Statistica Sinica **17**(2007), 177-197

ADAPTIVE VARYING-COEFFICIENT LINEAR MODELS FOR STOCHASTIC PROCESSES: ASYMPTOTIC THEORY

Zudi Lu[1,2], Dag Tjøstheim[3] and Qiwei Yao[1,4]

[1]*London School of Economics,* [2]*Chinese Academy of Sciences*
[3]*University of Bergen and* [4]*Peking University*

Abstract: We establish the asymptotic theory for the estimation of adaptive varying-coefficient linear models. More specifically, we show that the estimator of the index parameter is root-n-consistent. It differs from the locally optimal estimator that has been proposed in the literature with a prerequisite that the estimator is within a $n^{-\delta}$-distance of the true value. To this end, we establish two fundamental lemmas for the asymptotic properties of the estimators of parametric components in a general semiparametric setting. Furthermore, the estimation for the coefficient functions is asymptotically adaptive to the unknown index parameter. Asymptotic properties are derived using the empirical process theory for strictly stationary β-mixing processes.

Key words and phrases: Adaptive varying-coefficient model, asymptotic normality, β-mixing, empirical process, index parameter, root-n consistency, uniform convergence.

1. Introduction

We consider a class of adaptive varying-coefficient models of the form

$$Y_t = a_0\{\boldsymbol{\alpha}_0^\top \mathbf{X}_t\} + \mathbf{b}_0\{\boldsymbol{\alpha}_0^\top \mathbf{X}_t\}^\top \mathbf{X}_t + \varepsilon_t, \tag{1.1}$$

where t is time, $\mathbf{X}_t = (X_{t,1}, \ldots, X_{t,d})^\top$ is a $d \times 1$ predictor vector which may consist of some lagged values of Y_t and/or other exogenous variables, and $E(\varepsilon_t|\mathbf{X}_t) = 0$. In (1.1), the index parameter $\boldsymbol{\alpha}_0$ is unknown, and both $a_0(\cdot)$ and $\mathbf{b}_0(\cdot)$, which are $\mathbb{R}^1$- and $\mathbb{R}^d$-valued respectively, are also unknown. This model is coined as *adaptive* by Fan, Yao and Cai (2003) to indicate that the coefficients are functions of an unknown index variable $\boldsymbol{\alpha}_0^\top \mathbf{X}_t$, in contrast to, for example, the functional-coefficient models of Chen and Tsay (1993) and Cai, Fan and Yao (2000). This is a general nonlinear (dynamical) model. For example, if $\mathbf{X}_t = \{Y_{t-1}, Y_{t-2}, \ldots, Y_{t-d}\}^\top$, (1.1) reduces to *the adaptive varying-coefficient linear autoregressive* model (Tong (1990), Xia and Li (1999) and Fan, Yao and Cai (2003)). On the other hand, to include some financial econometrics models one could specify $\mathbf{X}_t = \{Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}, U_t, U_{t-1}, \ldots, U_{t-q}\}^{\mathsf{T}}$ for some exogenous process U_t; see Hannan (1970), Gourieroux and Jasiak (2001) and Hong and Lee (2003). Formally (1.1) also includes the popular single-index model and the generalized partially linear single-index models as special cases; see Chapter 8 of Fan and Yao (2003) and the references within. The major advantage of (1.1) is that it does not suffer from the curse of dimensionality often encountered in multivariate nonparametric settings, since both $a_0(\cdot), \mathbf{b}_0(\cdot)$ are functions of a univariate variable.

The estimation for model (1.1) with independent observations has been investigated in several papers. Ichimura (1993) proposed the form (1.1). Following the lead of Härdle, Hall and Ichimura (1993), Xia and Li (1999) estimated the index parameter $\boldsymbol{\alpha}_0$ by a computationally expensive cross-validation method. By assuming this cross-validation estimator to be within $n^{-\delta}$-distance of $\boldsymbol{\alpha}_0$ for some $\delta \in (3/10, 1/2)$, Xia and Li (1999) showed that the estimator is root-n consistent. More recently, Fan, Yao and Cai (2003) established a new computationally efficient procedure based on the profile least-squares local linear weighted regression. They also addressed the issue of deleting locally insignificant variables to avoid overfitting. While the proposed methodology was proved effective in numerical illustrations with both simulated and real data, the sampling properties of the estimators remain to be established. One of the primary goals of this paper is to fill this gap.

More precisely, we establish the asymptotic theory for the estimation of adaptive vary-coefficient linear modelling with the observations from a mixing processes, which is applicable to both independent data and time series. We show that the estimator for the global parameter $\boldsymbol{\alpha}_0$ is root-n consistent without assuming it to be within a $n^{-\delta}$-distance of the true value, a condition often imposed for problem of this nature; see, for example, Härdle et al. (1993), Carroll, Fan, Gijbels and Wand (1997) and Xia and Li (1999). Based on this result, we also show that the coefficient functions $a_0(\cdot)$ and $\mathbf{b}_0(\cdot)$ can be estimated asymptotically as well as if $\boldsymbol{\alpha}_0$ were given. Our asymptotic theory shows that two different bandwidths should be used in estimating the global parameter $\boldsymbol{\alpha}_0$ and the local parameters $a_0, \mathbf{b}_0$. This is in line with the common knowledge that a global parameter should be estimated in an undersmoothed manner.

At the technical level, our approach differs from that of Härdle et al. (1993) and Xia and Li (1999). Lemmas 4.1 and 4.2 in Section 4 below play a basic role in deriving the asymptotic properties of the estimators. They themselves are of independent interest, as they provide a general framework for establishing the root-n consistency and the asymptotic normality for profile M-estimators (such as profile maximum likelihood estimation or profile least squares estimation) for global parameters in semiparametric settings. They may be viewed as an analogue of the results of Chan, Linton and Keilegom (2003) which dealt with generalised method-of-moments estimation only. We validate the conditions of the lemmas using the empirical process theory of Doukhan, Massart and Rio (1995). In Lu, Steinskog, Tjøstheim and Yao (2006), we apply the theory developed in this paper to varying coefficient models for spatio-temporal processes, where further numerical illustrations are reported.

A short overview of the paper is as follows. The model and the estimation method are stated in Section 2. Asymptotic properties are presented in Section 3. Two general lemmas on the consistency and the asymptotic normality of profile M-estimation are established in Section 4. We prove the main results in Section 5. A uniform convergence rate of the profile kernel regression estimator is established in the Appendix.

2. Estimation Procedure

Fan, Yao and Cai (2003) showed that appropriate constraints on the form (1.1) should be imposed in order to make the model identifiable. In fact, Fan, Yao and Cai (2003, p.59) assumed the last component of $\boldsymbol{\alpha}_0 = (\alpha_{01}, \ldots, \alpha_{0d})^\top$ to be non-zero, resulting in the model

$$Y_t = a_0\{\boldsymbol{\alpha}_0^\top \mathbf{X}_t\} + \breve{\mathbf{X}}_t^\top \mathbf{b}_0\{\boldsymbol{\alpha}_0^\top \mathbf{X}_t\} + \varepsilon_t = \mathbb{X}_t^\top \mathbf{g}_0\{\boldsymbol{\alpha}_0^\top \mathbf{X}_t\} + \varepsilon_t. \tag{2.1}$$

Here $\breve{\mathbf{X}}_t = (X_{t,1}, \ldots, X_{t,d-1})^\top$ is the remaining vector of $\mathbf{X}_t$ with its dth component deleted, the notation $\mathbf{b}_0$ is retained but with only $d-1$ components from now on, $\mathbb{X}_t = (1, \breve{\mathbf{X}}_t^\top)^\top$, and $\mathbf{g}_0 = (a_0, \mathbf{b}_0^\top)^\top$. Note (1.1) may always be expressed in the form of (2.1) provided $\alpha_{0d} \neq 0$. Furthermore, we assume that $||\boldsymbol{\alpha}_0|| = 1$, the first non-zero component of $\boldsymbol{\alpha}_0$ is positive, that

$$E(Y_t|\mathbf{X}_t = \mathbf{x}) \neq \boldsymbol{\alpha}_0^\top \mathbf{x} \boldsymbol{\beta}^\top \mathbf{x} + \boldsymbol{\gamma}^\top \mathbf{x} + c$$

for some $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^d$ and $c \in \mathbb{R}^1$, and that $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_0$ are not parallel to each other. Then $\boldsymbol{\alpha}_0, a_0(\cdot)$ and $\mathbf{b}_0(\cdot)$ in (2.1) are all identifiable; see Theorem 1 of Fan, Yao and Cai (2003). Furthermore, we assume $|\alpha_{0d}| > \epsilon_0$ for a given constant $\epsilon_0 \in (0, 1)$. Put

$$\begin{aligned}\mathbb{B} = \{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top \in \mathbb{R}^d :\ \|\boldsymbol{\alpha}\| = 1, &\text{the first non-zero element is positive,} \\ &\text{and } |\alpha_d| \geq \epsilon_0\}.\end{aligned} \tag{2.2}$$

Then $\boldsymbol{\alpha}_0$ is an inner point of the compact set $\mathbb{B}$. Therefore we need only search for $\boldsymbol{\alpha}_0$ over $\mathbb{B}$.

Let $K(\cdot)$ be a kernel function, $K_h(\cdot) = h^{-1}K(\cdot/h)$, $h > 0$ be a bandwidth, and $w(\cdot) = I_{[-L,L]}(\cdot)$ $(L > 0)$ be a weight function controlling the edge effect in the estimation. Fan, Yao and Cai (2003) proposed the following iterative profile least squares estimation procedure.

1. For given $\boldsymbol{\alpha}$ and $Z_t = \boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t$, minimise

$$\sum_{t=1}^{n}\Big[Y_t - a - c(Z_t - z) - \{\mathbf{b} - \mathbf{d}(Z_t - z)\}^\mathsf{T}\breve{\mathbf{X}}_t\Big]^2 K_h\{Z_t - z\}w\{Z_t\} \tag{2.3}$$

over $\boldsymbol{\theta} = \boldsymbol{\theta}(z, \boldsymbol{\alpha}) = (a, \mathbf{b}, c, \mathbf{d})$, leading to the estimators

$$\begin{aligned}\widehat{\boldsymbol{\theta}}(z, \boldsymbol{\alpha}, h) \equiv \widehat{\boldsymbol{\theta}}(z, \boldsymbol{\alpha}) &= \{\widehat{a}(z, \boldsymbol{\alpha}, h),\ \widehat{\mathbf{b}}(z, \boldsymbol{\alpha}, h)^\mathsf{T},\ \widehat{\dot{a}}(z, \boldsymbol{\alpha}, h),\ \widehat{\dot{\mathbf{b}}}(z, \boldsymbol{\alpha}, h)^\mathsf{T}\}^\mathsf{T}\\ &= \{\widehat{a}(z, \boldsymbol{\alpha}),\ \widehat{\mathbf{b}}(z, \boldsymbol{\alpha})^\mathsf{T},\ \widehat{\dot{a}}(z, \boldsymbol{\alpha}),\ \widehat{\dot{\mathbf{b}}}(z, \boldsymbol{\alpha})^\mathsf{T}\}^\mathsf{T}\\ &= (\widehat{a},\ \widehat{\mathbf{b}}^\mathsf{T},\ \widehat{c},\ \widehat{\mathbf{d}}^\mathsf{T})^\mathsf{T},\end{aligned} \tag{2.4}$$

where $\dot{f}$ denote the derivative of a function f.

2. Let $\widetilde{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ minimises

$$R(\boldsymbol{\beta}) = \frac{1}{n}\sum_{t=1}^{n}\Big[Y_t - \widehat{a}\{\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_t, \boldsymbol{\alpha}\} - \widehat{\mathbf{b}}\{\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_t, \boldsymbol{\alpha}\}^\mathsf{T}\breve{\mathbf{X}}_t\Big]^2 w(\boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t). \tag{2.5}$$

3. Repeat the above two steps with $\boldsymbol{\alpha} = \widetilde{\boldsymbol{\alpha}}$ until the successive values of $R(\widetilde{\boldsymbol{\alpha}})$ differ insignificantly. Denote by $\widehat{\boldsymbol{\alpha}}$ the final estimator of the index parameter.

It is well known that if $\boldsymbol{\alpha}$ is known, the optimal bandwidth h used in (2.3) is of order $O(n^{-1/5})$. However, if $\boldsymbol{\alpha}$ is unknown, in order to ensure that the estimator $\widehat{\boldsymbol{\alpha}}$ is root-n consistent, the bandwidth h used in the above iteration should be smaller than $O(n^{-1/5})$ if we only assume $a(\cdot)$ and $\mathbf{b}(\cdot)$ are second order differentiable (see Theorem 3.2 below). Note that once the estimator $\widehat{\boldsymbol{\alpha}}$ is available, a different bandwidth $\hbar$ of order $O(n^{-1/5})$ should be used in the *final* estimators for $a(\cdot)$ and $\mathbf{b}(\cdot)$ (see Corollary 3.3 below).

For a fixed $\boldsymbol{\alpha}$, the sampling properties of the estimator $\widehat{\boldsymbol{\theta}}$ defined in (2.4) follows the standard asymptotic theory of local linear regression estimation (Fan and Gijbels (1996) and Fan and Yao (2003)). However it is more challenging to develop the asymptotic properties of the estimator $\widehat{\boldsymbol{\alpha}}$ as defined in (2.5). One fundamental difficulty is the lack of an explicit expression for such an $\widehat{\boldsymbol{\alpha}}$, since it is defined iteratively. To get around this difficulty, we slightly alter the definition of the estimator for $\boldsymbol{\alpha}$ and take

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{B}} R_n\{\widehat{a}(\cdot, \boldsymbol{\alpha}, h), \widehat{\mathbf{b}}(\cdot, \boldsymbol{\alpha}, h), \boldsymbol{\alpha}\}, \tag{2.6}$$

where $\mathbb{B}$ is defined in (2.2) and

$$R_n\{\widehat{a}(\cdot,\boldsymbol{\alpha},h),\widehat{\mathbf{b}}(\cdot,\boldsymbol{\alpha},h),\boldsymbol{\alpha}\} = \frac{1}{n}\sum_{t=1}^{n}\Big[Y_t - \widehat{a}\{\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha},h\} - \widehat{\mathbf{b}}\{\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha},h\}^{\mathsf{T}}\breve{\mathbf{X}}_t\Big]^2 w(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t). \quad (2.7)$$

It is easy to see the backfitting iteration of (2.3) – (2.5) is an approximate and computationally efficient way to evaluate $\widehat{\boldsymbol{\alpha}}$ defined in (2.6), while (2.6) itself is theoretically more tractable. We sketch below how we proceed with the theoretical investigation.

With $\boldsymbol{\alpha}$ and Z_t given, (2.3) divided by n is a consistent estimate of

$$\begin{aligned} R_z(a(\cdot),\mathbf{b}(\cdot),\boldsymbol{\alpha}) &= E\left\{\left(Y_t - a(Z_t) - \mathbf{b}(Z_t)^{\mathsf{T}}\breve{\mathbf{X}}_t\right)^2 w(Z_t) \mid Z_t = z\right\} \\ &= E\left\{\left(Y_t - a(z) - \mathbf{b}(z)^{\mathsf{T}}\breve{\mathbf{X}}_t\right)^2 w(z) \mid Z_t = z\right\}. \end{aligned} \quad (2.8)$$

Corresponding to (2.5), we define

$$R(a(\cdot),\mathbf{b}(\cdot),\boldsymbol{\alpha}) = E\left\{\left(Y_t - a(Z_t) - \mathbf{b}(Z_t)^{\mathsf{T}}\breve{\mathbf{X}}_t\right)^2 w(Z_t)\right\}, \quad (2.9)$$

which is related to $R_z(a(\cdot),\mathbf{b}(\cdot),\boldsymbol{\alpha})$ via

$$\begin{aligned} R(a(\cdot),\mathbf{b}(\cdot),\boldsymbol{\alpha}) &= \int R_z(a(\cdot),\mathbf{b}(\cdot),\boldsymbol{\alpha}) f_Z(z)dz \\ &= E\left\{\left(Y_t - a(Z_t) - \mathbf{b}(Z_t)^{\mathsf{T}}\breve{\mathbf{X}}_t\right)^2 w(Z_t)\right\}, \end{aligned} \quad (2.10)$$

where $f_Z(z) = f_Z(z,\boldsymbol{\alpha})$ is the density function of $Z_t = Z_t(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t$. Note that, with $\boldsymbol{\alpha}$ given, the minimiser of (2.8) is

$$\begin{pmatrix} a_0(z,\boldsymbol{\alpha}) \\ \mathbf{b}_0(z,\boldsymbol{\alpha}) \end{pmatrix} = \left[E\left(\mathbb{X}_t\,\mathbb{X}_t^{\mathsf{T}}\Big|\, Z_t(\boldsymbol{\alpha}) = z\right)\right]^{-1}\left[E(\mathbb{X}_t Y_t | Z_t(\boldsymbol{\alpha}) = z)\right], \quad (2.11)$$

where $\mathbb{X}_t = (1, \breve{\mathbf{X}}_t^{\mathsf{T}})^{\mathsf{T}}$, defined below (2.1). It is easy to see from (2.10) that $\{a_0(\cdot,\boldsymbol{\alpha}),\ \mathbf{b}_0(\cdot,\boldsymbol{\alpha})\}$ is also the minimizer of (2.9) for any fixed $\boldsymbol{\alpha}$. Now the true value of the index parameter should satisfy

$$\boldsymbol{\alpha}_0 = \arg\min_{\boldsymbol{\alpha}} R(a_0(\cdot,\boldsymbol{\alpha}),\mathbf{b}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}). \quad (2.12)$$

We see intuitively that $\widehat{a}(z,\boldsymbol{\alpha})$ and $\widehat{\mathbf{b}}(z,\boldsymbol{\alpha})$ defined in (2.4) are consistent estimators of $a_0(z,\boldsymbol{\alpha})$ and $\mathbf{b}_0(z,\boldsymbol{\alpha})$ (see Theorem 3.1). The estimator $\widehat{\boldsymbol{\alpha}}$ at (2.6) is a consistent estimator of $\boldsymbol{\alpha}_0$ given in (2.12) (Theorem 3.2), and (2.7) is a consistent estimator of $R(a_0(\cdot,\boldsymbol{\alpha}),\mathbf{b}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$. Finally, $\widehat{a}_0(z) \equiv \widehat{a}(z,\widehat{\boldsymbol{\alpha}},\hbar)$ and $\widehat{\mathbf{b}}_0(z) \equiv \widehat{\mathbf{b}}(z,\widehat{\boldsymbol{\alpha}},\hbar)$ (see (2.4)) are, respectively, the consistent estimators for $a_0(z) \equiv a_0(z,\boldsymbol{\alpha}_0)$ and $\mathbf{b}_0(z) \equiv \mathbf{b}_0(z,\boldsymbol{\alpha}_0)$ (Corollary 3.3).

3. Main Results

3.1. Regularity conditions and notations

We assume $\{(Y_t, \mathbf{X}_t)\}$ is a strictly stationary process. Put $\varepsilon_{t,\boldsymbol{\alpha}} = Y_t - a_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{b}_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \check{\mathbf{X}}_t$ for $\boldsymbol{\alpha} \in \mathbb{B}$ defined in (2.2). Then $\varepsilon_{t,\boldsymbol{\alpha}_0} = \varepsilon_t$ defined in (2.1). Note $\{\varepsilon_t\}$ may not be a process of *i.i.d.* variables. We denote the conditional probability density of $\boldsymbol{\xi}$ given $\boldsymbol{\eta}$ by $f_{\boldsymbol{\xi}|\boldsymbol{\eta}}(\cdot|\cdot)$. Some regularity conditions are now in order.

(C1) (Conditions on probability densities)

The density $f_Z(z, \boldsymbol{\alpha}) = f_{\boldsymbol{\alpha}^\top \mathbf{X}_t}(z)$ is continuous and bounded away from zero for $z \in [-L, L]$, uniformly for $\boldsymbol{\alpha} \in \mathbb{B}$. Furthermore, the joint probability density function of $(\boldsymbol{\alpha}^\top \mathbf{X}_{t_0}, \boldsymbol{\alpha}^\top \mathbf{X}_{t_1}, \ldots, \boldsymbol{\alpha}^\top \mathbf{X}_{t_s})$ exists and is bounded uniformly for any $t_0 < t_1 < \cdots < t_s$ and $0 \le s \le 2(r-1)$ and $\boldsymbol{\alpha} \in \mathbb{B}$, where $r \ge 1$ is some positive integer.

(C2) (Moment conditions)

$E|Y_t|^{\varrho r} < \infty$, $E\|\mathbf{X}_t\|^{\varrho r} < \infty$, $E|\varepsilon_t|^{\varrho r} < \infty$, and $\sup_{\boldsymbol{\alpha} \in \mathbb{B}} E|\varepsilon_{t,\boldsymbol{\alpha}}|^{\varrho r} < \infty$ for some real number $\varrho > 4 - 2/r$, with r given in (C1).

(C3) (Inverse matrix conditions)

The matrix function $A_1(z, \boldsymbol{\alpha}) \equiv E\left(\mathbb{X}_t\, \mathbb{X}_t^\top \,\middle|\, \boldsymbol{\alpha}^\top \mathbf{X}_t = z\right)$ is positive definite for $z \in [-L, L]$ and $\boldsymbol{\alpha} \in \mathbb{B}$, where $\mathbb{X}_t$ is defined in (2.1).

(C4) (Conditions on the nonparametric functions)

The functions $a_0(z, \boldsymbol{\alpha})$ and $\mathbf{b}_0(z, \boldsymbol{\alpha})$, defined in (2.11), are twice continuously differentiable with respect to z and continuously differentiable with respect to $\boldsymbol{\alpha}$. Also, the derivative of $R(a_0(\cdot, \boldsymbol{\alpha}), \mathbf{b}_0(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha})$ defined by (2.9) with respect to $\boldsymbol{\alpha}$, and the expectation involved, are exchangeable.

(C5) (Mixing conditions)

The process $\{(Y_t, \mathbf{X}_t)\}$ is β-mixing with the mixing coefficients $\beta(t) = O(t^{-b})$ for some $b > \max\{2(\varrho r + 1)/(\varrho r - 2), (r + a)/(1 - 2/\varrho)\}$, where r and ϱ are specified in (C1) and (C2), and $a \ge (r\varrho - 2)r/(2 + r\varrho - 4r)$.

(C6) (Conditions on the kernel function)

The kernel $K(\cdot)$ is a bounded and symmetric density function on $\mathbb{R}^1$ with bounded support S_K. Furthermore, $|K(x) - K(y)| \le C|x - y|$ for $x,\ y \in S_K$ and some $0 < C < \infty$.

(C7) (Conditions on the bandwidth)

The bandwidth h satisfies the conditions

$$\lim_{n\to\infty} h = 0 \quad \text{and} \quad \lim_{n\to\infty}\inf nh^{\frac{2(r-1)a+(\varrho r-2)}{(a+1)\varrho}} > 0 \tag{3.1}$$

for some integer $r \geq 3$. Furthermore, there exists a sequence of positive integers $s_n \to \infty$ such that $s_n = o((nh)^{1/2})$, $ns_n^{-b} \to 0$ and $s_n h^{2(\varrho r-2)/[2+b(\varrho r-2)]} > 1$ as $n \to \infty$.

Remark 1. Condition (C1) is satisfied under some mild conditions on the regressor process $\{\mathbf{X}_t\}$; see condition (C1′) and Lemma 3.1 below. Let $L_0 = L(1+\epsilon_0^{-1})$. The proof of Lemma 3.1 is given in the Appendix.

(C1′) The probability density $f_{\mathbf{X}_t}(x)$ of $\mathbf{X}_t$ is continuous and bounded away from zero on the compact set $[-L_0, L_0]^d \subset \mathbb{R}^d$. Furthermore, the joint probability density function of $(\mathbf{X}_{t_0}, \mathbf{X}_{t_1}, \ldots, \mathbf{X}_{t_s})$ and the conditional density function of $(X_{t_0,d}, X_{t_1,d}, \ldots, X_{t_s,d})$ given $(\breve{\mathbf{X}}_{t_0}, \breve{\mathbf{X}}_{t_1}, \ldots, \breve{\mathbf{X}}_{t_s})$ exist and are bounded uniformly for any $t_0 < t_1 < \cdots < t_s$ and $0 \leq s \leq 2(r-1)$, where $r \geq 1$ is some positive integer.

Lemma 3.1. *Condition* (C1) *is implied by* (C1′).

Remark 2. Condition (C2) is stronger than the standard ones imposed in nonparametric regression estimation. This is due to the fact that we need to establish uniform convergence for nonparametric regression estimators of $a(\cdot, \boldsymbol{\alpha})$ and $b(\cdot, \boldsymbol{\alpha})$ for all $\boldsymbol{\alpha} \in \mathbb{B}$. In fact the moment condition $E(e^{|\varepsilon_t|}) < \infty$ employed by Härdle et al. (1993) and Xia and Li (1999) is stronger than (C2). Many linear and nonlinear time series satisfy mixing condition (C5); see, for example, Section 2.6 of Fan and Yao (2003). The bandwidth condition (C7) is also standard for this type of problem in the context of time series (c.f., Fan and Yao (2003)). Note that (3.1) holds for $h = O(n^{-1/5})$ if $a > \{(r-5)\varrho - 2\}/\{5\varrho - 2r + 2\}$ with $\varrho > \max\{2(r-2)/5, 2/(r-5)\}$ and $r > 5$. It also holds for $h = O(n^{-1/4})$ if $a > \{(r-4)\varrho - 2\}/\{4\varrho - 2r + 2\}$ with $\varrho > \max\{(r-2)/2, 2/(r-4)\}$ and $r > 4$. Therefore condition (C7) is met for appropriate constants r, ϱ and a given in (C1), (C2) and (C5).

We now list the notation to be used in the rest of the paper. Let $X_{t0} \equiv 1$, $\mathcal{D}_n = \{1, \ldots, n\}$, $S_w = [-L, L]$, $\mu_{i,K} = \int u^i K(u)\, du$ and $\nu_{i,K} = \int u^i K^2(u)\, du$. Let $\mathbf{S}$ and $\widetilde{\mathbf{S}}$ be 2×2 matrices with, respectively, $\mu_{i+j-2,K}$ and $\nu_{i+j-2,K}$ as the (i,j)th elements. Let $\mathbf{s} = (\mu_{2,K}, \mu_{3,K})^{\mathsf{T}}$ be a 2×1 vector.

Put $\boldsymbol{\theta}_0(z,\boldsymbol{\alpha}) = (a_0(z,\boldsymbol{\alpha}), \mathbf{b}_0(z,\boldsymbol{\alpha})^{\mathsf{T}}, \dot{a}_0(z,\boldsymbol{\alpha}), \dot{\mathbf{b}}_0(z,\boldsymbol{\alpha})^{\mathsf{T}})^{\mathsf{T}}$, where $\dot{a}_0(z,\boldsymbol{\alpha}) = \partial a_0(z,\boldsymbol{\alpha})/\partial z$, $\dot{\mathbf{b}}_0(z,\boldsymbol{\alpha}) = \partial \mathbf{b}_0(z,\boldsymbol{\alpha})/\partial z$. Similarly, write $\ddot{a}_0(z,\boldsymbol{\alpha}) = \partial^2 a_0(z,\boldsymbol{\alpha})/\partial z^2$ and $\ddot{\mathbf{b}}_0(z,\boldsymbol{\alpha}) = \partial^2 \mathbf{b}_0(z,\boldsymbol{\alpha})/\partial z^2$.

For notational convenience, we write

$$\mathbf{g} = (\mathbf{g}^1, \ldots, \mathbf{g}^d)^{\mathsf{T}} = \mathbf{g}(z,\boldsymbol{\alpha}) = (a(z,\boldsymbol{\alpha}), \mathbf{b}(z,\boldsymbol{\alpha})^{\mathsf{T}})^{\mathsf{T}}, \tag{3.2}$$

and $\mathbf{g}_0 = \mathbf{g}_0(z,\boldsymbol{\alpha})$ and $\widehat{\mathbf{g}} = \widehat{\mathbf{g}}(z,\boldsymbol{\alpha})$ are defined in a similar manner. Assume $\mathbf{g}(z,\boldsymbol{\alpha})$ is second order differentiable. Denote by $\mathbf{g}_1 = \mathbf{g}_1(z,\boldsymbol{\alpha})$ the $d \times 1$ vector whose jth element is $\mathbf{g}_1^j = \mathbf{g}_1^j(z,\boldsymbol{\alpha}) = \partial \mathbf{g}^j(z,\boldsymbol{\alpha})/\partial z$, and $\mathbf{g}_2 = \mathbf{g}_2(z,\boldsymbol{\alpha})$ the $d \times d$ matrix whose (i,j)th element is $\mathbf{g}_2^{ij} = \mathbf{g}_2^{ij}(z,\boldsymbol{\alpha}) = \partial \mathbf{g}^i(z,\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^j$. Similarly, we define $\mathbf{g}_{01} = \mathbf{g}_{01}(z,\boldsymbol{\alpha}) = \partial \mathbf{g}_0(z,\boldsymbol{\alpha})/\partial z$, $\mathbf{g}_{02} = \mathbf{g}_{02}(z,\boldsymbol{\alpha}) = \partial \mathbf{g}_0(z,\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^{\mathsf{T}}$ and $\widehat{\mathbf{g}}_1 = \widehat{\mathbf{g}}_1(z,\boldsymbol{\alpha}) = \partial \widehat{\mathbf{g}}(z,\boldsymbol{\alpha})/\partial z$, $\widehat{\mathbf{g}}_2 = \widehat{\mathbf{g}}_2(z,\boldsymbol{\alpha}) = \partial \widehat{\mathbf{g}}(z,\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^{\mathsf{T}}$.

The Euclidean norm of $\mathbf{g}$ is denoted as before by $\|\mathbf{g}\| = (\mathbf{g}^{\mathsf{T}}\mathbf{g})^{1/2}$. We also use the notation $\|\mathbf{g}\|_{\mathcal{G}} = \sup_{|z|\le L, \boldsymbol{\alpha}\in\mathbb{B}} \|\mathbf{g}(z,\boldsymbol{\alpha})\|$ for a continuous function $\mathbf{g}$ defined on $S_w \times \mathbb{B}$ (c.f., § 5.2). Under (C4), such a norm can apply to $\mathbf{g}_0(z,\boldsymbol{\alpha})$ and its first order partial derivatives.

For $\boldsymbol{\alpha} \in \mathbb{B}$ fixed, we are also concerned with an alternative norm of $\mathbf{g}(z,\boldsymbol{\alpha})$ as a function of z. For any nonnegative integer κ and any smooth function $g : S_w \mapsto \mathbb{R}^d$, define the differential operator $\mathcal{D}^\kappa g(z) = d^\kappa g(z)/dz^\kappa$, note $S_w = [-L, L]$ is the support of $w(\cdot)$ and is a bounded, convex subset of $\mathbb{R}^1$ with nonempty interior. For some $\phi > 0$, let $[\phi]$ be the largest integer not greater than ϕ, and define (if it exists)

$$\|g\|_{\infty,\phi} = \max_{0\le\kappa\le[\phi]} \sup_{|z|\le L} \|\mathcal{D}^\kappa g(z)\| + \sup_{\substack{z\ne z' \\ |z|\le L}} \frac{\|\mathcal{D}^{[\phi]} g(z) - \mathcal{D}^{[\phi]} g(z')\|}{|z - z'|^{\phi-[\phi]}}.$$

Further, let $C_c^\phi(S_w)$ be the set of all continuous functions $g : S_w \mapsto \mathbb{R}^d$ with $\|g\|_{\infty,\phi} \le c$. With these notations at hand, we define a function space $\mathcal{G}$ in Section 5.2. Clearly, under (C4), such a norm may apply to the function $\mathbf{g}_0(z,\boldsymbol{\alpha})$ and its first order partial derivatives (with $\boldsymbol{\alpha}$ fixed) with $\phi = 2$ and $\phi = 1$, respectively.

3.2. Asymptotic properties

We state the asymptotic properties of our estimation procedure in two steps. First Theorem 3.1 states that $\widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha},h)$, defined in (2.4) with $h = O(n^{-1/5})$, is asymptotically normal for any $\boldsymbol{\alpha} \in \mathbb{B}$ fixed. Furthermore the same result still holds if $\boldsymbol{\alpha}$ is replaced by a root-n consistent estimator. Theorem 3.2 presents the asymptotic normality for the estimator $\widehat{\boldsymbol{\alpha}}$, given as (2.6), with the standard root-n convergence rate provided $h = o(n^{-1/4})$.

Theorem 3.1. *Let* (C1)–(C7) *hold and* $h = O(n^{-1/5})$. *Then it holds, for* $\boldsymbol{\alpha} \in \mathbb{B}$, *that*

$$\sqrt{nh}\left[H_n\left\{\widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha},h) - \boldsymbol{\theta}_0(z,\boldsymbol{\alpha})\right\} - \frac{1}{2}h^2\mathbf{B}(z)\right] \xrightarrow{D} N\{\mathbf{0},\mathbf{A}(z)\}, \tag{3.3}$$

where $H_n = \mathrm{diag}(1,h)\otimes I_{d\times d}$, *with* $I_{d\times d}$ *the* $d\times d$ *identity matrix and* $\otimes$ *the Kroneck product,* $\mathbf{B}(z) = \{(\mathbf{S}^{-1}\mathbf{s})\otimes I_{d\times d}\}(\ddot{a}_0(z,\boldsymbol{\alpha}),\ddot{\mathbf{b}}_0(z,\boldsymbol{\alpha})^{\mathsf{T}})^{\mathsf{T}} = ((\ddot{a}_0(z,\boldsymbol{\alpha}), \ddot{\mathbf{b}}_0(z,\boldsymbol{\alpha})^{\mathsf{T}})\mu_{2,K},0,\ldots,0)^{\mathsf{T}} \in \mathbb{R}^{2d}$, $\mathbf{A}(z) = \{f_Z(z)\}^{-1}(\mathbf{S}^{-1}\widetilde{\mathbf{S}}\mathbf{S}^{-1})\otimes(\mathbf{G}^{-1}(z)\widetilde{\mathbf{G}}(z)\,\mathbf{G}^{-1}(z))$, *and* $\mathbf{G}(z)$ *and* $\widetilde{\mathbf{G}}(z)$ *are two* $d\times d$ *matrices with, respectively,* $G_{ij}(z) = E(X_{t,i-1}X_{t,j-1}|Z_t = z)$ *and* $\widetilde{G}_{ij}(z) = E(\varepsilon^2_{t,\boldsymbol{\alpha}}X_{t,i-1}X_{t,j-1}|Z_t = z)$ *as the* (i,j)*th elements.*

Furthermore, (3.3) *still holds if* $\boldsymbol{\alpha}$ *is replaced by* $\check{\boldsymbol{\alpha}}$ *provided* $\check{\boldsymbol{\alpha}}-\boldsymbol{\alpha} = O_p(n^{-1/2})$.

Theorem 3.2. *Suppose* (C1)–(C7) *hold. Set* $Z^o_t = \boldsymbol{\alpha}_0^{\mathsf{T}}\mathbf{X}_t$. *Then if* $\varrho \ge 6$, $r > 3d$ *and* $nh^4 = O(1)$, $nh^{3+3d/r}\to\infty$ *as* $n\to\infty$, *it holds that*

$$\sqrt{n}\left\{\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 + \frac{1}{2}\Gamma_0^-\mathcal{B}h^2\right\} \xrightarrow{D} N\left(\mathbf{0},\ \Gamma_0^-\mathcal{V}_0(\Gamma_0^-)^{\mathsf{T}}\right), \tag{3.4}$$

where, setting $\mathbf{g}_{01t} = \mathbf{g}_{01}(Z^o_t,\boldsymbol{\alpha}_0)$ *and* $\mathbf{g}_{02t} = \mathbf{g}_{02}(Z^o_t,\boldsymbol{\alpha}_0)$,

$$\begin{aligned}
\mathcal{B} &= E\left\{(\mathbf{g}_{01t}\mathbf{X}_t^{\mathsf{T}} + \mathbf{g}_{02t})^{\mathsf{T}}\mathbb{X}_t\mathbb{X}_t^{\mathsf{T}}\ddot{g}_0(Z^o_t)w(Z^o_t)\right\}\mu_{2,K},\\
\Gamma_0 &= E\left\{(\mathbf{g}_{01t}\mathbf{X}_t^{\mathsf{T}} + \mathbf{g}_{02t})^{\mathsf{T}}\mathbb{X}_t\mathbb{X}_t^{\mathsf{T}}(\mathbf{g}_{01t}\mathbf{X}_t^{\mathsf{T}} + \mathbf{g}_{02t})w(Z^o_t)\right\},\\
\mathcal{V}_0 &= E\varepsilon_t^2\left[\Xi_t\Xi_t^{\mathsf{T}} - \{E(\Xi_t\mathbb{X}_t^{\mathsf{T}}|Z^o_t)\}\{E(\mathbb{X}_t\mathbb{X}_t^{\mathsf{T}}|Z^o_t)\}^{-1}\{E(\mathbb{X}_t\Xi_t^{\mathsf{T}}|Z^o_t)\}\right],
\end{aligned}$$

Γ_0^- *is a generalized inverse of* Γ_0, $\Xi_t = \mathbf{X}_t\left\{\dot{a}_0(Z^o_t) + \dot{\mathbf{b}}_0(Z^o_t)^{\mathsf{T}}\breve{\mathbf{X}}_t\right\}w(Z^o_t)$, *and* $\mathbf{G}_0(z)$ *is a* $d\times d$ *matrix with the* (i,j)*th elements* $G^0_{ij}(z) = E(X_{t,i-1}X_{t,j-1}|Z^o_t = z)$.

Furthermore, if $nh^4 = o(1)$, *then* (3.4) *reduces to*

$$\sqrt{n}\left\{\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right\} \xrightarrow{D} N\left(\mathbf{0},\ \Gamma_0^-\mathcal{V}_0(\Gamma_0^-)^{\mathsf{T}}\right). \tag{3.5}$$

Corollary 3.3. *Assume the conditions of Theorem* 3.2 *hold, with* $\widehat{\boldsymbol{\alpha}}$ *defined in* (2.6) *with* $h = o(n^{-1/4})$. *Then letting* $\hbar = O(n^{-1/5})$, *we have*

$$\sqrt{n\hbar}\left[\bar{H}_n\left\{\widehat{\boldsymbol{\theta}}(z,\widehat{\boldsymbol{\alpha}},\hbar) - \boldsymbol{\theta}_0(z,\boldsymbol{\alpha}_0)\right\} - \frac{1}{2}\hbar^2\mathbf{B}_0(z)\right] \xrightarrow{D} N\{\mathbf{0},\mathbf{A}_0(z)\}, \tag{3.6}$$

where $\bar{H}_n = \mathrm{diag}(1,\hbar)\otimes I_{d\times d}$, *and* $\mathbf{B}_0(z)$ *and* $\mathbf{A}_0(z)$ *are defined in the same way as* $\mathbf{B}(z)$ *and* $\mathbf{A}(z)$ *in Theorem* 3.1, *with* $\boldsymbol{\alpha}$ *replaced by* $\boldsymbol{\alpha}_0$.

This corollary follows directly from Theorems 3.1 and 3.2.

Remark 3. (i) The estimator $\widehat{\boldsymbol{\theta}}(z, \widehat{\boldsymbol{\alpha}}, \hbar)$ is asymptotically adaptive to unknown $\boldsymbol{\alpha}_0$ in the sense that $\widehat{\boldsymbol{\theta}}(z, \widehat{\boldsymbol{\alpha}}, \hbar)$ and $\widehat{\boldsymbol{\theta}}(z, \boldsymbol{\alpha}_0, \hbar)$ share the same (first order) asymptotic distribution.

(ii) For $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$, $E\{\varepsilon_{t,\boldsymbol{\alpha}}\} \neq 0$. However the estimator $\widehat{\boldsymbol{\theta}}(z, \boldsymbol{\alpha}, \hbar)$ is still asymptotic unbiased due to the least squares property; see Lemma 5.1 below.

4. Two Important Lemmas

To establish the asymptotic properties for the estimator $\widehat{\boldsymbol{\alpha}}$, we first establish two lemmas of independent interest. We do not make use of the specific forms of $\mathbf{g}(z, \boldsymbol{\alpha})$ and $\mathbb{B}$ in the proofs, so they are applicable to the estimators for parameter vectors in general semiparametric settings.

4.1. Consistency lemma

In this section, for generality, let $\mathbb{B}$ be a closed subset in $\mathbb{R}^d$, and $\mathcal{G}$ the space of functions of the form $\mathbf{g}(z, \boldsymbol{\alpha})$, defined on $S_w \times \mathbb{B}$, with a norm $\|\mathbf{g}\|_{\mathcal{G}}$. We are concerned with the functions $\mathbf{g}(z, \boldsymbol{\alpha})$, $\widehat{\mathbf{g}}(z, \boldsymbol{\alpha})$ and $\mathbf{g}_0(z, \boldsymbol{\alpha})$ in $\mathcal{G}$. Let $\mathbf{g}_0(z) = \mathbf{g}_0(z, \boldsymbol{\alpha}_0)$. In Section 5, we specify $\mathbb{B}$ and $\mathcal{G}$ with the norm $\|\mathbf{g}\|_{\mathcal{G}}$ in the context of the (2.1).

Lemma 4.1. *Suppose* $\boldsymbol{\alpha}_0 \in \mathbb{B}$ *satisfies* $R(\mathbf{g}_0(\cdot), \boldsymbol{\alpha}_0) = \inf_{\boldsymbol{\alpha} \in \mathbb{B}} R(\mathbf{g}_0(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha})$, *and that the following hold.*

(i) $R_n\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right) \leq \inf_{\boldsymbol{\alpha} \in \mathbb{B}} R_n\left(\widehat{\mathbf{g}}(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}\right) + o_P(1)$.

(ii) *For all* $\delta > 0$, *there exists* $\epsilon(\delta) > 0$ *such that*

$$\inf_{\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| > \delta} R\left(\mathbf{g}_0(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}\right) \geq R\left(\mathbf{g}_0(\cdot), \boldsymbol{\alpha}_0\right) + \epsilon(\delta).$$

(iii) *Uniformly for all* $\boldsymbol{\alpha} \in \mathbb{B}$, $R\left(\mathbf{g}(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}\right)$ *is continuous [with respect to the metric* $\|\cdot\|_{\mathcal{G}}$*] in* $\mathbf{g}(\cdot, \boldsymbol{\alpha})$ *at* $\mathbf{g}_0(\cdot, \boldsymbol{\alpha})$.

(iv) $\|\widehat{\mathbf{g}}(\cdot, \cdot) - \mathbf{g}_0(\cdot, \cdot)\|_{\mathcal{G}} = o_P(1)$.

(v) *For all* $\{\delta_n\}$ *with* $\delta_n = o(1)$,

$$\sup_{\boldsymbol{\alpha} \in \mathbb{B}} \sup_{\|\mathbf{g}(\cdot, \boldsymbol{\alpha}) - \mathbf{g}_0(\cdot, \boldsymbol{\alpha})\|_{\mathcal{G}} \leq \delta_n} \left|R_n\left(\mathbf{g}(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}\right) - R\left(\mathbf{g}(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}\right)\right| = o_P(1).$$

Then $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = o_P(1)$.

Proof. The proof is similar to that of Corollary 3.2 in Pakes and Pollard (1989) and Theorem 1 in Chan, Linton and Keilegom (2003). By (ii), for all $\delta > 0$,

$$P\{\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\| > \delta\} \leq P\left\{R\left(\mathbf{g}_0(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right) - R\left(\mathbf{g}_0(\cdot), \boldsymbol{\alpha}_0\right) \geq \epsilon(\delta)\right\},$$

hence it suffices to show that

$$R\left(\mathbf{g}_0(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right) - R\left(\mathbf{g}_0(\cdot), \boldsymbol{\alpha}_0\right) = o_P(1). \tag{4.1}$$

Note that

$$R\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) - R\left(\mathbf{g}_0(\cdot),\boldsymbol{\alpha}_0\right) = R\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) - R\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) \tag{4.2}$$

$$+R\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) - R_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) \tag{4.3}$$

$$+R_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) - R\left(\mathbf{g}_0(\cdot),\boldsymbol{\alpha}_0\right). \tag{4.4}$$

That the expression in (4.2) tends to 0 in probability clearly follows from conditions (iii) and (iv). The absolute value of the expression in (4.3) is bounded above by $\sup_{\boldsymbol{\alpha}\in\mathbb{B}} |R(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) - R_n(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}})| = o_P(1)$, which follows from (iv) and (v). Finally, we have to show that the expression in (4.4) tends to 0 in probability. As $R_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) = \inf_{\boldsymbol{\alpha}\in\mathbb{B}} R_n\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$, note that

$$\begin{aligned}
R_n\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) &= \left\{R_n\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) - R\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)\right\} \\
&\quad + \left\{R\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) - R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)\right\} + R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) \\
&\le \sup_{\boldsymbol{\alpha}\in\mathbb{B}} \left|R_n\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) - R\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)\right| \\
&\quad + \sup_{\boldsymbol{\alpha}\in\mathbb{B}} \left|R\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) - R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)\right| + R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right), \\
&\equiv R_1 + R_2 + R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right).
\end{aligned} \tag{4.5}$$

Then we have

$$R_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) \le R_1 + R_2 + \inf_{\boldsymbol{\alpha}\in\mathbb{B}} R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) = R_1 + R_2 + R\left(\mathbf{g}_0(\cdot),\boldsymbol{\alpha}_0\right). \tag{4.6}$$

It follows, from (iv) and (v) that $R_1 = o_P(1)$, and from (iii) and (iv) that $R_2 = o_P(1)$, and thus we deduce from (4.6) that, for any $\varepsilon > 0$ as $n \to \infty$,

$$P\left\{R_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right) \le \varepsilon + R\left(\mathbf{g}_0(\cdot),\boldsymbol{\alpha}_0\right)\right\} \to 1. \tag{4.7}$$

Similarly, by exchanging $R_n\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ and $R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ in (4.5), we can prove

$$P\left\{R\left(\mathbf{g}_0(\cdot),\boldsymbol{\alpha}_0\right) \le \varepsilon + R_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)\right\} \to 1. \tag{4.8}$$

Therefore it follows from (4.7) and (4.8) that (4.4) tends to 0 in probability, and hence (4.1) is proved.

4.2. Asymptotic normality lemma

Suppose $R\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ and $R_n\left(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ are differentiable with respect to $\boldsymbol{\alpha}$. Denote the derivatives of $R\left(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ and $R_n\left(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ with respect to $\boldsymbol{\alpha}$ by

$$\dot{R}\left(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) = \frac{dR\left(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)}{d\boldsymbol{\alpha}} \quad \text{and} \quad \dot{R}_n\left(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right) = \frac{dR_n\left(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)}{d\boldsymbol{\alpha}}.$$

Then as $\boldsymbol{\alpha}_0$ and $\widehat{\boldsymbol{\alpha}}$ are the minimizers of $R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ and $R_n(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$, respectively, we have

$$\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) = \dot{R}(\mathbf{g}_0(\cdot),\boldsymbol{\alpha}_0) = 0 \quad \text{and} \quad \dot{R}_n(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) = 0.$$

Define the ordinary derivative of $\dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ (if it exists) as

$$\Gamma_1(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) = \frac{d\dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})}{d\boldsymbol{\alpha}^{\mathsf{T}}} = \frac{d^2R(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})}{d\boldsymbol{\alpha}d\boldsymbol{\alpha}^{\mathsf{T}}},$$

and the functional derivative Γ_2 of $\dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ with respect to $\mathbf{g}(\cdot,\boldsymbol{\alpha})$ at $\mathbf{g}_0(\cdot,\boldsymbol{\alpha})$ in the direction $\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})$ by

$$\begin{aligned}&\Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})[\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})]\\ &\quad = \lim_{\tau\to 0}\Big[\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}) + \tau(\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\Big]\frac{1}{\tau} \qquad (4.9)\end{aligned}$$

(if the limit exists) for all $\mathbf{g}(\cdot,\boldsymbol{\alpha})$ satisfying $\mathbf{g}_0(\cdot,\boldsymbol{\alpha}) + \tau(\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})) \in \mathcal{G}$ with $\tau \in [0,1]$.

Now we assume that $\widehat{\boldsymbol{\alpha}}$ is consistent and $\boldsymbol{\alpha}_0 \in \mathbb{B}$. Therefore the parameter space $\mathbb{B}$ and $\mathcal{G}$ can be replaced by small or even shrinking sets. Define $\mathbb{B}_\delta = \{\boldsymbol{\alpha} \in \mathbb{B} : \ \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \le \delta\}$ and $\mathcal{G}_\delta = \{\mathbf{g} \in \mathcal{G} : \ \|\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})\|_{\mathcal{G}} \le \delta\}$.

Lemma 4.2. *Assume that $R_n(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ is differentiable with respect to $\boldsymbol{\alpha}$ with the derivative $\dot{R}_n(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$, and $R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ is second order differentiable with respect to $\boldsymbol{\alpha}$, with first order derivative $\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ and second order derivative $\Gamma_1(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$. Suppose that $\boldsymbol{\alpha}_0 \in \mathbb{B}_\delta$ satisfies $\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) = 0$, that $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = o_P(1)$, and that the following hold.*

(i) $\dot{R}_n(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) = o_P(n^{-1/2})$.

(ii) (1) $\Gamma_1(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ is continuous at $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$.
(2) $\Gamma_1 = \Gamma_1(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)$ has generalized inverse, Γ_1^-.

(iii) For all $\boldsymbol{\alpha} \in \mathbb{B}_\delta$, the pathwise derivative, $\Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})[\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})]$ (c.f., (4.9)), of $\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ exists in all directions $\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha}) \in \mathcal{G}_\delta$, and satisfies (1) uniformly for $\boldsymbol{\alpha} \in \mathbb{B}_\delta$, $\|\dot{R}(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})]\| = o_P(n^{-1/2})$;
(2) for all $(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) \in \mathcal{G}_{\delta_n} \times \mathbb{B}_{\delta_n}$ with a positive sequence $\delta_n = o(1)$: $\|\Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})[\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})] - \Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)[\mathbf{g}(\cdot,\boldsymbol{\alpha}_0) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]\| \le o(1)\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|$.

(iv) $\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}) \in \mathcal{G}$ *with probability tending to* 1, *and* $\|\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})\|_{\mathcal{G}} = o_P(1)$, $\|\widehat{\mathbf{g}}_1(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_{10}(\cdot,\boldsymbol{\alpha})\|_{\mathcal{G}} = o_P(1)$, *and* $\|\widehat{\mathbf{g}}_2(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_{20}(\cdot,\boldsymbol{\alpha})\|_{\mathcal{G}}] = o_P(1)$.

(v) *For all sequences of positive numbers* $\{\delta_n\}$ *with* $\delta_n = o(1)$,

$$\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\le\delta_n}\ \sup_{\|\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})\|_{\mathcal{G}}\le\delta_n}\Big\|\dot{R}_n(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \dot{R}_n(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\Big\| = o_P(n^{-\frac{1}{2}}).$$

(vi) *For some* $B_n = O(n^{-1/2})$ *and some finite matrix* V_1,

$$\sqrt{n}\left\{\dot{R}_n\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)+\Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]-B_n\right\} \xrightarrow{D} N(0,V_1).$$

Then $\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0+\Gamma_1^- B_n)\xrightarrow{D} N(0,\Omega)$, *where* $\Omega=\Gamma_1^- V_1(\Gamma_1^-)^{\mathsf{T}}$.

Remark 4. The objective function defining the semi-parametric estimators in Chen et al. (2003) is of a generalized method of moments (GMM) type, and hence their Theorem 2 does not apply directly to least-squares (in this paper) or maximum-likelihood-like semi-parametric estimators. Their argument is, however, helpful for the proof of this lemma. Note that the conditions (i), (ii) and (v) specified for the derivative of the objective function in this lemma are basically similar to those on the GMM type objective function in Theorem 2 of Chen et al. (2003), while the conditions (iii), (iv) and (vi) are different. In fact, condition (iv) is much weaker than that of Chen et al. (2003) which requires the convergence of rate $o_P(n^{-1/4})$, and condition (vi) allows a bias term B_n.

Proof. We only sketch the proof here. First we are establishing $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}_0$. Owing to $\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0=o_P(1)$ and condition (iv), we can choose a positive sequence $\delta_n=o(1)$ such that $P\{\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|\le\delta_n,\ \|\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})\|_{\mathcal{G}}\le\delta_n\}\to 1$. So in the following we need only to look at $(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\in\mathcal{G}_{\delta_n}\times\mathbb{B}_{\delta_n}$. In light of $\dot{R}\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)=0$ and condition (ii), we have by Taylor expansion that

$$\dot{R}\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)=\Gamma_1\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)(1+o_P(1)), \tag{4.10}$$

which implies that $\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0$ has the same convergence rate to 0 as that of $\dot{R}(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}})$. Similarly to (5) and (6) of Chen et al. (2003), it is obvious that

$$\begin{aligned}\|\dot{R}\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)\| &\le \|\dot{R}\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)-\dot{R}\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)\|\\ &\quad+\|\dot{R}\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)-\dot{R}_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)+\dot{R}_n\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\|\\ &\quad+\|\dot{R}_n\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)\|+\|-\dot{R}_n\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\|\\ &\equiv D_1+D_2+D_3+D_4,\end{aligned}$$

$$\begin{aligned}D_1\le&\ \|\dot{R}\left(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)-\dot{R}\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)-\Gamma_2\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)[\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}})-\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}})]\|\\ &+\|\Gamma_2\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)[\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}})-\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}})]-\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]\|\\ &+\|\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]\|\\ \equiv&\ D_{11}+D_{12}+D_{13}.\end{aligned}$$

Clearly, conditions (iii)(1) implies $D_{11}=o_P(n^{-1/2})$; condition (iii)(2) and (4.10) imply $D_{12}=\|\dot{R}\left(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}\right)\|\times o_P(1)$; condition (vi) implies $D_{13}=O_P(n^{-1/2})$

and $D_4 = O_P(n^{-1/2})$; condition (i) implies $D_3 = o_P(n^{-1/2})$; condition (v) implies $D_2 = o_P(n^{-1/2})$. Therefore it follows that $\|\dot{R}\,(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}})\,\| \times (1 - o_P(1)) = O_P(n^{-1/2})$, and hence $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = O_P(n^{-1/2})$.

Next, set $\mathcal{L}_n(\boldsymbol{\alpha}) = \dot{R}_n\,(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) + \Gamma_1(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0) + \Gamma_2\,(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\,[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]$. It is obvious that

$$
\begin{aligned}
&\|\mathcal{L}_n(\widehat{\boldsymbol{\alpha}})\| \\
&\quad \le \|\dot{R}_n\,(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) - \mathcal{L}_n(\widehat{\boldsymbol{\alpha}})\| + \|\dot{R}_n\,(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}})\,\| \\
&\quad \le \|\dot{R}\,(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) - \dot{R}\,(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) - \Gamma_2\,(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\,[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]\| \\
&\qquad + \|\dot{R}\,(\mathbf{g}_0(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) - \Gamma_1(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)\| \\
&\qquad + \|\dot{R}_n\,(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) - \dot{R}\,(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) - \dot{R}_n\,(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\,\| \\
&\qquad + \|\dot{R}_n\,(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}})\,\| \\
&\quad \equiv D_5 + D_6 + D_7 + D_8.
\end{aligned}
$$

Clearly conditions (iii) and (iv), together with $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = O_P(n^{-1/2})$, imply $D_5 = o_P(n^{-1/2})$; in view of $\dot{R}\,(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) = 0$, it follows by Taylor expansion, with condition (ii)(1) as well as $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = O_P(n^{-1/2})$, that $D_6 = o_P(n^{-1/2})$; condition (v) implies $D_7 = o_P(n^{-1/2})$; condition (i) implies $D_8 = o_P(n^{-1/2})$. Therefore $\mathcal{L}_n(\boldsymbol{\alpha}) = o_P(n^{-1/2})$, which leads to

$$
\begin{aligned}
&\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 + \Gamma_1^- B_n \\
&\quad = -\Gamma_1^- \left\{\dot{R}_n\,(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) + \Gamma_2\,(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\,[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)] - B_n\right\} \\
&\qquad + o_P(n^{-\frac{1}{2}}),
\end{aligned}
$$

and hence the lemma follows from condition (vi).

5. Proof of the Main Results

Recall that $\varepsilon_{t,\boldsymbol{\alpha}} = Y_t - a_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{b}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\check{\mathbf{X}}_t = Y_t - \mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top \mathbb{X}_t$, and $\mathbb{X}_t$ was defined in (2.1). The following lemma is basic and is used throughout.

Lemma 5.1. *For any measurable function* $\mathbf{g}(\cdot,\boldsymbol{\alpha}) = (a(\cdot,\boldsymbol{\alpha}),\mathbf{b}(\cdot,\boldsymbol{\alpha})^\top)^\top$ *on* $\mathbb{R}^1$, *we have*

$$
E\varepsilon_{t,\boldsymbol{\alpha}}\{a(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}) + \mathbf{b}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\check{\mathbf{X}}_t\} = E\varepsilon_{t,\boldsymbol{\alpha}}\mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t = 0. \qquad (5.1)
$$

Proof of Lemma 5.1. Note that the left hand side of (5.1) equals

$$
\int \left[E\left\{\left(Y_t - a_0(z,\boldsymbol{\alpha}) - \mathbf{b}_0(z,\boldsymbol{\alpha})^\top\check{\mathbf{X}}_t\right)\left(a(z) + \mathbf{b}(z)^\top\check{\mathbf{X}}_t\right)\Big|\,\boldsymbol{\alpha}^\top\mathbf{X}_t = z\right\}\right] f_Z(z)\,dz,
$$

and that by the definition of $a_0(\cdot,\boldsymbol{\alpha})$ and $\mathbf{b}_0(\cdot,\boldsymbol{\alpha})$ in (2.11),

$$E\left\{\left(Y_t - a_0(z,\boldsymbol{\alpha}) - \mathbf{b}_0(z,\boldsymbol{\alpha})^{\mathsf{T}}\check{\mathbf{X}}_t\right)\mathbb{X}_t\middle|\,\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t = z\right\} = 0.$$

Therefore (5.1) follows.

5.1. Proof of Theorem 3.1

It follows from (2.3) by least squares that

$$\begin{aligned}\widehat{\theta}(z,\boldsymbol{\alpha}) = \widehat{\theta}(z,\boldsymbol{\alpha},h) &= H_n^{-1}\left(\widehat{a}_z, \widehat{\mathbf{b}}_z^{\mathsf{T}}, \widehat{\dot{a}}_z h, \widehat{\dot{\mathbf{b}}}_z^{\mathsf{T}} h\right)^{\mathsf{T}} \\ &= H_n^{-1}\left\{\mathcal{X}(z)^{\mathsf{T}}\mathcal{W}(z)\mathcal{X}(z)\right\}^{-1}\left\{\mathcal{X}(z)^{\mathsf{T}}\mathcal{W}(z)\mathcal{Y}\right\}, \qquad (5.2)\end{aligned}$$

where $\mathcal{Y} = (Y_1,\ldots,Y_n)^{\mathsf{T}}$, $\mathcal{W}(z) = \mathcal{W}(z,\boldsymbol{\alpha})$ is an $n\times n$ diagonal matrix with $K_h\{Z_t - z\}w\{Z_t\}$ as its tth diagonal element, $\mathcal{X}(z) = \mathcal{X}(z,\boldsymbol{\alpha})$ is an $n \times 2d$ matrix with $(\mathbb{X}_t^{\mathsf{T}}, h^{-1}(Z_t - z)\mathbb{X}_t^{\mathsf{T}})$ as its tth row and $\mathbb{X}_t = (1, \check{\mathbf{X}}_t^{\mathsf{T}})^{\mathsf{T}}$, and $H_n = \operatorname{diag}(1,h)\otimes I_{d\times d}$.

Let $\widehat{\Phi} = \widehat{\Phi}(z;\boldsymbol{\alpha}) = n^{-1}\mathcal{X}(z)^{\mathsf{T}}\mathcal{W}(z)\mathcal{X}(z)$ and $\widehat{\Psi} = \widehat{\Psi}(z;\boldsymbol{\alpha}) = n^{-1}\mathcal{X}(z)^{\mathsf{T}}\mathcal{W}(z)\mathcal{Y}$, with the (i,j)th elements $\widehat{\Phi}_{i,j}$ and $\widehat{\Psi}_{i,j}$, respectively. Also, recall that $X_{t,0}\equiv 1$ for notational convenience. Then with the notations in (5.2), we have, for $i,j = 1,\ldots,d$,

$$\widehat{\Phi}_{i,j} = n^{-1}\sum_{t=1}^{n} X_{t,i-1}X_{t,j-1}K_h(Z_t - z)w(Z_t), \qquad (5.3)$$

$$\widehat{\Phi}_{i,d+j} = \widehat{\Phi}_{d+j,i} = n^{-1}\sum_{t=1}^{n} X_{t,i-1}X_{t,j-1}\frac{Z_t - z}{h}K_h(Z_t - z)w(Z_t), \qquad (5.4)$$

$$\widehat{\Phi}_{d+i,d+j} = n^{-1}\sum_{t=1}^{n} X_{t,i-1}X_{t,j-1}(\frac{Z_t - z}{h})^2K_h(Z_t - z)w(Z_t), \qquad (5.5)$$

$$\widehat{\Psi}_{i} = n^{-1}\sum_{t=1}^{n} Y_tX_{t,i-1}K_h(Z_t - z)w(Z_t), \qquad (5.6)$$

$$\widehat{\Psi}_{d+i} = n^{-1}\sum_{t=1}^{n} Y_tX_{t,i-1}\frac{Z_t - z}{h}K_h(Z_t - z)w(Z_t). \qquad (5.7)$$

Let $\boldsymbol{\theta}_0 \equiv \boldsymbol{\theta}_0(z,\boldsymbol{\alpha}) = (a_0(z,\boldsymbol{\alpha}), \mathbf{b}_0(z,\boldsymbol{\alpha})^{\mathsf{T}}, \dot{a}_0(z,\boldsymbol{\alpha}), \dot{\mathbf{b}}_0(z,\boldsymbol{\alpha})^{\mathsf{T}})^{\mathsf{T}}$. Then by (5.3)–(5.7), we have

$$\widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha}) - \boldsymbol{\theta}_0(z,\boldsymbol{\alpha}) = \widehat{\Phi}^{-1}(\widehat{\Psi} - \widehat{\Phi}\boldsymbol{\theta}_0) \equiv \widehat{\Phi}^{-1}\widehat{\mathbf{W}}, \qquad (5.8)$$

where $\widehat{\mathbf{W}} = \widehat{\mathbf{W}}(z;\boldsymbol{\alpha})$ is a $2d$-dimensional vector with elements

$$\widehat{W}_i = n^{-1}\sum_{t=1}^{n} Y_t^* X_{t,i-1} K_h(Z_t - z) w(Z_t). \tag{5.9}$$

Moreover,

$$\widehat{W}_{d+i} = n^{-1}\sum_{t=1}^{n} Y_t^* X_{t,i-1} \frac{Z_t - z}{h} K_h(Z_t - z) w(Z_t) \tag{5.10}$$

for $i = 1, 2, \ldots, d$, with

$$Y_t^* = Y_t^*(z,\boldsymbol{\alpha}) = Y_t - \Big\{a_0(z,\boldsymbol{\alpha}) + \breve{\mathbf{X}}_t^{\mathsf{T}}\mathbf{b}_0(z,\boldsymbol{\alpha})\Big\} - \Big\{\dot{a}_0(z,\boldsymbol{\alpha}) + \breve{\mathbf{X}}_t^{\mathsf{T}}\dot{\mathbf{b}}_0(z,\boldsymbol{\alpha})\Big\}(Z_t - z).$$

With (2.1) and $Z_t = \boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t$ in mind, we then have, by a Taylor expansion of order 2,

$$Y_t^* = \frac{1}{2}\Big\{\ddot{a}_0(\xi,\boldsymbol{\alpha}) + \breve{\mathbf{X}}_t^{\mathsf{T}}\ddot{\mathbf{b}}_0(\xi,\boldsymbol{\alpha})\Big\}(Z_t - z)^2 + \varepsilon_{t,\boldsymbol{\alpha}}, \tag{5.11}$$

where $\xi = z + \eta_1\,(Z_t - z)$ with $|\eta_1| < 1$.

With $\bar{g}_{ij}(z) = \bar{g}_{ij}(z,\boldsymbol{\alpha}) = E\{X_{t,i}X_{t,j}\ |Z_t(\boldsymbol{\alpha}) = z\}$ for $i, j = 0, 1, \ldots, d$, we take $\bar{g}_i(z) = \bar{g}_{i0}(z)$, and $G_i(z) = (\bar{g}_{i1}(z), \ldots, \bar{g}_{i,d-1}(z))^{\mathsf{T}}$ (a $(d-1)$-dimensional vector). As $\mu_{i,K} = \int u^i K(u)\,du$, using time series asymptotics (see e.g., Lu and Cheng (1997)) it follows from (5.3)$-$(5.5) and (5.9)$-$(5.10), together with (5.11), that for $i, j = 1, \ldots, d$,

$$\widehat{\Phi}_{i,j} = \bar{g}_{i-1,j-1}(z) f_Z(z) w(z)\mu_{0,K}(1 + o_P(1)), \tag{5.12}$$

$$\widehat{\Phi}_{i,d+j} = \widehat{\Phi}_{d+j,i} = \bar{g}_{i-1,j-1}(z) f_Z(z) w(z)\mu_{1,K}(1 + o_P(1)) = 0 \text{ (owing to } \mu_{1,K} = 0), \tag{5.13}$$

$$\widehat{\Phi}_{d+i,d+j} = \bar{g}_{i-1,j-1}(z) f_Z(z) w(z)\mu_{2,K}(1 + o_P(1)), \tag{5.14}$$

$$\widehat{W}_i = B_{i-1}(z)\mu_{2,K}h^2(1 + o_P(1)) + n^{-1}\sum_{t=1}^{n}\varepsilon_{t,\boldsymbol{\alpha}} X_{t,i-1} K_h(Z_t - z) w(Z_t), \tag{5.15}$$

$$\begin{aligned}\widehat{W}_{d+i} &= B_{i-1}(z)\mu_{3,K}h^2(1 + o_P(1)) + n^{-1}\sum_{t=1}^{n}\varepsilon_{t,\boldsymbol{\alpha}} X_{t,i-1}\frac{Z_t - z}{h} K_h(Z_t - z) w(Z_t)\\ &= n^{-1}\sum_{t=1}^{n}\varepsilon_{t,\boldsymbol{\alpha}} X_{t,i-1}\frac{Z_t - z}{h} K_h(Z_t - z) w(Z_t) \quad (\text{owing to } \mu_{3,K} = 0),\end{aligned} \tag{5.16}$$

where $B_{i-1}(z) = (1/2)\{\ddot{a}(z)\bar{g}_{i-1}(z) + \ddot{b}(z)^{\mathsf{T}} G_{i-1}(z)\} w(z) f_Z(z)$.

Now it follows from (5.12)−(5.14) that

$$\widehat{\Phi} = \begin{pmatrix} \mu_{0,K}\mathbf{G}(z) & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \mu_{2,K}\mathbf{G}(z) \end{pmatrix} w(z) f_Z(z)(1+o_P(1))$$
$$= (\mathbf{S}\otimes\mathbf{G}(z)) w(z) f_Z(z)(1+o_P(1)) \equiv \Phi(1+o_P(1)), \tag{5.17}$$

where $\mathbf{0}_{d\times d}$ is a $d\times d$ matrix of elements 0, and $\mathbf{G}(z)$ is a $d\times d$ matrix with (i,j)th element $G_{ij}(z) = \bar{g}_{i-1,j-1}(z)$ for $i,j = 1,\ldots,d$.

Recall that $\nu_{i,K} = \int u^i K^2(u)\, du$, and denote the second terms on the right hand sides of (5.15) and (5.16) by $\widehat{W}_{i,2}$ and $\widehat{W}_{i+d,2}$, respectively. Moreover, let $\widehat{W}_{c,2} = \sum_{i=1}^d (c_i \widehat{W}_{i,2} + c_{i+d}\widehat{W}_{i+d,2})$ for any real constants c_i. Then, under the assumptions of this theorem, we have

$$E(\widehat{W}_{c,2})^2 = E\left\{ n^{-1}\sum_{t=1}^n \varepsilon_{t,\boldsymbol{\alpha}} \sum_{i=1}^d \left(c_i X_{t,i-1} + c_{i+d} X_{t,i-1}\frac{Z_t - z}{h}\right) K_h(Z_t - z) w(Z_t)\right\}^2$$
$$= (nh)^{-1} V_c^2(z)(1+o(1)), \tag{5.18}$$

where

$$V_c^2(z) = \Big\{ \sum_{i=1}^d \sum_{j=1}^d \widetilde{G}_{i,j}(z)(c_i c_j \nu_{0,K} + c_{i+d}c_{j+d}\nu_{2,K})\Big\} w^2(z) f_Z(z,\boldsymbol{\alpha}) \equiv \mathbf{c}^{\mathsf{T}}\mathbf{V}^{(2)}(z)\mathbf{c}, \tag{5.19}$$

with $\mathbf{c} = (c_1,\ldots,c_d,c_{d+1},\ldots,c_{2d})^{\mathsf{T}}$, and

$$\mathbf{V}^{(2)}(z) = \begin{pmatrix} \nu_{0,K}\widetilde{\mathbf{G}}(z) & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \nu_{2,K}\widetilde{\mathbf{G}}(z) \end{pmatrix} w^2(z) f_Z(z,\boldsymbol{\alpha}) = (\widetilde{\mathbf{S}}\otimes\widetilde{\mathbf{G}}(z)) w^2(z) f_Z(z,\boldsymbol{\alpha}),$$

where $\widetilde{\mathbf{G}}(z)$ is a $d\times d$ matrix with (i,j)th element $\widetilde{G}_{i,j}(z) = E(\varepsilon^2_{t,\boldsymbol{\alpha}} X_{t,i-1}X_{t,j-1} \mid Z_t = z)$ for $i,j = 1,\ldots,d$. Therefore, it follows from (5.15),(5.16), (5.18) and (5.19) that

$$\widehat{\mathbf{W}} = \frac{1}{2}h^2\mathbf{U}(z)(1+o_P(1)) + \left(\frac{1}{nh}\right)^{\frac{1}{2}} \mathbf{V}(z)\xi_N(1+o_P(1)). \tag{5.20}$$

Here

$$\mathbf{U}(z) = \begin{pmatrix} \mu_{2,K}\mathbf{G}(z) \\ \mathbf{0}_{d\times d} \end{pmatrix} \begin{pmatrix} \ddot{a}(z) \\ \ddot{\mathbf{b}}(z) \end{pmatrix} w(z) f_Z(z,\boldsymbol{\alpha}) = (\mathbf{s}\otimes\mathbf{G}(z)) \begin{pmatrix} \ddot{a}(z) \\ \ddot{\mathbf{b}}(z) \end{pmatrix} w(z) f_Z(z,\boldsymbol{\alpha}),$$

$$\mathbf{V}(z) = \Big\{\mathbf{V}^{(2)}(z)\Big\}^{\frac{1}{2}} = \Big(\widetilde{\mathbf{S}}\otimes\widetilde{\mathbf{G}}(z)\Big)^{\frac{1}{2}} w(z) f_Z^{\frac{1}{2}}(z),$$

with $\mathbf{V}$ a $d\times d$ matrix such that $\mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{V}^{(2)}$. Here ξ_N is a $(2d)$-dimensional standard normal vector. The proof is routine because of the CLT for strong mixing processes based on the Bernstein blocking technique (see, e.g., Hallin, Lu and Tran (2004) Theorem 3.1, and Lu and Linton (2004)), and therefore the details are omitted.

Finally, (3.3) in Theorem 3.1 follows from (5.17) and (5.20) with

$$\begin{aligned}
\mathbf{B}(z) &= \Phi^{-1}(z)\mathbf{U}(z) = \left\{\left(\mathbf{S}^{-1}\mathbf{s}\right)\otimes I_{d\times d}\right\}(\ddot{a}(z), \ddot{\mathbf{b}}(z)^{\mathsf{T}})^{\mathsf{T}},\\
\mathbf{A}(z) &= \Phi^{-1}(z)\mathbf{V}(z)\left(\Phi^{-1}(z)\mathbf{V}(z)\Phi^{-1}(z)\right)^{\mathsf{T}} = \Phi^{-1}(z)\mathbf{V}^{(2)}(z)\Phi^{-1}(z)\\
&= \{f_Z(z)\}^{-1}\left(\mathbf{S}^{-1}\widetilde{\mathbf{S}}\mathbf{S}^{-1}\right)\otimes\left(\mathbf{G}^{-1}(z)\widetilde{\mathbf{G}}(z)\mathbf{G}^{-1}(z)\right).
\end{aligned}$$

When $\boldsymbol{\alpha}$ is replaced by $\check{\boldsymbol{\alpha}}$ with $\check{\boldsymbol{\alpha}}-\boldsymbol{\alpha} = O_P(n^{-1/2})$, then $\check{Z}_t = \check{\boldsymbol{\alpha}}^{\mathsf{T}}\mathbf{X}_t$ satisfies $\check{Z}_t - Z_t = (\check{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^{\mathsf{T}}\mathbf{X}_t = O_P(n^{-1/2})\mathbf{X}_t$. It is easily seen that the proof can be modified to prove the last statement of this theorem. The details are omitted.

5.2. Outline of proof of Theorem 3.2

The proof of Theorem 3.2 is technically involved. The key is in the application of Lemmas 4.1 and 4.2, for which we need to verify the various regularity conditions. To that end, we apply the uniform consistency results collected in Lemma A.1 in the Appendix below, and the empirical process theory for β-mixing processes due to Doukhan, Massart and Rio (1995, p.405). The detailed proof is given in Lu, Tjøstheim and Yao (2006).

Acknowledgement

We thank the Guest-Editor and the two referees for the very helpful comments and suggestions. This project was partially supported by a Leverhulme Trust research grant. Lu's work was also partially supported by the National Natural Science Foundation of China (70271003 and 70221001).

Appendix

Proof of Lemma 3.1 As $\boldsymbol{\alpha} \in \mathbb{B}$, we have the last component $|\alpha_d| \geq \epsilon_0 > 0$, with $|\alpha_j| \leq 1$ for $j = 1, \ldots, d$. Let $z = \boldsymbol{\alpha}^{\mathsf{T}}\mathbf{x} = \sum_{j=1}^{d}\alpha_j x_j$. Then it is easily seen that

$$\begin{aligned}
f_{\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t}(z) &= |\alpha_d|^{-1}\int\cdots\int f_{\mathbf{X}_t}\left(x_1, \ldots, x_{d-1}, \frac{z-\sum_{j=1}^{d-1}\alpha_j x_j}{\alpha_d}\right)dx_1\ \cdots dx_{d-1}\\
&\geq \int_{-\tilde{L}}^{\tilde{L}}\cdots\int_{-\tilde{L}}^{\tilde{L}} f_{\mathbf{X}_t}\left(x_1, \ldots, x_{d-1}, \frac{z-\sum_{j=1}^{d-1}\alpha_j x_j}{\alpha_d}\right)dx_1\ \cdots dx_{d-1}, \quad \text{(A.1)}
\end{aligned}$$

where $\tilde{L} = L\epsilon_0/d \leq L$. Note that when $x_i \in [-\tilde{L}, \tilde{L}]$ for $i = 1, \ldots, d-1$, and $z \in [-L, L]$,

$$\left| \frac{z - \sum_{j=1}^{d-1} \alpha_j x_j}{\alpha_d} \right| \leq L(1 + \epsilon_0^{-1}) = L_0,$$

and therefore $[-\tilde{L}, \tilde{L}]^{d-1} \times [-L_0, L_0] \subset [-L_0, L_0]^d$. Now by condition (C1′), $f_{\mathbf{X}_t}(\mathbf{x})$ is continuous and bounded away from zero on $[-L_0, L_0]^d$, and hence it easily follows that $f_{\boldsymbol{\alpha}^\intercal \mathbf{X}_t}(z)$ is bounded away from zero on $[-L, L]$. The second part of (C1) can be obtained by using similar equalities to (A.1), and the details are omitted.

A uniform convergence lemma

Lemma A.1. *Suppose the conditions of Theorem* 3.2 *hold, and let* $\lim_{n\to\infty} n^{2r+1}h^{3(d-1)} > 0$. *Then for* $\widehat{\Phi}_{i,j}(z, \boldsymbol{\alpha})$, $\widehat{\Phi}_{i+d,j+d}(z, \boldsymbol{\alpha})$ *and* $G_{ij}(z, \boldsymbol{\alpha}) = \bar{g}_{i-1,j-1}(z, \boldsymbol{\alpha})$ *as defined in Subsection* 5.1*,*

$$\sup_{z \in S_w, \boldsymbol{\alpha} \in \mathbb{B}} |\widehat{\Phi}_{i,j}(z, \boldsymbol{\alpha}) - \mu_{0,K} G_{ij}(z, \boldsymbol{\alpha}) w(z) f_Z(z, \boldsymbol{\alpha})| = O_P\left[\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}} + h^2\right], \tag{A.2}$$

$$\sup_{z \in S_w, \boldsymbol{\alpha} \in \mathbb{B}} |\widehat{\Phi}_{i+d,j+d}(z, \boldsymbol{\alpha}) - \mu_{2,K} G_{ij}(z, \boldsymbol{\alpha}) w(z) f_Z(z, \boldsymbol{\alpha})| = O_P\left[h^{-1}\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}} + h\right], \tag{A.3}$$

for $i, j = 1, \ldots, d$.

$$\sup_{z \in S_w, \boldsymbol{\alpha} \in \mathbb{B}} \|\widehat{\mathbf{g}}(z, \boldsymbol{\alpha}) - \mathbf{g}_0(z, \boldsymbol{\alpha})\| = O_P\left[\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}} + h^2\right], \tag{A.4}$$

$$\sup_{z \in S_w, \boldsymbol{\alpha} \in \mathbb{B}} \|\widehat{\mathbf{g}}_1(z, \boldsymbol{\alpha}) - \mathbf{g}_{01}(z, \boldsymbol{\alpha})\| = O_P\left[h^{-1}\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}} + h\right], \tag{A.5}$$

$$\sup_{z \in S_w, \boldsymbol{\alpha} \in \mathbb{B}} \|\widehat{\mathbf{g}}_2(z, \boldsymbol{\alpha}) - \mathbf{g}_{02}(z, \boldsymbol{\alpha})\| = O_P\left[h^{-1}\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}} + h\right]. \tag{A.6}$$

The proof of this lemma is given in Lu, Tjøstheim and Yao (2006).

References

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series models. *J. Amer. Statist. Assoc.* **95**, 888-902.

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.

Chan, R. and Tsay, R. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88**, 298-308.

Chan, X., Linton, O. and Keilegom, I. V. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71**, 1591-1608.

Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.* **31**, 393-427.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*. Chapman and Hall, London.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. Roy. Statist. Soc. Ser. B* **65**, 57-80.

Gourieroux, C. and Jasiak, J. (2001). *Financial Econometrics*. Princeton University, Princeton.

Hallin, M., Lu, Z. and Tran, L. T. (2004). Local linear spatial regression. *Ann. Statist.* **32**, 2469-2500.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157-178.

Hannan, E. (1970). *Multiple Time Series*. Wiley, New York.

Hong, Y. and Lee, T.-H. (2003). Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Rev. Econom. Statist.* **85**, 1048-1062.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58**, 71-120.

Lu, Z. and Cheng, P. (1997). Distribution-free strong consistency for nonparametric kernel regression involving nonlinear time series. *J. Statist. Plann. Inference* **65**, 67-86.

Lu, Z. and Linton, O. (2004). Local linear fitting for econometric nonlinear processes. Unpublished technical report, Department of Economics, London School of Economics and Political science. Available at http://www.lse.ac.uk/ olinton/ll.pdf.

Lu, Z., Steinskog, D.J., Tjøstheim, D. and Yao, Q. (2006). Adaptive varying coefficient spatio-temporal models. Manuscript.

Lu, Z., Tjøstheim, D. and Yao, Q. (2006). Additional proofs for "Adaptive varying-coefficient linear models for stochastic processes: asymptotic theory". Available at http://www3.stat.sinica.edu.tw/statistica/

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Xia, Y. and Li, W. K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94**, 1275-1285.

Department of Mathematics and Statistics, Curtin University of Technology, GPO Box U1987, Perth, WA6845, Australia.

Department of Statistics, London School of Economics, London, WC2A 2AE, U.K.

Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China.

E-mail: z.lu@curtin.edu.au

Department of Mathematics, University of Bergen, 5007 Bergen, Norway.

E-mail: dag.tjostheim@mi.uib.no

Department of Statistics, London School of Economics, Houghton Street, London, WC2A 2AE, U.K.

Guanghua School of Management, Peking University, Beijing 100871, China.

E-mail: q.yao@lse.ac.uk

(Received January 2006; accepted April 2006)

Statistica Sinica **17**(2007): Supplement, S19∼S35

ADAPTIVE VARYING-COEFFICIENT LINEAR MODELS FOR STOCHASTIC PROCESSES: ASYMPTOTIC THEORY

Zudi Lu[1,2], Dag Tjøstheim[3] and Qiwei Yao[1,4]

[1] *London School of Economics,* [2] *Chinese Academy of Sciences*
[3] *University of Bergen and* [4] *Peking University*

Supplementary Material

We provide the proofs for Theorem 3.2 and Lemma A.1 in "Adaptive Varying -Coefficient Linear Models for Stochastic Processes: Asymptotic Theory" by Lu, Tjøstheim and Yao.

A. Proof of Theorem 3.2

In this section, we are establishing in detail the asymptotics for $\widehat{\boldsymbol{\alpha}}$ defined in Section 2, using the two general lemmas developed in Section 4. We first derive the preliminary quantities used, under the adaptive varying-coefficient modelling of (2.1).

A.1. Derivation of preliminary quantities

We define, for some $c_0 > 0$,

$$\begin{aligned}\mathcal{G} = \{\mathbf{g} : S_w \times \mathbb{B} \mapsto \mathbb{R}^d \mid \; &\text{For any fixed } \boldsymbol{\alpha} \in \mathbb{B},\ \mathbf{g}(\cdot,\boldsymbol{\alpha}) \in C^2_{c_0}(S_w),\ \mathbf{g}_1(\cdot,\boldsymbol{\alpha}) \in \\ &C^1_{c_0}(S_w) \text{ and } \mathbf{g}_2(\cdot,\boldsymbol{\alpha}) \in C^1_{c_0}(S_w), \text{ and for any } z \in S_w,\ \|\mathbf{g}(z,\boldsymbol{\alpha}) - \\ &\mathbf{g}(z,\boldsymbol{\alpha}')\| \le C\|\boldsymbol{\alpha}-\boldsymbol{\alpha}'\|, \|\mathbf{g}_1(z,\boldsymbol{\alpha}) - \mathbf{g}_1(z,\boldsymbol{\alpha}')\| \le C\|\boldsymbol{\alpha}-\boldsymbol{\alpha}'\| \text{ and} \\ &\|\mathbf{g}_2(z,\boldsymbol{\alpha}) - \mathbf{g}_2(z,\boldsymbol{\alpha}')\| \le C\|\boldsymbol{\alpha}-\boldsymbol{\alpha}'\| \text{ for any } \boldsymbol{\alpha},\boldsymbol{\alpha}' \in \mathbb{B}\}, \qquad \text{(A.1)}\end{aligned}$$

where $C^j_{c_0}(S_w)$ for $j = 1$ and 2 was defined at the end of Subsection 3.1.

As defined in (2.9),

$$\begin{aligned}R(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) &= E\left(Y_t - \mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t,\boldsymbol{\alpha})^\top \mathbb{X}_t\right)^2 w(\boldsymbol{\alpha}^\top \mathbf{X}_t) \\ &= \int \left(y - a(\boldsymbol{\alpha}^\top \mathbf{x},\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}^\top \mathbf{x},\boldsymbol{\alpha})^\top \breve{\mathbf{x}}\right)^2 w(\boldsymbol{\alpha}^\top \mathbf{x}) f_{Y,\mathbf{X}}(y,\mathbf{x})dyd\mathbf{x},\end{aligned}$$

where $\breve{\mathbf{x}}$ is the $(d-1)$-dimensional vector obtained by deleting the d-th component of $\mathbf{x}$, and $f_{Y,\mathbf{X}}(y,\mathbf{x})$ is the joint probability density function of $(Y_t, \mathbf{X}_t)$; and

$$\begin{aligned}R_n(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) &= \frac{1}{n}\sum_{t=1}^{n}\Big(Y_t - \mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t\Big)^2 w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\\ &= \frac{1}{n}\sum_{t=1}^{n}\Big(Y_t - a(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\breve{\mathbf{X}}_t\Big)^2 w(\boldsymbol{\alpha}^\top\mathbf{X}_t).\end{aligned}$$

Then we can deduce (the notation having been explained in section 4.2, for simplicity, we assume $dw(z)/dz = 0$)

$$\begin{aligned}\dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) =& -2E\Big(Y_t - \mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t\Big)\Big\{\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top + \mathbf{g}_2(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\Big\}^\top\\ &\times\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t),\\ \dot{R}_n(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) =& -\frac{2}{n}\sum_{t=1}^{n}\Big(Y_t - \mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t\Big)\Big\{\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top + \mathbf{g}_2(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\Big\}^\top\\ &\times\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t). \qquad\qquad (A.2)\end{aligned}$$

Therefore, by Lemma 5.1, the ordinary derivative of $R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$,

$$\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) = -2E\Big(Y_t - \mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t\Big)\Big\{\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top\Big\}^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t),$$

and further the derivative of $\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ equals (as assumed in condition (C4), the derivative and the expectation are exchangeable)

$$\begin{aligned}&\Gamma_1(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\\ &\quad= 2E\Big[\Big\{\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top + \mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\Big\}^\top\mathbb{X}_t\mathbb{X}_t^\top\\ &\qquad\quad\times\Big\{\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top + \mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\Big\}\Big] w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\\ &\qquad - 2E\Big(Y_t - \mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t\Big)\Big\{\mathbf{g}_{0,11}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t\mathbf{X}_t^\top + \mathbf{g}_{0,12}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top\Big\}^\top\\ &\qquad\quad\times\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t),\end{aligned}$$

where $\mathbf{g}_{0,11}(z,\boldsymbol{\alpha}) = \partial\mathbf{g}_{01}(z,\boldsymbol{\alpha})/\partial z$ and $\mathbf{g}_{0,12}(z,\boldsymbol{\alpha}) = \partial\mathbf{g}_{01}(z,\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$. As $\varepsilon_{t,\boldsymbol{\alpha}_0} = Y_t - \mathbf{g}_0(Z_t^o,\boldsymbol{\alpha}_0)^\top\mathbb{X}_t = \varepsilon_t$, as assumed, satisfies $E(\varepsilon_t|\mathbf{X}_t) = 0$, hence

$$\begin{aligned}\Gamma_1 &= \Gamma_1(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\\ &= 2E\Big[\Big\{\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top + \mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\Big\}^\top\mathbb{X}_t\mathbb{X}_t^\top\\ &\qquad\times\Big\{\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top + \mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\Big\}\Big] w(Z_t^o). \qquad\qquad (A.3)\end{aligned}$$

Furthermore, note that, by Lemma 5.1 with some algebraic calculations,

$$
\begin{aligned}
&\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha})+\tau(\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})),\boldsymbol{\alpha})-\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\\
&\quad=-2E\left\{Y_t-\Big(\mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})+\tau(\mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\Big)^\top\mathbb{X}_t\right\}\\
&\qquad\times\left\{\Big(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})+\tau(\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\Big)\mathbf{X}_t^\top\right.\\
&\qquad+\left.\Big(\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})+\tau(\mathbf{g}_2(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\Big)\right\}^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\\
&\qquad+2E\left\{Y_t-\mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t\right\}\left\{\Big(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top\Big)+\Big(\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\Big)\right\}^\top\\
&\qquad\times\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\\
&\quad=-2\tau E\left\{\varepsilon_{t,\boldsymbol{\alpha}}\Big((\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\mathbf{X}_t^\top\Big)\right.\\
&\qquad-\left.\Big((\mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))^\top\mathbb{X}_t\Big)\Big(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top+\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\Big)\right\}^\top\\
&\qquad\times\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t)+2\tau^2E\left\{(\mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))^\top\mathbb{X}_t\right\}\\
&\qquad\times\Big((\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\mathbf{X}_t^\top+(\mathbf{g}_2(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\Big)^\top\\
&\qquad\times\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t).
\end{aligned}\tag{A.4}
$$

Therefore the functional derivative of $\dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ with respect to $\mathbf{g}(\cdot,\boldsymbol{\alpha})$ at $\mathbf{g}_0(\cdot,\boldsymbol{\alpha})$ in the direction $\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})$ satisfies

$$
\begin{aligned}
&\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)[\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})]\\
&\quad=\lim_{\tau\to0}\frac{\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha})+\tau(\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})),\boldsymbol{\alpha})-\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})}{\tau}\\
&\quad=-2E\left\{\varepsilon_{t,\boldsymbol{\alpha}}\Big((\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\mathbf{X}_t^\top\Big)\right.\\
&\qquad-\left.\Big((\mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))^\top\mathbb{X}_t\Big)\Big(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top+\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\Big)\right\}^\top\\
&\qquad\times\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t),
\end{aligned}\tag{A.5}
$$

and therefore

$$
\begin{aligned}
&\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)[\mathbf{g}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]\\
&\quad=2E\Big((\mathbf{g}(Z_t^o,\boldsymbol{\alpha}_0)-\mathbf{g}_0(Z_t^o,\boldsymbol{\alpha}_0))^\top\mathbb{X}_t\Big)\Big(\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top+\mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\Big)^\top\mathbb{X}_t w(Z_t^o)\\
&\quad=2\int\widetilde{\Phi}_0(z)\left(\mathbf{g}(z,\boldsymbol{\alpha}_0)-\mathbf{g}_0(z,\boldsymbol{\alpha}_0)\right)w(z)f_{Z_t^o}(z)dz,
\end{aligned}\tag{A.6}
$$

where $\widetilde{\Phi}_0(z)=E\left\{\left(\mathbf{g}_{01}(z,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top+\mathbf{g}_{02}(z,\boldsymbol{\alpha}_0)\right)^\top\mathbb{X}_t\mathbb{X}_t^\top\,\Big|\,Z_t^o=z\right\}$.

Next, we are establishing in detail the consistency of $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}_0$ by Lemma 4.1.

A.2. Derivation of consistency of $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}_0$

The consistency of $\widehat{\boldsymbol{\alpha}}$ can be proved by checking the conditions in Lemma 4.1 step by step: As $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$ are the minimizers of $R_n(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ and $R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$, respectively, (i) and (ii) hold obviously. (iii) also holds clearly by the following fact: noting Lemma 5.1 as well as the boundedness of $w(\cdot)$,

$$
\begin{aligned}
&\sup_{\boldsymbol{\alpha}\in\mathbb{B}}\Big|R(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})-R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\Big|\\
&\quad\le \sup_{\boldsymbol{\alpha}\in\mathbb{B}}\Big|E\left(2\varepsilon_{t,\boldsymbol{\alpha}}-(\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t\right)\\
&\qquad\qquad\times(\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t w(\boldsymbol{\alpha}^{\top}\mathbf{X}_t)\Big|\\
&\quad\le \sup_{\boldsymbol{\alpha}\in\mathbb{B}}\Big|E\left((\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t\right)\\
&\qquad\qquad\times(\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t w(\boldsymbol{\alpha}^{\top}\mathbf{X}_t)\Big|\\
&\quad\le C\|\mathbf{g}-\mathbf{g}_0\|_{\mathcal{G}}\|E\mathbb{X}_t\mathbb{X}_t^{\top}\|,
\end{aligned}
\tag{A.7}
$$

where the final inequality follows from the definition of norm $\|\cdot\|_{\mathcal{G}}$ in Subsection 3.1. (iv) follows clearly from Lemma A.1 in the Appendix. For (v), letting $\delta_n=o(1)$ and $\|\mathbf{g}-\mathbf{g}_0\|_{\mathcal{G}}\le\delta_n$, we notice that

$$
\begin{aligned}
&R_n(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})-R(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\\
&\quad=\{R_n(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})-R_n(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\}+\{R_n(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})-R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\}\\
&\qquad+\{R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})-R(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\}\\
&\quad=I+II+III,
\end{aligned}
$$

where by (A.7) III tends to 0, uniformly for $\boldsymbol{\alpha}\in\mathbb{B}$ and with $\mathbf{g}$ satisfying $\|\mathbf{g}-\mathbf{g}_0\|_{\mathcal{G}}\le\delta_n$. That I tends to 0, uniformly for $\boldsymbol{\alpha}\in\mathbb{B}$ and $\mathbf{g}$ with $\|\mathbf{g}-\mathbf{g}_0\|_{\mathcal{G}}\le\delta_n$, can be proved in the same way as for III, because in fact $E[I]=III$; II can also be proved easily to tend to zero.

Finally, we are finishing the proof by Lemma 4.2.

A.3. Derivation of asymptotic normality of $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}_0$

As we have proved that $\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0=o_P(1)$, and from Lemma A.1 in the Appendix, $\|\widehat{\mathbf{g}}-\mathbf{g}_0\|_{\mathcal{G}}=o_P(1)$ as well as $\|\widehat{\mathbf{g}}_1-\mathbf{g}_{01}\|_{\mathcal{G}}=o_P(1)$ and $\|\widehat{\mathbf{g}}_2-\mathbf{g}_{02}\|_{\mathcal{G}}=o_P(1)$, we can assume that $\boldsymbol{\alpha}$ and $\mathbf{g}=(a,\mathbf{b}^{\top})^{\top}$ lie in $\mathbb{B}_\delta$ and $\mathcal{G}_\delta$, respectively, with $\delta = \delta_n \to 0$, where

$$\mathbb{B}_\delta = \{\boldsymbol{\alpha} \in \mathbb{B} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \le \delta\},$$
$$\mathcal{G}_\delta = \{\mathbf{g} \in \mathcal{G} : \|\mathbf{g} - \mathbf{g}_0\|_{\mathcal{G}} \le \delta, \|\mathbf{g}_1 - \mathbf{g}_{01}\|_{\mathcal{G}} \le \delta, \|\mathbf{g}_2 - \mathbf{g}_{02}\|_{\mathcal{G}} \le \delta\}. \tag{A.8}$$

As $\boldsymbol{\alpha}_0$ is the minimizers of $R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ which is differentiable with respect to $\boldsymbol{\alpha}$, $\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) = 0$.

We proceed to check the conditions (i)−(vi) in Lemma 4.2:

(i) This is clear, as $\widehat{\boldsymbol{\alpha}}$ is the minimizers of $R_n(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})$ which is differentiable with respect to $\boldsymbol{\alpha}$, and hence $\dot{R}_n(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\alpha}}) = 0$.

(ii) Both (ii)(1)−(2) are clear from Assumption (C4) in Section 3.

(iii) It follows from (A.4) with $\tau = 1$ and (A.6) that

$$\begin{aligned}
&\mathbf{c}^{\mathsf{T}}\left\{\dot{R}(\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)\left[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})\right]\right\}\\
&=2\mathbf{c}^{\mathsf{T}}E\left\{(\widehat{\mathbf{g}}(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{g}_0(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}))^{\mathsf{T}}\mathbb{X}_t\right\}\\
&\quad\times\left((\widehat{\mathbf{g}}_1(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{g}_{01}(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}))\mathbf{X}_t^{\mathsf{T}} + (\widehat{\mathbf{g}}_2(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{g}_{02}(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}))\right)^{\mathsf{T}}\\
&\quad\times\mathbb{X}_t w(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t)\\
&=2\int(\widehat{\mathbf{g}}(z,\boldsymbol{\alpha}) - \mathbf{g}_0(z,\boldsymbol{\alpha}))^{\mathsf{T}}E\{\mathbb{X}_t\mathbf{c}^{\mathsf{T}}\mathbf{X}_t\mathbb{X}_t^{\mathsf{T}}|Z_t(\boldsymbol{\alpha}) = z\}(\widehat{\mathbf{g}}_1(z,\boldsymbol{\alpha}) - \mathbf{g}_{01}(z,\boldsymbol{\alpha}))\\
&\qquad\times w(z)f_Z(z,\boldsymbol{\alpha})dz + 2\int(\widehat{\mathbf{g}}(z,\boldsymbol{\alpha}) - \mathbf{g}_0(z,\boldsymbol{\alpha}))^{\mathsf{T}}E\{\mathbb{X}_t\mathbb{X}_t^{\mathsf{T}}|Z_t(\boldsymbol{\alpha}) = z\}\\
&\qquad\times(\widehat{\mathbf{g}}_2(z,\boldsymbol{\alpha}) - \mathbf{g}_{02}(z,\boldsymbol{\alpha}))\mathbf{c}w(z)f_Z(z,\boldsymbol{\alpha})dz\\
&\quad\equiv\widetilde{D}_1 + \widetilde{D}_2,
\end{aligned}\tag{A.9}$$

from which (iii) (1) can be deduced as follows.

Set $\gamma = (I_{d\times d}, \mathbf{0}_{d\times d})$ a $d \times (2d)$ matrix. Note that it follows from (5.8), the uniform consistency lemma (Lemma A.1) and then (5.17) that

$$\begin{aligned}
\widehat{\mathbf{g}}(z,\boldsymbol{\alpha}) - \mathbf{g}_0(z,\boldsymbol{\alpha}) &= \gamma(\widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha}) - \boldsymbol{\theta}_0(z,\boldsymbol{\alpha}))\\
&= \gamma\widehat{\Phi}^{-1}\widehat{\mathbf{W}} = (1 + o_P(1))\gamma\Phi^{-1}\widehat{\mathbf{W}}\\
&= (1 + o_P(1))(\mu_{0,K}w(z)f_Z(z,\boldsymbol{\alpha}))^{-1}\mathbf{G}^{-1}(z,\boldsymbol{\alpha})\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}),
\end{aligned}\tag{A.10}$$

where $o_P(1)$ is uniform with respect to $z \in S_w$ and $\boldsymbol{\alpha} \in \mathbb{B}$, and $\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha})$ is the vector consisting of the first d components of $\widehat{\mathbf{W}}$ defined in (5.9), that is

$$\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}) = n^{-1}\sum_{t=1}^{n} Y_t^*(z,\boldsymbol{\alpha})\mathbb{X}_t K_h(Z_t(\boldsymbol{\alpha}), -z)w(Z_t(\boldsymbol{\alpha})), \tag{A.11}$$

where $Y_t^*(z,\boldsymbol{\alpha})$ is as defined in (5.11) in the notation of this section as

$$Y_t^*(z,\boldsymbol{\alpha}) = \varepsilon_{t,\boldsymbol{\alpha}} + (\mathbf{g}_0(Z_t(\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \mathbf{g}_0(z,\boldsymbol{\alpha}) - \dot{\mathbf{g}}_0(z,\boldsymbol{\alpha})(Z_t(\boldsymbol{\alpha})-z))^\mathsf{T}\,\mathbb{X}_t$$
$$= \varepsilon_{t,\boldsymbol{\alpha}} + \frac{1}{2}\left(\ddot{\mathbf{g}}_0(z+\eta_1(Z_t(\boldsymbol{\alpha})-z),\boldsymbol{\alpha})(Z_t(\boldsymbol{\alpha})-z)^2\right)^\mathsf{T}\mathbb{X}_t \qquad \text{(A.12)}$$

with $|\eta_1|<1$. Thus, setting $G_{\mathbf{c}}(z,\boldsymbol{\alpha}) = E\{\mathbb{X}_t\mathbf{c}^\mathsf{T}\mathbf{X}_t\mathbb{X}_t^\mathsf{T}|Z_t(\boldsymbol{\alpha})=z\}$, with uniformity of $o_P(1)$ with respect to $z\in S_w$ and $\boldsymbol{\alpha}\in\mathbb{B}$ in (A.10), together with (A.12)

$$\widetilde{D}_1 = 2\int((1+o_P(1))(\mu_{0,K}w(z)f_Z(z,\boldsymbol{\alpha}))^{-1}\mathbf{G}^{-1}(z,\boldsymbol{\alpha})\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}))^\mathsf{T}G_{\mathbf{c}}(z,\boldsymbol{\alpha})$$
$$\times(\widehat{\mathbf{g}}_1(z,\boldsymbol{\alpha})-\mathbf{g}_{01}(z,\boldsymbol{\alpha}))w(z)f_Z(z,\boldsymbol{\alpha})dz$$
$$= 2(1+o_P(1))(\mu_{0,K})^{-1}\int\mathbf{G}^{-1}(z,\boldsymbol{\alpha})\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}))^\mathsf{T}$$
$$\times G_{\mathbf{c}}(z,\boldsymbol{\alpha})(\widehat{\mathbf{g}}_1(z,\boldsymbol{\alpha})-\mathbf{g}_{01}(z,\boldsymbol{\alpha}))dz$$
$$= \frac{1}{\sqrt{n}}(\nu_n(\widehat{\mathbf{g}}_1,\boldsymbol{\alpha})-\nu_n(\mathbf{g}_{01},\boldsymbol{\alpha})) + O_P(h^2)\|\widehat{\mathbf{g}}_1-\mathbf{g}_{01}\|_{\mathcal{G}}, \qquad \text{(A.13)}$$

where $\nu_n(\mathbf{g}_1,\boldsymbol{\alpha}) = n^{-1/2}\sum_{i=1}^n\varepsilon_{t,\boldsymbol{\alpha}}\,\mathbb{X}^\mathsf{T}\,w(Z_t(\boldsymbol{\alpha}))\,\mathbf{G}^{-1}(Z_t(\boldsymbol{\alpha}),\boldsymbol{\alpha})\,G_{\mathbf{c}}(Z_t(\boldsymbol{\alpha}),\boldsymbol{\alpha})\mathbf{g}_1(Z_t(\boldsymbol{\alpha}),\boldsymbol{\alpha})$. Using the empirical process techniques, similarly to the proof of (v) below, we can show the stochastic equicontinuity of $\nu_n(\mathbf{g}_1,\boldsymbol{\alpha})$, and hence $n^{-1/2}(\nu_n(\widehat{\mathbf{g}}_1,\boldsymbol{\alpha})-\nu_n(\mathbf{g}_{01},\boldsymbol{\alpha})) \le n^{-1/2}\sup_{\|\mathbf{g}_1-\mathbf{g}_{01}\|_{\mathcal{G}}\le\delta}\|\nu_n(\mathbf{g}_1,\boldsymbol{\alpha})-\nu_n(\mathbf{g}_{01},\boldsymbol{\alpha})\| = o_P(n^{-1/2})$; for detail, see the proof of (v) below as the proof there is more complex. Also, as $nh^4=O(1)$ is assumed as in a condition in Theorem 3.2 and $\|\widehat{\mathbf{g}}_1-\mathbf{g}_{01}\|_{\mathcal{G}} = o_P(1)$, we have $O_P(h^2)\|\widehat{\mathbf{g}}_1-\mathbf{g}_{01}\|_{\mathcal{G}} = o_P(n^{-1/2})$. Therefore $\widetilde{D}_1 = o_P(n^{-1/2})$. Similarly, we can prove $\widetilde{D}_2 = o_P(n^{-1/2})$, and thus (iii)(1) follows from (A.9).

In addition, it follows from (A.6) together with Lemma 5.1 and $E(\varepsilon_{t,\boldsymbol{\alpha}_0}|\mathbf{X}_t)=0$ that

$$\Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})[\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})] - \Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)[\mathbf{g}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)]$$
$$= -2E(\varepsilon_{t,\boldsymbol{\alpha}}-\varepsilon_{t,\boldsymbol{\alpha}_0})\left((\mathbf{g}_1(\boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{01}(\boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t,\boldsymbol{\alpha}))\mathbf{X}_t^\mathsf{T}\right)^\mathsf{T}\mathbb{X}_t w(\boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t)$$
$$+ 2E\left\{\tilde{\delta}_t(\boldsymbol{\alpha})\left(\mathbf{g}_{01}(\boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\mathsf{T}+\mathbf{g}_{02}(\boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t,\boldsymbol{\alpha})\right)^\mathsf{T}\mathbb{X}_t w(\boldsymbol{\alpha}^\mathsf{T}\mathbf{X}_t)\right.$$
$$\left. -\tilde{\delta}_t(\boldsymbol{\alpha}_0)\left(\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\mathsf{T}+\mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\right)^\mathsf{T}\mathbb{X}_t w(Z_t^o)\right\}$$
$$\equiv \Omega_1+\Omega_2,$$

where $\tilde{\delta}_t(\boldsymbol{\alpha}) = \mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})$. By condition (C4),

$$|\varepsilon_t(\boldsymbol{\alpha}) - \varepsilon_{t,\boldsymbol{\alpha}_0}| = |\left(\mathbf{g}_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_0(Z_t^o, \boldsymbol{\alpha}_0)\right)^\top \mathbb{X}_t| \\ \leq C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|(1 + \|\mathbf{X}_t\|)\|\mathbb{X}_t\|,$$

and $\|\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_{01}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\|_{\mathcal{G}} = o(1)$, therefore $\Omega_1 = o(1)\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|$. For Ω_2, it is obvious by condition (C4) and $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| = o(1)$ that

$$\begin{aligned}\tilde{\delta}_t(\boldsymbol{\alpha}) - \tilde{\delta}_t(\boldsymbol{\alpha}_0) &= \mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}(Z_t^o, \boldsymbol{\alpha}_0) - \left(\mathbf{g}_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_0(Z_t^o, \boldsymbol{\alpha}_0)\right) \\ &= \mathbf{g}_1(Z_t^o, \boldsymbol{\alpha}_0)(1 + o(1))(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^\top \mathbf{X}_t + \mathbf{g}_2(Z_t^o, \boldsymbol{\alpha}_0)^\top (1 + o(1))(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \\ &\quad - \left(\mathbf{g}_{01}(Z_t^o, \boldsymbol{\alpha}_0)(1 + o(1))(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^\top \mathbf{X}_t + \mathbf{g}_{02}(Z_t^o, \boldsymbol{\alpha}_0)^\top (1+o(1))(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)\right) \\ &= o_P(1)\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|,\end{aligned}$$

which follows from $\|\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_{01}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\|_{\mathcal{G}} = o_P(1)$ and $\|\mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_{02}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\|_{\mathcal{G}} = o_P(1)$; and

$$\Omega_3 \equiv \left(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\mathbf{X}_t^\top + \mathbf{g}_{02}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\right) - \left(\mathbf{g}_{01}(Z_t^o, \boldsymbol{\alpha}_0)\mathbf{X}_t^\top + \mathbf{g}_{02}(Z_t^o, \boldsymbol{\alpha}_0)\right) \\ \leq C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|(1 + \|\mathbf{X}_t\|).$$

Therefore it easily follows that $\Omega_2 = o_P(1)\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|$. Hence (iii)(2) follows.

(iv) It is clear from the uniform convergence lemma, Lemma A.1, that

$$\|\widehat{\mathbf{g}} - \mathbf{g}_0\|_{\mathcal{G}} = O_P\left[\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}}\right] + O(h^2),$$

$$\|\widehat{\mathbf{g}}_1 - \mathbf{g}_{10}\|_{\mathcal{G}} = O_P\left[h^{-1}\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}}\right] + O(h),$$

$$\|\widehat{\mathbf{g}}_2 - \mathbf{g}_{20}\|_{\mathcal{G}} = O_P\left[h^{-1}\left(nh^{1+\frac{2d}{r}}\right)^{-\frac{r}{2r+d}}\right] + O(h),$$

and therefore $\|\widehat{\mathbf{g}} - \mathbf{g}_0\|_{\mathcal{G}} \to 0$, $\|\widehat{\mathbf{g}}_1 - \mathbf{g}_{10}\|_{\mathcal{G}} \to 0$, and $\|\widehat{\mathbf{g}}_2 - \mathbf{g}_{20}\|_{\mathcal{G}} \to 0$ if $nh^{3+3d/r} \to \infty$ with $r > 3d$ as $n \to \infty$. Hence (iv) follows.

(v) For notational convenience, let $F_t = (Y_t, \mathbf{X}_t)$, $m(F_t, \mathbf{g}, \boldsymbol{\alpha}) = m_{1t}(\mathbf{g}, \boldsymbol{\alpha})m_{2t}(\mathbf{g}, \boldsymbol{\alpha})m_{3t}(\boldsymbol{\alpha})$ with $m_{1t}(\mathbf{g}, \boldsymbol{\alpha}) = Y_t - \mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbb{X}_t$, $m_{2t}(\mathbf{g}, \boldsymbol{\alpha}) = \{\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbf{X}_t - \mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\}^\top$ and $m_{3t}(\boldsymbol{\alpha}) = \mathbb{X}_t w(\boldsymbol{\alpha}^\top \mathbf{X}_t)$, and define the empirical process

$$\nu_n(\mathbf{g}, \boldsymbol{\alpha}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{m(F_t, \mathbf{g}, \boldsymbol{\alpha}) - Em(F_t, \mathbf{g}, \boldsymbol{\alpha})\}.$$

Then it is obvious that

$$\dot{R}_n(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) = -\frac{2}{\sqrt{n}}\nu_n(\mathbf{g},\boldsymbol{\alpha}),$$

and as $\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) = 0$, we clearly have

$$\begin{aligned}&\dot{R}_n(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}) - \dot{R}_n(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\\ &\quad= -\frac{2}{\sqrt{n}}\{\nu_n(\mathbf{g},\boldsymbol{\alpha}) - \nu_n(\mathbf{g}_0,\boldsymbol{\alpha}_0)\}.\end{aligned}$$

Therefore for (v), it suffices to prove the stochastic equicontinuity of the empirical process $\{\nu_n(\mathbf{g},\boldsymbol{\alpha}):\ \mathbf{g}\in\mathcal{G}_1,\ \mathbf{b}\in\mathbb{B}_1\}$, where $\mathbb{B}_1$ and $\mathcal{G}_1$ are defined in (A.8) with $\delta = 1$, which are subsets of $\mathbb{B}$ and $\mathcal{G}$, respectively, and suffices for our proof of (v) as $\delta_n < 1$ for n large enough by $\delta_n \to 0$. This stochastic equicontinuity follows by checking the following conditions, due to Doukhan, Massart and Rio (1995, p.405):

(a) $\{F_t:\ t\geq 1\}$ is a stationary absolutely regular sequence with mixing coefficient $\beta(s)\leq Cs^{-b}$ for some $b > r/(r-1)$ and some $r > 1$,

(b) $E[\widetilde{m}^{2r}(F_t)] < \infty$ for r as in (a), where $\widetilde{m}(\cdot)$ is the envelope of $\mathcal{M} = \{m(\cdot,\mathbf{g},\boldsymbol{\alpha}) : \mathbf{g}\in\mathcal{G}_1,\ \boldsymbol{\alpha}\in\mathbb{B}_1\}$, that is $|m(\cdot,\mathbf{g},\boldsymbol{\alpha})|\leq|\widetilde{m}(\cdot)|$ for any $\mathbf{g}\in\mathcal{G}_1,\ \boldsymbol{\alpha}\in\mathbb{B}_1$.

(c) For any $\varepsilon > 0$, $\log N_2(\varepsilon,\mathcal{M})\leq C\varepsilon^{-2\eta}$ for some $\eta > 0$, with $b(1-\eta) > r/(r-1)$ for r as in (a), where $N_2(\varepsilon,\mathcal{M})$ is the $\mathcal{L}_2$-bracketing cover number of $\mathcal{M}$ in (b).

We check those conditions as follows. Here, (a) holds by the condition (C5). To show (b), notice that for $\boldsymbol{\alpha}\in\mathbb{B}_1$ and $\mathbf{g}\in\mathcal{G}_1$, we have $\|\boldsymbol{\alpha}\|\leq\|\boldsymbol{\alpha}_0\|+1\equiv C_0$, $\|\mathbf{g}\|_{\mathcal{G}}\leq\|\mathbf{g}_0\|_{\mathcal{G}}+1\equiv C_1$, $\|\mathbf{g}_1\|_{\mathcal{G}}\leq\|\mathbf{g}_{01}\|_{\mathcal{G}}+1\equiv C_2$ and $\|\mathbf{g}_2\|_{\mathcal{G}}\leq\|\mathbf{g}_{02}\|_{\mathcal{G}}+1\equiv C_3$, and therefore for $m\in\mathcal{M}$, $|m(F_t,\mathbf{g},\boldsymbol{\alpha})|\leq(|Y_t|+C_1\|\mathbb{X}\|)(C_2\|\mathbf{X}_t\|+C_3)\|\mathbb{X}\|w_0$, where $w_0 = \sup_{z\in S_w} w(z)$. So we can take $\widetilde{m}(F_t) = (|Y_t|+C_1\|\mathbb{X}\|)(C_2\|\mathbf{X}_t\| + C_3)\|\mathbb{X}\|w_0$, and hence (b) holds by condition (C2). Finally for (c), as $\mathbb{B}$ is a bounded subset in $\mathbb{R}^d$, for any $\varepsilon > 0$, we can cover $\mathbb{B}$ by finite number, $N_1 = C\varepsilon^{-(d-1)}$, of balls of radius ε with centers $\boldsymbol{\alpha}_j$, $j = 1,\ldots,N_1$, in $\mathbb{R}^d$, say, $\mathbb{B}_j$, $j = 1,\ldots,N_1$, such that

$$\forall\,\boldsymbol{\alpha}\in\mathbb{B},\ \exists\boldsymbol{\alpha}_j,\ \text{such that}\ \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_j\|\leq\varepsilon/(2C). \tag{A.14}$$

Then for each given $\boldsymbol{\alpha}_j$ and for $\mathbf{g}\in\mathcal{G}$, by the definition of $\mathcal{G}$ in this section, $\mathbf{g}(\cdot,\boldsymbol{\alpha}_j)\in C^2_{c_0}(S_w)$, and $\mathbf{g}_1(\cdot,\boldsymbol{\alpha}_j)\in C^1_{c_0}(S_w)$ and $\mathbf{g}_2(\cdot,\boldsymbol{\alpha}_j)\in C^1_{c_0}(S_w)$. Therefore, with the norm imposed on $C^2_{c_0}(S_w)$ by the sup norm $\|\mathbf{g}\|_\infty = \sup_{z\in S_w}\|\mathbf{g}(z)\|$ for $\mathbf{g}\in C^2_{c_0}(S_w)$, and similarly for $C^1_{c_0}(S_w)$, it is well known (c.f., van der Vaart and Wellner, 1996, Theorem 2.7.1) that we can cover $C^2_{c_0}(S_w)$ by finite number $N_2 =$

$N(\varepsilon, C_{c_0}^2(S_w), \|\cdot\|_\infty)$, of balls of functions centered at, say, $\mathbf{g}^{\ell,j}(\cdot)$, $\ell = 1, \ldots, N_2$, in $C_{c_0}^2(S_w)$, such that

$$\log N(\varepsilon, C_{c_0}^2(S_w), \|\cdot\|_\infty) \leq \text{const.} \times \varepsilon^{-1/2},$$

and

$$\forall\, \mathbf{g}(\cdot, \boldsymbol{\alpha}_j) \in C_{c_0}^2(S_w),\ \exists \mathbf{g}^{\ell,j}(\cdot),\ \text{such that } \|\mathbf{g}(\cdot, \boldsymbol{\alpha}_j) - \mathbf{g}^{\ell,j}(\cdot)\| \leq \varepsilon.$$

Similarly $C_{c_0}^1(S_w)$ can be covered by a finite number $N_3 = N(\varepsilon, C_{c_0}^1(S_w), \|\cdot\|_\infty)$, balls of functions centered at $\mathbf{g}_1^{\ell,j}(\cdot)$ and $\mathbf{g}_2^{\ell,j}(\cdot)$, respectively, $\ell = 1, \ldots, N_3$, in $C_{c_0}^1(S_w)$, such that

$$\log N(\varepsilon, C_{c_0}^1(S_w), \|\cdot\|_\infty) \leq \text{const.} \times \varepsilon^{-1},$$

with

$$\forall\, \mathbf{g}_1(\cdot, \boldsymbol{\alpha}_j) \in C_{c_0}^1(S_w),\ \exists\ \mathbf{g}_1^{\ell,j}(\cdot),\ \text{such that } \|\mathbf{g}_1(\cdot, \boldsymbol{\alpha}_j) - \mathbf{g}_1^{\ell,j}(\cdot)\| \leq \varepsilon,$$

and

$$\forall\, \mathbf{g}_2(\cdot, \boldsymbol{\alpha}_j) \in C_{c_0}^1(S_w),\ \exists\ \mathbf{g}_2^{\ell,j}(\cdot),\ \text{such that } \|\mathbf{g}_2(\cdot, \boldsymbol{\alpha}_j) - \mathbf{g}_2^{\ell,j}(\cdot)\| \leq \varepsilon.$$

Thus we can cover $\mathcal{G}_1 \subset \mathcal{G}$ by finite number of $N_1 N_2$ balls of centers $\mathbf{g}^{\ell,j}(\cdot)$, $j = 1, \ldots, N_1$, $\ell = 1, \ldots, N_2$, since for any $\mathbf{g}(z, \boldsymbol{\alpha}) \in \mathcal{G}$, we can suitably choose $\boldsymbol{\alpha}_j$ and $\mathbf{g}^{\ell,j}(\cdot)$ such that

$$\begin{aligned}
&\sup_{z \in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}) - \mathbf{g}^{\ell,j}(z)\| \\
&\qquad \leq \sup_{z \in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}) - \mathbf{g}(z, \boldsymbol{\alpha}_j)\| + \sup_{z \in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}_j) - \mathbf{g}^{\ell,j}(z)\| \\
&\qquad \leq C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_j\| + \sup_{z \in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}_j) - \mathbf{g}^{\ell,j}(z)\| \leq \frac{3}{2}\varepsilon,
\end{aligned} \tag{A.15}$$

and similarly, we can cover $\mathcal{G}_1^{(1)} = \{\mathbf{g}_1 : S_w \times \mathbb{B} \mapsto \mathbb{R}^d \mid \mathbf{g} \in \mathcal{G}_1\}$ and $\mathcal{G}_1^{(2)} = \{\mathbf{g}_2 : S_w \times \mathbb{B} \mapsto \mathbb{R}^{d \times d} \mid \mathbf{g} \in \mathcal{G}_1\}$ by finite number of $N_1 N_3$ balls of centers $\mathbf{g}_1^{\ell,j}(\cdot)$ and $\mathbf{g}_2^{\ell,j}(\cdot)$, $j = 1, \ldots, N_1$, $\ell = 1, \ldots, N_3$, respectively, since for any $\mathbf{g}_1(z, \boldsymbol{\alpha}) \in \mathcal{G}_1^{(1)}$ and $\mathbf{g}_2(z, \boldsymbol{\alpha}) \in \mathcal{G}_1^{(2)}$, we can suitably choose $\boldsymbol{\alpha}_j$ and $\mathbf{g}_1^{\ell,j}(\cdot)$ and $\mathbf{g}_2^{\ell,j}(\cdot)$, respectively, such that, as in (A.15),

$$\sup_{z \in S_w} \|\mathbf{g}_1(z, \boldsymbol{\alpha}) - \mathbf{g}_1^{\ell,j}(z)\| \leq \varepsilon, \quad \sup_{z \in S_w} \|\mathbf{g}_2(z, \boldsymbol{\alpha}) - \mathbf{g}_2^{\ell,j}(z)\| \leq \varepsilon. \tag{A.16}$$

Therefore, with $\boldsymbol{\alpha}^\top \mathbf{X}_t \in S_w$ and $\mathbf{g}^{\ell,j}(\cdot) \in C^2_{c_0}(S_w)$, it follows from (A.14), (A.15) and (A.16) that

$$\begin{aligned}
&\|\mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t)\| \\
&\quad \le \|\mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}^{\ell,j}(\boldsymbol{\alpha}^\top \mathbf{X}_t)\| + \|\mathbf{g}^{\ell,j}(\boldsymbol{\alpha}^\top \mathbf{X}_t) - \mathbf{g}^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t)\| \\
&\quad \le \varepsilon + C\|\mathbf{X}_t\| \ \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_j\| \le \varepsilon(1 + C\|\mathbf{X}_t\|),
\end{aligned}$$

similarly,

$$\begin{aligned}
&\|\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_1^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t, \boldsymbol{\alpha}_j)\| \le \varepsilon(1 + C\|\mathbf{X}_t\|), \\
&\|\mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_2^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t, \boldsymbol{\alpha}_j)\| \le \varepsilon(1 + C\|\mathbf{X}_t\|);
\end{aligned}$$

and with $\mathbf{g} \in \mathcal{G}_1$, it follows that

$$\begin{aligned}
&\|\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\| \le \|\mathbf{g}_{01}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\| + 1 \le \|\mathbf{g}_{01}\|_{\mathcal{G}} + 1, \\
&\|\mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\| \le \|\mathbf{g}_{02}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\| + 1 \le \|\mathbf{g}_{02}\|_{\mathcal{G}} + 1,
\end{aligned}$$

Note, for any $m \in \mathcal{M}$,

$$E|m(F_t, \mathbf{g}, \boldsymbol{\alpha}) - m(F_t, \mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j)|^2 \le C(M_1 + M_2 + M_3), \tag{A.17}$$

where

$$\begin{aligned}
M_1 &= E\Big|(m_{1t}(\mathbf{g}, \boldsymbol{\alpha}) - m_{1t}(\mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j)) m_{2t}(\mathbf{g}, \boldsymbol{\alpha}) m_{3t}(\boldsymbol{\alpha})\Big|^2 \\
&\le E\Big[\|\mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t)\| \|\mathbb{X}_t\| \\
&\quad \times \Big\| \Big\{\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbf{X}_t - \mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\Big\}^\top \Big\| \|\mathbb{X}_t\| \ w(\boldsymbol{\alpha}^\top \mathbf{X}_t)\Big]^2 \\
&\le C\varepsilon E\Big[\|\mathbb{X}_t\| \| \Big\{\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbf{X}_t - \mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\Big\}^\top \| \ \|\mathbb{X}_t\| \ w(\boldsymbol{\alpha}^\top \mathbf{X}_t)\Big]^2 \\
&\le C\varepsilon^2 E\Big[\|\mathbb{X}_t\|^2 \ (C\|\mathbf{X}_t\| + C)^2 \ \|\mathbb{X}_t\|^2\Big] \le C\varepsilon^2,
\end{aligned} \tag{A.18}$$

$$\begin{aligned}
M_2 &= E\Big|m_{1t}(\mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j)(m_{2t}(\mathbf{g}, \boldsymbol{\alpha}) - m_{2t}(\mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j)) m_{3t}(\boldsymbol{\alpha})\Big|^2 \\
&\le E\Big[\|Y_t - \mathbf{g}^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t)^\top \mathbb{X}_t\| \{\|\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_1^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t, \boldsymbol{\alpha}_j)\| \ \|\mathbf{X}_t\| \\
&\qquad + \|\mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{g}_2\ell, j(\boldsymbol{\alpha}_j^\top \mathbf{X}_t, \boldsymbol{\alpha}_j)\|\} \ \|\mathbb{X}_t\| \ w(\boldsymbol{\alpha}^\top \mathbf{X}_t)\Big]^2 \\
&\le C\varepsilon^2 E\Big[(|Y_t| + c_0\|\mathbb{X}_t\|)^2 (1 + C\|\mathbf{X}_t\|)^2 \|\mathbb{X}_t\|^2\Big] \le C\varepsilon^2,
\end{aligned} \tag{A.19}$$

$$M_3 = E\Big[m_{1t}(\mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j) m_{2t}(\mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j)(m_{3t}(\boldsymbol{\alpha}) - m_{3t}(\boldsymbol{\alpha}_j))\Big]^2 \le C\varepsilon^2, \tag{A.20}$$

and where C is allowed to change in value from line to line. Then it follows from (A.17) together with (A.18), (A.19) and (A.20) that

$$\|m(F_t, \mathbf{g}, \boldsymbol{\alpha}) - m(F_t, \mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j)\|_{\mathcal{L}_2} \leq C\varepsilon,$$

and thus $N(C\varepsilon, \mathcal{M}, \|\cdot\|_{\mathcal{L}_2}) \leq (N_1N_2)(N_1N_3)N_1$, which leads to

$$\log N(C\varepsilon, \mathcal{M}, \|\cdot\|_{\mathcal{L}_2}) \leq C(\log N_1 + \log N_2 + \log N_3) \leq C\varepsilon^{-1}.$$

Now (c) holds easily.

(vi) Finally we are in a position to establish (vi) of Lemma 4.2. Note that it follows from (A.10) with $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ that

$$\begin{aligned}
&\widehat{\mathbf{g}}(z, \boldsymbol{\alpha}_0) - \mathbf{g}_0(z, \boldsymbol{\alpha}_0) \\
&\quad = \gamma(\widehat{\boldsymbol{\theta}}(z, \boldsymbol{\alpha}_0) - \boldsymbol{\theta}_0(z, \boldsymbol{\alpha}_0)) \\
&\quad = \gamma\widehat{\Phi}^{-1}(z, \boldsymbol{\alpha}_0)\widehat{\mathbf{W}}(z, \boldsymbol{\alpha}_0) = (1 + o_P(1))\gamma\Phi^{-1}(z, \boldsymbol{\alpha}_0)\widehat{\mathbf{W}}(z, \boldsymbol{\alpha}_0) \\
&\quad = (1 + o_P(1))(\mu_{0,K}w(z)f_Z(z, \boldsymbol{\alpha}_0))^{-1}\mathbf{G}^{-1}(z, \boldsymbol{\alpha}_0)\widehat{\mathbf{W}}^{(1)}(z, \boldsymbol{\alpha}_0),
\end{aligned} \tag{A.21}$$

where $o_P(1)$ is uniform with respect to $z \in S_w$, and $\widehat{\mathbf{W}}^{(1)}(z, \boldsymbol{\alpha}_0)$ is defined in (A.11). Then (A.6) together with Lemma A.1 and (A.21) then leads to

$$\begin{aligned}
&\Gamma_2\left(\mathbf{g}_0(\cdot, \boldsymbol{\alpha}_0), \boldsymbol{\alpha}_0\right)\left[\widehat{\mathbf{g}}(\cdot, \boldsymbol{\alpha}_0) - \mathbf{g}_0(\cdot, \boldsymbol{\alpha}_0)\right] \\
&\quad = 2\int \widetilde{\Phi}_0(z)(\widehat{\mathbf{g}}(z, \boldsymbol{\alpha}_0) - \mathbf{g}_0(z, \boldsymbol{\alpha}_0))w(z)f_0(z, \boldsymbol{\alpha}_0)dz \\
&\quad = (1 + o_P(1))2\mu_{0,K}^{-1}\int \widetilde{\Phi}_0(z)G^{-1}(z, \boldsymbol{\alpha}_0)\widehat{\mathbf{W}}^{(1)}(z, \boldsymbol{\alpha}_0)dz \\
&\quad = (1 + o_P(1))2\mu_{0,K}^{-1}n^{-1}\sum_{t=1}^{n}\int \widetilde{\Phi}_0(z)G^{-1}(z, \boldsymbol{\alpha}_0) \\
&\qquad \times \left\{\varepsilon_t + \frac{1}{2}\left(\ddot{\mathbf{g}}_0(z + \eta_1(Z_t^o - z))(Z_t^o - z)^2\right)^{\mathsf{T}}\mathbb{X}_t\right\}\mathbb{X}_tK_h(Z_t^o - z)w(Z_t^o)dz \\
&\quad = (1 + o_P(1))2\mu_{0,K}^{-1}n^{-1}\sum_{t=1}^{n}\widetilde{\Phi}_0(Z_t^o)G^{-1}(Z_t^o, \boldsymbol{\alpha}_0) \\
&\qquad \times \left\{\varepsilon_t\mu_{0,K} + \frac{1}{2}\left(\ddot{\mathbf{g}}_0(Z_t^o)\mu_{2,K}h^2(1 + o(1))\right)^{\mathsf{T}}\mathbb{X}_t\right\}\mathbb{X}_tw(Z_t^o) \\
&\quad = (1 + o_P(1))2\left\{n^{-1}\sum_{t=1}^{n}\varepsilon_t\mathbf{U}_t + \frac{1}{2}h^2\mu_{0,K}^{-1}\mu_{2,K}E\left(\ddot{\mathbf{g}}_0(Z_t^o))^{\mathsf{T}}\mathbb{X}_t\mathbf{U}_t\right)\right\} \\
&\qquad + o_P(n^{-\frac{1}{2}}),
\end{aligned} \tag{A.22}$$

as $n^{-1}\sum_{t=1}^n \left\{\ddot{\mathbf{g}}_0(Z_t^o))^\top \mathbb{X}_t \mathbf{U}_t - E\left(\ddot{\mathbf{g}}_0(Z_t^o))^\top \mathbb{X}_t \mathbf{U}_t\right)\right\} = O_P(n^{-1/2})$ according to the CLT for a strongly mixing strictly stationary process, where

$$\begin{aligned}\mathbf{U}_t &= \widetilde{\Phi}_0(Z_t^o)G^{-1}(Z_t^o,\boldsymbol{\alpha}_0)\mathbb{X}_t w(Z_t^o)\\ &= E\left(\mathbf{X}_t\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)^\top \mathbb{X}_t\mathbb{X}_t^\top | Z_t^o\right)G^{-1}(Z_t^o,\boldsymbol{\alpha}_0)\mathbb{X}_t w(Z_t^o)\\ &\quad + \mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)^\top \mathbb{X}_t w(Z_t^o). \end{aligned} \tag{A.23}$$

Now we have from (A.2) and (A.22) and then from (A.23) that

$$\begin{aligned}&\sqrt{n}\left\{\dot{R}_n(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) + \Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\left[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)\right]\right\}\\ &\quad = \sqrt{n}\Big\{-\frac{2}{n}\sum_{t=1}^n \varepsilon_t\Big(\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top + \mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\Big)^\top \mathbb{X}_t w(Z_t^o)\\ &\qquad + (1+o_P(1))2\Big[n^{-1}\sum_{t=1}^n \varepsilon_t \mathbf{U}_t + \frac{1}{2}h^2\mu_{0,K}^{-1}\mu_{2,K}E\Big(\ddot{\mathbf{g}}_0(Z_t^o))^\top \mathbb{X}_t\mathbf{U}_t\Big)\Big]\\ &\qquad + o_P(n^{-\frac{1}{2}})\Big\}\\ &\quad = \sqrt{n}\Big\{-\frac{2}{n}\sum_{t=1}^n \varepsilon_t \mathbf{V}_t + (1+o_P(1))h^2\mu_{0,K}^{-1}\mu_{2,K}E\Big(\ddot{\mathbf{g}}_0(Z_t^o))^\top \mathbb{X}_t\mathbf{U}_t\Big)\\ &\qquad + o_P(n^{-\frac{1}{2}})\Big\},\end{aligned}$$

where $\mathbf{V}_t = \left[\mathbf{X}_t\mathbf{g}_{01}^\top(Z_t^o,\boldsymbol{\alpha}_0) - \{E\left(\mathbf{X}_t\mathbf{g}_{01}^\top(Z_t^o,\boldsymbol{\alpha}_0)\mathbb{X}_t\mathbb{X}_t^\top|Z_t^o\right)\}G^{-1}(Z_t^o,\boldsymbol{\alpha}_0)\right]$ $\mathbb{X}_t w(Z_t^o)$. Therefore, by CLT for mixing stationary process,

$$\begin{aligned}&\sqrt{n}\Big\{\dot{R}_n(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0) + \Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\left[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)\right]\\ &\qquad -(1+o_P(1))h^2\mu_{0,K}^{-1}\mu_{2,K}E\left(\ddot{\mathbf{g}}_0(Z_t^o))^\top \mathbb{X}_t\mathbf{U}_t\right)\Big\} \xrightarrow{D} N(\mathbf{0},\mathcal{V}),\end{aligned} \tag{A.24}$$

where

$$\mathcal{V} = 4E\varepsilon_t^2\mathbf{V}_t\mathbf{V}_t^\top = 4E\varepsilon_t^2\{\Xi_t\Xi_t^\top - E(\Xi_t\mathbb{X}_t^\top|Z_t^o)G_t^{-1}E(\mathbb{X}_t\Xi_t^\top|Z_t^o)\}$$

with $\Xi_t = \mathbf{X}_t\mathbf{g}_{01}^\top(Z_t^o,\boldsymbol{\alpha}_0)\mathbb{X}_t w(Z_t^o)$ and $G_t = G(Z_t^o,\boldsymbol{\alpha}_0) = E(\mathbb{X}_t\mathbb{X}_t^\top|Z_t^o)$.

Finally, letting $\Gamma_0 = \Gamma_1/2$ and $\mathcal{V}_0 = \mathcal{V}/4$, we get the desired result of (3.4) following from Lemma 4.2. The proof is completed.

B. Proof of Lemma A.1

We collect and prove the uniform convergence results which were used in the proof for Section 3.2. All limits are taken as $n \to \infty$ unless stated otherwise.

B.1. Technical lemmas

For the proof of uniform-consistency lemmas, we need to repeatedly use the following moment inequalities, which are stated for reference below.

Lemma B.1.(Cox and Kim (1995)'s moment inequality) *Let* $\{\xi_t\}$ *be a strongly mixing process with* $E\xi_t = 0$*, and* r *a positive integer. Assume that for some* $q > 2$*,*

$$M_{qr} = \sup_t\{\|\xi_t\|_{qr}\} = \sup_t\{(E|\xi_t|^{qr})^{\frac{1}{qr}}\} \leq 1,$$

and that there is a constant ν *not depending on* t *such that*

$$E[|\xi_t|^k] \leq \nu, \quad 2 \leq k \leq 2r,$$

and that the mixing coefficients satisfy

$$\sum_{i=1}^{\infty} i^{r-1}\beta(i)^{1-\frac{2}{q}} < \infty.$$

Then there exists a constant C *depending on* r *but not depending on the distribution of* ξ_t *nor on* ν*,* n*, nor* $\tilde{P}$ *such that*

$$E\Big(\sum_{t=1}^{n}\xi_t\Big)^{2r} \leq C\Big\{n^r M_{qr}^{2r}\sum_{i=\tilde{P}}^{\infty} i^{r-1}\beta(i)^{1-\frac{2}{q}} + \sum_{j=1}^{r} n^j \tilde{P}^{2r-j}\nu^j\Big\}$$

for any integers n *and* $\tilde{P}$ *with* $0 < \tilde{P} < n$*.*

Proof. This is Theorem 1 of Cox and Kim (1995, p.152).

Lemma B.2.(Gao, Lu and Tjøstheim (2004)'s moment inequality) *Assume that the process* $\{(X_t, Y_t) : t \in \mathbb{Z}^1\}$ *is* β*-mixing and strictly stationary with* Y_t *and* X_t *being* $\mathbb{R}^1$*-valued respectively. Let* $\xi_t = K_t\boldsymbol{\theta}_t = K((X_t - x)/h)\boldsymbol{\theta}_t$ *with* $E[\xi_t] = 0$*, where* $\boldsymbol{\theta}_t = \boldsymbol{\theta}(X_t, Y_t)$ *and* $K(\cdot)$ *is a bounded kernel function defined on* $\mathbb{R}^1$*. The joint probability density* $f_s(x_1, \ldots, x_s)$ *of* $(X_{t_1}, \ldots, X_{t_s})$ *exists and is bounded uniformly for* $s = 1, \ldots, 2r-1$*, where* r *is some positive integer. Assume further that* $E\left[|\boldsymbol{\theta}_t|^{qr}\right] < \infty$ *for some* $q > 2$*. The mixing coefficient* β *satisfies*

$$\lim_{T\to\infty} T^a \sum_{t=T}^{\infty} t^{r-1}\beta(t)^{\frac{qr-2}{qr}} = 0$$

for some constant $a \geq (rq-2)r/(2+rq-4r)$ *with* $q > (4r-2)/r$*. The probability kernel function* $K(x)$ *is a symmetric and bounded density function on* $\mathbb{R}^1$ *with compact support,* C_K*, and finite variance such that* $|K(x) - K(y)| \leq M_K|x-y|$ *for* $x,\ y \in C_K$ *and* $0 < M_K < \infty$*. The bandwidth* $h = h_n$ *satisfies that*

$$\lim_{n\to\infty} h_n = 0 \quad \textit{and} \quad \lim\inf_{n\to\infty} nh_n^{\frac{2(r-1)a+(qr-2)}{(a+1)q}} > 0$$

for some integer $r \geq 3$. Then there exists a constant $C = C(r)$ depending on r but not depending on the distribution of ξ_t nor on h, n such that

$$E\Big[\Big(\sum_{i=1}^{n}\xi_i\Big)^{2r}\Big] \leq C\,(nh)^r. \tag{B.1}$$

Proof. It is a special case of Theorem 1.1 of Gao, Lu and Tjøstheim (2004) with $N = 1$ there.

B.2. Proof of Lemma A.1

Proof of Lemma A.1. As the proofs of (A.2)−(A.6) are similar, so we only sketch the proof of (A.4) below. It follows from conditions (C1) and (C3) that $f_Z(z,\boldsymbol{\alpha})$, which equals $f_{\boldsymbol{\alpha}^\top \mathbf{X}_t}(z)$, and $\mathbf{G}(z,\boldsymbol{\alpha})$, which is equal to $E(\mathbb{X}_t\mathbb{X}_t^\top \mid \boldsymbol{\alpha}^\top\mathbf{X}_t = z)$, are bounded away from zero over $z \in S_w, \boldsymbol{\alpha} \in \mathbb{B}$. Therefore, it is derived from (A.10) that $\widehat{\mathbf{g}}(z,\boldsymbol{\alpha}) - \mathbf{g}_0(z,\boldsymbol{\alpha})$ tending to 0 uniformly is equivalent to $\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha})$ (see (A.11)) tending to 0 uniformly, where $\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha})$ can be separated into two parts of the bias term and the error term owing to (A.12). As the bias term is easily taken care of, so we are only concerned with the uniform convergence rate, for the error term, of

$$\widehat{W}_2(z,\boldsymbol{\alpha}) = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_{t,\boldsymbol{\alpha}}\mathbb{X}_t K_h(\boldsymbol{\alpha}^\top\mathbf{X}_t - z)w(\boldsymbol{\alpha}^\top\mathbf{X}_t)$$

below. It follows from Lemma 5.1 that $E\widehat{W}_2(z,\boldsymbol{\alpha}) = 0$. When $\boldsymbol{\alpha}$ is fixed, the uniform convergence rate of $\widehat{W}_2(z,\boldsymbol{\alpha})$ with respect to z was established by Masry and Tjøstheim (1995). Here we establish the lemma with convergence rate also uniform with respect to $\boldsymbol{\alpha} \in \mathbb{B}$, by

$$\sup_{\boldsymbol{\alpha}\in\mathbb{B}}\sup_{z\in S_w}\|\widehat{W}_2(z,\boldsymbol{\alpha})\| = \Big(nh^{1+\frac{2d}{r}}\Big)^{-\frac{r}{2r+d}}, \tag{B.2}$$

where we note $\mathbb{B} \subset \{\boldsymbol{\alpha} \in \mathbb{R}^d : \|\boldsymbol{\alpha}\| = 1\}$ and without loss of generality let $\mathbb{B}$ be compact.

Because $\mathbb{B}$ and S_w are compact, we can cover $\mathbb{B}$ and S_w by a finite number $M = M_n$ of cubes $I_k \subset \mathbb{B}$ with centers $\boldsymbol{\alpha}_k$ in $\mathbb{B}$, satisfying $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \leq$ const.$/M^{1/(d-1)}$ for any $\boldsymbol{\alpha} \in I_k$, and a finite number $N = N_n$ of cubes $J_\ell \subset S_w$ with centers z_ℓ in S_w, satisfying $|z - z_\ell| \leq$ const.$/N$ for $z \in J_\ell$, respectively, where M and N are to be specified later. Therefore

$$\begin{aligned}&\sup_{\boldsymbol{\alpha}\in\mathbb{B}}\sup_{z\in S_w}\|\widehat{W}_2(z,\boldsymbol{\alpha})\| \\ &\quad\leq \max_{1\leq k\leq M}\sup_{z\in S_w}\|\widehat{W}_2(z,\boldsymbol{\alpha}_k)\| + \max_{1\leq k\leq M}\sup_{\boldsymbol{\alpha}\in I_k}\sup_{z\in S_w}\|\widehat{W}_2(z,\boldsymbol{\alpha}) - \widehat{W}_2(z,\boldsymbol{\alpha}_k)\| \\ &\quad\equiv W_{21} + W_{22}.\end{aligned} \tag{B.3}$$

We first consider W_{22}. Note that

$$\widehat{W}_2(z,\boldsymbol{\alpha}) - \widehat{W}_2(z,\boldsymbol{\alpha}_k) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{X}_t\Big\{\varepsilon_{t,\boldsymbol{\alpha}} K_h(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t - z) w(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t) - \varepsilon_{t,\boldsymbol{\alpha}_k} K_h(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t - z) w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t)\Big\}, \tag{B.4}$$

and that

$$\begin{aligned}|\varepsilon_{t,\boldsymbol{\alpha}} - \varepsilon_{t,\boldsymbol{\alpha}_k}| &\le |a_0(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}) - a_0(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}_k)| \\ &\quad + \|\mathbf{b}_0(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{b}_0(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t,\boldsymbol{\alpha}_k)\| \ \|\check{\mathbf{X}}_t\| \\ &\le C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|(1 + \|\mathbf{X}_t\|)^2,\end{aligned}$$

and

$$\begin{aligned}&|K_h(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t - z) w(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t) - K_h(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t - z) w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t)| \\ &\qquad \le Ch^{-2}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \ \|\mathbf{X}_t\|.\end{aligned}$$

Thus

$$\begin{aligned}&\widehat{W}_2(z,\boldsymbol{\alpha}) - \widehat{W}_2(z,\boldsymbol{\alpha}_k) \\ &\quad \le \frac{1}{n}\sum_{t=1}^{n} \|\mathbb{X}_t\|\{h^{-1}(1 + \|\mathbf{X}_t\|)^2 + |\varepsilon_{t,\boldsymbol{\alpha}_k}|h^{-2} \ \|\mathbf{X}_t\|\}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \\ &\quad \le \frac{1}{n}\sum_{t=1}^{n} \left\{h^{-1}(1 + \|\mathbf{X}_t\|)^3 + |\varepsilon_{t,\boldsymbol{\alpha}_k}|h^{-2} \ \|\mathbf{X}_t\|^2\right\} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|,\end{aligned} \tag{B.5}$$

and it follows from (B.3) and (B.5) that

$$\begin{aligned}W_{22} &\le \max_{1\le k\le M} \sup_{\boldsymbol{\alpha}\in I_k} \frac{1}{n}\sum_{t=1}^{n} \left\{h^{-1}(1 + \|\mathbf{X}_t\|)^3 + |\varepsilon_{t,\boldsymbol{\alpha}_k}|h^{-2} \ \|\mathbf{X}_t\|^2\right\} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \\ &\le CM^{-\frac{1}{d-1}} \max_{1\le k\le M} \frac{1}{n}\sum_{t=1}^{n} \left\{h^{-1}(1 + \|\mathbf{X}_t\|)^3 + |\varepsilon_{t,\boldsymbol{\alpha}_k}|h^{-2} \ \|\mathbf{X}_t\|^2\right\} \\ &\le CM^{-\frac{1}{d-1}} \left\{h^{-1}\frac{1}{n}\sum_{t=1}^{n}(1 + \|\mathbf{X}_t\|)^3 + \max_{1\le k\le M} \frac{1}{n}\sum_{t=1}^{n} |\varepsilon_{t,\boldsymbol{\alpha}_k}|h^{-2}\|\mathbf{X}_t\|^2\right\} \\ &\le CM^{-\frac{1}{d-1}} \left\{h^{-1}O_P(1) + h^{-2}W_{222} + h^{-2}W_{223}\right\},\end{aligned} \tag{B.6}$$

where $O_P(1)$ is uniform with respect to $z \in S_w$ and $\boldsymbol{\alpha} \in \mathbb{B}$ as $n \to \infty$, and

$$\begin{aligned}W_{222} &= \max_{1\le k\le M} \left|\frac{1}{n}\sum_{t=1}^{n} \left(|\varepsilon_{t,\boldsymbol{\alpha}_k}| \ \|\mathbf{X}_t\|^2 - E|\varepsilon_{t,\boldsymbol{\alpha}_k}| \ \|\mathbf{X}_t\|^2\right)\right|, \\ W_{223} &= \max_{1\le k\le M} E|\varepsilon_{t,\boldsymbol{\alpha}_k}| \ \|\mathbf{X}_t\|^2.\end{aligned}$$

Clearly, by the condition that $\sup_{\boldsymbol{\alpha}\in\mathbb{B}} E|\varepsilon_{t,\boldsymbol{\alpha}}|^2 < \infty$,

$$W_{223} \le \max_{1\le k\le M} \left\{E|\varepsilon_{t,\boldsymbol{\alpha}_k}|^2\right\}^{\frac{1}{2}} \left\{E\|\mathbf{X}_t\|^4\right\}^{\frac{1}{2}} = O(1), \tag{B.7}$$

which is uniform with respect to $z \in S_w$ and $\boldsymbol{\alpha} \in \mathbb{B}$ as $n \to \infty$. Further, we consider W_{222}. Set $u_{t,k} = |\varepsilon_{t,\boldsymbol{\alpha}_k}|\ \|\mathbf{X}_t\|^2$ and $\Delta_k = \frac{1}{n}\sum_{t=1}^n (u_{t,k} - Eu_{t,k})$, and therefore $W_{222} = \max_{1\le k\le M} |\Delta_k|$. Applying Lemma B.1. with $P = 1$ leads to $E|\Delta_k|^{2r} \le C_r n^{-r}$, where C_r only depends on r. Thus, if $M = O(n^r)$, then

$$\begin{aligned} P\{W_{222} > 2A\} &= P\left\{\max_{1\le k\le M} |\Delta_k| > 2A\right\} \le \sum_{k=1}^M P\{|\Delta_k| > A\} \\ &= C_r M A^{-r} n^{-r} = CA^{-r} \to 0 \end{aligned} \tag{B.8}$$

as $A \to \infty$, which leads to $W_{222} = O_P(1)$. This together with (B.6) and (B.7) implies

$$W_{22} = O_P(M^{-\frac{1}{d-1}} h^{-2}) = O_P\left\{\zeta_n\right\}, \tag{B.9}$$

where we take $M = (h^2\zeta_n)^{-(d-1)}$ with ζ_n to be specified later, and $O_P(\cdot)$ is uniform with respect to $z \in S_w$ and $\boldsymbol{\alpha} \in \mathbb{B}$ as $n \to \infty$.

Next, we consider W_{21} in (B.3). As $\widehat{W}_2(z,\boldsymbol{\alpha}_k) = \widehat{W}_2(z_\ell,\boldsymbol{\alpha}_k) + (\widehat{W}_2(z,\boldsymbol{\alpha}_k) - \widehat{W}_2(z_\ell,\boldsymbol{\alpha}_k))$, we can break W_{21} into two parts:

$$\begin{aligned} W_{21} &\le \max_{1\le k\le M}\max_{1\le \ell\le N} \widehat{W}_2(z_\ell,\boldsymbol{\alpha}_k) \\ &\quad + \max_{1\le k\le M}\max_{1\le \ell\le N}\sup_{z\in J_\ell} \|\widehat{W}_2(z,\boldsymbol{\alpha}_k) - \widehat{W}_2(z_\ell,\boldsymbol{\alpha}_k)\| \\ &\equiv W_{211} + W_{212}. \end{aligned} \tag{B.10}$$

For W_{212}, note that, using the Lipschitz continuity of $K(\cdot)$ and the boundedness of $w(\cdot)$,

$$\begin{aligned} &\|\widehat{W}_2(z,\boldsymbol{\alpha}_k) - \widehat{W}_2(z_\ell,\boldsymbol{\alpha}_k)\| \\ &\quad = \|\frac{1}{n}\sum_{t=1}^n \varepsilon_{t,\boldsymbol{\alpha}_k}\ \mathbb{X}_t \left\{K_h(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t - z) - K_h(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t - z_\ell)\right\} w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t)\| \\ &\quad \le C\frac{1}{n}\sum_{t=1}^n |\varepsilon_{t,\boldsymbol{\alpha}_k}|\ \|\mathbb{X}_t\| h^{-2}|z - z_\ell| w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t) \\ &\quad = Ch^{-2}|z - z_\ell|\ \frac{1}{n}\sum_{t=1}^n |\varepsilon_{t,\boldsymbol{\alpha}_k}|\ \|\breve{\mathbf{X}}_t\| w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t), \end{aligned}$$

therefore, noting $\|\breve{\mathbf{X}}_t\| \le \|\mathbb{X}_t\|$,

$$\begin{aligned} W_{212} &= \max_{1\le k\le M}\max_{1\le \ell\le N}\sup_{z\in J_\ell}\|\widehat{W}_2(z,\boldsymbol{\alpha}_k)-\widehat{W}_2(z_\ell,\boldsymbol{\alpha}_k)\| \\ &= C\max_{1\le k\le M}\max_{1\le \ell\le N}\sup_{z\in J_\ell} h^{-2}|z-z_\ell|\ \frac{1}{n}\sum_{t=1}^{n}|\varepsilon_{t,\boldsymbol{\alpha}_k}|\ \|\mathbb{X}_t\| w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t) \\ &\le Ch^{-2}N^{-1}\max_{1\le k\le M}\frac{1}{n}\sum_{t=1}^{n}|\varepsilon_{t,\boldsymbol{\alpha}_k}|\ \|\mathbb{X}_t\| w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t) \\ &= O_P\left(h^{-2}N^{-1}\right) = O_P(\zeta_n), \end{aligned} \tag{B.11}$$

where in the final equality of (B.11), we take $N=(h^2\zeta_n)^{-1}$, and $O_P(\cdot)$ is uniform with respect to $z\in S_w$ and $\boldsymbol{\alpha}\in\mathbb{B}$ as $n\to\infty$, the argument being the same as that for $W_{222}=O_P(1)$ and $W_{223}=O(1)$ in (B.6) in the above. Now we consider W_{211} in (B.10). With $\xi_t=\varepsilon_{t,\boldsymbol{\alpha}_k}\ \mathbb{X}_t K((\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t-z_\ell)/h)w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t)$ and $\boldsymbol{\theta}_t=\varepsilon_{t,\boldsymbol{\alpha}_k}\ \mathbb{X}_t w(\boldsymbol{\alpha}_k^{\mathsf{T}}\mathbf{X}_t)$ in Lemma B.2, it follows from Lemma B.2 that

$$\begin{aligned} &P\{W_{211}\ge\varepsilon\} \\ &\quad\le \sum_{k=1}^{M}\sum_{\ell=1}^{N}P\{\|W_2(z_\ell,\boldsymbol{\alpha}_k)\|\ge\varepsilon\} = \sum_{k=1}^{M}\sum_{\ell=1}^{N}P\{\|(nh)^{-1}\sum_{i=1}^{n}\xi_i\|\ge\varepsilon\} \\ &\quad\le \sum_{k=1}^{M}\sum_{\ell=1}^{N}\varepsilon^{-2r}(nh)^{-2r}E\|\sum_{i=1}^{n}\xi_i\|^{2r} \le \varepsilon^{-2r}(nh)^{-2r}MNC(nh)^r \\ &\quad= C\varepsilon^{-2r}(nh)^{-r}MN = C\varepsilon^{-2r}(nh)^{-r}(h^2\zeta_n)^{-d}. \end{aligned}$$

Therefore

$$W_{211} = O_P\left((nh)^{-\frac{1}{2}}(h^2\zeta_n)^{-\frac{d}{2r}}\right), \tag{B.12}$$

where $O_P(\cdot)$ is uniform with respect to $z\in S_w$ and $\boldsymbol{\alpha}\in\mathbb{B}$ as $n\to\infty$.

Finally, taking $\zeta_n=(nh^{1+2d/r})^{-r/(2r+d)}$, then $N=(h^2\zeta_n)^{-1}=(nh^{-3})^{r/(2r+d)}$, and $M=(h^2\zeta_n)^{-(d-1)}=(nh^{-3})^{(d-1)r/(2r+d)}=O(n^r)$ as $\lim_{n\to\infty}n^{2r+1}h^{3(d-1)}>0$. For such a ζ_n, (B.8), (B.9) and (B.11) hold simultaneously. Thus the result of (B.2) follows from (B.3), (B.9), (B.10), (B.11) and (B.12).

References

Cox, D. D. and Kim, T. Y. (1995). Moment bounds for mixing random variables useful in nonparametric function estimation. *Stochastic Process. Appl.* **56**, 151-158.

Gao, J., Lu, Z. and Tjøstheim, D. (2004). Moment inequality for spatial processes. Unpublished technical report, School of Mathematics and Statistics, The University of Western Australia. Available at www.maths.uwa.edu.au/˜jiti/glt04.pdf.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. *Econom. Theory* **11**, 258-289.