_j} \right| = O\big(\Delta^a\big),$$

and that

$$\inf_{\boldsymbol{\theta}\in\Theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\omega, \boldsymbol{\theta}_0)}{f(\omega, \boldsymbol{\theta})}\, dw = 1, \qquad f(\omega, \boldsymbol{\theta}) \not\equiv f(\omega, \boldsymbol{\theta}_0) \qquad \text{if } \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > 0.$$

Here, $\boldsymbol{\theta}_0$ is the true values of the parameters; see Lemma 2 of Hannan (1973).

(A6) Let $F(u)$ and $F_N(u)$ be respectively the theoretical and empirical cumulative distribution function (CDF) of X_{t-d}, and $A_N = \max_{a \leq u \leq b} |F_N(u) - F(u)|$. Let $F_j(u, x)$ and $F_{N,j}(u, x)$, $j = 1, \dots, r$, be the theoretical and empirical joint CDF of $(X_{t-d}, X_{t-j}) \in \Re_j$ when $j \neq d$, and $B_N = \max_{(u,x)\in\Re_j} |F_{N,j}(u, x) -$

$F_j(u,x)|$. The number of knots, k_0, satisfies $k_0/N^{1/2} \to 0$. For any square-integrable function $s(u)$, that is, $\int_a^b s^2(u)\,dF(u) < \infty$, there exists constant $0 < C_{U_j} < \infty$ such that

$$\begin{aligned} &\iint_{\Re_j} s^2(u)x^2\,dF_j(u,x) = C_{U_j}\int_a^b s^2(u)\,dF(u), \\ &\qquad\qquad A_N = o_p(k_0^{-1}), \qquad B_N = o_p(k_0^{-1}). \end{aligned} \tag{3.2}$$

(A7) Sequence $\{X_t, \xi_t\}$ be a strictly stationary and α-mixing process, with mixing rate $\alpha(j)$ satisfying $\sum_{j\geq 1}\alpha(j)^{1-2/\tau} < \infty$ for some constant $\tau > 2$, $E|X_t|^{2\tau} < \infty$, $E|\xi_t|^{2\tau} < \infty$.

Assumption (A6) was commonly used in the literature; see, for example, Zhou, Shen and Wolfe (1998). Actually, Yu (1994) proved that if the mixing rate defined in (A7) is $O(n^{-\alpha})$, then

$$\sup_u |\hat{F}_N(u) - F(u)| = O_p(n^{-s/(1+s)}), \tag{3.3}$$

where $s < \max(\alpha, 1)$. Thus, when $k_0/N^{1/2} \to 0$, (3.3) guarantees assumption (A6) hold. If $\{X_{t-d}, X_{t-j}\}$ are independent, then C_{U_j} will be equal to $E|X_{t-j}|^2$, which is not related to $U = X_{t-d}$, but in model (2.1) $\{X_{t-d}, X_{t-}j\}$ are dependent, thus C_{U_j} is indexed by U. In (A7), the ergodicity of X_t can be ensured by Proposition 2.8 of Fan and Yao (2003), it stems from the fact that an α-mixing process is mixing in the sense of ergodic theorem.

THEOREM 3.1. *Suppose assumptions* (A1) *to* (A7) *hold*, *for any fixed* $u \in [a,b]$. *Then*

$$\sqrt{N}\{\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\} \xrightarrow{d} N(\mathbf{0}, \Phi^{-1}A_{\mathbb{I}}\Omega_{\boldsymbol{\beta\theta}}A_{\mathbb{I}}^\top\Phi^{-1}), \tag{3.4}$$

here

$$A_{\mathbb{I}} = \begin{pmatrix} \mathbb{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbb{I}_q & \mathbb{I}_q \end{pmatrix}, \qquad \Omega_{\boldsymbol{\beta\theta}} = \begin{pmatrix} \Omega_{\boldsymbol{\beta}} & \Omega_{\boldsymbol{\beta\theta}\mathrm{I}} & \Omega_{\boldsymbol{\beta\theta}\mathrm{II}} \\ \Omega_{\boldsymbol{\beta\theta}\mathrm{I}}^\top & \Omega_{\boldsymbol{\theta}\mathrm{I}} & \Omega_{\boldsymbol{\theta}\mathrm{I,II}} \\ \Omega_{\boldsymbol{\beta\theta}\mathrm{II}}^\top & \Omega_{\boldsymbol{\theta}\mathrm{I,II}}^\top & \Omega_{\boldsymbol{\theta}\mathrm{II}} \end{pmatrix},$$

$$\Phi = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial^2(f_z(\omega,\boldsymbol{\alpha}_0)f^{-1}(\omega,\boldsymbol{\theta}_0))}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top}\,d\omega,$$

$\mathbb{I}_p$ *and* $\mathbb{I}_q$ *are identity matrices*, p *is the number of linear lagged variables*, *that is*, *the length of* $\boldsymbol{\beta}$, *and* $q = q_1 + q_2$ *is the length of* $\boldsymbol{\theta}$. *Blocks of* $\Omega_{\boldsymbol{\beta\theta}}$ *are defined respectively by* (A.42), (A.44), (A.46), (A.48), (A.50) *and* (A.52) *in the proof.* $\hat{\mathbf{g}}(u,\hat{\boldsymbol{\alpha}}) = (\hat{g}_1(u,\hat{\boldsymbol{\alpha}}),\ldots,\hat{g}_r(u,\hat{\boldsymbol{\alpha}}))^\top$ *is asymptotically normal*, *that is*,

$$\sqrt{Nh}\{\hat{\mathbf{g}}(u,\hat{\boldsymbol{\alpha}}) - (\mathbf{g}(u) + \mathbf{b}(u) + \boldsymbol{\mu}_v^*)\} \xrightarrow{d} N(0, \Sigma_v(u)), \tag{3.5}$$

where $\mathbf{b}(u)$, $\boldsymbol{\mu}_v^*$ *and* $\Sigma_v(u)$ *are defined in Lemma* A.2.

Because Φ and $\Omega_{\boldsymbol{\beta\theta}}$ are functions of α and the coefficient functions, they can also be estimated according to the definitions of Φ above and of $\Omega_{\boldsymbol{\beta\theta}}$ in (A.42), (A.44), (A.46), (A.48), (A.50) and (A.52) based on the estimated α and the coefficient functions. Similarly, we can estimate $\Sigma_v(u)$ in (3.5).

4. Model selection. This section includes the selection of lagged variables, the order of the ARMA for the residuals, the number of knots and their positions. We first modify the BIC criterion in order to select significant lagged variables and to distinguish the varying and constant coefficients. For these two types of coefficients, the corresponding variables are called varying-coefficient variables and linear variables, respectively. As the varying coefficients are approximated by B-spline bases, Huang and Yang (2004) suggested using BIC to select them. Recall that the original BIC is defined as

$$\mathrm{BIC} = \log(\mathrm{MSE}) + \frac{P}{N}\log N,$$

where P is the number of parameters used in the model, and MSE is an estimator for σ_0^2. Because for $\mathbb{Q}_N(\hat{\boldsymbol{\alpha}})$ in (2.13), Lemma A.3 shows its consistency $\mathbb{Q}_N(\hat{\boldsymbol{\alpha}}) \stackrel{\text{a.s.}}{\to} \sigma_0^2$, thus we can use $\mathbb{Q}_N(\hat{\boldsymbol{\alpha}})$ to replace MSE.

As our selection also needs to distinguish varying coefficients from constant coefficients. We thus define BIC_w as

$$\mathrm{BIC}_w = \log(Q_N) + \frac{P_{vc}}{N}\log(N), \tag{4.1}$$

where Q_N is defined in (2.6), $P_{vc} = P_v + P_c$ is the total number of parameters, P_v is the number of parameters for the B-spines approximation of the varying coefficients, P_c is the number of parameters for the constant coefficients and those in ξ_t. For model (2.1), we also have $P_v = \sum_j^r P_{vj}$, $j = 1, \dots, r$, $P_{vj} = k + m - 2$, here k is number of knots and m is the smooth degree in (A3), r is the number of varying coefficients. Then it is easily seen that (4.1) should be written as

$$\mathrm{BIC}_w = \log(Q_N) + \log(N)\frac{r(k+m-2)+p+q}{N}, \tag{4.2}$$

here p is the number of linear variables, $q = q_1 + q_2$ is the number of coefficients for ARMA(q_1, q_2) of ξ_t. For varying coefficient variables, $\log(N)(k+m-2)/N$ is the penalty of adding one variable. In contrast, for linear variable, the penalty is $\log(N)/N$.

Finally, we write the procedure of model selection as follows. Let $S_{\max}$ denote the number of candidate variables to be considered. In our calculation, we fix $S_{\max} = \log(N)$. Assume the best model is $S_{0c} \cup S_{0v} = S_0 \subset \{1, 2, \dots, S_{\max}\}$, S_{0c} contains linear variables, S_{0v} contains varying-coefficient variables. For any number of knots k and threshold variable's lag $d \le S_{\max}$, the knots are equally placed between the 0.5 and 99.5 percentiles of the threshold variable. Generally, there are two ways to choose the knots, one is equally spaced, the other one is sample quantiles of the threshold variable. For our SVCARMA models, the threshold variable is usually a lagged variable which does not spread evenly and are very sparse near the boundaries, sample quantiles will choose too many knots in the middle of the data range, thus we use equally spaced knots in our calculation. The following four steps are summarized for our model selection procedure:

Step 1. Begin with two empty sets S_c and S_v, add one lagged variable at a time, and decide whether it is varying coefficient or linear by comparing BIC_w. Thus, there will be a total of $2S_{\max}$ different BIC_w to be compared in the first step. If $j \in \{1, 2, \ldots, S_{\max}\}$ as varying coefficient leads to the smallest BIC_w, then put j into set $S_v = \{j\}$; otherwise put j into set $S_c = \{j\}$.

Step 2. Begin with the updated S_c and S_v, repeat step 1 by comparing the rest candidate variables, and update S_c and S_v.

Step 3. Repeat steps 1 and 2 until the BIC_w could not be smaller.

Step 4. After step 3, we have S_c and S_v. As the final selection result might be affected by the order of entrance, it is necessary to check whether some variables in S_v could be moved into S_c using similar procedure as steps 1 and 2. Finally, a model with $S_w = S_c \cup S_v$ is generated.

Denote the selected set of variables as $S_w(d) = S_c(d) \cup S_v(d)$ when X_{t-d} is used as threshold variable. Denote the corresponding BIC_w by $\mathrm{BIC}_w(d)$. Then the $d^* = \arg\min_d \mathrm{BIC}_w(d)$ is the lag selected for the threshold variable. For the number of knots, we use $k \asymp k_c N^{1/5}$ where k_c is a tuning constant whose default value is 2 as Huang and Yang (2004) suggested. In our calculation, similar to d^*, we apply BIC_w to select the number of knots around $k_c N^{1/5}$ to control model complexity and model fitting. The selected number of knots $\hat{k}$ will still satisfy Assumption (A4). Consequently, Theorem 3.1 remains true when the selected number of knots and order are used.

5. Numerical studies. In this section, three numerical examples will be studied. Example 1 is a simulation study to check the performance of our estimation method and model selection procedure. Examples 2 and 3 demonstrate the usefulness of our modeling in practical application. We first use BIC_w to choose an appropriate model, including the number of knots, and then use the method in Section 2 to estimate the selected model, and finally out-of-sample prediction is made and compared with those prediction by other existing models.

EXAMPLE 1 (Simulation study of estimation consistency and model selection consistency). Consider the following five semi-varying coefficient models:

M.1: $X_t = 0.5X_{t-2} + \varepsilon_t + \theta\varepsilon_{t-1}$,

M.2: $X_t = 0.5X_{t-1} - 0.5X_{t-2} + \varepsilon_t + \theta\varepsilon_{t-1}$, M.3: $\quad X_t = 0.6\cos(X_{t-3})X_{t-1} - 0.4X_{t-2} + 0.2X_{t-4} + \varepsilon_t + \theta\varepsilon_{t-1},$

M.4: $\quad X_t = \bigl(0.8e^{-X_{t-2}^2} - 0.5\bigr)X_{t-1} + \bigl(-0.6e^{-0.5X_{t-2}^2+0.3}\bigr)X_{t-3} + \varepsilon_t + \theta\varepsilon_{t-1},$

M.5: $\quad X_t = 0.5\cos(X_{t-2})X_{t-2} - 0.3X_{t-3} + \varepsilon_t + \theta\varepsilon_{t-1},$

where ε_t are i.i.d. N(0, 1) and $\theta = 0.3$. For each model, samples of size $N = 200, 400, 600$ are generated. For each sample of size N, we use B-spline with degree of $m = 3$ to approximate varying coefficient functions and to estimate the model, estimators of varying and constant coefficients are denoted by $\hat{g}_t(u)$, $\hat{\beta}_t$ and $\hat{\theta}_t$, respectively, $t = 1, \dots, T$, with $T = 1000$ replications. We define

$$\mathrm{mse}(\hat{\boldsymbol{\beta}}) = T^{-1}\sum_{t=1}^{T}\|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}\|^2/p, \qquad \mathrm{mse}(\hat{\theta}) = T^{-1}\sum_{t=1}^{T}(\hat{\theta}_t - \theta)^2,$$

$$\mathrm{mse}\bigl(\hat{g}(u)\bigr) = T^{-1}\sum_{t=1}^{T}\sum_{u\in u_c}\bigl\|\hat{g}_t(u) - g(u)\bigr\|^2/L_u, \qquad \mathrm{bias}^2(\hat{\boldsymbol{\beta}}) = \|\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2/p,$$

$$\mathrm{bias}^2(\hat{\theta}) = \|\bar{\theta} - \theta\|^2/p, \qquad \mathrm{bias}^2\bigl(\hat{g}(u)\bigr) = \sum_{u\in u_c}\bigl\|\bar{g}_t(u) - g(u)\bigr\|^2/L_u,$$

where $u_c = (u_{0.025}, u_{0.025} + 0.05, \dots, u_{0.975})$, and $u_{0.025}$ and $u_{0.975}$ are sample quantiles of threshold variable X_{t-d}, L_u is the length of u_c. Note that ignorance of the serial correlation will cause estimation bias, thus we also define the square of bias as bias^2 in the above equations, where $\bar{\boldsymbol{\beta}}$ is the mean of $\{\hat{\boldsymbol{\beta}}_t, t = 1, \dots, T\}$, and similarly for $\bar{\theta}$ and $\bar{g}(u)$.

For comparison, we estimate the model without considering the serial correlation in the residuals, treating them as i.i.d. residuals. Now the corresponding mse and bias^2 defined above for the estimators are denoted as mse$_0$ and bias$_0^2$, respectively. Table 1 summarizes the results for all models. It can be seen that for M.2, M.3 and M.4 models, both mse and bias^2 become smaller as N increases. However, mse$_0$ and bias$_0^2$ remain big as N grows. Because for models M.1 and M.5, (1.2) can be fulfilled, the difference between mse and mse$_0$, and that between bias^2 and bias$_0^2$ are smaller than those of models M.2, M.3 and M.4, but mse and bias^2 are still smaller than mse$_0$ and bias$_0^2$.

Let V_0 be the percentage of $S_v \supset S_{0v}$, where S_{0v} represents the actual collection of varying coefficient variables for each model, S_v is the selected collection of varying coefficient variables. Let V_1 be the percentage of $S_{0v} = S_v$. Thus, V_0 depicts the probability that S_v covers S_{0v}, and V_1 means S_v is exactly the same as S_{0v}. For the constant coefficients, the percentages C_0 and C_1 are similarly defined as V_0 and V_1. Define PER to be the percentage of $\{S_{0c} = S_c\}\&\{S_{0v} = S_v\}$, which describes the probability that both varying coefficient variables and linear variables are accurately selected out. In this study, we first assume the threshold

TABLE 1
Estimation results of Example 1

Models	Criteria (10^{-3})	$N = 200$ β	θ	$g(u)$	$N = 400$ β	θ	$g(u)$	$N = 600$ β	θ	$g(u)$
M.1	mse	4.6	5.2	–	2.1	2.5	–	1.6	1.6	–
	bias^2	0.3	0.2	–	0.1	0.1	–	0.0	0.0	–
	mse_0	5.1	–	–	2.4	–	–	1.8	–	–
	bias_0^2	0.4	–	–	0.0	–	–	0.0	–	–
M.2	mse	8.2	14.4	–	4.1	6.5	–	2.3	3.9	–
	bias^2	0.0	0.2	–	0.0	0.0	–	0.0	0.0	–
	mse_0	23.8	–	–	22.3	–	–	22.4	–	–
	bias_0^2	20.5	–	–	20.8	–	–	21.5	–	–
M.3	mse	5.1	16.8	45.6	2.1	5.5	7.4	1.4	4.0	5.5
	bias^2	0.1	0.3	23.7	0.0	0.0	0.1	0.0	0.0	0.1
	mse_0	5.5	–	71.1	3.8	–	34.3	3.4	–	31.6
	bias_0^2	1.8	–	56.1	2.2	–	28.6	2.2	–	27.6
M.4	mse	–	20.0	43.0	–	6.7	10.5	–	4.2	5.9
	bias^2	–	1.5	30.0	–	0.1	3.3	–	0.0	0.2
	mse_0	–	–	66.4	–	–	27.4	–	–	21.3
	bias_0^2	–	–	56.5	–	–	20.9	–	–	15.8
M.5	mse	4.1	5.2	50.0	2.2	2.7	10.9	1.4	1.7	6.2
	bias^2	0.0	0.2	32.8	0.0	0.0	1.0	0.0	0.0	0.4
	mse_0	4.3	–	50.1	2.3	–	12.1	1.4	–	6.6
	bias_0^2	0.0	–	33.2	0.0	–	1.0	0.0	–	0.4

variable X_{t-d} to be known in calculating $\{C_0, C_1, V_0, V_1\}$, then assume X_{t-d} to be unknown and select it by comparing BIC_w. Let R_d denotes the percentage of accurately detecting threshold variable X_{t-d}. Because M.1 and M.2 have no varying coefficients, R_d means the percentage of selecting linear models. With 200 replications, the percentages of accurate model selection are calculated and summarized in Table 2. It shows that when sample size N grows to 600, the percentage of correctly detecting varying-coefficient variable is almost 100% for all models, and the percentage of correctly choosing linear variables is also greater than 95%. The fifth columns of each sub-table display the percentages of accurately selecting out both varying coefficient variables and linear variables, which is also quite satisfactory.

EXAMPLE 2 (The sunspots data). Tong (1990) modeled the square root transformed series $y_t = 2 \times (\sqrt{1 + x_t} - 1)$ of annual number of sunspots, x_t, for the period 1700–1979, by the threshold autoregressive (TAR) model with 11 lagged

TABLE 2
Model selection results of Example 1

	$N = 200$						$N = 400$						$N = 600$					
Models: %	V_0	V_1	C_0	C_1	**PER**	R_d	V_0	V_1	C_0	C_1	**PER**	R_d	V_0	V_1	C_0	C_1	**PER**	R_d
M.1	–	99.5	99.0	94.0	94.0	98.0	–	100	99.5	97.0	97.0	100	–	100	100	97.5	97.5	100
M.2	–	97.5	90.0	86.0	86.0	97.0	–	100	97.0	94.0	94.0	100	–	100	98.0	95.5	95.5	100
M.3	95.0	92.0	67.5	66.5	65.0	92.5	99.5	99.0	98.5	95.0	94.0	99.5	100	100	99.0	98.0	98.0	100
M.4	58.0	57.5	–	67.0	49.0	89.5	96.5	96.5	–	92.0	91.5	97.5	100	100	–	97.0	97.0	100
M.5	87.0	87.0	98.0	89.5	79.5	91.5	99.5	99.5	99.5	98.0	98.0	98.5	100	100	100	99.0	98.5	100

variables, that is, TAR(11),

$$y_t = \begin{cases} \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + \beta_5 y_{t-5} + \beta_6 y_{t-6} \\ \qquad + \beta_7 y_{t-7} + \beta_8 y_{t-8} + \beta_9 y_{t-9} + \beta_{10} y_{t-10} + \beta_{11} y_{t-11} + \varepsilon_t^1, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad y_{t-8} \le r, \\ \beta_{*0} + \beta_{*1} y_{t-1} + \beta_{*2} y_{t-2} + \beta_{*3} y_{t-3} + \varepsilon_t^2, \qquad y_{t-8} > r. \end{cases}$$

Chen and Tsay (1993) proposed a functional-coefficient autoregressive (FAR) model to fit the data, that is, FAR(8),

$$y_t = g_0(y_{t-3}) + g_1(y_{t-3}) y_{t-1} + g_2(y_{t-3}) y_{t-2} + g_8(y_{t-3}) y_{t-8} + \varepsilon_t.$$

We speculate the phenomenon that TAR needs 11 lagged variables may be caused by ignoring the autocorrelation of random errors, adding high order lagged variables is just like using truncated AR(p) process with large p to approximate an ARMA process. The FAR(8) model indirectly shows that the high lag in the TAR(11) is redundant. The functional coefficients might also be biased due to ignorance of the autocorrelation of random errors, as shown in Example 1. Therefore, a SVCARMA model should be applied to the sunspot data, to test whether a model with lower order lagged variables is sufficient if ARMA process for the residuals is used, and thus to improve the out-of-sample prediction.

We consider candidates of different smooth degrees $m = \{2, 3, 4\}$, and different number of knots $k = \{2, 3, 4, 5, 6, 7, 8\}$. For each pair of (m, k), the model selection procedure is executed and the BIC_w is calculated. Moreover, threshold variables, y_{t-8} and y_{t-3}, are also compared as they were used in the above TAR(11) and FAR(8).

Table 3 summaries all the results for comparison. Obviously, $\{m = 2, k = 4, d = 3\}$ leads to the minimum BIC_w. Thus, the most appropriate model in terms of BIC_w is the one taking y_{t-3} as threshold variable, and $y_{t-1}, y_{t-2}, y_{t-8}$ as varying coefficient variable and y_{t-5} as linear variable. Our final selected SVCARMA model with residuals being MA(1) takes the following form:

$$y_t = \beta_0 + g_1(y_{t-3}) y_{t-1} + g_2(y_{t-3}) y_{t-2} + g_8(y_{t-3}) y_{t-8} + \beta_5 y_{t-5} + \xi_t,$$
$$\text{where } \xi_t = \varepsilon_t + \theta \varepsilon_{t-1}.$$

TABLE 3
BIC_w *of different* m, k *and* d

$m \setminus k$	2	3	4	5	6	7	8
BIC_w				$d = 3$			
2	1.5324	1.5464	1.5247	1.5501	1.5744	1.5600	1.5640
3	1.5507	1.5321	1.5600	1.5600	1.5485	1.5485	1.5485
4	1.5510	1.5600	1.5600	1.5485	1.5485	1.5485	1.5485
BIC_w				$d = 8$			
2	1.5425	1.5600	1.5600	1.5600	1.5600	1.5600	1.5600
3	1.5600	1.5600	1.5485	1.5600	1.5600	1.5600	1.5600
4	1.5600	1.5485	1.5600	1.5600	1.5600	1.5600	1.5600

This model has the same lagged variables as FAR(8). However, we will show later that by assuming the random errors to be MA(1), the out-of-sample prediction could be improved.

Figure 1 shows the estimated varying coefficient functions for lagged variables y_{t-1}, y_{t-2} and y_{t-8}. We also carry out model diagnostics for the proposed model. Applying the Ljung–Box Q-test [see, Ljung and Box (1978) and McLeod and Li (1983)] to $\{\xi_t, t = 1, 2, \dots, N\}$, we reject (with the p-value < 0.0001) the hypothesis that ξ_t is white-noise. However, using the estimated $\hat{\theta} = -0.377$, we calculate the innovations $\{\varepsilon_t, t = 1, 2, \dots, N\}$. The Ljung–Box Q-test accepts with p-value 0.2217 that $\varepsilon_t, t = 1, 2, \dots, N$ is white-noise. We also carry out the Generalized Variance Portmanteau Test [see Mahdi and McLeod (2012)], the corresponding p-

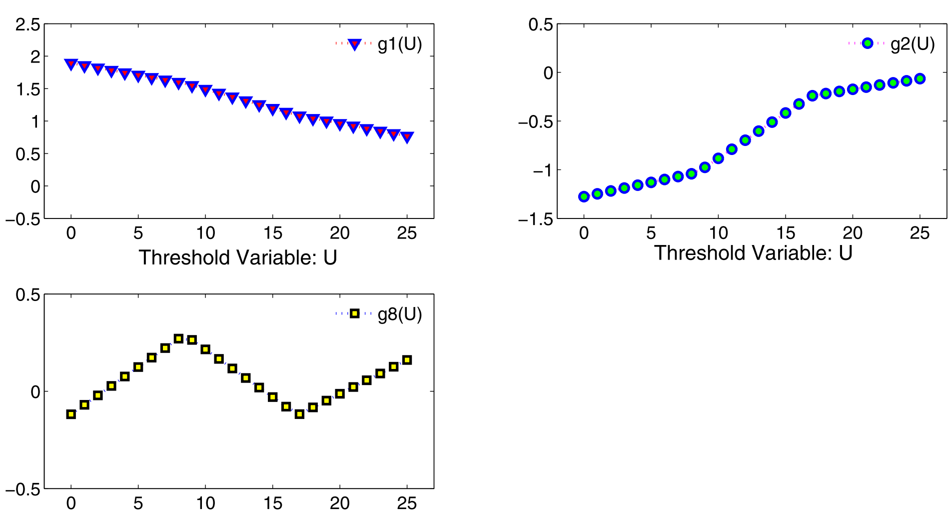

FIG. 1. *Estimated varying coefficient functions.*

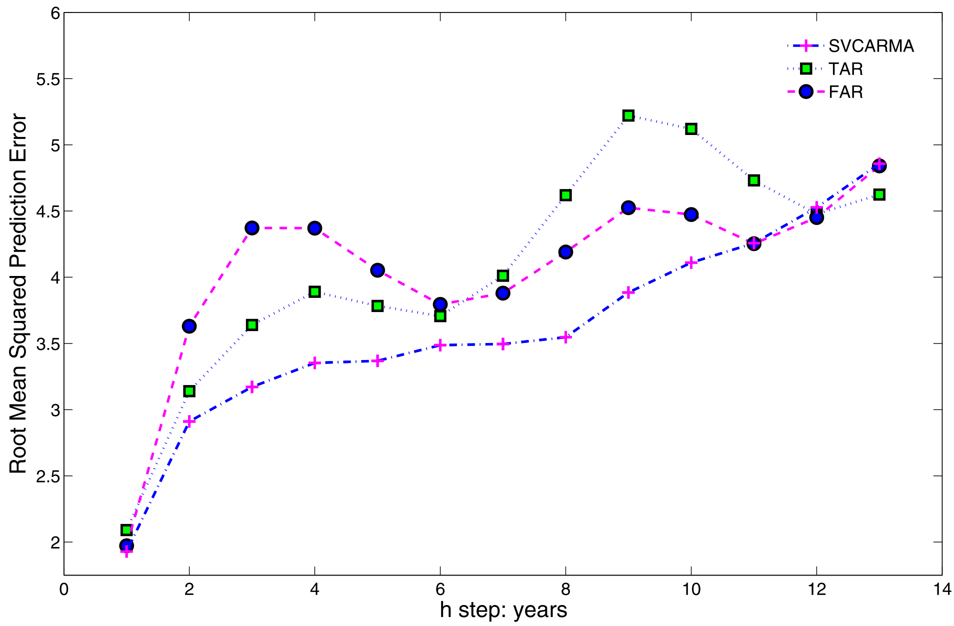

FIG. 2. *Out of sample RMSPE for Sun spots data.*

values for ξ_t is smaller than 0.0001, while that for ε_t is $= 0.1413$, suggesting the same conclusion for ξ_t and ε_t as the Ljung–Box Q-test does.

Finally, prediction ability is compared for the above three models. The 1-step ahead prediction is defined as $E(y_{t+1}|\mathbb{F}_t)$. For h ($\geq$2) steps ahead prediction, it can be calculated recursively. In this example, we set the largest $h = 13$ as Chen and Tsay (1993) did. After calculating all $h = 1, \dots, 13$ step ahead prediction for period 1700–1979, we shift forward the training data set by one unit of time, and calculate the 13 steps ahead predictions again, until no more data available. For each $h = \{1, 2, \dots, 13\}$, if we make K predictions, and the root mean squared prediction error (RMSPE) is calculated by

$$\mathrm{RMSPE}(h) = \sqrt{\frac{1}{K} \sum (\hat{y}_{t+h} - y_{t+h})^2}.$$

Figure 2 displays the out-of-sample RMSPE of the three competitive models. We can see that the FAR(8) is not better than TAR(11) when $h \leq 6$. However, our model can almost give better predictions for all $h < 11$ steps. This phenomenon shows that by adding a MA(1) structure to the random errors, the model can improve the accuracy of prediction of the data.

EXAMPLE 3 (Sea surface temperature data analysis). The El Niño Southern Oscillation (ENSO) and its economic effects have been analyzed by a lot of studies; see Adams et al. (1999), Glantz (2001) and Ubilava and Helmers (2013). The ENSO is represented by an abnormal increase (El Niño) or decrease (El Naño) of the Sea Surface Temperatures (SST). The SST for ESNO anomaly, marked as Niño 3.4, is defined by the Climate Prediction Center at the National Oceanic and Atmospheric Administration. This index measures the difference of SST in the area of the Pacific Ocean between 5°N–5°S and 170°W–120°W; see Trenberth and Stepaniak (2001). Consequently, SST anomaly is the deviation of the Niño 3.4 monthly measure from the average historic measure of that particular month from 1971–2000. For the SST anomaly, Ubilava and Helmers (2013) proposed a smooth transition autoregressive model to fit the data as follows:

$$\begin{aligned} y_t = {}& \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_6 y_{t-6} + \delta_1^\top M_t \\ & + (\beta_{*0} + \beta_{*1} y_{t-1} + \cdots + \beta_{*6} y_{t-6} + \delta_2^\top M_t) G_t + \varepsilon_t, \end{aligned} \tag{5.1}$$

where $G_t = (1 + \exp(-1.196/0.835(y_{t-1} + 0.447)))^{-1}$ and $M_t = (M_{t,1}, \ldots, M_{t,11})^T$ is a vector of dummy variables for different months of the year.

In this study, we consider SST anomaly data from January 1950 to December 2013. Our conjecture is that by using an ARMA$(1, 1)$ process for serial correlation of the random errors, the complexity of their model could be reduced and prediction ability can be improved. In the modeling, we have 6 lagged variables and 11 dummy variables to select. As is well known, when prediction is concerned, AIC usually performs better than BIC. Thus, we change the BIC penalty $\log(n)$ into AIC penalty 2 in (4.2). Because varying coefficient function G_t in model (5.1) has high order of smoothness, we use $m = 3$ for B-spline with threshold variable y_{t-1}. It is suggested by the AIC criterion of (4.2) that $k = 3$ is the most appropriate.

Thus, we set $\{m = 3, k = 3, d = 1\}$ in the modeling. The corresponding selected varying coefficient variables are $\{y_{t-6}, M_{t,1}, M_{t,2}, M_{t,7}, M_{10}\}$, and the linear variables $\{y_{t-1}, y_{t-2}, y_{t-3}, y_{t-5}, M_{t,6}, M_{t,8}\}$. The speculated SVCARMA model is thus

$$\begin{aligned} y_t = {}& \beta_0 + g(y_{t-1}) y_{t-6} + g_{m1}(y_{t-1}) M_{t,1} + g_{m2}(y_{t-1}) M_{t,2} \\ & + g_{m7}(y_{t-1}) M_{t,7} + g_{m10}(y_{t-1}) M_{t,10} \\ & + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_5 y_{t-5} + \beta_{m6} M_{t,6} + \beta_{m8} M_{t,8} + \xi_t, \\ & \text{where } \xi_t = \theta_a \xi_{t-1} + \varepsilon_t + \theta_m \varepsilon_{t-1}. \end{aligned} \tag{5.2}$$

The estimated varying coefficient functions are shown in Figure 3. Because $m = 3$ and $k = 3$, there are 4 B-splines coefficients for each varying coefficient function, thus model (5.2) has total number of parameters $4 \times 5 + 7 + 2 = 29$, that is, 5 varying coefficient functions, 7 constant coefficients and two ARMA$(1, 1)$ coefficients. In comparison, model (5.1) has $(7 + 11) \times 2 + 2 = 38$ parameters. This

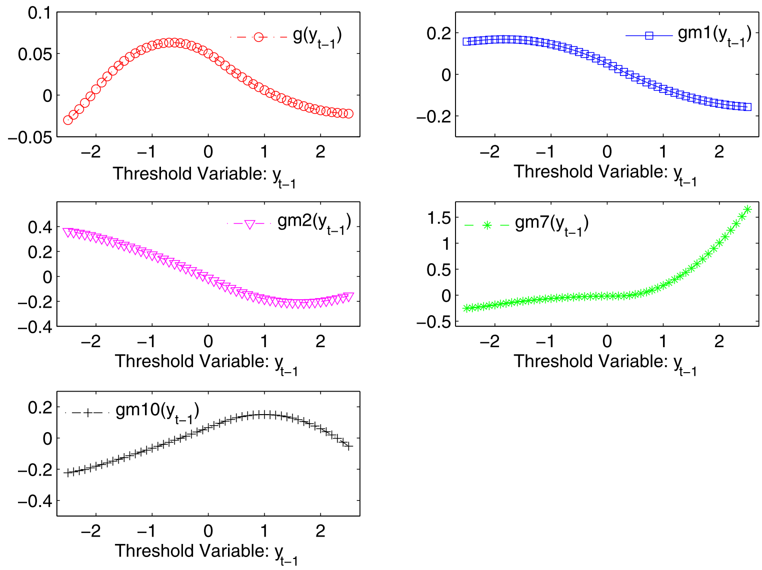

FIG. 3. *Estimated varying-coefficient functions.*

means model (5.2) is more concise than (5.1) in terms of the number of parameters. In other words, after considering the serial correlation in the residuals, the model can be simplified.

We estimate $\hat{\theta}_a = 0.238$, $\hat{\theta}_m = -0.842$ in (5.2). For model diagnostics, the Ljung–Box Q-test gives a p-value smaller than 0.0001 under the hypothesis that $\{\xi_t, t = 1, \dots, N\}$ is white noise; while the corresponding p-value for $\{\varepsilon_t, t = 1, \dots, N\}$ is 0.8462. The Generalized Variance Portmanteau Test gives p-values <0.0001 and 0.9386, respectively, for the two sequences. These p-values give strong support to our modeling that the serial correlation in the residuals exist and can be modeled by an ARMA(1, 1) process.

Finally, the out-of-sample prediction for model (5.1), denoted by LSTAR, and model (5.2), denoted by SVCARMA, is carried out. We also calculate the prediction of model (5.2) assuming i.i.d. errors, denoted by SVC, and the modified model of (5.1) in Wang and Xia (2014), denoted by LSTAR-MA. For this purpose, we split the whole data into two parts. The first part is from January 1950 to December 2006, used as the training data; the second part is from January 2007 to December 2013, used for the out-of-sample prediction. Figure 4 shows the RMSPE versus h-steps for all the four models. It can be easily seen that by adding an ARMA(1, 1) structure into the residuals, the out-of-sample prediction performs better, which is in consistency with Wang and Xia (2014). By allowing lagged variables and the dummy variables to have different varying coefficient functions, rather than taking all of them to be G_t in (5.1), the out-of-sample prediction is further improved.

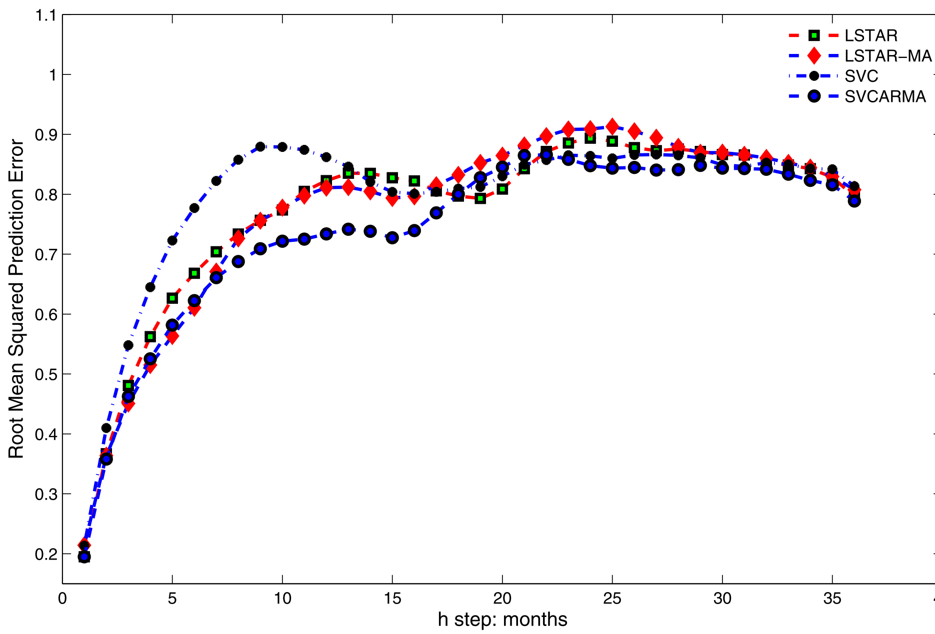

FIG. 4. *Out of sample RMSPE for SST anomaly.*

6. Conclusion. Serial correlation in regression residuals have long and commonly been noticed in the literature. However, for dynamical nonlinear models, the serial correlated residuals causes many problems in both model identification and estimation, and thus were not well addressed before. In this paper, a profile based Whittle Likelihood estimation is proposed and studied in detail for a SVCARMA model. Asymptotical properties have been built and a model selection procedure has been proposed, which can be easily applied to other semiparametric models; numerical studies also demonstrated some merits of our model and method in nonlinear time series analysis. However, some issues such as the consistency of the BIC_w criterion need further investigation.

APPENDIX: PROOFS OF THEOREMS

PROOF OF LEMMA 2.1. We first consider the special case when $\xi_t = \theta(B)\varepsilon_t$ with $\theta(B) = 1 + \theta B$, that is, ξ_t is MA(1). As $\varepsilon_t = X_t - g(X_{t-\mathbb{S}_g}) - \theta\varepsilon_{t-1}$, thus $\varepsilon_{t-1} = X_{t-1} - g(X_{t-1-\mathbb{S}_g}) - \theta\varepsilon_{t-2}$, after infinite iterations, we have

$$X_t = g(X_{t-\mathbb{S}_g}) + \varepsilon_t + \sum_{j=1}^{\infty}(-\theta)^j\{g(X_{t-j-\mathbb{S}_g}) - X_{t-j}\} + \lim_{n\to\infty}(-\theta)^n\varepsilon_{t-n}.$$

Under assumption (A1), $(-\theta)^n \varepsilon_{t-n} \overset{\text{a.s.}}{\to} 0$, thus by the dominated convergence theorem,

$$E(X_t|\mathbb{F}_{t-1}) = g(X_{t-\mathbb{S}_g}) + \sum_{j=1}^{\infty}(-\theta)^j\{g(X_{t-j-\mathbb{S}_g}) - X_{t-j}\}. \tag{A.1}$$

As $\xi_t = X_t - g(X_{t-\mathbb{S}_g})$, (A.1) is equivalent to

$$E\left(\xi_t + \sum_{j=1}^{\infty}(-\theta)^j \xi_{t-j}\middle|\mathbb{F}_{t-1}\right) = 0$$

and in back-shift polynomial,

$$E(\theta(B)^{-1}\xi_t|\mathbb{F}_{t-1}) = E(\varepsilon_t|\mathbb{F}_{t-1}) = 0. \tag{A.2}$$

When $\xi_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$, that is, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$, we can extend (A.2) in the same way. Suppose the uniqueness is not true, that is, there exist $g(X_{t-\mathbb{S}_g})$, $g'(X_{t-\mathbb{S}_g})$, $\theta(B)$ and $\theta'(B)$ in model (2.2) such that

$$X_t = g(X_{t-\mathbb{S}_g}) + \xi_t \qquad \text{where } \xi_t = \theta(B)\varepsilon_t \tag{A.3}$$

and

$$X_t = g'(X_{t-\mathbb{S}_g}) + \xi'_t \qquad \text{where } \xi'_t = \theta'(B)\varepsilon'_t. \tag{A.4}$$

According to (A.1) and (A.2), we have

$$\mu = E(X_t|\mathbb{F}_{t-1}) = g(X_{t-\mathbb{S}_g}) - \sum_{j=1}^{\infty}\varphi_j\xi_{t-j}, \tag{A.5}$$

and

$$\mu = E(X_t|\mathbb{F}_{t-1}) = g'(X_{t-\mathbb{S}_g}) - \sum_{j=1}^{\infty}\varphi'_j\xi'_{t-j}, \tag{A.6}$$

from (A.3) and (A.4), respectively, where φ_j is the corresponding coefficient of polynomial $\theta(B)^{-1} = (1 + \theta_1 B + \cdots + \theta_q B^q)^{-1} = 1 + \sum_{j=1}^{\infty}\varphi_j B^j$, and φ'_j is similarly defined. By (A.5) and (A.6), calculate their difference

$$\begin{aligned}
0 &= g(X_{t-\mathbb{S}_g}) - g'(X_{t-\mathbb{S}_g}) - \sum_{j=1}^{\infty}\varphi_j\xi_{t-j} + \sum_{j=1}^{\infty}\varphi'_j\xi'_{t-j} \\
&= g(X_{t-\mathbb{S}_g}) - g'(X_{t-\mathbb{S}_g}) + \sum_{j=1}^{\infty}\varphi_j\{g(X_{t-j-\mathbb{S}_g}) - X_{t-j}\} \\
&\quad - \sum_{j=1}^{\infty}\varphi'_j\{g'(X_{t-j-\mathbb{S}_g}) - X_{t-j}\}
\end{aligned} \tag{A.7}$$

$$= g(X_{t-\mathbb{S}_g}) - g'(X_{t-\mathbb{S}_g}) + \sum_{j=1}^{\infty} \varphi_j \{g(X_{t-j-\mathbb{S}_g}) - g'(X_{t-j-\mathbb{S}_g})\}$$

$$+ \sum_{j=1}^{\infty} \varphi_j \{g'(X_{t-j-\mathbb{S}_g}) - X_{t-j}\} - \sum_{j=1}^{\infty} \varphi'_j \{g'(X_{t-j-\mathbb{S}_g}) - X_{t-j}\}.$$

Let $\eta_{t-1} = g(X_{t-\mathbb{S}_g}) - g'(X_{t-\mathbb{S}_g})$, and thus $B\eta_{t-1} = \eta_{t-2} = g(X_{t-1-\mathbb{S}_g}) - g'(X_{t-1-\mathbb{S}_g})$. Together with (A.3), the above equations can be written as

$$\theta(B)^{-1}\eta_{t-1} = \theta(B)^{-1}\xi'_t - \theta'(B)^{-1}\xi'_t = \frac{\theta'(B) - \theta(B)}{\theta(B)\theta'(B)}\xi'_t.$$

Because $\theta'(B)^{-1}\xi'_t = \varepsilon'_t$, the above equations lead to

$$\eta_{t-1} = (\theta'(B) - \theta(B))\varepsilon'_t. \tag{A.8}$$

Similarly, equation (A.7) with a different way of transformation leads to

$$\eta_{t-1} = (\theta'(B) - \theta(B))\varepsilon_t. \tag{A.9}$$

Equations (A.8) and (A.9) with assumption (A1) conclude that $\varepsilon_t = \varepsilon'_t$, because otherwise $\theta(B) = \theta'(B)$ will be deduced and proof of this lemma will be completed. When $\varepsilon_t = \varepsilon'_t$, model (A.3) and (A.4) can be written as

$$X_t = g(X_{t-\mathbb{S}_g}) + \theta(B)\varepsilon_t, \tag{A.10}$$

and

$$X_t = g'(X_{t-\mathbb{S}_g}) + \theta'(B)\varepsilon_t. \tag{A.11}$$

Let $X_t - g(X_{t-\mathbb{S}_g}) = G(X_{t-\mathbb{S}_g})$ and $X_t - g'(X_{t-\mathbb{S}_g}) = G'(X_{t-\mathbb{S}_g})$. Under assumption (A1) and (A2), we have

$$\theta'(B)G(X_{t-\mathbb{S}_g}) = \theta(B)G'(X_{t-\mathbb{S}_g}).$$

Therefore,

$$G(X_{t-\mathbb{S}_g}) = \theta'(B)^{-1}\theta(B)G'(X_{t-\mathbb{S}_g})$$

which contradicts with the assumption (A2). Thus, $\theta_a(B)$ and $G(X_{t-\mathbb{S}_g})$ are both unique.

Now for general ARMA(q_1, q_2) errors, that is, $\xi_t = \theta_a(B)^{-1}\theta_m(B)$, (2.2) can be written as

$$\theta_a(B)G(X_{t-\mathbb{S}_g}) = \theta_m(B)\varepsilon_t.$$

The above argument has proved that $\theta_m(B)$ and $\theta_a(B)G(X_{t-\mathbb{S}_g})$ are unique. Suppose there exists $\theta_a^*(B)G^*(X_{t-\mathbb{S}_g}) = \theta_a(B)G(X_{t-\mathbb{S}_g})$, then

$$G(X_{t-\mathbb{S}_g}) = \theta_a(B)^{-1}\theta_a^*(B)G^*(X_{t-\mathbb{S}_g})$$

which contradicts to assumption (A2). Therefore, $\theta_a(B)$ and $G(X_{t-\mathbb{S}_g})$ are both unique. $\square$

LEMMA A.1. *Let $G_{Nk} = \mathbb{D}^\top E_N \mathbb{D}/N$ where $\mathbb{D}$ is defined in* (2.9). *Let $\lambda_{rk}^{G_{Nk}}$ and $\lambda_1^{G_{Nk}}$ be the minimum and maximum eigenvalues of G_{Nk}. Under assumptions* (A5) *and* (A6), *the eigenvalues are bounded in probability by*

$$(C_{11} + o_p(1))h \le \lambda_{rk}^{G_{Nk}} \le \lambda_1^{G_{Nk}} \le (C_{12} + o_p(1))h, \tag{A.12}$$

where $0 < C_{11} \le C_{22} < \infty$, h is the maximum distance between two adjacent knots in (A4).

PROOF. Let $\underline{a}^\top = \{\underline{a}_1^\top, \dots, \underline{a}_r^\top\}$ with $\underline{a}_j = \{a_{j1}, \dots, a_{jk}\}^\top$, $j = 1, \dots, r$ be column vectors such that $\sum_{j=1}^r \sum_{\ell=1}^k a_{j\ell}^2 = 1$. By the definition of eigenvalues, the minimum eigenvalue should be

$$\lambda_{rk}^{G_{Nk}} = \min_{\sum\sum a_{j\ell}^2=1} \left\{ \frac{1}{N} \underline{a}^\top \mathbb{D}^\top E_N \mathbb{D}\underline{a} \right\} = \min_{\sum\sum a_{j\ell}^2=1} \operatorname{Tr}\left\{ \frac{1}{N} (\mathbb{D}\underline{a})^\top E_N \mathbb{D}\underline{a} \right\}.$$

Lemma 6.5 of Zhou, Shen and Wolfe (1998) shows that, for positive semidefinite matrices A and B, $\lambda_{\min}^A \operatorname{Tr}(B) \le \operatorname{Tr}(AB) \le \lambda_{\max}^A \operatorname{Tr}(B)$. It is easy to see that E_N and $\mathbb{D}\underline{a}(\mathbb{D}\underline{a})^\top$ are both positive semidefinite matrices. Thus, the following inequality follows:

$$\lambda_{rk}^{G_{Nk}} = \min_{\sum\sum a_{j\ell}^2=1} \operatorname{Tr}\left\{ \frac{1}{N} E_N \mathbb{D}\underline{a}(\mathbb{D}\underline{a})^\top \right\} \ge \min_{\sum\sum a_{j\ell}^2=1} \lambda_{\min}^{E_N} \operatorname{Tr}\left\{ \frac{1}{N} \mathbb{D}\underline{a}(\mathbb{D}\underline{a})^\top \right\}.$$

By (2.9), $\mathbb{D}\underline{a} = \{D(\mathbf{X}_{t-1})B_{Nk}(\mathbf{X}_{t-d})\underline{a}_1, \dots, D(\mathbf{X}_{t-r})B_{Nk}(\mathbf{X}_{t-d})\underline{a}_r\}$, hence

$$\begin{aligned}
\lambda_{rk}^{G_{Nk}} &\ge \lambda_{\min}^{E_N} \min_{\sum\sum a_{j\ell}^2=1} \sum_{j=1}^r \operatorname{Tr}\left\{ \frac{1}{N} D(\mathbf{X}_{t-j}) B_{Nk}(\mathbf{X}_{t-d}) \right. \\
&\qquad \left. \times \underline{a}_j \big(D(\mathbf{X}_{t-j}) B_{Nk}(\mathbf{X}_{t-d})\underline{a}_j\big)^\top \right\} \\
&= \lambda_{\min}^{E_N} \min_{\sum\sum a_{j\ell}^2=1} \sum_{j=1}^r \left\{ \iint_{\Re_j} s_j^2(u) x^2 \, dF_{Nj}(u, x) \right\},
\end{aligned}$$

where $s_j(u) = \sum_{\ell=1}^k a_{j\ell} B_\ell(u) \in S(m, \underline{c})$. Under (A6), directly applying Lemma 6.1 of Zhou, Shen and Wolfe (1998), we have

$$\begin{aligned}
\lambda_{rk}^{G_{Nk}} &\ge \lambda_{\min}^{E_N} \min_{\sum a_{j\ell}^2=1} \sum_{j=1}^r C_{Uj} \int_a^b s_j^2(u) \, dF(u) \qquad \text{by (3.2)} \\
&\ge \lambda_{\min}^{E_N} \min_{\sum a_{j\ell}^2=1} \sum_{j=1}^r C_{Uj} \big(C_j^* + o_p(1)\big) h \sum_{\ell=1}^k a_{j\ell}^2 \\
&\ge \lambda_{\min}^{E_N} \big(C_{\min} + o_p(1)\big) h \sum_{j=1}^r \sum_{\ell=1}^k a_{j\ell}^2 = \big(C_{11} + o_p(1)\big) h,
\end{aligned}$$

where $C_{\min} = \min_{1\le j\le r} C_{Uj} \times C_j^* > 0$ and $C_{11} = \lambda_{\min}^{E_N} C_{\min}$. By Lemma 1 of Xiao and Wu (2012), it can be derived that any eigenvalue of E_N is bounded between $\min_\omega f^{-1}(\omega, \boldsymbol{\theta})$ and $\max_\omega f^{-1}(\omega, \boldsymbol{\theta})$. Then (A5) with inequations (3.1) leads to that λ^{E_N} is positive and bounded away from zero. Therefore, $C_{11} > 0$. Furthermore, the other side of (A.12) can be shown with the same process. □

LEMMA A.2. *Suppose assumptions* (A3)–(A7) *hold. Then for any fixed* $u \in [a, b]$, *estimator* $\hat{\mathbf{g}}(u, \boldsymbol{\alpha}_0) = \{\hat{g}_1(u, \boldsymbol{\alpha}_0), \dots, \hat{g}_r(u, \boldsymbol{\alpha}_0)\}^\top$ *is asymptotically normal*

$$\sqrt{Nh}\{\hat{\mathbf{g}}(u, \boldsymbol{\alpha}_0) - (\mathbf{g}(u) + \mathbf{b}(u) + \boldsymbol{\mu}_\nu^*)\} \xrightarrow{d} N(\mathbf{0}, \Sigma_u), \tag{A.13}$$

where $\boldsymbol{\mu}_\nu^* = O(h^m)$ *and* $\mathbf{b}(u) = \{b_1(u), \dots, b_r(u)\}^\top$ *are biases due to B-spline approximation*:

$$b_j(u) = -\frac{g_j^{(m)}(u) h_i^m}{m!} P_m\left(\frac{u - c_i}{h_i}\right),$$

here $P_m(\cdot)$ *is the mth Bernoulli polynomial which is recursively defined as*

$$P_0(u) = 1, \qquad P_m(u) = \int_a^u m P_{m-1}(v)\, dv + b_m,$$

where $u \in [a, b]$, $b_m = -m \int_a^b \int_a^u P_{m-1}(v)\, dv\, du$ *is the mth Bernoulli number defined in Barrow and Smith* (1978/79).

PROOF. By the result of Barrow and Smith (1978/79), there exists $s_j(u) \in S(m, \underline{c})$ for each $j = 1, \dots, r$, such that

$$\inf_{s_j(u)\in S(m,\underline{c})} \|g_j(u) + b_j^*(u) - s_j(u)\|_{L_\infty} = o(h^m),$$

$$b_j^*(u) = -\frac{g_j^{(m)}(c_i) h_i^m}{m!} P_m\left(\frac{u - c_i}{h_i}\right),$$

where $\|\cdot\|_{L_\infty}$ is the maximum norm. Under (A3), there exist $b_j(u) = b_j^*(u) + o(h^m)$ and $s_{gj}(u) \in S(m, \underline{c})$ such that

$$g_j(u) + b_j(u) - s_{gj}(u) = o(h^m). \tag{A.14}$$

Note that

$$\begin{aligned} &D(B(u)) G_{Nk}^{-1} \frac{1}{N} \mathbb{D}^\top E_N \{D(\mathbf{X}_{t-1}) s_{g1}(\mathbf{X}_{t-d}) + \dots + D(\mathbf{X}_{t-r}) s_{gr}(\mathbf{X}_{t-d})\} \\ &\quad = D(B(u)) G_{Nk}^{-1} \frac{1}{N} \mathbb{D}^\top E_N \mathbb{D} \boldsymbol{\gamma} = D(B(u)) \boldsymbol{\gamma} = \mathbf{s_g}(u). \end{aligned}$$

We have the following decomposition:

$$
\begin{aligned}
&\hat{\mathbf{g}}(u,\boldsymbol{\alpha}_0)-\big(\mathbf{g}(u)+\mathbf{b}(u)\big)\\
&\quad=\bigg\{\hat{\mathbf{g}}(u,\boldsymbol{\alpha}_0)-D\big(B(u)\big)G_{Nk}^{-1}\frac{1}{N}\mathbb{D}^\top E_N\{D(\mathbf{X}_{t-1})g_1(\mathbf{X}_{t-d})+\cdots\\
&\qquad+D(\mathbf{X}_{t-r})g_r(\mathbf{X}_{t-d})\}\bigg\}\\
&\qquad+\bigg\{D\big(B(u)\big)G_{Nk}^{-1}\frac{1}{N}\mathbb{D}^\top E_N\{D(\mathbf{X}_{t-1})(g_1-s_{g1})+\cdots\\
&\qquad+D(\mathbf{X}_{t-r})(g_1-s_{gr})\}\bigg\}\\
&\qquad+\big\{\mathbf{s}_{\mathbf{g}}(u)-\big(\mathbf{g}(u)+\mathbf{b}(u)\big)\big\}\\
&\quad\overset{\text{def}}{=}\text{part (i)}+\text{part (ii)}+\text{part (iii)},
\end{aligned}
\tag{A.15}
$$

where $\mathbf{s}_{\mathbf{g}}(u)=\{s_{g1}(u),\dots,s_{gr}(u)\}^\top$. Next, we investigate the above three parts separately. First, by equation (A.14), part (iii) is equal to

$$\mathbf{s}_{\mathbf{g}}(u)-\big(\mathbf{g}(u)+\mathbf{b}(u)\big)=\mathbf{o}\big(h^m\big).$$

For part (ii), write

$$
\begin{aligned}
&D\big(B(u)\big)G_{Nk}^{-1}\frac{1}{N}\mathbb{D}^\top E_N\{D(\mathbf{X}_{t-1})(g_1-s_{g1})+\cdots+D(\mathbf{X}_{t-r})(g_1-s_{gr})\}\\
&\quad=D\big(B(u)\big)G_{Nk}^{-1}\underline{\boldsymbol{\tau}},
\end{aligned}
$$

where $\underline{\boldsymbol{\tau}}=(\tau_{ji},\ j=1,\dots,r,i=1,\dots,k)_{rk\times 1}^\top$ and

$$\tau_{ji}=\frac{1}{N}B_i(\mathbf{X}_{t-d})D(X_{t-j})E_N\big\{D(\mathbf{X}_{t-1})(g_1-s_{g1})+\cdots+D(\mathbf{X}_{t-r})(g_1-s_{gr})\big\}.$$

Under (A3), $b_j(u)=O(h^m)$ according to its formula. Besides, under (A7), each $D(\mathbf{X}_{t-j})(g_j-s_{gj})$ is $O_p(h^m)$, thus each $\tau_{ji}=O_p(h^m/\sqrt{N})=o_p(h^m)$. By Lemma A.1, G_{Nk}'s eigenvalues are within interval $[C_{11}h,C_{22}h]$, and $0\le B_j(u)\le 1$. Thus, part (ii) is converging to 0 in probability with order $\mathbf{o}_p(h^m)$. Lastly, by model assumption (2.3) and estimator (2.15), part (i) turns out to be

$$
\begin{aligned}
\text{part (i)}&=D\big(B(u)\big)G_{Nk}^{-1}\frac{1}{N}\mathbb{D}^\top E_N\big\{\mathbf{Y}-\mathbf{X}_{Np}\boldsymbol{\beta}_0\\
&\qquad-\big\{D(\mathbf{X}_{t-1})g_1(\mathbf{X}_{t-d})+\cdots+D(\mathbf{X}_{t-r})g_r(\mathbf{X}_{t-d})\big\}\big\}\\
&=D\big(B(u)\big)G_{Nk}^{-1}\frac{1}{N}\mathbb{D}^\top E_N\boldsymbol{\xi}_0.
\end{aligned}
\tag{A.16}
$$

Therefore, it follows from (A.15) that

$$\text{(A.17)}\qquad \begin{aligned}&\hat{\mathbf{g}}(u,\boldsymbol{\alpha}_0)-\big(\mathbf{g}(u)+\mathbf{b}(u)\big)\\ &\quad = D\big(B(u)\big)G_{Nk}^{-1}\mathbb{D}^\top\frac{1}{N}E_N(\boldsymbol{\xi}_0)+\mathbf{o}\big(h^m\big)+\mathbf{o}_p\big(h^m\big),\end{aligned}$$

where $\boldsymbol{\xi}_0=\{\xi_{01},\dots,\xi_{0N}\}^\top$ is the ARMA process with the true parameters. Multiplying $\sqrt{Nh}$ to both sides of (A.17), we have

$$\begin{aligned}&\sqrt{Nh}\big\{\hat{\mathbf{g}}(u,\boldsymbol{\alpha}_0)-\big(\mathbf{g}(u)+\mathbf{b}(u)\big)\big\}\\ &\quad=\sqrt{Nh}D\big(B(u)\big)G_{Nk}^{-1}\mathbb{D}^\top\frac{1}{N}E_N\boldsymbol{\xi}_0+\sqrt{Nh}\mathbf{o}_p\big(h^m\big).\end{aligned}$$

Under (A3) and (A4), $h=O(N^{-1/(2m+1)})$, then $\sqrt{Nh}\mathbf{o}(h^m)=\sqrt{N}\mathbf{o}(h^{m+1/2})=\mathbf{o}(1)$. Similarly, we have $\sqrt{Nh}\mathbf{o}_p(h^m)=\mathbf{o}_p(1)$. Therefore, applying Slutsky's theorem, the rest of the proof is to show that

$$\text{(A.18)}\qquad \sqrt{Nh}(\phi_1,\dots,\phi_r)^\top=\sqrt{Nh}D\big(B(u)\big)G_{Nk}^{-1}\mathbb{D}^\top\frac{1}{N}E_N\boldsymbol{\xi}_0$$

converge to normal distribution, where each $\phi_j=V_{\mathrm{ec}}(\mathbf{B}_j)G_{Nk}^{-1}\mathbb{D}^\top\frac{1}{N}E_N\boldsymbol{\xi}_0$, here $V_{\mathrm{ec}}(\mathbf{B}_j)=\{\mathbf{0},\dots,\mathbf{0}_{j-1},B(u),\mathbf{0}_{j+1},\dots,\mathbf{0}\}$, and $\mathbf{0}=\{0,\dots,0\}_{1\times k}$. Write

$$\text{(A.19)}\qquad \begin{aligned}\sqrt{Nh}\phi_1&=\sqrt{Nh}V_{\mathrm{ec}}(\mathbf{B}_1)G_{Nk}^{-1}\mathbb{D}^\top\frac{1}{N}E_N\boldsymbol{\xi}_0\\ &=\frac{1}{\sqrt{N}}\boldsymbol{\eta}_1^\top\boldsymbol{\zeta},\end{aligned}$$

where $\boldsymbol{\eta}_1^\top=\sqrt{h}V_{\mathrm{ec}}(\mathbf{B}_1)G_{Nk}^{-1}\mathbb{D}^\top$, and $\boldsymbol{\zeta}=E_N\boldsymbol{\xi}_0$. Similarly, with Lemma A.1, $\mathrm{Var}(\sqrt{Nh}\phi_1)=hV_{\mathrm{ec}}(\mathbf{B}_1)G_{Nk}^{-1}\mathbb{D}^\top\frac{1}{N}E_N\Sigma_0E_N\mathbb{D}G_{Nk}^{-1}V_{\mathrm{ec}}(\mathbf{B}_1)^\top$. First, consider the covariance of $\boldsymbol{\zeta}$,

$$\mathrm{Cov}(\boldsymbol{\zeta},\boldsymbol{\zeta})=E_N\Sigma_0E_N,$$

where Σ_0 is the covariance matrix of ξ_t,

$$\Sigma_0=\begin{pmatrix}\frac{1}{2\pi N}\sum\frac{1}{f^{-1}(\omega_n,\boldsymbol{\theta})} & \frac{1}{2\pi N}\sum\frac{\cos(\omega_n)}{f^{-1}(\omega_n,\boldsymbol{\theta})} & \cdots & \frac{1}{2\pi N}\sum\frac{\cos((N-1)\omega_n)}{f^{-1}(\omega_n,\boldsymbol{\theta})}\\ \vdots & \cdots & \cdots & \vdots\\ \vdots & \ddots & \ddots & \vdots\\ \frac{1}{2\pi N}\sum\frac{\cos((N-1)\omega_n)}{f^{-1}(\omega_n,\boldsymbol{\theta})} & \cdots & \cdots & \frac{1}{2\pi N}\sum\frac{1}{f^{-1}(\omega_n,\boldsymbol{\theta})}\end{pmatrix}.$$

As $\{1,\cos(\omega),\dots,\cos((N-1)\omega)\}$ is orthogonal, that is, $\int\cos(j_2\omega)\times\cos(j_1\omega)\,d\omega=0$ if $j_1\neq j_2$, it is easy to check $\Sigma_0E_N=\mathbb{I}$, an identity matrix. Thus, $\mathrm{Cov}(\boldsymbol{\zeta},\boldsymbol{\zeta})=E_N$. Under assumption (A5), $1/f(\omega,\boldsymbol{\theta})$ could be considered as a theoretical SDF of an ARMA$(q_2,q_1,\boldsymbol{\theta}^*)$ process, with reversed

$\boldsymbol{\theta}^* = \{\theta_{m1}, \dots, \theta_{mq2}, \theta_{a1}, \dots, \theta_{aq_1}\}$. The process is also a stationary and invertible process with $\theta_a(B)$ and $\theta_m(B)$ interchanged when compared to (2.1). Thus,

$$E(\zeta_t) = 0. \tag{A.20}$$

This also means that $\{\zeta_1, \dots, \zeta_N\}$ is α-mixing. For each k, let $G_k \overset{\text{def}}{=} \lim_{N\to\infty} G_{Nk}$. Following Lemma 6.4 of Zhou, Shen and Wolfe (1998), it can be derived that

$$\max_{i,j} \left| G_{Nk}^{-1}(i,j) - G_k^{-1}(i,j) \right| = o_p(h^{-1}).$$

Therefore,

$$\begin{aligned}
\sqrt{Nh}\phi_1 &= \sqrt{Nh} V_{\mathrm{ec}}(\mathbf{B}_1) G_{Nk}^{-1} \mathbb{D}^\top \frac{1}{N} E_N \boldsymbol{\xi}_0 \\
&= V_{\mathrm{ec}}(\mathbf{B}_1) G_k^{-1} \frac{\sqrt{h}}{\sqrt{N}} \mathbb{D}^\top E_N \boldsymbol{\xi}_0 + V_{\mathrm{ec}}(\mathbf{B}_1) \left(G_{Nk}^{-1} - G_k^{-1}\right) \frac{\sqrt{h}}{\sqrt{N}} \mathbb{D}^\top E_N \boldsymbol{\xi}_0 \\
&= V_{\mathrm{ec}}(\mathbf{B}_1) G_k^{-1} \frac{\sqrt{h}}{\sqrt{N}} \mathbb{D}^\top E_n \boldsymbol{\xi}_0 + o_p(1) \\
&= \frac{\boldsymbol{\eta}_1^\top \boldsymbol{\zeta}}{\sqrt{N}} + o_p(1) \overset{\text{def}}{=} \frac{1}{\sqrt{N}} \sum_{t=1}^N \nu_{1t} + o_p(1),
\end{aligned} \tag{A.21}$$

where, $\nu_{1t} = \eta_{1t}\zeta_t$,

$$\boldsymbol{\eta}_1^\top \overset{\text{def}}{=} \{\eta_{11}, \dots, \eta_{1N}\}^\top = V_{\mathrm{ec}}(\mathbf{B}_1) G_k^{-1} \sqrt{h} \mathbb{D}^\top, \qquad \eta_{1t} = V_{\mathrm{ec}}(\mathbf{B}_1) G_k^{-1} \sqrt{h} \mathbb{D}(t,:)^\top,$$

and

$$\begin{aligned}
\mathbb{D}(t,:) = \{&X_{t-1}B_1(X_{t-d}), \dots, X_{t-1}B_k(X_{t-d}), \dots, \\
&X_{t-r}B_1(X_{t-d}), \dots, X_{t-r}B_k(X_{t-d})\}.
\end{aligned}$$

Under (A7) that $\{X_t, t = 1, \dots, N\}$ is α-mixing, thus $\{\eta_{1t}, t = 1, \dots, N\}$ which can be considered as a measurable function of $\{X_{t-d}, X_{t-1}, \dots, X_{t-r}\}$ with finite r, is also α-mixing, and $\nu_{1t}, t = 1, \dots, N$ is also α-mixing. It can be checked that

$$\begin{aligned}
|E(\nu_{1t})| = |E(\eta_{1t}\zeta_t)| &\le \{E(\eta_{1t}^2)\}^{1/2}\{E(\zeta_t^2)\}^{1/2} \\
&\le C_0\{E(\sqrt{h}\,\mathrm{Vec}(\mathbf{B}_1) G_k^{-1} \mathbb{D}(t,:)^\top \mathbb{D}(t,:) G_k^{-1} \mathrm{Vec}(\mathbf{B}_1)' \sqrt{h})\}^{1/2} \\
&= C_0\{\sqrt{h}\,\mathrm{Vec}(\mathbf{B}_1) G_k^{-1} E(\mathbb{D}(t,:)^\top \mathbb{D}(t,:)) G_k^{-1} \mathrm{Vec}(\mathbf{B}_1)' \sqrt{h}\}^{1/2}.
\end{aligned}$$

Under condition (A6), similar to Lemma A.1, entries of matrix $E(\mathbb{D}(t,:)^\top \mathbb{D}(t,:)) \overset{\text{def}}{=} \Sigma_D$ should be $C_j h$, $j = 1, \dots, rk$ with $C_j > 0$. Therefore, eigenvalues of Σ_D is $O(h)$, which implies that eigenvalues of $G_k^{-1}\Sigma_D G_k^{-1}$ are $O(h^{-1})$. Consequently,

$$|\mu_{1\nu}| = |E(\nu_{1t})| \le C_0 h\,\mathrm{Vec}(\mathbf{B}_1)\{G_k^{-1}\Sigma_D G_k^{-1}\}\mathrm{Vec}(\mathbf{B}_1)' < M_B < \infty,$$

where M_B is a finite constant, which means $|\mu_{1\nu}|$ is uniformly bounded over u by M_B because entries of $\mathrm{Vec}(\mathbf{B}_1)$ are all positive and their summation is 1 due to the definition of B-spline. By Theorem 2.21 of Fan and Yao (2003), $1/\sqrt{N}\sum \nu_{1t}$ converges to normal distribution $N(\mu_{1\nu}, \Sigma_1)$. Convergence of $\sqrt{Nh}\phi_2, \dots, \sqrt{Nh}\phi_r$ can be similarly proved. As any linear combination of $\sqrt{Nh}\phi_1, \dots, \sqrt{Nh}\phi_r$ could be considered as a linear combination of α-mixing process whose asymptotic distribution can be built in the same way as for $\sqrt{Nh}\phi_1$, we finally have

$$\sqrt{Nh}(\phi_1, \dots, \phi_r)^\top \xrightarrow{d} N(\boldsymbol{\mu}_\nu, \Sigma_u),$$

where

$$\boldsymbol{\mu}_\nu = E(\boldsymbol{\nu}_t) \stackrel{\text{def}}{=} E\{\nu_{1t}, \dots, \nu_{rt}\}^\top, \qquad \Sigma_u = \sum_{j=-\infty}^{\infty} E(\boldsymbol{\nu}_t \boldsymbol{\nu}_{t+j}^\top),$$

where Σ_u exists and is finite under (A7). Finally, we have

$$\sqrt{Nh}\{\hat{\mathbf{g}}(u, \boldsymbol{\alpha}_0) - (\mathbf{g}(u) + \mathbf{b}(u) + \boldsymbol{\mu}_\nu^*)\} \xrightarrow{d} N(\mathbf{0}, \Sigma_u),$$

where $\boldsymbol{\mu}_\nu^* \stackrel{\text{def}}{=} \boldsymbol{\mu}_{\boldsymbol{\nu}}/\sqrt{Nh} = O(h^m)$ is bias caused by autocorrelation of regressors and random errors in model (2.1). Lemma A.2 is thus verified. □

LEMMA A.3. *Under assumptions* (A1) *to* (A7), *the estimator* $\hat{\boldsymbol{\alpha}} \to \boldsymbol{\alpha}_0$ *and* $\hat{\sigma}^2 = \mathbb{Q}_N(\hat{\boldsymbol{\alpha}}) \to \sigma_0^2$ *almost surely.*

PROOF. We first study the limit of $\mathbb{Q}_N(\boldsymbol{\alpha})$. Let $\xi_t(\boldsymbol{\theta}_0) = \theta_{0a}^{-1}(B)\theta_{0m}(B)\varepsilon_t$ be ARMA(q_1, q_2) errors under the true parameters $\boldsymbol{\theta}_0$. It has been shown by Lemma 2 of Hannan (1973) under (A5), that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_*(\omega, \boldsymbol{\theta}_0)}{f(\omega, \boldsymbol{\theta})}\, d\omega > \sigma_0^2, \qquad \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \tag{A.22}$$

where $f_*(\omega, \boldsymbol{\theta}_0)$ and $f(\omega, \boldsymbol{\theta})$ are respectively the theoretical SDF and standardized SDF defined in (2.5). Let $\mathbf{Z}(\boldsymbol{\alpha}) = \{Z_t(\boldsymbol{\alpha}), t = 1, \dots, N\}^\top = \mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta} - \mathbb{D}\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha})$, which approximates $\xi_t(\boldsymbol{\theta}_0)$ at $\boldsymbol{\alpha}$. Write

$$Z_t(\boldsymbol{\alpha}_0) = \xi_t(\boldsymbol{\theta}_0) + \mathrm{bias}_t,$$

where $\mathrm{bias}_t = \{1, X_{t-1}, \dots, X_{t-r}\}(\mathbf{b}(u) + \boldsymbol{\mu}_\nu^*)$, as shown in Lemma A.2 $\mathbf{b}(u) + \boldsymbol{\mu}_\nu^*$ is uniformly $\mathbf{O}(h^m)$. Under assumption (A7) the second moment of bias_t is bounded. Thus,

$$E(|\mathrm{bias}_t|^2) < C_b h^{2m}, \qquad \text{where } C_b < \infty.$$

By the definition of ACVF,

$$\lambda_{\xi_0}(\kappa) = E(\xi_t(\boldsymbol{\theta}_\mathbf{0})\xi_{t+\kappa}(\boldsymbol{\theta}_\mathbf{0})).$$

Comparing the ACVF of $Z_t(\boldsymbol{\alpha}_0)$ and $\xi_t(\boldsymbol{\theta}_0)$, we have

$$\begin{aligned}\lambda_{Z_0}(s) &= \mathrm{Cov}\big(Z_t(\boldsymbol{\alpha}_0), Z_{t+\kappa}(\boldsymbol{\alpha}_0)\big)\\ &= E\big(\{Z_t(\boldsymbol{\alpha}_0) - E\big(Z_t(\boldsymbol{\alpha}_0)\big)\}\{Z_{t+\kappa}(\boldsymbol{\alpha}_0) - E\big(Z_{t+\kappa}(\boldsymbol{\alpha}_0)\big)\}\big)\\ &= E\big(\{\xi_t(\boldsymbol{\theta}_0) + \mathrm{bias}_t - E(\mathrm{bias}_t)\}\{\xi_{t+\kappa}(\boldsymbol{\theta}_0) + \mathrm{bias}_{t+\kappa} - E(\mathrm{bias}_t)\}\big)\\ &= E\big(\xi_t(\boldsymbol{\theta}_0)\xi_{t+\kappa}(\boldsymbol{\theta}_0)\big) + E\big(\xi_{t+\kappa}\big(\mathrm{bias}_t - E(\mathrm{bias}_t)\big)\big)\\ &\quad + E\big(\xi_t\big(\mathrm{bias}_{t+\kappa} - E(\mathrm{bias}_{t+\kappa})\big)\big)\\ &\quad + E\big(\{\mathrm{bias}_t - E(\mathrm{bias}_t)\}\{\mathrm{bias}_{t+\kappa} - E(\mathrm{bias}_{t+\kappa})\}\big).\end{aligned}$$

By Hölder inequality, we have

$$\lambda_{Z_0}(\kappa) = \lambda_{\xi_0}(\kappa) + C_0 h^m.$$

Note here $Z_t(\boldsymbol{\alpha}_0)$ is a process related to sample size N, the above calculation shows that as N gets larger, $Z_t(\boldsymbol{\alpha}_0)$ approximates $\xi_t(\boldsymbol{\theta}_0)$ better. Because $\xi_t(\boldsymbol{\theta}_0)$ is a stationary and invertible process, for any $\epsilon > 0$ there exists $M > 0$ such that

$$\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_*(\omega,\boldsymbol{\theta}_0)}{f(\omega,\boldsymbol{\theta})}\,d\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{\kappa=-M}^{M}\lambda_{\xi_0}(\kappa)e^{-i\kappa\omega}}{f(\omega,\boldsymbol{\theta})}\,d\omega\right| < \frac{\epsilon}{3}. \tag{A.23}$$

As $\lambda_{Z_0}(\kappa) = \lambda_{\xi_0}(\kappa) + C_0 h^m$ and $h \to 0$ as $N \to \infty$, there exists $N_1 > 0$, such that when $N > N_1$,

$$\begin{aligned}&\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{\kappa=-M}^{M}\lambda_{Z_0}(\kappa)e^{-i\kappa\omega}}{f(\omega,\boldsymbol{\theta})}\,d\omega\right.\\ &\qquad\left. - \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{\kappa=-M}^{M}\lambda_{\xi_0}(\kappa)e^{-i\kappa\omega}}{f(\omega,\boldsymbol{\theta})}\,d\omega\right| < \frac{\epsilon}{3}.\end{aligned} \tag{A.24}$$

Since $\lambda_{\xi_0}(\kappa) \to 0$ exponentially as $\kappa \to \infty$, thus there exists N_2 such that for any $N > N_2$,

$$\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{|\kappa|>M}\lambda_{Z_0}(\kappa)e^{-i\kappa\omega}}{f(\omega,\boldsymbol{\theta})}\,d\omega\right| < \frac{\epsilon}{3}. \tag{A.25}$$

Combining (A.22), (A.23), (A.24) and (A.25), we get

$$\lim_{N\to\infty}\mathbb{Q}_N(\boldsymbol{\alpha}_0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_z(\omega,\boldsymbol{\alpha}_0)}{f(\omega,\boldsymbol{\theta}_0)}\,d\omega = \sigma_0^2. \tag{A.26}$$

Moreover, when $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$, that is, $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| > 0$,

$$\begin{aligned}Z_t(\boldsymbol{\alpha}) &= \xi_t(\boldsymbol{\theta}_0) + \mathrm{bias}_t + X_{tp}(\beta_0 - \beta) + \mathbb{D}_t\big(\hat{\gamma}(\boldsymbol{\alpha}_0) - \hat{\gamma}(\boldsymbol{\alpha})\big)\\ &\overset{\mathrm{def}}{=} \xi_t(\boldsymbol{\theta}_0) + \mathrm{Bias}_t,\end{aligned}$$

where X_{tp} and $\mathbb{D}_t$ are corresponding tth rows of $\mathbf{X}_{Np}$ and $\mathbb{D}$, and Bias_t is a linear combination of $\{\text{bias}_t, X_{t-1}, \dots, X_{t-r-p}, X_{t-d}\}$. Thus, $f_z(\omega, \boldsymbol{\alpha})/f(\omega, \boldsymbol{\theta})$ can be regarded as the SDF of following process:

$$Z_t(\boldsymbol{\alpha}) - \varphi\big(Z_{t-1}(\boldsymbol{\alpha}), Z_{t-2}(\boldsymbol{\alpha}), \dots\big) = e_t^*, \tag{A.27}$$

where φ is set of constant coefficients related to $\boldsymbol{\theta}$. By Theorem 2.12 in Fan and Yao (2003),

$$Z_t(\boldsymbol{\alpha}) - \varphi\big(Z_{t-1}(\boldsymbol{\alpha}), Z_{t-2}(\boldsymbol{\alpha}), \dots\big) = \left(\sum_{j=0}^{\infty} \phi_j B^j\right) Z_t(\boldsymbol{\alpha}) = \theta_m^{-1}(B)\theta_a(B) Z_t(\boldsymbol{\alpha}).$$

As $1 = e^{-0i\omega}$, the following integration is equal to its ACVF $\lambda(0) = \text{Var}(e_t^*)$:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_z(\omega, \boldsymbol{\alpha})}{f(\omega, \boldsymbol{\theta})}\, d\omega = \text{Var}(e_t^*) = \sigma_*^2.$$

Note that $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$ means Case one $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, or Case two $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. For Case one, under (A5), we have

$$\lim_{N\to\infty} \mathbb{Q}_N(\boldsymbol{\alpha}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_z(\omega, \boldsymbol{\alpha})}{f(\omega, \boldsymbol{\theta})}\, d\omega = \sigma_*^2 \geq \sigma_0^2. \tag{A.28}$$

If the equality above satisfied, by (A5), $f_z(\omega, \boldsymbol{\alpha}) = \sigma_0^2 f(\omega, \boldsymbol{\theta})$, which means that there exist another model for X_t,

$$X_t = g'(X_{t-\mathbb{S}_g}) + \xi_t', \qquad \text{where } \xi_t' = \theta_a'(B)^{-1}\theta_m'(B)\varepsilon_t'.$$

This is a contradiction to Lemma 2.1, thus $\sigma_*^2 > \sigma_0^2$. For Case two, if the equality of (A.28) is satisfied, similar to the proof of (A.26), we have

$$Z_t(\boldsymbol{\alpha}) = \xi_t(\boldsymbol{\theta}_0) + o_p(1),$$

which will lead to $X_{tp}(\beta_0 - \beta) = o_p(1)$, and this is impossible when $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Therefore, we have shown that for any $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$,

$$\lim_{N\to\infty} \mathbb{Q}_N(\boldsymbol{\alpha}) > \sigma_0^2. \tag{A.29}$$

If $\hat{\boldsymbol{\alpha}}$ does not converge to $\boldsymbol{\alpha}_0$, then there exists a subsequence $\hat{\boldsymbol{\alpha}}_l$ converging to $\boldsymbol{\alpha}_* \in \mathbb{A}$ as $l \to \infty$, and $\boldsymbol{\alpha}_* \neq \boldsymbol{\alpha}_0$. By equation (A.29) with $\boldsymbol{\alpha}_* \neq \boldsymbol{\alpha}_0$, we have

$$\underline{\lim}_{l\to\infty} \mathbb{Q}_l(\hat{\boldsymbol{\alpha}}_l) = \underline{\lim}_{l\to\infty} \mathbb{Q}_l(\boldsymbol{\alpha}_*) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_z(\omega, \boldsymbol{\alpha}_*)}{f(\omega, \boldsymbol{\theta}_*)}\, d\omega > \sigma_0^2. \tag{A.30}$$

On the other hand, because $\hat{\boldsymbol{\alpha}}$ minimizes $\mathbb{Q}_N$, for any $\boldsymbol{\alpha} \in \mathbb{A}$, we have $\mathbb{Q}_l(\hat{\boldsymbol{\alpha}}) \leq \mathbb{Q}_l(\boldsymbol{\alpha})$ and

$$\overline{\lim}_{l\to\infty} \mathbb{Q}_l(\hat{\boldsymbol{\alpha}}_l) \leq \overline{\lim}_{l\to\infty} \mathbb{Q}_l(\boldsymbol{\alpha}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_z(\omega, \boldsymbol{\alpha})}{f(\omega, \boldsymbol{\theta})}\, d\omega.$$

Therefore, by (A.26),

$$\overline{\lim_{l\to\infty}}\, \mathbb{Q}_l(\hat{\boldsymbol{\alpha}}_l) \le \inf_{\boldsymbol{\alpha}\in\mathbb{A}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_z(\omega,\boldsymbol{\alpha})}{f(\omega,\boldsymbol{\theta})}\,d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_z(\omega,\boldsymbol{\alpha}_0)}{f(\omega,\boldsymbol{\theta}_0)}\,d\omega = \sigma_0^2. \tag{A.31}$$

Equation (A.31) contradicts with (A.30). Therefore, we must have $\boldsymbol{\alpha}_* = \boldsymbol{\alpha}_0$, and thus $\hat{\boldsymbol{\alpha}} \to \boldsymbol{\alpha}_0$ almost surely. The above arguments also lead to

$$\underline{\lim_{l\to\infty}}\, \mathbb{Q}_l(\hat{\boldsymbol{\alpha}}) \ge \sigma_0^2 \ge \overline{\lim_{l\to\infty}}\, \mathbb{Q}_l(\hat{\boldsymbol{\alpha}}),$$

and thus $\mathbb{Q}_N(\hat{\boldsymbol{\alpha}}) \to \sigma_0^2$ almost surely. $\square$

PROOF OF THEOREM 3.1. We apply Taylor expansion to each derivative of $\mathbb{Q}_N(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}_0$. There exists $\boldsymbol{\alpha}_*^j$, $j=1,\dots,J$ such that

$$\sqrt{N}\left(\frac{\partial \mathbb{Q}_N(\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\alpha}} - \frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}\right) = \sqrt{N}\left(\frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_*^1)}{\partial \alpha_1 \partial \boldsymbol{\alpha}^\top}, \dots, \frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_*^J)}{\partial \alpha_J \partial \boldsymbol{\alpha}^\top}\right)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0), \tag{A.32}$$

where $J = p + q_1 + q_2$ is length of $\boldsymbol{\alpha}$, and each $\boldsymbol{\alpha}_*^j$ is closer to $\boldsymbol{\alpha}_0$ than $\hat{\boldsymbol{\alpha}}_0$ in terms of Euclidean distance,

$$\|\boldsymbol{\alpha}_*^j - \boldsymbol{\alpha}_0\| \le \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|.$$

When $N \to \infty$, by Lemma A.3, $\hat{\boldsymbol{\alpha}} \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0$, thus $\boldsymbol{\alpha}_*^j \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0$, $j = 1,\dots,J$. Therefore, $\boldsymbol{\alpha}_*^j$, $j=1,\dots,J$ could be represented by a single $\boldsymbol{\alpha}_* \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0$ for convenience when $N \to \infty$, where $\|\boldsymbol{\alpha}_* - \boldsymbol{\alpha}_0\| \le \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|$. Equation (A.32) becomes

$$\sqrt{N}\left(\frac{\partial \mathbb{Q}_N(\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\alpha}} - \frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}\right) = \sqrt{N}\frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^\top}\big(1 + \mathbf{o}_p(1)\big)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0). \tag{A.33}$$

Because $\partial \mathbb{Q}_N(\hat{\boldsymbol{\alpha}})/\partial \boldsymbol{\alpha} = \mathbf{0}$, it suffices to prove

$$\sqrt{N}\frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \xrightarrow{d} N(\mathbf{0}, \Omega) \quad \text{and} \quad \frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^\top} \xrightarrow{p} \Phi.$$

Write

$$\sqrt{N}\frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} = \sqrt{N}\left(\frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\beta}^\top}, \frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\theta}^\top}\right)^\top \stackrel{\text{def}}{=} \left(Q_{\sqrt{N}\boldsymbol{\beta}}^\top, Q_{\sqrt{N}\boldsymbol{\theta}}^\top\right)^\top.$$

We first study $Q_{\sqrt{N}\boldsymbol{\beta}}$. Under assumption (A5), the difference between summation and integration could be controlled by $\mathbf{O}(N^{1/2-a})$,

$$Q_{\sqrt{N}\boldsymbol{\beta}} = \sqrt{N}\frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\beta}} = \frac{1}{\sqrt{N}} \sum_{\omega_n} \frac{\partial I(\omega_n, \mathbf{Z}(\boldsymbol{\alpha}_0))/\partial \boldsymbol{\beta}}{f(\omega_n, \boldsymbol{\theta}_0)}$$

$$= \sqrt{N}\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0))/\partial \boldsymbol{\beta}}{f(\omega, \boldsymbol{\theta}_0)}\, d\omega + \mathbf{O}(N^{1/2-a}).$$

By Lemma A.3, as the integration reaches the minimum at $\boldsymbol{\alpha}_0$, thus

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial f_z(\omega, \boldsymbol{\alpha}_0)/\partial \boldsymbol{\beta}}{f(\omega, \boldsymbol{\theta}_0)} = 0$$

and because $a > 1/2$ as assumed in (A5), we have $\mathbf{O}(N^{1/2-a}) = \mathbf{o}(1)$. Thus,

$$Q_{\sqrt{N}\boldsymbol{\beta}} = \sqrt{N}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\partial I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0))/\partial \boldsymbol{\beta}}{f(\omega, \boldsymbol{\theta}_0)} - \frac{\partial f_z(\omega, \boldsymbol{\alpha}_0)/\partial \boldsymbol{\beta}}{f(\omega, \boldsymbol{\theta}_0)}\right) d\omega + \mathbf{o}(1), \tag{A.34}$$

where the periodogram and SDF in the above equation have the following formulas by their definition:

$$I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0)) = \frac{1}{2\pi}\sum_{-N+1}^{N-1} c_z(\kappa) e^{-i\kappa\omega}, \qquad c_z(\kappa) = \frac{1}{N}\sum_{t=1}^{N-|\kappa|} Z_t(\boldsymbol{\alpha}_0) Z_{t+|\kappa|}(\boldsymbol{\alpha}_0),$$

$$f_z(\omega, \boldsymbol{\alpha}_0) = \frac{1}{2\pi}\sum_{-\infty}^{\infty} \lambda(\kappa) e^{-i\kappa\omega}, \qquad \lambda(\kappa) = \frac{1}{N-|\kappa|}\sum_{t=1}^{N-|\kappa|} E\{Z_t(\boldsymbol{\alpha}_0) Z_{t+|\kappa|}(\boldsymbol{\alpha}_0)\}.$$

Let $f'_{Nz}(\omega, \boldsymbol{\alpha}_0)$ be the Cesàro summation of $\partial f_z(\omega, \boldsymbol{\alpha}_0)/\partial \boldsymbol{\beta}$ up to N terms, so

$$f'_{Nz}(\omega, \boldsymbol{\alpha}_0) = \sum_{-N+1}^{N-1} \left(1 - \frac{|\kappa|}{N}\right) \frac{\partial \lambda(\kappa)}{\partial \boldsymbol{\beta}} e^{-i\kappa\omega}.$$

By assumption (A5), the difference between the Cesàro summation and the original function is bounded by $O(N^{-a})$, that is,

$$\sup_{\omega} \left| f'_{Nz}(\omega, \boldsymbol{\alpha}_0) - \frac{\partial f_z(\omega, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\beta}} \right| = O(N^{-a}).$$

See, for example, page 91 of Zygmund (1959). Thus, equation (A.34) reduces to

$$Q_{\sqrt{N}\boldsymbol{\beta}} = \sqrt{N}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{\partial I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0))/\partial \boldsymbol{\beta}}{f(\omega, \boldsymbol{\theta}_0)} - \frac{f'_{Nz}(\omega, \boldsymbol{\alpha}_0)}{f(\omega, \boldsymbol{\theta}_0)}\right) d\omega + \mathbf{o}_p(1)$$

$$= \sqrt{N}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{1}{2\pi}\sum_{-N+1}^{N-1} \left(\frac{\partial c_z(\kappa)}{\partial \boldsymbol{\beta}_0}\right.\right.$$

$$- \left(1 - \frac{|\kappa|}{N}\right) \frac{\partial \lambda_z(\kappa)}{\partial \boldsymbol{\beta}_0}\right) e^{-i\kappa\omega} f^{-1}(\omega, \boldsymbol{\theta}_0)\Bigg) \, d\omega + \mathbf{o}_p(1)$$

$$= \sqrt{N} \sum_{-N+1}^{N-1} \left(\frac{\partial c_z(\kappa)}{\partial \boldsymbol{\beta}_0} - \left(1 - \frac{|\kappa|}{N}\right) \frac{\partial \lambda_z(\kappa)}{\partial \boldsymbol{\beta}_0} \right) \rho_0(\kappa, \boldsymbol{\theta}_0) + \mathbf{o}_p(1),$$

where

$$\rho_0(\kappa, \boldsymbol{\theta}_0) = 1/(2\pi)^2 \int_{-\pi}^{\pi} f^{-1}(\omega, \boldsymbol{\theta}_0) e^{-i\kappa\omega} \, d\omega. \tag{A.35}$$

It is easy to see that $\rho_0(\kappa, \boldsymbol{\theta}_0)$ is the ACVF of ARMA$(q_2, q_1, \boldsymbol{\theta}^*)$ multiplied by $1/(2\pi)$, where $\boldsymbol{\theta}^*$ is the AR coefficients and MA coefficients being interchanged. For convenience, as $\mathbf{o}_p(1)$ is negligible according to Slutsky's theorem, we remove it symbolically afterward. Consequently,

$$Q_{\sqrt{N}\boldsymbol{\beta}} = \frac{1}{\sqrt{N}} \sum_{-N+1}^{N-1} \left(\sum_{t=1}^{N-|\kappa|} \frac{\partial (Z_t Z_{t+|\kappa|})}{\partial \boldsymbol{\beta}} - \sum_{t=1}^{N-|\kappa|} \frac{\partial (E\{Z_t Z_{t+|\kappa|}\})}{\partial \boldsymbol{\beta}} \right) \rho_0(\kappa, \boldsymbol{\theta}_0)$$

$$= \sum_{-N+1}^{N-1} Z'_{\Pi}(\kappa) \rho_0(\kappa, \boldsymbol{\theta}_0),$$

where $Z'_{\Pi}(\kappa)$ stands for the following formula:

$$Z'_{\Pi}(\kappa) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N-|\kappa|} \left(\frac{\partial (Z_t Z_{t+|\kappa|})}{\partial \boldsymbol{\beta}} - \frac{\partial (E\{Z_t Z_{t+|\kappa|}\})}{\partial \boldsymbol{\beta}} \right),$$

where $\kappa = -N+1, \dots, N-1$. In $Q_{\sqrt{N}\boldsymbol{\beta}}$, $Z'_{\Pi}(\kappa)$ is a function of Z_t, while $\rho_0(\kappa, \boldsymbol{\theta}_0)$ does not depend on Z_t. Because for any κ_1 and $\kappa_2 \in [-N+1, N-1]$,

$$E(Z'_{\Pi}(\kappa_1)) = E(Z'_{\Pi}(\kappa_2)) = 0,$$

we have

$$\begin{aligned}
&\operatorname{Cov}(Z'_{\Pi}(\kappa_1), Z'_{\Pi}(\kappa_2)) \\
&\quad = E(Z'_{\Pi}(\kappa_1) Z'_{\Pi}(\kappa_2)) \\
&\quad = \frac{1}{N} \sum_{t_1 \in \mathbb{T}_1} \sum_{t_2 \in \mathbb{T}_2} E\left(\left(\frac{\partial (Z_{t_1} Z_{t_1+|\kappa_1|})}{\partial \boldsymbol{\beta}} - \frac{\partial E(Z_{t_1} Z_{t_1+|\kappa_1|})}{\partial \boldsymbol{\beta}} \right) \right. \\
&\qquad \left. \times \left(\frac{\partial (Z_{t_2} Z_{t_2+|\kappa_2|})}{\partial \boldsymbol{\beta}^\top} - \frac{\partial E(Z_{t_2} Z_{t_2+|\kappa_2|})}{\partial \boldsymbol{\beta}^\top} \right) \right) \\
&\quad = \frac{1}{N} \sum_{t_1 \in \mathbb{T}_1} \sum_{t_2 \in \mathbb{T}_2} \left(E\left\{ \frac{\partial (Z_{t_1} Z_{t_1+|\kappa_1|})}{\partial \boldsymbol{\beta}} \frac{\partial (Z_{t_2} Z_{t_2+|\kappa_2|})}{\partial \boldsymbol{\beta}^\top} \right\} \right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{\partial E(Z_{t_1} Z_{t_1+|\kappa_1|})}{\partial \boldsymbol{\beta}} \frac{\partial E(Z_{t_2} Z_{t_2+|\kappa_2|})}{\partial \boldsymbol{\beta}^\top}\Big) \\
&= \frac{1}{N} \sum_{t_1 \in \mathbb{T}_1} \sum_{t_2 \in \mathbb{T}_2} \Bigg(E\Bigg\{ \bigg(\frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|} + \frac{\partial Z_{t_1+|\kappa_1|}}{\partial \boldsymbol{\beta}} Z_{t_1} \bigg) \\
&\quad \times \bigg(\frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} + \frac{\partial Z_{t_2+|\kappa_2|}}{\partial \boldsymbol{\beta}^\top} Z_{t_2} \bigg) \Bigg\} \\
&\quad - \bigg\{ E\bigg(\frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|} \bigg) + E\bigg(\frac{\partial Z_{t_1+|\kappa_1|}}{\partial \boldsymbol{\beta}} Z_{t_1} \bigg) \bigg\} \\
&\quad \times \bigg\{ E\bigg(\frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} \bigg) + E\bigg(\frac{\partial Z_{t_2+|\kappa_2|}}{\partial \boldsymbol{\beta}^\top} Z_{t_2} \bigg) \bigg\} \Bigg) \\
&= \frac{1}{N} \sum_{t_1 \in \mathbb{T}_1} \sum_{t_2 \in \mathbb{T}_2} \Bigg(E\bigg(\frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|} \frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} \bigg) \\
&\quad - E\bigg(\frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|} \bigg) E\bigg(\frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} \bigg) \\
&\quad + E\bigg(\frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|} \frac{\partial Z_{t_2+|\kappa_2|}}{\partial \boldsymbol{\beta}^\top} Z_{t_2} \bigg) - E\bigg(\frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|} \bigg) E\bigg(\frac{\partial Z_{t_2+|\kappa_2|}}{\partial \boldsymbol{\beta}^\top} Z_{t_2} \bigg) \\
&\quad + E\bigg(\frac{\partial Z_{t_1+|\kappa_1|}}{\partial \boldsymbol{\beta}} Z_{t_1} \frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} \bigg) - E\bigg(\frac{\partial Z_{t_1+|\kappa_1|}}{\partial \boldsymbol{\beta}} Z_{t_1} \bigg) E\bigg(\frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} \bigg) \\
&\quad + E\bigg(\frac{\partial Z_{t_1+|\kappa_1|}}{\partial \boldsymbol{\beta}} Z_{t_1} \frac{\partial Z_{t_2+|\kappa_2|}}{\partial \boldsymbol{\beta}^\top} Z_{t_2} \bigg) - E\bigg(\frac{\partial Z_{t_1+|\kappa_1|}}{\partial \boldsymbol{\beta}} Z_{t_1} \bigg) E\bigg(\frac{\partial Z_{t_2+|\kappa_2|}}{\partial \boldsymbol{\beta}^\top} Z_{t_2} \bigg) \Bigg) \\
&= \frac{1}{N} \sum_{t_1 \in \mathbb{T}_1} \sum_{t_2 \in \mathbb{T}_2} (\mathrm{Part}_{\mathrm{I}} + \mathrm{Part}_{\mathrm{II}} + \mathrm{Part}_{\mathrm{III}} + \mathrm{Part}_{\mathrm{IV}}),
\end{aligned}
\tag{A.36}
$$

where $\mathbb{T}_1 = \{1, \dots, N - |\kappa_1|\}$ and $\mathbb{T}_2 = \{1, \dots, N - |\kappa_2|\}$. Applying Proposition 2.5 of Fan and Yao (2003) which gives the bound of the covariance, we have

$$
\begin{aligned}
&\bigg| \frac{1}{N} \sum_{t_1 \in \mathbb{T}_1} \sum_{t_2 \in \mathbb{T}_2} \mathrm{Part}_{\mathrm{I}} \bigg| \\
&\quad \le \frac{N - |\kappa_1|}{N} \max_{t_1 \in \mathbb{T}_1} \sum_{\mathbb{T}_2} |\mathrm{Part}_{\mathrm{I}}| \\
&\quad = \frac{N - |\kappa_1|}{N} \max_{t_1 \in \mathbb{T}_1} \sum_{\mathbb{T}_2} \bigg| \mathrm{Cov}\bigg(\frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|}, \frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} \bigg) \bigg| \\
&\quad \le \frac{N - |\kappa_1|}{N} \max_{t_1 \in \mathbb{T}_1} \sum_{\mathbb{T}_2} 8 \alpha_{d_{12}}^{1-2/\tau} \bigg\{ E\bigg| \frac{\partial Z_{t_1}}{\partial \boldsymbol{\beta}} Z_{t_1+|\kappa_1|} \bigg|^\tau E\bigg| \frac{\partial Z_{t_2}}{\partial \boldsymbol{\beta}^\top} Z_{t_2+|\kappa_2|} \bigg|^\tau \bigg\}^{1/\tau},
\end{aligned}
\tag{A.37}
$$

where $d_{12} = \min(|t_1 - t_2|, |t_1 - t_2 - |\kappa_2||, |t_1 + |\kappa_1| - t_2|, |t_1 + |\kappa_1| - t_2 - |\kappa_2||)$. Under (A7), there exists $R_{\mathrm{I}} < \infty$ such that $1/N \sum_{t_1 \in \mathbb{T}_1} \sum_{t_2 \in \mathbb{T}_2} \mathrm{Part}_{\mathrm{I}} < R_{\mathrm{I}}$. Similar results hold for parts $\mathrm{Part}_{\mathrm{II}}$, $\mathrm{Part}_{\mathrm{III}}$ and $\mathrm{Part}_{\mathrm{IV}}$. As a consequence, there exists R_0 such that

$$\left|\mathrm{Cov}\big(Z'_{\Pi}(\kappa_1), Z'_{\Pi}(\kappa_2)\big)\right| \le R_0. \tag{A.38}$$

As $\rho_0(\kappa, \boldsymbol{\theta}_0)$ is proportional to ACVF of invertible and stationary time series ARMA$(q_2, q_1, \boldsymbol{\theta}^*)$, $\rho_0(\kappa, \boldsymbol{\theta}_0)$ decreases to zero exponentially when $\kappa \to \infty$. Therefore, there exists $M > 0$, such that $\sum_{\kappa \in \mathbb{S}_{N/M}} |\rho_0(\kappa, \boldsymbol{\theta}_0)| < \epsilon$ for any small $\epsilon > 0$; here $\mathbb{S}_{N/M} = \{\kappa : N - 1 \ge |\kappa| \ge M\}$. We further have

$$\begin{aligned}
E&\left|\sum_{\kappa \in \mathbb{S}_{N/M}} Z'_{\Pi}(\kappa)\rho_0(\kappa, \boldsymbol{\theta}_0)\right| \\
&\le \left[E\left(\sum_{\kappa \in \mathbb{S}_{N/M}} Z'_{\Pi}(\kappa)\rho_0(\kappa, \boldsymbol{\theta}_0)\right)^2\right]^{1/2} \\
&\le \left[E \sum_{\kappa_1 \in \mathbb{S}_{N/M}} \sum_{\kappa_2 \in \mathbb{S}_{N/M}} Z'_{\Pi}(\kappa_1) Z'_{\Pi}(\kappa_2)\rho_0(\kappa_1, \boldsymbol{\theta}_0)\rho_0(\kappa_2, \boldsymbol{\theta}_0)\right]^{1/2} \\
&\le \left[\sum_{\kappa_1 \in \mathbb{S}_{N/M}} \sum_{\kappa_2 \in \mathbb{S}_{N/M}} \left|E\big(Z'_{\Pi}(\kappa_1) Z'_{\Pi}(\kappa_2)\big)\right| \left|\rho_0(\kappa_1, \boldsymbol{\theta}_0)\rho_0(\kappa_2, \boldsymbol{\theta}_0)\right|\right]^{1/2} \qquad \text{(A.39)} \\
&\le \left[R_0 \sum_{\kappa_1 \in \mathbb{S}_{N/M}} \sum_{\kappa_2 \in \mathbb{S}_{N/M}} \left|\rho_0(\kappa_1, \boldsymbol{\theta}_0)\rho_0(\kappa_2, \boldsymbol{\theta}_0)\right|\right]^{1/2} \\
&= \left[R_0 \left(\sum_{\kappa_1 \in \mathbb{S}_{N/M}} \left|\rho_0(\kappa_1, \boldsymbol{\theta}_0)\right|\right)\left(\sum_{\kappa_2 \in \mathbb{S}_{N/M}} \left|\rho_0(\kappa_2, \boldsymbol{\theta}_0)\right|\right)\right]^{1/2} \\
&\le \left[R_0 \epsilon^2\right]^{1/2} \le \sqrt{R_0}\epsilon.
\end{aligned}$$

In the above and following calculations, operators such as $|\,|$, $\sqrt{\ }$, $\le$, $(\,)^2$ and $\sum$ are applied to matrices and vectors elementwise. By (A.39), $Q_{\sqrt{B}\boldsymbol{\beta}}$ could be reduced to $\widetilde{Q}_{\sqrt{B}\boldsymbol{\beta}}$ with M terms, that is,

$$\begin{aligned}
\widetilde{Q}_{\sqrt{N}\boldsymbol{\beta}} &= \sum_{-M}^{M} Z'_{\Pi}(\kappa)\rho_0(\kappa, \boldsymbol{\theta}_0) \\
&= \frac{1}{\sqrt{N}} \sum_{-M}^{M} \sum_{t=1}^{N-|\kappa|} \left(\frac{\partial(Z_t Z_{t+|\kappa|})}{\partial \boldsymbol{\beta}} - \frac{\partial(E\{Z_t Z_{t+|\kappa|}\})}{\partial \boldsymbol{\beta}}\right)\rho_0(\kappa, \boldsymbol{\theta}_0)
\end{aligned}$$

$$= \frac{1}{\sqrt{N}} \sum_{-M}^{M} \sum_{t=1}^{N-|\kappa|} \left(\frac{\partial Z_t}{\partial \boldsymbol{\beta}} Z_{t+|\kappa|} + \frac{\partial Z_{t+|\kappa|}}{\partial \boldsymbol{\beta}} Z_t - E\left(\frac{\partial Z_t}{\partial \boldsymbol{\beta}} Z_{t+|\kappa|} \right) - E\left(\frac{\partial Z_{t+|\kappa|}}{\partial \boldsymbol{\beta}} Z_t \right) \right) \rho_0(\kappa, \boldsymbol{\theta}_0).$$

Because the derivative of $Z_t(\boldsymbol{\alpha}_0)$ takes the following forms:

$$\frac{\partial Z_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\beta}} = X_{t,p}^\top - \left(\frac{1}{N} \mathbb{D}(t,:) G_{Nk}^{-1} \mathbb{D}^\top E_N \mathbf{X}_{Np} \right)^\top = X_{t,p}^\top + \mathbf{o}_p(1),$$

and

$$Z_t(\boldsymbol{\alpha}_0) = Y_t - X_{t,p}\boldsymbol{\beta}_0 - \mathbb{D}(t,:)\hat{\gamma}_0(\boldsymbol{\alpha}_0) = \xi_{0,t} + o_p(1),$$

where $X_{t,p}$ and $\mathbb{D}(t,:)$ are corresponding row vectors of $\mathbf{X}_{Np}$ and $\mathbb{D}$, respectively, by changing the subscripts, it follows that

$$\begin{aligned}
\widetilde{Q}_{\sqrt{N}\boldsymbol{\beta}} &= \frac{1}{\sqrt{N}} \sum_{-M}^{M} \sum_{t=1}^{N-|\kappa|} \big(X_{t,p}^\top \xi_{0,t+|\kappa|} + X_{t+|\kappa|,p}^\top \xi_{0,t} \\
&\qquad - E\big(X_{t,p}^\top \xi_{0,t+|\kappa|}\big) - E\big(X_{t+|\kappa|,p}^\top \xi_{0,t}\big)\big) \rho_0(\kappa, \boldsymbol{\theta}_0) \\
&= \frac{1}{\sqrt{N}} \sum_{\kappa=-M}^{M} \sum_{t=1}^{N-|\kappa|} \mathbb{K}(\kappa, t) \rho_0(\kappa, \boldsymbol{\theta}_0) \\
&= \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \sum_{\kappa=-\min\{M,N-t\}}^{\min\{M,N-t\}} \mathbb{K}(\kappa, t) \rho_0(\kappa, \boldsymbol{\theta}_0) \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \mathbb{V}(t, M).
\end{aligned} \tag{A.40}$$

Let

$$\begin{aligned}
\Omega_{\boldsymbol{\beta}}^{\#}(\kappa_1, \kappa_2) &= \operatorname{Cov}\big(Z'_\Pi(\kappa_1), Z'_\Pi(\kappa_2)\big) \\
&= \frac{1}{N} \operatorname{Cov}\left(\sum_{t=1}^{N-|\kappa_1|} X_{t,p}^\top \xi_{0,t+|\kappa_1|} + X_{t+|\kappa_1|,p}^\top \xi_{0,t}, \right. \\
&\qquad \left. \sum_{t=1}^{N-|\kappa_2|} X_{t,p} \xi_{0,t+|\kappa_2|} + X_{t+|\kappa_2|,p} \xi_{0,t} \right).
\end{aligned} \tag{A.41}$$

Inequality (A.38) ensures that $\Omega_{\boldsymbol{\beta}}^{\#}(\kappa_1, \kappa_2)$ converges and exists. Let

$$\Omega_{\boldsymbol{\beta}} = \sum_{\kappa_1,\kappa_2=-\infty}^{\infty} \rho_0(\kappa_1, \boldsymbol{\theta}_0) \rho_0(\kappa_2, \boldsymbol{\theta}_0) \Omega_{\boldsymbol{\beta}}^{\#}(\kappa_1, \kappa_2). \tag{A.42}$$

Since $\rho_0(\kappa, \boldsymbol{\theta}_0) \to 0$ exponentially when $\kappa \to \infty$, thus $\Omega_{\boldsymbol{\beta}}$ also converges and exists. From equation (A.40) and (A7), it is easy to verify that $\mathbb{V}(t, M)$ satisfies the conditions of Theorem 2.21 of Fan and Yao (2003), and that $E(\mathbb{V}(t, M)) = 0$. Thus,

$$Q_{\sqrt{N}\boldsymbol{\beta}} \overset{d}{\simeq} \widetilde{Q}_{\sqrt{N}\boldsymbol{\beta}} \xrightarrow{d} N(\mathbf{0}, \Omega_{\boldsymbol{\beta}}), \tag{A.43}$$

where $\overset{d}{\simeq}$ means the same asymptotic distribution. Next, consider $Q_{\sqrt{N}\boldsymbol{\theta}}$,

$$\begin{aligned} Q_{\sqrt{N}\boldsymbol{\theta}} &= \sqrt{N}\frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\theta}} \\ &= \frac{1}{\sqrt{N}}\sum_{\omega_n} I(\omega_n, \mathbf{Z}(\boldsymbol{\alpha}_0))\frac{\partial f^{-1}(\omega_n, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{1}{\sqrt{N}}\sum_{\omega_n}\frac{\partial I(\omega_n, \mathbf{Z}(\boldsymbol{\alpha}_0))/\partial \boldsymbol{\theta}}{f(\omega_n, \boldsymbol{\theta}_0)} \\ &= \sqrt{N}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0))\frac{\partial f^{-1}(\omega, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0))/\partial \boldsymbol{\theta}}{f(\omega, \boldsymbol{\theta}_0)}\right) d\omega \\ &\quad + \mathbf{O}(N^{1/2-a}). \end{aligned}$$

From Lemma A.3, we have

$$1/(2\pi)\int_{-\pi}^{\pi}\left(f_z(\omega, \boldsymbol{\alpha}_0)\frac{\partial f^{-1}(\omega, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial f_z(\omega, (\boldsymbol{\alpha}_0))/\partial \boldsymbol{\theta}}{f(\omega, \boldsymbol{\theta}_0)}\right) d\omega = 0.$$

Thus,

$$\begin{aligned} Q_{\sqrt{N}\boldsymbol{\theta}} &= \sqrt{N}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigg(\big[I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0)) - f_z(\omega, \boldsymbol{\alpha}_0)\big]\frac{\partial f^{-1}(\omega, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \\ &\qquad + \left[\frac{\partial I(\omega, \mathbf{Z}(\boldsymbol{\alpha}_0))/\partial \boldsymbol{\theta}}{f(\omega, \boldsymbol{\theta}_0)} - \frac{\partial f_z(\omega, (\boldsymbol{\alpha}_0))/\partial \boldsymbol{\theta}}{f(\omega, \boldsymbol{\theta}_0)}\right]\Bigg) d\omega + \mathbf{O}(N^{1/2-a}) \\ &= Q_{\sqrt{N}\boldsymbol{\theta}\mathrm{I}} + Q_{\sqrt{N}\boldsymbol{\theta}\mathrm{II}} + \mathbf{O}(N^{1/2-a}). \end{aligned}$$

It follows from Theorem 2 of Hannan (1973) that

$$Q_{\sqrt{N}\boldsymbol{\theta}\mathrm{I}} \overset{d}{\simeq} \frac{1}{\sqrt{N}}\sum_{\omega_n} I(\omega_n, \boldsymbol{\xi}_0)\frac{\partial f^{-1}(\omega_n, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \rightarrow N(\mathbf{0}, \Omega_{\boldsymbol{\theta}\mathrm{I}}), \tag{A.44}$$

where

$$\Omega_{\boldsymbol{\theta}\mathrm{I}} = \frac{\sigma_0^4}{\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial \log f(\omega, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\}\left\{\frac{\partial \log f(\omega, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\}^{\top} d\omega.$$

Now, consider $Q_{\sqrt{N}\boldsymbol{\theta}\mathrm{II}}$. It is easy to check

$$\begin{aligned} \frac{\partial \mathbf{Z}(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\theta}^{\top}} &= \frac{1}{N}\mathbb{D}G_{Nk}^{-1}\mathbb{D}^{\top}\frac{\partial E_N}{\partial \boldsymbol{\theta}^{\top}}\{\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta}_0\} \\ &\qquad + \frac{1}{N^2}\mathbb{D}G_{Nk}^{-1}\mathbb{D}\frac{\partial E_N}{\partial \boldsymbol{\theta}^{\top}}\mathbb{D}^{\top}G_{Nk}^{-1}\mathbb{D}^{\top}E_N\{\mathbf{Y} - \mathbf{X}_{Np}\boldsymbol{\beta}_0\} \end{aligned}$$

$$= \frac{1}{N}\mathbb{D}G_{Nk}^{-1}\mathbb{D}^\top \frac{\partial E_N}{\partial \boldsymbol{\theta}^\top}(\mathbf{G}_{xu} + \boldsymbol{\xi}_0) + \mathbf{o}_p(1)$$

$$\stackrel{\text{def}}{=} [\mathbf{G}'_\xi]^\top,$$

where $\mathbf{G}_{xu} = D(\mathbf{X}_{t-1})g_1(\mathbf{X}_{t-d}) + \cdots + D(\mathbf{X}_{t-r})g_r(\mathbf{X}_{t-d})$. Because $\partial f^{-1}(\omega, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$ is continuous, thus $\mathbf{G}'_\xi = \{[G'_\xi(1)]^\top, \dots, [G'_\xi(N)]^\top\}$ is a measurable function of $\{X_{t-1}, \dots, X_{t-r}, X_{t-d}\}$, $t = 1, \dots, N$, and each $G'_\xi(t)$ is a row vector. Similar to (A.41), define

$$\begin{aligned}\Omega^{\#}_{\boldsymbol{\theta}\mathrm{II}}(\kappa_1, \kappa_2) = \frac{1}{N}\operatorname{Cov}\Bigg(&\sum_{t=1}^{N-|\kappa_1|} G'_\xi(t)^\top \xi_{0,t+|\kappa_1|} + G'_\xi\big(t+|\kappa_1|\big)^\top \xi_{0,t}, \\ &\sum_{t=1}^{N-|\kappa_2|} G'_\xi(t)\xi_{0,t+|\kappa_2|} + G'_\xi\big(t+|\kappa_2|\big)\xi_{0,t}\Bigg). \end{aligned} \tag{A.45}$$

Following the same procedure of proving $Q_{\sqrt{B}\boldsymbol{\beta}}$, we have

$$Q_{\sqrt{N}\boldsymbol{\theta}\mathrm{II}} \xrightarrow{d} N(\mathbf{0}, \Omega_{\boldsymbol{\theta}\mathrm{II}}), \tag{A.46}$$

where

$$\Omega_{\boldsymbol{\theta}\mathrm{II}} = \sum_{\kappa_1,\kappa_2=-\infty}^{\infty} \rho_0(\kappa_1, \boldsymbol{\theta}_0)\rho_0(\kappa_2, \boldsymbol{\theta}_0)\Omega^{\#}_{\boldsymbol{\theta}\mathrm{II}}(\kappa_1, \kappa_2)$$

and

$$\begin{aligned}\Omega^{\#}_{\boldsymbol{\beta\theta}\mathrm{II}}(\kappa_1, \kappa_2) = \frac{1}{N}\operatorname{Cov}\Bigg(&\sum_{t=1}^{N-|\kappa_1|} X_{t,p}^\top \xi_{0,t+|\kappa_1|} + X_{t+|\kappa_1|,p}^\top \xi_{0,t}, \\ &\sum_{t=1}^{N-|\kappa_2|} G'_\xi(t)\xi_{0,t+|\kappa_2|} + G'_\xi\big(t+|\kappa_2|\big)\xi_{0,t}\Bigg)\end{aligned} \tag{A.47}$$

and the covariance matrix is

$$\Omega_{\boldsymbol{\beta\theta}\mathrm{II}} = \sum_{\kappa_1,\kappa_2=-\infty}^{\infty} \rho_0(\kappa_1, \boldsymbol{\theta}_0)\rho_0(\kappa_2, \boldsymbol{\theta}_0)\Omega^{\#}_{\boldsymbol{\beta\theta}\mathrm{II}}(\kappa_1, \kappa_2). \tag{A.48}$$

Similarly, we have

$$\begin{aligned}&\Omega^{\#}_{\boldsymbol{\beta\theta}\mathrm{I}}(\kappa_1, \kappa_2) \\ &\quad = \frac{1}{N}\operatorname{Cov}\Bigg(\sum_{t=1}^{N-|\kappa_1|} X_{t,p}^\top \xi_{0,t+|\kappa_1|} + X_{t+|\kappa_1|,p}^\top \xi_{0,t}, \sum_{t=1}^{N-|\kappa_2|} \xi_{0,t}\xi_{0,t+|\kappa_2|}\Bigg),\end{aligned} \tag{A.49}$$

and

$$\Omega_{\boldsymbol{\beta\theta}\mathrm{I}} = \sum_{\kappa_1,\kappa_2=-\infty}^{\infty} \rho_0(\kappa_1, \boldsymbol{\theta}_0)\Omega^{\#}_{\boldsymbol{\beta\theta}\mathrm{I}}(\kappa_1, \kappa_2)\rho'_0(\kappa_2, \boldsymbol{\theta}_0)^\top, \tag{A.50}$$

where

$$\rho_0'(\kappa_2, \boldsymbol{\theta}_0) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{\partial f^{-1}(\omega, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} e^{-i\kappa_2 \omega}\, d\omega$$

and $\rho_0'(\kappa_2, \boldsymbol{\theta}_0)$ is similarly calculated as $\rho_0(\kappa_2, \boldsymbol{\theta}_0)$ in (A.35) except that it is a column vector. For covariance matrix between $Q_{\sqrt{N}\boldsymbol{\theta}\mathrm{I}}$ and $Q_{\sqrt{N}\boldsymbol{\theta}\mathrm{II}}$, let

$$\begin{aligned}&\Omega^{\#}_{\boldsymbol{\theta}\mathrm{I},\mathrm{II}}(\kappa_1, \kappa_2)\\ &\quad= \frac{1}{N} \operatorname{Cov}\left(\sum_{t=1}^{N-|\kappa_1|} \xi_{0,t}\xi_{0,t+|\kappa_1|} \sum_{t=1}^{N-|\kappa_2|} G_\xi'(t)\xi_{0,t+|\kappa_2|} + G_\xi'\big(t+|\kappa_2|\big)\xi_{0,t} \right)\end{aligned} \tag{A.51}$$

and

$$\Omega_{\boldsymbol{\theta}\mathrm{I},\mathrm{II}} = \sum_{\kappa_1,\kappa_2=-\infty}^{\infty} \rho_0'(\kappa_1, \boldsymbol{\theta}_0)\Omega^{\#}_{\boldsymbol{\theta}\mathrm{I},\mathrm{II}}(\kappa_1, \kappa_2)\rho_0(\kappa_2, \boldsymbol{\theta}_0)^\top. \tag{A.52}$$

Construct matrix

$$A_{\mathbb{I}} = \begin{pmatrix} \mathbb{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbb{I}_q & \mathbb{I}_q \end{pmatrix}, \qquad \Omega_{\boldsymbol{\beta\theta}} = \begin{pmatrix} \Omega_{\boldsymbol{\beta}} & \Omega_{\boldsymbol{\beta\theta}\mathrm{I}} & \Omega_{\boldsymbol{\beta\theta}\mathrm{II}} \\ \Omega_{\boldsymbol{\beta\theta}\mathrm{I}}^\top & \Omega_{\boldsymbol{\theta}\mathrm{I}} & \Omega_{\boldsymbol{\theta}\mathrm{I},\mathrm{II}} \\ \Omega_{\boldsymbol{\beta\theta}\mathrm{II}}^\top & \Omega_{\boldsymbol{\theta}\mathrm{I},\mathrm{II}}^\top & \Omega_{\boldsymbol{\theta}\mathrm{II}} \end{pmatrix},$$

where $\mathbb{I}_p$ and $\mathbb{I}_q$ are identity matrices, and p is the number of linear lagged variables, $q = q_1 + q_2$ is the length of $\boldsymbol{\theta} = \{\theta_{a1}, \dots, \theta_{aq_1}, \theta_{m1}, \dots, \theta_{mq_2}\}^\top$. By these notation, we can write

$$\sqrt{N}\frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} = A_{\mathbb{I}} \times \big(Q_{\sqrt{N}\boldsymbol{\beta}}^\top, Q_{\sqrt{N}\boldsymbol{\theta}\mathbf{I}}^\top, Q_{\sqrt{N}\boldsymbol{\theta}\mathbf{II}}^\top\big)^\top.$$

All the entries of the $\Omega_{\boldsymbol{\beta\theta}}$ are defined by (A.42), (A.44), (A.46), (A.48), (A.50) and (A.52). Recall that we have proved

$$\big(Q_{\sqrt{N}\boldsymbol{\beta}}^\top, Q_{\sqrt{N}\boldsymbol{\theta}\mathbf{I}}^\top, Q_{\sqrt{N}\boldsymbol{\theta}\mathbf{II}}^\top\big)^\top \xrightarrow{d} N(\mathbf{0}, \Omega_{\boldsymbol{\beta\theta}}).$$

Combining all the results above, we have

$$\sqrt{N}\frac{\partial \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}} \xrightarrow{d} N\big(\mathbf{0}, A_{\mathbb{I}}\Omega_{\boldsymbol{\beta\theta}}A_{\mathbb{I}}^\top\big). \tag{A.53}$$

Next, we need to show $\partial^2\mathbb{Q}_N(\boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top \xrightarrow{p} \Phi$ for some finite matrix Φ. According to (A.34),

$$\begin{aligned}&\frac{\partial^2\mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} - \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial^2 f_z(\omega, \boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} f^{-1}(\omega, \boldsymbol{\theta}_0)\, d\omega\\ &\quad= \frac{1}{N}\sum_{\omega_n} \frac{\partial^2 I(\omega_n, \mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} f^{-1}(\omega, \boldsymbol{\theta}_0)\end{aligned} \tag{A.54}$$

$$
\begin{aligned}
&- \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial^2 f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} f^{-1}(\omega,\boldsymbol{\theta}_0)\,d\omega \\
&= \frac{\sqrt{N}}{\sqrt{N}}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial^2 I(\omega,\mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} - \frac{\partial^2 f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right) f^{-1}(\omega,\boldsymbol{\theta}_0)\,d\omega \\
&\quad + \mathbf{o}_p(1).
\end{aligned}
$$

Following the same steps as for $Q_{\sqrt{N}\boldsymbol{\beta}}$ in (A.43), for each entry above it can be proved that

$$
\sqrt{N}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial^2 I(\omega,\mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\beta_{j1}\partial\beta_{j2}} - \frac{\partial^2 f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\beta_{j1}\partial\beta_{j2}}\right) f^{-1}(\omega,\boldsymbol{\theta}_0)\,d\omega \xrightarrow{d} N(0,\Sigma_{ij}).
$$

Thus,

$$
\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial^2 I(\omega,\mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\beta_{j1}\partial\beta_{j2}} - \frac{\partial^2 f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\beta_{j1}\partial\beta_{j2}}\right) f^{-1}(\omega,\boldsymbol{\theta}_0)\,d\omega \xrightarrow{p} 0.
$$

Therefore,

$$
\Phi_{\boldsymbol{\beta}} \stackrel{\text{def}}{=} \frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} \xrightarrow{p} \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial^2 f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} f^{-1}(\omega,\boldsymbol{\theta}_0)\,d\omega \tag{A.55}
$$

and

$$
\begin{aligned}
\frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^\top} &= \frac{1}{N}\sum_{\omega_n}\frac{\partial I(\omega,\mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\boldsymbol{\beta}}\frac{\partial f^{-1}(\omega,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top} \\
&\quad + \frac{1}{N}\sum_{\omega_n}\frac{\partial^2 I(\omega,\mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^\top} f^{-1}(\omega,\boldsymbol{\theta}_0).
\end{aligned}
$$

In a similar manner, we can show that

$$
\begin{aligned}
&\frac{1}{N}\sum_{\omega_n}\frac{\partial I(\omega,\mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\boldsymbol{\beta}}\frac{\partial f^{-1}(\omega,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top} + \frac{1}{N}\sum_{\omega_n}\frac{\partial^2 I(\omega,\mathbf{Z}(\boldsymbol{\alpha}_0))}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^\top} f^{-1}(\omega,\boldsymbol{\theta}_0) \\
&\quad - \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial^2 f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^\top} f^{-1}(\omega,\boldsymbol{\theta}_0)\,d\omega \\
&\quad - \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}}\frac{\partial f^{-1}(\omega,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\,d\omega \xrightarrow{p} 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial^2 f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^\top} f^{-1}(\omega,\boldsymbol{\theta}_0)\,d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial f_z(\omega,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\beta}}\frac{\partial f^{-1}(\omega,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\,d\omega \\
&\quad = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial^2 (f_z(\omega,\boldsymbol{\alpha}_0) f^{-1}(\omega,\boldsymbol{\theta}_0))}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^\top}\,d\omega.
\end{aligned}
$$

Thus,

$$\frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}^\top} \xrightarrow{p} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial^2 (f_z(\omega, \boldsymbol{\alpha}_0) f^{-1}(\omega, \boldsymbol{\theta}_0))}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}^\top} \, d\omega. \tag{A.56}$$

Similarly,

$$\frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \xrightarrow{p} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial^2 (f_z(\omega, \boldsymbol{\alpha}_0) f^{-1}(\omega, \boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \, d\omega. \tag{A.57}$$

Based (A.55), (A.56) and (A.57), it follows that

$$\frac{\partial^2 \mathbb{Q}_N(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} \xrightarrow{p} \Phi = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial^2 (f_z(\omega, \boldsymbol{\alpha}_0) f^{-1}(\omega, \boldsymbol{\theta}_0))}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} \, d\omega. \tag{A.58}$$

Based on (A.53) and (A.58), we immediately have

$$\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} N\big(\mathbf{0}, \Phi^{-1} A_{\mathbb{I}} \Omega_{\boldsymbol{\beta}\boldsymbol{\theta}} A_{\mathbb{I}}^\top \Phi^{-1}\big). \tag{A.59}$$

This completes the proof of (3.4). Note that

$$\sqrt{Nh}\big(\hat{\mathbf{g}}(u, \hat{\boldsymbol{\alpha}}) - \hat{\mathbf{g}}(u, \boldsymbol{\alpha}_0)\big) = \sqrt{h} \frac{\partial \hat{\mathbf{g}}(u, \boldsymbol{\alpha}_*)}{\partial \boldsymbol{\alpha}} \sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0),$$

where $\hat{\boldsymbol{\alpha}} \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0$ leads to $\boldsymbol{\alpha}_* \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0$, where $\|\boldsymbol{\alpha}_* - \boldsymbol{\alpha}_0\| \le \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|$. (3.4) implies that $\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = \mathbf{O}_p(1)$. In the proof of (3.4), it has been shown that $\partial \hat{\mathbf{g}}(u, \boldsymbol{\alpha}_0)/\partial \boldsymbol{\alpha} = \mathbf{O}_p(1)$, thus the above difference is of order $\mathbf{o}_p(1)$ as $h \to 0$ when $N \to \infty$. (3.5) follows from Lemma A.2. $\square$

Acknowledgments. The authors would like to thank the Associate Editor and two referees for thoughtful comments which lead to substantial improvement of the paper.

REFERENCES

ADAMS, R., CHEN, C., MCCARL, B. and WEIHER, R. (1999). The economic consequences of ENSO events for agriculture. *Clim. Res.* **3** 165–172.

BARROW, D. L. and SMITH, P. W. (1978/79). Asymptotic properties of best $L_2[0, 1]$ approximation by splines with variable knots. *Quart. Appl. Math.* **36** 293–304. MR0508773

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York. MR1093459

CAI, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136** 163–188. MR2328589

CAI, Z., FAN, J. and YAO, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95** 941–956. MR1804449

CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842

CHEN, Z., LI, R. and LI, Y. (2015). Varying coefficient models for data with auto-correlated error process. *Statist. Sinica* **25** 709–723. MR3379095

CHEN, R. and TSAY, R. S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88** 298–308. MR1212492

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. MR1964455

GAO, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods. Monographs on Statistics and Applied Probability* **108**. Chapman & Hall/CRC, Boca Raton, FL. MR2297190

GIRAITIS, L. and ROBINSON, P. M. (2001). Whittle estimation of ARCH models. *Econometric Theory* **17** 608–631. MR1841822

GLANTZ, M. (2001). *Currents of Change: Impacts of El Niño and La Niña on Climate and Society*. Cambridge Univ Press, Cambridge.

HALL, P. and VAN KEILEGOM, I. (2003). Using difference-based methods for inference in nonparametric regression with time series errors. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 443–456. MR1983757

HANNAN, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probab.* **10** 130–145, corrections, ibid. 10 (1973), 913. MR0365960

HART, J. D. (1991). Kernel regression estimation with time series errors. *J. Roy. Statist. Soc. Ser. B* **53** 173–187. MR1094279

HASTIE, T. J. and TIBSHIRANI, R. J. (1993). Varying-coefficient models (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 757–796. MR1229881

HUANG, J. Z. and YANG, L. (2004). Identification of non-linear additive autoregressive models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 463–477. MR2062388

LI, Q., HUANG, C. J., LI, D. and FU, T. (2002). Semiparametric smooth coefficient models. *J. Bus. Econom. Statist.* **20** 412–422. MR1939909

LIU, J. M., CHEN, R. and YAO, Q. (2010). Nonparametric transfer function models. *J. Econometrics* **157** 151–164. MR2652287

LJUNG, G. M. and BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65** 297–303.

MA, S. and YANG, L. (2011). Spline-backfitted kernel smoothing of partially linear additive model. *J. Statist. Plann. Inference* **141** 204–219. MR2719488

MAHDI, E. and MCLEOD, A. I. (2012). Improved multivariate portmanteau test. *J. Time Series Anal.* **33** 211–222. MR2902459

MCLEOD, A. I. and LI, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Series Anal.* **4** 269–273. MR0738587

OPSOMER, J., WANG, Y. and YANG, Y. (2001). Nonparametric regression with correlated errors. *Statist. Sci.* **16** 134–153. MR1861070

PIERCE, D. A. (1971). Least squares estimation in the regression model with autoregressive-moving average errors. *Biometrika* **58** 299–312. MR0329169

RAY, B. K. and TSAY, R. S. (1997). Bandwidth selection for kernel regression with long-range dependent errors. *Biometrika* **84** 791–802. MR1625031

SEVERINI, T. A. and WONG, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20** 1768–1802. MR1193312

SU, L. and ULLAH, A. (2006). More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. *Econometric Theory* **22** 98–126. MR2212694

TJØSTHEIM, D. and AUESTAD, B. H. (1994). Nonparametric identification of nonlinear time series: Projections. *J. Amer. Statist. Assoc.* **89** 1398–1409. MR1310230

TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach. Oxford Statistical Science Series* **6**. The Clarendon Press, New York. MR1079320

TRENBERTH, K. and STEPANIAK, D. (2001). Indices of El Niño evolution. *J. Climate* **14** 1697–1701.

UBILAVA, D. and HELMERS, C. G. (2013). Forecasting ENSO with a smooth transition autoregressive model. *Environ. Model. Softw.* **40** 181–190.

WANG, T. and XIA, Y. (2014). Whittle likelihood estimation of nonlinear autoregressive models with moving average residuals. *J. Amer. Statist. Assoc.* **110** 1083–1099. MR3420686

WHITTLE, P. (1953). The analysis of multiple stationary time series. *J. Roy. Statist. Soc. Ser. B.* **15** 125–139. MR0056902

XIAO, H. and WU, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. MR3014314

XIAO, Z., LINTON, O. B., CARROLL, R. J. and MAMMEN, E. (2003). More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *J. Amer. Statist. Assoc.* **98** 980–992. MR2041486

XUE, L. and YANG, L. (2006). Additive coefficient modeling via polynomial spline. *Statist. Sinica* **16** 1423–1446. MR2327498

YAO, Q. and BROCKWELL, P. J. (2006). Gaussian maximum likelihood estimation for ARMA models. I. Time series. *J. Time Series Anal.* **27** 857–875. MR2328545

YU, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22** 94–116. MR1258867

ZHANG, W., LEE, S. and SONG, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** 166–188. MR1918619

ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. MR1673277

ZYGMUND, A. (1959). *Trigonometric Series*. Cambridge Univ. Press, Cambridge, UK.

H. LEI
SOUTHWEST JIAOTONG UNIVERSITY
BLOCK 2, ROOM 414A
HIGH-TECH DISTRICT WEST ZONE
CHENGDU 610000
P.R. CHINA
E-MAIL: stahl@swjtu.edu.cn

Y. XIA
NATIONAL UNIVERSITY OF SINGAPORE
BLOCK S16, LEVEL 6
SCIENCE DRIVE 2
SINGAPORE 117546
AND
UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY
NO. 2006, XIYUAN AVE
WEST HI-TECH ZONE 611731
CHENGDU, SICHUAN
P.R. CHINA
E-MAIL: staxyc@nus.edu.sg

X. QIN
UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY
NO. 2006, XIYUAN AVE
WEST HI-TECH ZONE 611731
CHENGDU, SICHUAN
P.R. CHINA
E-MAIL: qinxu@uestc.edu.cn