Zudi Lu, Dag Tjøstheim and Qiwei Yao

Adaptive varying-coefficient linear models for stochastic processes: asymptotic theory

Article (Accepted version) (Refereed)

Adaptive Varying-Coefficient Linear Models for Stochastic Processes: Asymptotic Theory *

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

[1]Department of Statistics, London School of Economics, London, WC2A 2AE, UK

[2]Institute of Systems Science, Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing 100080, China

[3]Department of Mathematics, University of Bergen, 5007 Bergen, Norway

[4]Guanghua School of Management, Peking University, Beijing 100871, China

Abstract

We have established the asymptotic theory for the estimation of adaptive varying-coefficient linear models. More specifically we have shown that the estimator for the global index parameter is root-n consistent without imposing, as a prerequisite, that the estimator is within $n^{-\delta}$-distance from the true value. To this end, we have established two fundamental lemmas for the asymptotic properties of the estimators for parametric components in general semi-parametric settings. Furthermore, the estimation for the coefficient functions is asymptotically adaptive to the unknown index parameter in the sense that the first order of the asymptotic distribution is the same as if the index parameter were known. The asymptotic properties were derived for the observations from a strictly stationary β-mixing process, which includes both independent observations and time series as special cases.

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

*Partially supported by a Leverhulme Trust research grant. Lu's work was also partially supported by the National Natural Science Foundation of China (70271003, 70221001).

1 Introduction

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

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

where t is time, $\mathbf{X}_t$ is a $d \times 1$ predictor vector which may consist of some lagged values of Y_t or/and other exogenous variables, and $E(\varepsilon_t|\mathbf{X}_t) = 0$. In model (1.1), the index parameter $\boldsymbol{\alpha}_0$ is unknown, and both functions $a_0(\cdot)$ 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 unknown index variable $\boldsymbol{\alpha}_0^\top \mathbf{X}_t$, in contrast to, for example, the functional-coefficient models of Chan and Tsay (1993), and Cai, Fan and Yao (2000). This is a quite general form of nonlinear dynamical model. For example, for $\mathbf{X}_t = \{Y_{t-1}, Y_{t-2}, \cdots, Y_{t-d}\}^\top$, (1.1) reduces to *the adaptive varying-coefficient linear autoregressive* model (Tong 1990, Xia and Li 1999, Fan, Yao and Cai 2003). On the other hand, some financial econometrics models specify $\mathbf{X}_t = \{Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}, U_t, U_{t-1}, \cdots, U_{t-q}\}^\top$ for some exogenous process U_t; see Hannan (1970), Gourieroux and Jasiak (2001) and Hong and Lee (2003). Formally model (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 model (1.1) is that it does not suffer from the curse of dimensionality encountered often in multivariate nonparametric modelling, since both $a_0(\cdot), \mathbf{b}_0(\cdot)$ are functions of univariate variables.

The estimation for model (1.1) with independent observations has been investigated in several papers. Ichimura (1993) proposed the form of the model (1.1). Following the lead of Härdle *et al* (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 is within $n^{-\delta}$-distance from $\boldsymbol{\alpha}_0$ 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. But no asymptotic properties of their estimator have been established.

The main purpose of this paper is to 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 from the true value, which is a condition often imposed for the problem of this nature; see, for example, Härdle *et al* (1993), Carroll *et al.* (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 global parameter $\boldsymbol{\alpha}_0$ and local parameters $a_0, \mathbf{b}_0$. This is consistent with the common knowledge that a global parameter should be estimated in an undersmoothed manner.

At the technical level, our approach is also different from that of Härdle *et al* (1993) and Xia and Li (1999). Although Lemmas 4.1 and 4.2 in section 4 below played a fundamental role in deriving the asymptotic properties of the estimators, they themselves are of independent interest. 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, and may be view as an analogue of the results of Chen *et al.* (2003) which dealt with generalised method-of-moments estimation only. We validated the conditions of those two lemmas under adaptive varying-coefficient linear model (1.1) in terms of the empirical process theory of Doukhan *et al.* (1995).

A short overview of the paper is as follows: The model and the estimation method are stated in Section 2. Its asymptotic properties is 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 in section 5 the main results. A uniform convergence rate of the profile kernel regression estimator is established in the Appendix.

2 Estimation procedure

It is easy to see that model (1.1) is not identifiable, as we may replace the $(a_0, \mathbf{b}_0)$ by $(a_0 + c\,\boldsymbol{\alpha}_0^{\top}\mathbf{X}_t, \mathbf{b}_0 - c\,\boldsymbol{\alpha}_0)$ for any $c \in \mathbb{R}$. To overcome this problem, we represent the model in a reduced form

$$Y_t = a_0\{\boldsymbol{\alpha}_0^\top \mathbf{X}_t\} + \mathbf{X}_{t,-d}^\top \mathbf{b}_0\{\boldsymbol{\alpha}_0^\top \mathbf{X}_t\} + \varepsilon_t, \tag{2.1}$$

where $\mathbf{X}_{t,-d}$ is the remaining vector of $\mathbf{X}_t$ with its d-th component deleted. Note (1.1) may always be expressed in the form of (2.1) provided the last component of $\boldsymbol{\alpha}_0$ is non-zero. Furthermore, we assume that $||\boldsymbol{\alpha}_0|| = 1$, the first non-zero component of α_0 is positive, and

$$E(Y_t|\mathbf{X}_t = \mathbf{x}) \neq \alpha_0^\top \mathbf{x}\boldsymbol{\beta}^\top \mathbf{x} + \boldsymbol{\gamma}^\top \mathbf{x} + c$$

for some $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^d$ and $c \in \mathbb{R}^1$. Then $\boldsymbol{\alpha}_0, a_0(\cdot), \mathbf{b}_0(\cdot)$ in (2.1) are all identifiable; see Theorem 1 of Fan *et al.* (2003). From now on, we always assume those conditions.

With observations $\{(Y_t, \mathbf{X}_t),\ 1 \le t \le n\}$, Fan *et al.* (2003) proposed an iterative profile least squares estimation as follows.

1. With given $\boldsymbol{\alpha}$ and $Z_t = \boldsymbol{\alpha}^\top \mathbf{X}_{t,-d}$, minimise

$$\sum_{t=1}^{n} \left[Y_t - a - c(Z_t - z) - \{\mathbf{b} - \mathbf{d}(Z_t - z)\}^\top \mathbf{X}_{t,-d}\right]^2 K_h\{Z_t - z\} w\{Z_t\} \tag{2.2}$$

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

$$\begin{aligned} \widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha},h) \ &\equiv\ \widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha}) \equiv \{\widehat{a}(z,\boldsymbol{\alpha},h),\ \widehat{\mathbf{b}}(z,\boldsymbol{\alpha},h)^\top,\ \hat{\dot{a}}(z,\boldsymbol{\alpha},h),\ \hat{\dot{\mathbf{b}}}(z,\boldsymbol{\alpha},h)^\top\}^\top \\ &\equiv\ \{\widehat{a}(z,\boldsymbol{\alpha}),\ \widehat{\mathbf{b}}(z,\boldsymbol{\alpha})^\top,\ \hat{\dot{a}}(z,\boldsymbol{\alpha}),\ \hat{\dot{\mathbf{b}}}(z,\boldsymbol{\alpha})^\top\}^\top \ =\ (\widehat{a},\ \widehat{\mathbf{b}}^\top,\ \widehat{c},\ \widehat{\mathbf{d}}^\top)^\top, \end{aligned} \tag{2.3}$$

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

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

$$R(\boldsymbol{\beta}) = \frac{1}{n}\sum_{t=1}^{n} \left[Y_t - \widehat{a}\{\boldsymbol{\beta}^\top \mathbf{X}_t, \boldsymbol{\alpha}\} - \widehat{\mathbf{b}}\{\boldsymbol{\beta}^\top \mathbf{X}_t, \boldsymbol{\alpha}\}^\top \mathbf{X}_{t,-d}\right]^2 w(\boldsymbol{\alpha}^\top \mathbf{X}_t). \tag{2.4}$$

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

In the above expressions, $K(\cdot)$ is a kernel function, $K_h(\cdot) = h^{-1}K(\cdot/h)$, $h > 0$ is a bandwidth, $w(\cdot) = I_{[-L,L]}(\cdot)$ $(L > 0)$ is a weight function controling the edge effect in the estimation.

In order to ensure that the estimator $\widehat{\boldsymbol{\alpha}}$ is root-n consistent, the bandwidth h used in the iteration should be smaller than $O(n^{-1/5})$; see Theorem 3.1 below. Such a small h is not optimal for estimating coefficient functions a and $\mathbf{b}$ for which a different bandwidth $\bar{h}$ should be used.

For fixed α, the sampling properties of the estimator $\widehat{\boldsymbol{\theta}}$ defined in (2.3) follows the standard sampling theory of local linear regression estimation (Fan and Gijbels 1996, Fan and Yao 2003). However it is more challenging to develop the asymptotic properties of estimator $\widehat{\boldsymbol{\alpha}}$. One fundamental difficulty underlying the complexity is the lack of an explicit expression for $\widehat{\boldsymbol{\alpha}}$ which is defined in an iterative manner. To get around this difficulty, we slightly alter the definition of the estimator for $\boldsymbol{\alpha}$ and let

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} R_n\{\widehat{a}(\cdot,\boldsymbol{\alpha}), \widehat{\mathbf{b}}(\cdot,\boldsymbol{\alpha}), \boldsymbol{\alpha}\}, \tag{2.5}$$

where

$$R_n\{\widehat{a}(\cdot,\boldsymbol{\alpha}), \widehat{\mathbf{b}}(\cdot,\boldsymbol{\alpha}), \boldsymbol{\alpha}\} = \frac{1}{n}\sum_{t=1}^{n}\Big[Y_t - \widehat{a}\{\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}\} - \widehat{\mathbf{b}}\{\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}\}^\top \mathbf{X}_{t,-d}\Big]^2 w(\boldsymbol{\alpha}^\top \mathbf{X}_t). \tag{2.6}$$

It is easy to see the backfitting iteration of (2.2) – (2.4) is an approximate and computationally efficient way to evaluate $\widehat{\boldsymbol{\alpha}}$ defined in (2.5) while the definition (2.5) itself is theoretically more tractable. We sketch below how we will proceed with the theoretical investigation.

With $\boldsymbol{\alpha}$ given, the formula (2.2) divided by n is a consistent estimate of

$$\begin{aligned} R_z(a(\cdot), \mathbf{b}(\cdot), \boldsymbol{\alpha}) &= E\left\{\Big(Y_t - a(Z_t) - \mathbf{b}(Z_t)^\top \mathbf{X}_{t,-d}\Big)^2 w(Z_t) \,|\, Z_t = z\right\} \\ &= E\left\{\Big(Y_t - a(z) - \mathbf{b}(z)^\top \mathbf{X}_{t,-d}\Big)^2 w(z) \,|\, Z_t = z\right\}. \end{aligned} \tag{2.7}$$

Corresponding to (2.4), we define

$$R(a(\cdot), \mathbf{b}(\cdot), \boldsymbol{\alpha}) = E\left\{\Big(Y_t - a(Z_t) - \mathbf{b}(Z_t)^\top \mathbf{X}_{t,-d}\Big)^2 w(Z_t)\right\}, \tag{2.8}$$

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

$$R(a(\cdot), \mathbf{b}(\cdot), \boldsymbol{\alpha}) = \int R_z(a(\cdot), \mathbf{b}(\cdot), \boldsymbol{\alpha}) f_Z(z)\, dz = E\left\{\Big(Y_t - a(Z_t) - \mathbf{b}(Z_t)^\top \mathbf{X}_{t,-d}\Big)^2 w(Z_t)\right\}, \tag{2.9}$$

where $f_Z(z) = f_Z(z, \boldsymbol{\alpha})$ is the density function of $Z_t = Z_t(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \mathbf{X}_t$. Note that with $\boldsymbol{\alpha}$ given, the minimiser of (2.7) is

$$\begin{pmatrix} a_0(z,\boldsymbol{\alpha}) \\ \mathbf{b}_0(z,\boldsymbol{\alpha}) \end{pmatrix} = \Big[E\Big(\mathbb{X}_t\, \mathbb{X}_t^\top \Big| Z_t(\boldsymbol{\alpha}) = z\Big)\Big]^{-1} \left[E(\mathbb{X}_t Y_t | Z_t(\boldsymbol{\alpha}) = z)\right], \tag{2.10}$$

where

$$\mathbb{X}_t = (1, \mathbf{X}_{t,-d}^\top)^\top \text{with } \mathbf{X}_{t,-d} \text{ defined right after (2.1).} \tag{2.11}$$

It is easy to see from (2.9) that $\{a_0(\cdot, \boldsymbol{\alpha}),\ \mathbf{b}_0(\cdot, \boldsymbol{\alpha})\}$ is also the minimizer of (2.8) for any fixed $\boldsymbol{\alpha}$. Now the true value of the index parameter should satisfy

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

We may see intuitively that $\widehat{a}(z, \boldsymbol{\alpha})$ and $\widehat{\mathbf{b}}(z, \boldsymbol{\alpha})$ defined in (2.3) are consistent estimators of $a_0(z, \boldsymbol{\alpha})$ and $\mathbf{b}_0(z, \boldsymbol{\alpha})$ (see Theorem 3.2). The estimator $\widehat{\boldsymbol{\alpha}}$ defined in (2.5) is a consistent estimator of $\boldsymbol{\alpha}_0$ given in (2.12) (Theorem 3.1), as (2.7)is a consistent estimator of $R(a_0(\cdot, \boldsymbol{\alpha}), \mathbf{b}_0(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha})$. Finally, $\widehat{a}_0(z) \equiv \widehat{a}(z, \widehat{\boldsymbol{\alpha}}, \bar{h})$ and $\widehat{\mathbf{b}}_0(z) \equiv \widehat{\mathbf{b}}(z, \widehat{\boldsymbol{\alpha}}, \bar{h})$ (see (2.3)) are, respectively, the consistent estimators for $a_0(z) \equiv a_0(z, \boldsymbol{\alpha}_0)$ and $\mathbf{b}_0(z) \equiv \mathbf{b}_0(z, \boldsymbol{\alpha}_0)$ (Theorem 3.2).

3 Main results

3.1 Regularity conditions and notations.

We always assume $\{(Y_t, \mathbf{X}_t)\}$ is a strictly stationary process. Put

$$\mathbb{B} = \{\boldsymbol{\alpha} \in \mathbb{R}^d : \ \|\boldsymbol{\alpha}\| = 1, \text{the first non-zero element is positive, and the last element is non-zero}\},$$

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

(C1) (Moment conditions)

$E|Y_t|^{\varrho r} < \infty$, $E\|\mathbf{X}_t\|^{\varrho r} < \infty$ and $E|\varepsilon_t|^{\varrho r} < \infty$ for some integer $r > 1$ and some real number $\varrho > 4 - 2/r$. Furthermore, $\sup_{\boldsymbol{\alpha} \in \mathbb{B}} E|\varepsilon_t(\boldsymbol{\alpha})|^{\varrho r} < \infty$.

(C2) (Conditions on probability densities)

The density $f_{\boldsymbol{\alpha}^\top \mathbf{X}_t}(z)$ is continuous and bounded away from zero uniformly for $\boldsymbol{\alpha} \in \mathbb{B}$. Furthermore, the joint probability density function of $(\boldsymbol{\alpha}^\top \mathbf{X}_{t_1}, \cdots, \boldsymbol{\alpha}^\top \mathbf{X}_{t_s})$ exists and is bounded uniformly for any $t_1 < \cdots < t_s$ and $1 \le s \le 2r - 1$ and $\boldsymbol{\alpha} \in \mathbb{B}$, where r is 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 positively definite for $|z| \leq L$ and $\boldsymbol{\alpha} \in \mathbb{B}$, where $\mathbb{X}_t$ is defined in (2.10).

(C4) (Conditions on the nonparametric functions)

The functions $a_0(z, \boldsymbol{\alpha})$ and $\mathbf{b}_0(z, \boldsymbol{\alpha})$, defined in (2.10), 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 (2.8) 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 $a \geq (r\varrho - 2)r/(2 + r\varrho - 4r)$.

(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, it has a finite variance such that $|K(x) - K(y)| \leq C|x - y|$ for $x,\ y \in S_K$ and some $0 < C < \infty$.

(C7) (Conditions on the bandwidth)

The bandwidth h satisfies the conditions

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

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

Remark 1. Conditions (C1) and (C2) may appear to be stronger than the standard ones imposed for nonparametric regression estimation. This is due to the fact that we need to establish the uniform convergence for nonparametric regression estimators for $a(\cdot, \boldsymbol{\alpha})$ and $b(\cdot, \boldsymbol{\alpha})$ for given $\boldsymbol{\alpha} \in \mathbb{B}$ in order to obtain the root-n consistency for $\widehat{\boldsymbol{\alpha}}$. In fact the moment condition $E(e^{|\varepsilon_t|}) < \infty$ employed by Härdle *et al.* (1993) and Xia and Li (1999) is stronger than the moment conditions in (C1). The β–mixing condition (C5) is not very strong either. Many linear and nonlinear time series satisfy this condition; see, for example, section 2.6 of Fan and Yao (2003). The bandwidth condition (C7) is also standard for this type of problem. Note (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 $r > 4$.

Before we end this section, we define some notation that will be used in the rest of the paper. Let $X_{t0} \equiv 1$, $\mathcal{D}_n = \{1, 2, \cdots, n\}$, $S_w = [-L, L]$, $\mu_{i,K} = \int u^i K(u)\, du$ 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 $\mathbf{s} = (\mu_{2,K}, \mu_{3,K})^\top$ be a 2×1 vector.

Put $\boldsymbol{\theta}_0(z, \boldsymbol{\alpha}) = (a_0(z, \boldsymbol{\alpha}), \mathbf{b}_0(z, \boldsymbol{\alpha})^\top, \dot{a}_0(z, \boldsymbol{\alpha}), \dot{\mathbf{b}}_0(z, \boldsymbol{\alpha})^\top)^\top$, 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, we 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

$$\mathbf{g} = (\mathbf{g}^1, \cdots, \mathbf{g}^d)^\top = \mathbf{g}(z, \boldsymbol{\alpha}) = (a(z, \boldsymbol{\alpha}), \mathbf{b}(z, \boldsymbol{\alpha})^\top)^\top. \tag{3.2}$$

$\mathbf{g}_0 = \mathbf{g}_0(z, \boldsymbol{\alpha})$ and $\widehat{\mathbf{g}} = \widehat{\mathbf{g}}(z, \boldsymbol{\alpha})$ are defined in the 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 $\mathbf{g}_1^j = \mathbf{g}_1^j(z, \boldsymbol{\alpha}) = \partial \mathbf{g}^j(z, \boldsymbol{\alpha})/\partial z$, and $\mathbf{g}_2 = \mathbf{g}_2(z, \boldsymbol{\alpha})$ the $d \times d$ matrix whose (i,j)-th element $\mathbf{g}_2^{ij} = \mathbf{g}_2^{ij}(z, \boldsymbol{\alpha}) = \partial \mathbf{g}^i(z, \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^j$. Similarly, we define $\mathbf{g}_{01} = \mathbf{g}_{01}(z, \boldsymbol{\alpha}) = \partial \mathbf{g}_0(z, \boldsymbol{\alpha})/\partial z$, $\mathbf{g}_{02} = \mathbf{g}_{02}(z, \boldsymbol{\alpha}) = \partial \mathbf{g}_0(z, \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^\top$ and $\widehat{\mathbf{g}}_1 = \widehat{\mathbf{g}}_1(z, \boldsymbol{\alpha}) = \partial \widehat{\mathbf{g}}(z, \boldsymbol{\alpha})/\partial z$, $\widehat{\mathbf{g}}_2 = \widehat{\mathbf{g}}_2(z, \boldsymbol{\alpha}) = \partial \widehat{\mathbf{g}}(z, \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^\top$.

The Euclidean norm of $\mathbf{g}$ is denoted as before by $\|\mathbf{g}\| = (\mathbf{g}^\top \mathbf{g})^{1/2}$. 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 assumption (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 will define a function space $\mathcal{G}$ in Subsection 5.2. Clearly, under assumption (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 (2.3) 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 $\widehat{\boldsymbol{\alpha}}$, defined in (2.5), with the standard root-n convergence rate provided $h = o(n^{-1/4})$.

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

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

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 sign of Kroneck product,

$$\mathbf{B}(z) = \left\{\left(\mathbf{S}^{-1}\mathbf{s}\right) \otimes I_{d\times d}\right\}(\ddot{a}_0(z, \boldsymbol{\alpha}), \ddot{\mathbf{b}}_0(z, \boldsymbol{\alpha})^\top)^\top = \Big((\ddot{a}_0(z, \boldsymbol{\alpha}), \ddot{\mathbf{b}}_0(z, \boldsymbol{\alpha})^\top)\mu_{2,K}, 0, \cdots, 0\Big)^\top \in \mathbb{R}^{2d}$$

and

$$\mathbf{A}(z) = \{f_Z(z)\}^{-1}\left(\mathbf{S}^{-1}\widetilde{\mathbf{S}}\mathbf{S}^{-1}\right) \otimes \left(\mathbf{G}^{-1}(z)\widetilde{\mathbf{G}}(z)\mathbf{G}^{-1}(z)\right),$$

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 $\widetilde{G}_{ij}(z) = E(\varepsilon_t(\boldsymbol{\alpha})^2 X_{t,i-1}X_{t,j-1}|Z_t = z)$ 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 *Let conditions (C1)-(C7) hold. Set $Z_t^o = \boldsymbol{\alpha}_0^\top \mathbf{X}_t$. 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*

$$\sqrt{n}\left\{\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 + \Gamma_1^- \mathcal{B} h^2(1 + o_P(1))\right\} \xrightarrow{D} N\left(\mathbf{0}, \Gamma_1^- \mathcal{V}(\Gamma_1^-)^\top\right), \tag{3.4}$$

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

$$\mathcal{B} = E\left(\left\{\ddot{a}_0(Z_t^o) + \ddot{\mathbf{b}}_0(Z_t^o)^\top \mathbf{X}_{t,-d}\right\}\mathbf{U}_t\right)\mu_{0,K}^{-1}\mu_{2,K},$$
$$\Gamma_1 = 2E\left[\left\{\mathbf{g}_{01t}\mathbf{X}_t^\top + \mathbf{g}_{02t}\right\}^\top \mathbb{X}_t\mathbb{X}_t^\top \left\{\mathbf{g}_{01t}\mathbf{X}_t^\top + \mathbf{g}_{02t}\right\}\right] w(Z_t^o),$$
$$\mathcal{V} = E\varepsilon_t^2\left[\Xi_t\Xi_t^\top - \{E(\Xi_t\mathbb{X}_t^\top|Z_t^o)\}\{E(\mathbb{X}_t\mathbb{X}_t^\top|Z_t^o)\}^{-1}\{E(\mathbb{X}_t\Xi_t^\top|Z_t^o)\}\right],$$

and Γ_1^- *is a generalized inverse of* Γ_1, *with* $\mathbf{U}_t = E\left(\mathbf{X}_t\mathbf{g}_{01t}^\top\mathbb{X}_t\mathbb{X}_t^\top|Z_t^o\right)\mathbf{G}_0^{-1}(Z_t^o)\mathbb{X}_t w(Z_t^o) + \mathbf{g}_{02t}\mathbb{X}_t w(Z_t^o)$, $\Xi_t = \mathbf{X}_t\left\{\dot{a}_0(Z_t^o) + \dot{\mathbf{b}}_0(Z_t^o)^\top\mathbf{X}_{t,-d}\right\}w(Z_t^o)$, *and* $\mathbf{G}_0(z)$ *is a* $d \times d$ *matrix whose* (i,j)*th elements* $G_{ij}^0(z) = E(X_{t,i-1}X_{t,j-1}|Z_t^o = z)$.

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

$$\sqrt{n}\,\{\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\} \xrightarrow{D} N\left(\mathbf{0}, \Gamma_1^-\mathcal{V}(\Gamma_1^-)^\top\right). \tag{3.5}$$

Corollary 3.3 *Under the conditions of Theorem 3.2, with* $\widehat{\boldsymbol{\alpha}}$ *defined in (2.5) with* $h = o(n^{-1/4})$ *as the estimator of* $\boldsymbol{\alpha}$ *in Theorem 3.1, we have*

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

where $\bar{h} = O(n^{-1/5})$, $\bar{H}_n = \mathrm{diag}(1, \bar{h}) \otimes I_{d\times d}$, *and* $\mathbf{B}_0(z)$ *and* $\mathbf{A}_0(z)$ *are defined in the same way as* $\mathbf{B}(z)$ *and* $\mathbf{A}(z)$ *with* $\boldsymbol{\alpha}$ *replaced by* $\boldsymbol{\alpha}_0$.

This corollary easily follows from Theorems 3.1 and 3.2.

Remark 2. (i) The estimator $\widehat{\boldsymbol{\theta}}(z, \widehat{\boldsymbol{\alpha}}, \bar{h})$ is asymptotically adaptive to unknown $\boldsymbol{\alpha}_0$ in the sense that $\widehat{\boldsymbol{\theta}}(z, \widehat{\boldsymbol{\alpha}}, \bar{h})$ and $\widehat{\boldsymbol{\theta}}(z, \boldsymbol{\alpha}_0, \bar{h})$ 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 $\widehat{\boldsymbol{\theta}}(z, \boldsymbol{\alpha}, \bar{h})$ 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 important lemmas. Those two lemmas are of independent interest as we do not make use of the specific forms of $\mathbf{g}(z, \boldsymbol{\alpha})$ and $\mathbb{B}$ in the proofs. Therefore 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 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 will specify $\mathbb{B}$ and $\mathcal{G}$ with the norm $\|\mathbf{g}\|_{\mathcal{G}}$ in the context of the model (2.1).

Lemma 4.1 *Suppose that* $\boldsymbol{\alpha}_0 \in \mathbb{B}$ *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:*

(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 Chen *et al* (2003). By condition (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

$$
\begin{aligned}
& R\left(\mathbf{g}_0(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)-R\left(\mathbf{g}_0(\cdot), \boldsymbol{\alpha}_0\right) \\
& \quad=R\left(\mathbf{g}_0(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)-R\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right) \qquad (4.2)\\
& \qquad+R\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)-R_n\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right) \qquad (4.3)\\
& \qquad+R_n\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)-R\left(\mathbf{g}_0(\cdot), \boldsymbol{\alpha}_0\right). \qquad (4.4)
\end{aligned}
$$

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

$$
\sup_{\boldsymbol{\alpha} \in \mathbb{B}}\left|R\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)-R_n\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)\right|=o_P(1),
$$

which follows from conditions (iv) and (v). Finally, we have to show that the expression in (4.4) tends to 0 in probability. As $R_n\left(\widehat{\mathbf{g}}(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)=\inf_{\boldsymbol{\alpha} \in \mathbb{B}} R_n\left(\widehat{\mathbf{g}}(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}\right)$ and note that

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

we have

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

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

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

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

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

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

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

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

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

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

$$\Gamma_2(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})[\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})] = \lim_{\tau\to 0}\left[\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha})+\tau(\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})\right]/\tau \tag{4.9}$$

(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 the 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:*

(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) $\Gamma_1=\Gamma_1(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)$ *is of a 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* $\delta_n = o(1)$,

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

(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 3. The objective function defining the semi-parametric estimators in Chen *et al.* (2003) is of a GMM (generalized method of moments) type and hence their general Theorem 2 does not apply directly to the least-squares (in this paper) or maximum-likelihood-like semi-parametric estimators. Their argument however can be 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 and modified from theirs: 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 only need 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 as that of $\dot{R}\left(\mathbf{g}_0(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)$ tending to 0. Similarly to (5) and (6) of Chen *et al.* (2003), it is obvious that

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

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

Clearly, conditions (iii)(1) imply $D_{11} = o_P(n^{-1/2})$; 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 $D_3 = o_P(n^{-1/2})$, and condition (v) implies $D_2 = o_P(n^{-1/2})$. Therefore it follows that $\|\dot{R}\left(\mathbf{g}_0(\cdot, \widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right)\| \times (1 - o_P(1)) = O_P(n^{-1/2})$, and hence $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = O_P(n^{-1/2})$.

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

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

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 $\dot{R}\left(\mathbf{g}_0(\cdot, \boldsymbol{\alpha}_0), \boldsymbol{\alpha}_0\right) = 0$, 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 $D_7 = o_P(n^{-1/2})$; and condition (i) implies $D_8 = o_P(n^{-1/2})$. Therefore $\mathcal{L}_n(\boldsymbol{\alpha}) = o_P(n^{-1/2})$, which leads to

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

and hence the lemma follows from condition (vi). □.

5 Proof of main results

The following lemma is basic and is used throughout.

Lemma 5.1 *Let* $\varepsilon_t(\boldsymbol{\alpha}) = Y_t - a_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{b}_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbf{X}_{t,-d} = Y_t - \mathbf{g}_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbb{X}_t$. *Then 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*

$$E\varepsilon_t(\boldsymbol{\alpha})\{a(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) + \mathbf{b}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbf{X}_{t,-d}\} = E\varepsilon_t(\boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbb{X}_t = 0, \tag{5.1}$$

where $\mathbb{X}_t$ was defined in (2.11).

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

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

and that by the definition of $a_0(\cdot, \boldsymbol{\alpha})$ and $\mathbf{b}_0(\cdot, \boldsymbol{\alpha})$ in (2.10),

$$E\left\{\left(Y_t - a_0(z, \boldsymbol{\alpha}) - \mathbf{b}_0(z, \boldsymbol{\alpha})^\top \mathbf{X}_{t,-d}\right)\mathbb{X}_t\Big|\, \boldsymbol{\alpha}^\top \mathbf{X}_t = z\right\} = 0.$$

Therefore (5.1) follows. □

5.1 Proof for Theorem 3.1.

It follows from (2.2) by least squares that

$$\widehat{\theta}(z, \boldsymbol{\alpha}) = \widehat{\theta}(z, \boldsymbol{\alpha}, h) = H_n^{-1}\left(\widehat{a}_z, \widehat{\mathbf{b}}_z^\top, \widehat{\dot{a}}_z h, \widehat{\dot{\mathbf{b}}}_z^\top h\right)^\top = H_n^{-1}\left\{\mathcal{X}(z)^\top \mathcal{W}(z)\mathcal{X}(z)\right\}^{-1}\left\{\mathcal{X}(z)^\top \mathcal{W}(z)\mathcal{Y}\right\}, \tag{5.2}$$

where $\mathcal{Y} = (Y_1, \cdots, Y_n)^\top$, $\mathcal{W}(z) = \mathcal{W}(z, \boldsymbol{\alpha})$ 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 $(\mathbb{X}_t^\top, h^{-1}(Z_t - z)\mathbb{X}_t^\top)$ as its tth row and $\mathbb{X}_t = (1, \mathbf{X}_{t,-d}^\top)^\top$, and $H_n = \mathrm{diag}(1, h) \otimes I_{d\times d}$.

Denote by $\widehat{\Phi} = \widehat{\Phi}(z; \boldsymbol{\alpha}) = n^{-1}\mathcal{X}(z)^\top \mathcal{W}(z)\mathcal{X}(z)$ and $\widehat{\Psi} = \widehat{\Psi}(z; \boldsymbol{\alpha}) = n^{-1}\mathcal{X}(z)^\top \mathcal{W}(z)\mathcal{Y}$ with (i, j)-th elements $\widehat{\Phi}_{i,j}$ and $\widehat{\Psi}_{i,j}$, respectively. Also, recall $X_{t,0} \equiv 1$ for notational convenience.

Then with the notations in (5.2), we have, for $i, j = 1, \cdots, d$,

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

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

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

and

$$\widehat{\Psi}_i = n^{-1} \sum_{t=1}^{n} Y_t X_{t,i-1} K_h(Z_t - z) w(Z_t), \tag{5.6}$$

$$\widehat{\Psi}_{d+i} = n^{-1} \sum_{t=1}^{n} Y_t X_{t,i-1} ((Z_t - z)/h) K_h(Z_t - z) w(Z_t). \tag{5.7}$$

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

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

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,

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

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

$$Y_t^* = Y_t^*(z, \boldsymbol{\alpha}) = Y_t - \left\{ a_0(z, \boldsymbol{\alpha}) + \mathbf{X}_{t,-d}^\top \mathbf{b}_0(z, \boldsymbol{\alpha}) \right\} - \left\{ \dot{a}_0(z, \boldsymbol{\alpha}) + \mathbf{X}_{t,-d}^\top \dot{\mathbf{b}}_0(z, \boldsymbol{\alpha}) \right\} (Z_t - z).$$

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

$$Y_t^* = \frac{1}{2} \left\{ \ddot{a}_0(\xi, \boldsymbol{\alpha}) + \mathbf{X}_{t,-d}^\top \ddot{\mathbf{b}}_0(\xi, \boldsymbol{\alpha}) \right\} (Z_t - z)^2 + \varepsilon_t(\boldsymbol{\alpha}), \tag{5.11}$$

where $\varepsilon_t(\boldsymbol{\alpha}) = Y_t - a_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{b}_0(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbf{X}_{t,-d}$, $\ddot{a}_0(z, \boldsymbol{\alpha})$ and $\ddot{\mathbf{b}}_0(z, \boldsymbol{\alpha})$ are the second partial derivatives of $a_0(z, \boldsymbol{\alpha})$ and $\mathbf{b})_0(z, \boldsymbol{\alpha})$ with respect to z, respectively, and $\xi = z + \eta\ (Z_t - z)$ with $|\eta| < 1$.

With $\bar{g}_{ij}(z) = \bar{g}_{ij}(z, \boldsymbol{\alpha}) = E\{X_{t,i} X_{t,j} \mid Z_t(\boldsymbol{\alpha}) = z\}$ for $i, j = 0, 1, \cdots, d$, we denote by $\bar{g}_i(z) = \bar{g}_{i0}(z)$, and $G_i(z) = (\bar{g}_{i1}(z), \cdots, \bar{g}_{i,d-1}(z))^\top$ (a $(d-1)$-dimensional vector) . Also, as $\mu_{i,K} =$

$\int u^i K(u)\, du$, then, 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 $i, j = 1, \cdots, d$,

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

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

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

and

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

and

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

where

$$B_{i-1}(z) = \frac{1}{2} \left\{ \ddot{a}(z) \bar{g}_{i-1}(z) + \ddot{b}(z)^{\top} G_{i-1}(z) \right\} w(z) f_Z(z).$$

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

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

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 equal to $G_{ij}(z) = \bar{g}_{i-1,j-1}(z)$ for $i, j = 1, 2, \cdots, d$.

Further, recall $\nu_{i,K} = \int u^i K^2(u)\, du$, 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

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

where

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

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

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

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

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

Here

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

$\mathbf{V}(z)$ is the root matrix of $\mathbf{V}^{(2)}(z)$, i.e.

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

with $\mathbf{V}$ a $d \times d$ matrix such that $\mathbf{V}^\top \mathbf{V} = \mathbf{V}^{(2)}$; ξ_N is a $(2d)$-dimensional random vector of standard multivariate normal distribution, the proof of which is a routine by the argument of CLT for strong mixing processes based on the Bernstein blocking technique: see, e.g., Hallin *et al.* (2004, Theorem 3.1) and Lu and Linton (2004), and therefore the detail is omitted.

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

$$\mathbf{B}(z) = \Phi^{-1}(z) \mathbf{U}(z) = \left\{ \left(\mathbf{S}^{-1} \mathbf{s} \right) \otimes I_{d\times d} \right\} (\ddot{a}(z), \ddot{\mathbf{b}}(z)^\top)^\top$$

and

$$\begin{aligned} \mathbf{A}(z) &= \Phi^{-1}(z) \mathbf{V}(z) \left(\Phi^{-1}(z) \mathbf{V}(z) \Phi^{-1}(z) \right)^\top = \Phi^{-1}(z) \mathbf{V}^{(2)}(z) \Phi^{-1}(z) \\ &= \{f_Z(z)\}^{-1} \left(\mathbf{S}^{-1} \widetilde{\mathbf{S}} \mathbf{S}^{-1} \right) \otimes \left(\mathbf{G}^{-1}(z) \widetilde{\mathbf{G}}(z) \mathbf{G}^{-1}(z) \right). \end{aligned}$$

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

5.2 Proof for Theorem 3.2.

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

5.2.1 Preliminaries.

With the notations defined in Subsection 3.1, we are in a position to define $\mathcal{G}$. For some $c_0 > 0$,

$$\begin{aligned}\mathcal{G} = \{\mathbf{g} : & S_w \times \mathbb{B} \mapsto \mathbb{R}^d \mid \text{ For any fixed } \boldsymbol{\alpha} \in \mathbb{B},\ \mathbf{g}(\cdot, \boldsymbol{\alpha}) \in C^2_{c_0}(S_w),\ \mathbf{g}_1(\cdot, \boldsymbol{\alpha}) \in C^1_{c_0}(S_w) \\ & \text{and } \mathbf{g}_2(\cdot, \boldsymbol{\alpha}) \in C^1_{c_0}(S_w), \text{ and for any } z \in S_w,\ \|\mathbf{g}(z, \boldsymbol{\alpha}) - \mathbf{g}(z, \boldsymbol{\alpha}')\| \le C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|, \\ & \|\mathbf{g}_1(z, \boldsymbol{\alpha}) - \mathbf{g}_1(z, \boldsymbol{\alpha}')\| \le C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| \text{ and } \|\mathbf{g}_2(z, \boldsymbol{\alpha}) - \mathbf{g}_2(z, \boldsymbol{\alpha}')\| \le C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| \\ & \text{for any } \boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathbb{B}\},\end{aligned} \tag{5.21}$$

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

As defined in (2.8),

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

where $\mathbf{x}^{(-d)}$ 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

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

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

$$\begin{aligned}\dot{R}(\mathbf{g}(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}) &= -2E\left(Y_t - \mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbb{X}_t\right) \left\{\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) \mathbf{X}_t^\top + \mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\right\}^\top \mathbb{X}_t w(\boldsymbol{\alpha}^\top \mathbf{X}_t), \\ \dot{R}_n(\mathbf{g}(\cdot, \boldsymbol{\alpha}), \boldsymbol{\alpha}) &= -\frac{2}{n} \sum_{t=1}^n \left(Y_t - \mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})^\top \mathbb{X}_t\right) \left\{\mathbf{g}_1(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) \mathbf{X}_t^\top + \mathbf{g}_2(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha})\right\}^\top \mathbb{X}_t w(\boldsymbol{\alpha}^\top \mathbf{X}_t).\end{aligned} \tag{5.22}$$

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

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

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)

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

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 $\varepsilon(\boldsymbol{\alpha}_0) = Y_t - \mathbf{g}_0(Z_t^o,\boldsymbol{\alpha}_0)^\top\mathbb{X}_t = \varepsilon_t$, as assumed, satisfies $E(\varepsilon_t|\mathbf{X}_t) = 0$, hence

$$\begin{aligned}
\Gamma_1 &= \Gamma_1(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)\\
&= 2E\left[\left\{\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top + \mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\right\}^\top\mathbb{X}_t\mathbb{X}_t^\top\left\{\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top + \mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\right\}\right] w(Z_t^o). \qquad (5.23)
\end{aligned}$$

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 - \left(\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}))\right)^\top\mathbb{X}_t\right\}\\
&\qquad\times\left\{\left(\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}))\right)\mathbf{X}_t^\top\right.\\
&\qquad\left. + \left(\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}))\right)\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\{\left(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top\right) + \left(\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\right)\right\}^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\\
&\quad= -2\tau E\left\{\varepsilon_t(\boldsymbol{\alpha})\left((\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\right)\right.\\
&\qquad\left. - \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)\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)\right\}^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\\
&\qquad + 2\tau^2 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\times\left((\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}))\right)^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t), \qquad (5.24)
\end{aligned}$$

where $\varepsilon_t(\boldsymbol{\alpha}) = Y_t - \mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})^\top\mathbb{X}_t$. 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

$$\begin{aligned}
&\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)\left[\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})\right] = \lim_{\tau\to 0}\frac{\dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}) + \tau(\mathbf{g}(\cdot,\boldsymbol{\alpha}) - \mathbf{g}_0(\cdot,\boldsymbol{\alpha})),\boldsymbol{\alpha}) - \dot{R}(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})}{\tau}\\
&\quad= -2E\left\{\varepsilon_t(\boldsymbol{\alpha})\left((\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\mathbf{X}_t^\top\right)\right.\\
&\qquad\left. - \left((\mathbf{g}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))^\top\mathbb{X}_t\right)\left(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top + \mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\right)\right\}^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t), \qquad (5.25)
\end{aligned}$$

and therefore

$$
\begin{aligned}
&\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\left[\mathbf{g}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)\right]\\
&\quad= 2E\left((\mathbf{g}(Z_t^o,\boldsymbol{\alpha}_0)-\mathbf{g}_0(Z_t^o,\boldsymbol{\alpha}_0))^{\top}\mathbb{X}_t\right)\left(\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^{\top}+\mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\right)^{\top}\mathbb{X}_t w(Z_t^o)\\
&\quad= 2\int \widetilde{\Phi}_0(z)\left(\mathbf{g}(z,\boldsymbol{\alpha}_0)-\mathbf{g}_0(z,\boldsymbol{\alpha}_0)\right)w(z)f_{Z_t^o}(z)dz, \qquad (5.26)
\end{aligned}
$$

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 the consistency of $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}_0$ by Lemma 4.1.

5.2.2 Proof 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 $w(\cdot)$,

$$
\begin{aligned}
&\sup_{\boldsymbol{\alpha}\in\mathbb{B}}|R(\mathbf{g}(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})-R(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha})|\\
&\quad\le \sup_{\boldsymbol{\alpha}\in\mathbb{B}}\left|E\left(2\varepsilon(\boldsymbol{\alpha})-(\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t\right)(\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t w(\boldsymbol{\alpha}^{\top}\mathbf{X}_t)\right|\\
&\quad\le \sup_{\boldsymbol{\alpha}\in\mathbb{B}}\left|E\left((\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t\right)(\mathbf{g}(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(\boldsymbol{\alpha}^{\top}\mathbf{X}_t,\boldsymbol{\alpha}))^{\top}\mathbb{X}_t w(\boldsymbol{\alpha}^{\top}\mathbf{X}_t)\right|\\
&\quad\le C\|\mathbf{g}-\mathbf{g}_0\|_{\mathcal{G}}\|E\mathbb{X}_t\mathbb{X}_t^{\top}\|, \qquad (5.27)
\end{aligned}
$$

where the final inequality follows from the definition of norm $\|\cdot\|_{\mathcal{G}}$ in Subsection 3.1. (iv) follows clearly from Lemma 6.3 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 (5.27) 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.

5.2.3 Proof 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 6.3 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 $\mathbf{g}=(a,\mathbf{b}^{\mathsf T})^{\mathsf T}$ lie in $\mathbb{B}_\delta$ and $\mathcal{G}_\delta$, respectively, with $\delta=\delta_n\to 0$, where

$$\begin{aligned}&\mathbb{B}_\delta=\{\boldsymbol\alpha\in\mathbb{B}:\|\boldsymbol\alpha-\boldsymbol\alpha_0\|\le\delta\},\\ &\mathcal{G}_\delta=\{\mathbf{g}\in\mathcal{G}:\|\mathbf{g}-\mathbf{g}_0\|_{\mathcal{G}}\le\delta,\|\mathbf{g}_1-\mathbf{g}_{01}\|_{\mathcal{G}}\le\delta,\|\mathbf{g}_2-\mathbf{g}_{02}\|_{\mathcal{G}}\le\delta\}.\end{aligned}\tag{5.28}$$

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:

<u>(i)</u>: 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 $\dot R_n(\widehat{\mathbf{g}}(\cdot,\widehat{\boldsymbol\alpha}),\widehat{\boldsymbol\alpha})=0$.

<u>(ii)</u>: Both (ii)(1)-(2) are clear from Assumption (C4) in Section 3.

<u>(iii)</u>: It follows from (5.24) with $\tau=1$ and (5.26) that

$$\begin{aligned}&\mathbf{c}^{\mathsf T}\left\{\dot R(\widehat{\mathbf{g}}(\cdot,\boldsymbol\alpha),\boldsymbol\alpha)-\dot R(\mathbf{g}_0(\cdot,\boldsymbol\alpha),\boldsymbol\alpha)-\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol\alpha),\boldsymbol\alpha\right)\left[\widehat{\mathbf{g}}(\cdot,\boldsymbol\alpha)-\mathbf{g}_0(\cdot,\boldsymbol\alpha)\right]\right\}\\ &\quad=2\mathbf{c}^{\mathsf T}E\left\{(\widehat{\mathbf{g}}(\boldsymbol\alpha^{\mathsf T}\mathbf{X}_t,\boldsymbol\alpha)-\mathbf{g}_0(\boldsymbol\alpha^{\mathsf T}\mathbf{X}_t,\boldsymbol\alpha))^{\mathsf T}\mathbb{X}_t\right\}\\ &\qquad\times\Big((\widehat{\mathbf{g}}_1(\boldsymbol\alpha^{\mathsf T}\mathbf{X}_t,\boldsymbol\alpha)-\mathbf{g}_{01}(\boldsymbol\alpha^{\mathsf T}\mathbf{X}_t,\boldsymbol\alpha))\mathbf{X}_t^{\mathsf T}+(\widehat{\mathbf{g}}_2(\boldsymbol\alpha^{\mathsf T}\mathbf{X}_t,\boldsymbol\alpha)-\mathbf{g}_{02}(\boldsymbol\alpha^{\mathsf T}\mathbf{X}_t,\boldsymbol\alpha))\Big)^{\mathsf T}\mathbb{X}_t w(\boldsymbol\alpha^{\mathsf T}\mathbf{X}_t)\\ &\quad=2\int(\widehat{\mathbf{g}}(z,\boldsymbol\alpha)-\mathbf{g}_0(z,\boldsymbol\alpha))^{\mathsf T}E\{\mathbb{X}_t\mathbf{c}^{\mathsf T}\mathbf{X}_t\mathbb{X}_t^{\mathsf T}|Z_t(\boldsymbol\alpha)=z\}(\widehat{\mathbf{g}}_1(z,\boldsymbol\alpha)-\mathbf{g}_{01}(z,\boldsymbol\alpha))w(z)f_Z(z,\boldsymbol\alpha)dz\\ &\qquad+2\int(\widehat{\mathbf{g}}(z,\boldsymbol\alpha)-\mathbf{g}_0(z,\boldsymbol\alpha))^{\mathsf T}E\{\mathbb{X}_t\mathbb{X}_t^{\mathsf T}|Z_t(\boldsymbol\alpha)=z\}(\widehat{\mathbf{g}}_2(z,\boldsymbol\alpha)-\mathbf{g}_{02}(z,\boldsymbol\alpha))\mathbf{c}w(z)f_Z(z,\boldsymbol\alpha)dz\\ &\quad\equiv\widetilde D_1+\widetilde D_2,\end{aligned}\tag{5.29}$$

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 6.3) and then (5.17) that

$$\begin{aligned}
\widehat{\mathbf{g}}(z,\boldsymbol{\alpha})-\mathbf{g}_0(z,\boldsymbol{\alpha}) &= \gamma(\widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha})-\boldsymbol{\theta}_0(z,\boldsymbol{\alpha}))\\
&= \gamma\widehat{\Phi}^{-1}\widehat{\mathbf{W}} = (1+o_P(1))\gamma\Phi^{-1}\widehat{\mathbf{W}}\\
&= (1+o_P(1))(\mu_{0,K}w(z)f_Z(z,\boldsymbol{\alpha}))^{-1}\mathbf{G}^{-1}(z,\boldsymbol{\alpha})\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}), \qquad (5.30)
\end{aligned}$$

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

$$\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}) = n^{-1}\sum_{t=1}^{n} Y_t^*(z,\boldsymbol{\alpha})\mathbb{X}_t K_h(Z_t(\boldsymbol{\alpha}),-z)w(Z_t(\boldsymbol{\alpha})), \qquad (5.31)$$

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

$$\begin{aligned}
Y_t^*(z,\boldsymbol{\alpha}) &= \varepsilon_t(\boldsymbol{\alpha}) + (\mathbf{g}_0(Z_t(\boldsymbol{\alpha}),\boldsymbol{\alpha})-\mathbf{g}_0(z,\boldsymbol{\alpha})-\dot{\mathbf{g}}_0(z,\boldsymbol{\alpha})(Z_t(\boldsymbol{\alpha})-z))^{\mathsf{T}}\,\mathbb{X}_t\\
&= \varepsilon_t(\boldsymbol{\alpha}) + \frac{1}{2}\left(\ddot{\mathbf{g}}_0(z+\eta(Z_t(\boldsymbol{\alpha})-z),\boldsymbol{\alpha})(Z_t(\boldsymbol{\alpha})-z)^2\right)^{\mathsf{T}}\mathbb{X}_t \qquad (5.32)
\end{aligned}$$

with $|\eta|<1$. 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 (5.30), together with (5.32)

$$\begin{aligned}
\widetilde{D}_1 &= 2\int((1+o_P(1))(\mu_{0,K}w(z)f_Z(z,\boldsymbol{\alpha}))^{-1}\mathbf{G}^{-1}(z,\boldsymbol{\alpha})\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}))^{\mathsf{T}}G_{\mathbf{c}}(z,\boldsymbol{\alpha})\\
&\quad\times(\widehat{\mathbf{g}}_1(z,\boldsymbol{\alpha})-\mathbf{g}_{01}(z,\boldsymbol{\alpha}))w(z)f_Z(z,\boldsymbol{\alpha})dz\\
&= 2(1+o_P(1))(\mu_{0,K})^{-1}\int\mathbf{G}^{-1}(z,\boldsymbol{\alpha})\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}))^{\mathsf{T}}G_{\mathbf{c}}(z,\boldsymbol{\alpha})(\widehat{\mathbf{g}}_1(z,\boldsymbol{\alpha})-\mathbf{g}_{01}(z,\boldsymbol{\alpha}))dz\\
&= \frac{1}{\sqrt{n}}(\nu_n(\widehat{\mathbf{g}}_1,\boldsymbol{\alpha})-\nu_n(\mathbf{g}_{01},\boldsymbol{\alpha}))+O_P(h^2)\|\widehat{\mathbf{g}}_1-\mathbf{g}_{01}\|_{\mathcal{G}} \qquad (5.33)
\end{aligned}$$

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 $\frac{1}{\sqrt{n}}(\nu_n(\widehat{\mathbf{g}}_1,\boldsymbol{\alpha})-\nu_n(\mathbf{g}_{01},\boldsymbol{\alpha}))\le\frac{1}{\sqrt{n}}\sup_{\|\mathbf{g}_1-\mathbf{g}_{01}\|_{\mathcal{G}}\le\delta}\|\nu_n(\mathbf{g}_1,\boldsymbol{\alpha})-\nu_n(\mathbf{g}_{01},\boldsymbol{\alpha})\| = o_P(n^{-1/2})$; 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 (5.29).

In addition, it follows from (5.26) together with Lemma 5.1 and $E(\varepsilon_t(\boldsymbol{\alpha}_0)|\mathbf{X}_t) = 0$ that

$$\begin{aligned}
&\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)\left[\mathbf{g}(\cdot,\boldsymbol{\alpha})-\mathbf{g}_0(\cdot,\boldsymbol{\alpha})\right]-\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\left[\mathbf{g}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)\right]\\
&\quad=-2E\left(\varepsilon_t(\boldsymbol{\alpha})-\varepsilon_t(\boldsymbol{\alpha}_0)\right)\left((\mathbf{g}_1(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha}))\mathbf{X}_t^\top\right)^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\\
&\qquad+2E\left\{\delta_t(\boldsymbol{\alpha})\left(\mathbf{g}_{01}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\mathbf{X}_t^\top+\mathbf{g}_{02}(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})\right)^\top\mathbb{X}_t w(\boldsymbol{\alpha}^\top\mathbf{X}_t)\right.\\
&\qquad\quad\left.-\,\delta_t(\boldsymbol{\alpha}_0)\left(\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)\mathbf{X}_t^\top+\mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\right)^\top\mathbb{X}_t w(Z_t^o)\right\}\\
&\quad\equiv\Omega_1+\Omega_2,
\end{aligned}$$

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

$$\begin{aligned}
|\varepsilon_t(\boldsymbol{\alpha})-\varepsilon_t(\boldsymbol{\alpha}_0)| &= |\left(\mathbf{g}_0(\boldsymbol{\alpha}^\top\mathbf{X}_t,\boldsymbol{\alpha})-\mathbf{g}_0(Z_t^o,\boldsymbol{\alpha}_0)\right)^\top\mathbb{X}_t|\\
&\le C\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|(1+\|\mathbf{X}_t\|)\|\mathbb{X}_t\|,
\end{aligned}$$

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

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

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

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

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 6.3, that

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

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

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

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.

<u>(v)</u>: 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 $m_{1t}(\mathbf{g}, \boldsymbol{\alpha}) = Y_t - \mathbf{g}(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t, \boldsymbol{\alpha})^{\mathsf{T}}\mathbb{X}_t$, $m_{2t}(\mathbf{g}, \boldsymbol{\alpha}) = \left\{\mathbf{g}_1(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t, \boldsymbol{\alpha})^{\mathsf{T}}\mathbf{X}_t - \mathbf{g}_2(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t, \boldsymbol{\alpha})\right\}^{\mathsf{T}}$ and $m_{3t}(\boldsymbol{\alpha}) = \mathbb{X}_t w(\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}_t)$, and define the empirical process

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

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

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

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 (5.28) 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, page 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 $m \in \mathcal{M}$,

$|m(F_t, \mathbf{g}, \boldsymbol{\alpha})| \le (|Y_t| + C_1\|\mathbb{X}\|)(C_2\|\mathbf{X}_t\| + C_3)\|\mathbb{X}\|w_0$, 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 (C1). 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 $\boldsymbol{\alpha}_j$, $j = 1, \cdots, N_1$, in $\mathbb{R}^d$, say, $\mathbb{B}_j$, $j = 1, \cdots, N_1$, such that

$$\forall \boldsymbol{\alpha} \in \mathbb{B},\ \exists \boldsymbol{\alpha}_j,\ \text{such that } \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_j\| \le \varepsilon/(2C). \tag{5.34}$$

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

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

and

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

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

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

with

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

and

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

Thus we can cover $\mathcal{G}_1 \subset \mathcal{G}$ by finite number of N_1N_2 balls of centers $\mathbf{g}^{\ell,j}(\cdot)$, $j = 1, \cdots, N_1$, $\ell = 1, \cdots, N_2$, 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

$$\begin{aligned}\sup_{z\in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}) - \mathbf{g}^{\ell,j}(z)\| &\le \sup_{z\in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}) - \mathbf{g}(z, \boldsymbol{\alpha}_j)\| + \sup_{z\in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}_j) - \mathbf{g}^{\ell,j}(z)\| \\ &\le C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_j\| + \sup_{z\in S_w} \|\mathbf{g}(z, \boldsymbol{\alpha}_j) - \mathbf{g}^{\ell,j}(z)\| \le \frac{3}{2}\varepsilon,\end{aligned} \tag{5.35}$$

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 $\mathbf{g}_2^{\ell,j}(\cdot)$, $j = 1, \cdots, N_1$, $\ell = 1, \cdots, N_3$, 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 (5.35),

$$\sup_{z\in S_w} \|\mathbf{g}_1(z, \boldsymbol{\alpha}) - \mathbf{g}_1^{\ell,j}(z)\| \le \varepsilon, \quad \sup_{z\in S_w} \|\mathbf{g}_2(z, \boldsymbol{\alpha}) - \mathbf{g}_2^{\ell,j}(z)\| \le \varepsilon. \tag{5.36}$$

Therefore, with $\boldsymbol{\alpha}^\top \mathbf{X}_t \in S_w$ and $\mathbf{g}^{\ell,j}(\cdot) \in C_{c_0}^2(S_w)$, it follows from (5.34), (5.35) and (5.36) 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,

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

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

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

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

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

$$E|m(F_t, \mathbf{g}, \boldsymbol{\alpha}) - m(F_t, \mathbf{g}^{\ell,j}, \boldsymbol{\alpha}_j)|^2 \le C(M_1 + M_2 + M_3), \tag{5.37}$$

where

$$
\begin{aligned}
M_1 &= E|(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})|^2 \\
&\le E[\|\mathbf{g}(\boldsymbol{\alpha}^\top \mathbf{X}_t,\boldsymbol{\alpha}) - \mathbf{g}^{\ell,j}(\boldsymbol{\alpha}_j^\top \mathbf{X}_t)\|)\ \|\mathbb{X}_t\|\ \|\left\{\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})\right\}^\top \|\ \|\mathbb{X}_t\|\ w(\boldsymbol{\alpha}^\top \mathbf{X}_t)]^2 \\
&\le C\varepsilon E[\|\mathbb{X}_t\|\ \|\left\{\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})\right\}^\top \|\ \|\mathbb{X}_t\|\ w(\boldsymbol{\alpha}^\top \mathbf{X}_t)]^2 \\
&\le C\varepsilon^2 E[\|\mathbb{X}_t\|^2\ (C\|\mathbf{X}_t\| + C)^2\ \|\mathbb{X}_t\|^2] \le C\varepsilon^2, \qquad (5.38)\\
M_2 &= E|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})|^2 \\
&\le E[\|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)]^2 \\
&\le C\varepsilon^2 E[(|Y_t| + c_0\|\mathbb{X}_t)^2(1 + C\|\mathbf{X}_t\|)^2\|\mathbb{X}_t\|^2] \le C\varepsilon^2, \qquad (5.39)\\
M_3 &= E[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))]^2 \le C\varepsilon^2, \qquad (5.40)
\end{aligned}
$$

and where C is allowed to change in value from line to line. Then it follows from (5.37) together with (5.38), (5.39) and (5.40) that

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

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

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

Now (c) holds easily.

(vi): Finally we are in a position to establish (vi) of Lemma 4.2. Note that it follows from (5.30) with $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ that

$$
\begin{aligned}
\widehat{\mathbf{g}}(z,\boldsymbol{\alpha}_0) - \mathbf{g}_0(z,\boldsymbol{\alpha}_0) &= \gamma(\widehat{\boldsymbol{\theta}}(z,\boldsymbol{\alpha}_0) - \boldsymbol{\theta}_0(z,\boldsymbol{\alpha}_0)) \\
&= \gamma\widehat{\Phi}^{-1}(z,\boldsymbol{\alpha}_0)\widehat{\mathbf{W}}(z,\boldsymbol{\alpha}_0) = (1 + o_P(1))\gamma\Phi^{-1}(z,\boldsymbol{\alpha}_0)\widehat{\mathbf{W}}(z,\boldsymbol{\alpha}_0) \\
&= (1 + o_P(1))(\mu_{0,K}w(z)f_Z(z,\boldsymbol{\alpha}_0))^{-1}\mathbf{G}^{-1}(z,\boldsymbol{\alpha}_0)\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}_0), \qquad (5.41)
\end{aligned}
$$

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 (5.31). Then (5.26) together with Lemma 6.3 and (5.41) then leads to

$$\begin{aligned}
&\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\left[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)\right] \\
&\quad= 2\int \widetilde{\Phi}_0(z)(\widehat{\mathbf{g}}(z,\boldsymbol{\alpha}_0)-\mathbf{g}_0(z,\boldsymbol{\alpha}_0))w(z)f_0(z,\boldsymbol{\alpha}_0)dz \\
&\quad= (1+o_P(1))2\mu_{0,K}^{-1}\int \widetilde{\Phi}_0(z)G^{-1}(z,\boldsymbol{\alpha}_0)\widehat{\mathbf{W}}^{(1)}(z,\boldsymbol{\alpha}_0)dz \\
&\quad= (1+o_P(1))2\mu_{0,K}^{-1}n^{-1}\sum_{t=1}^{n}\int \widetilde{\Phi}_0(z)G^{-1}(z,\boldsymbol{\alpha}_0) \\
&\qquad\quad\times\left\{\varepsilon_t+\frac{1}{2}\left(\ddot{\mathbf{g}}_0(z+\eta(Z_t^o-z))(Z_t^o-z)^2\right)^{\mathsf{T}}\mathbb{X}_t\right\}\mathbb{X}_tK_h(Z_t^o-z)w(Z_t^o)dz \\
&\quad= (1+o_P(1))2\mu_{0,K}^{-1}n^{-1}\sum_{t=1}^{n}\widetilde{\Phi}_0(Z_t^o)G^{-1}(Z_t^o,\boldsymbol{\alpha}_0) \\
&\qquad\quad\times\left\{\varepsilon_t\mu_{0,K}+\frac{1}{2}\left(\ddot{\mathbf{g}}_0(Z_t^o)\mu_{2,K}h^2(1+o(1))\right)^{\mathsf{T}}\mathbb{X}_t\right\}\mathbb{X}_tw(Z_t^o) \\
&\quad= (1+o_P(1))2\left\{n^{-1}\sum_{t=1}^{n}\varepsilon_t\mathbf{U}_t+\frac{1}{2}h^2\mu_{0,K}^{-1}\mu_{2,K}E\left(\ddot{\mathbf{g}}_0(Z_t^o))^{\mathsf{T}}\mathbb{X}_t\mathbf{U}_t\right)\right\}+o_P(n^{-1/2}), \qquad (5.42)
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{U}_t &= \widetilde{\Phi}_0(Z_t^o)G^{-1}(Z_t^o,\boldsymbol{\alpha}_0)\mathbb{X}_tw(Z_t^o) \\
&= E\left(\mathbf{X}_t\mathbf{g}_{01}(Z_t^o,\boldsymbol{\alpha}_0)^{\mathsf{T}}\mathbb{X}_t\mathbb{X}_t^{\mathsf{T}}|Z_t^o\right)G^{-1}(Z_t^o,\boldsymbol{\alpha}_0)\mathbb{X}_tw(Z_t^o)+\mathbf{g}_{02}(Z_t^o,\boldsymbol{\alpha}_0)\mathbb{X}_tw(Z_t^o). \qquad (5.43)
\end{aligned}$$

Now we have from (5.22) and (5.42) and then from (5.43) that

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

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

$$\begin{aligned}
&\sqrt{n}\left\{\dot{R}_n(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0)+\Gamma_2\left(\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0),\boldsymbol{\alpha}_0\right)\left[\widehat{\mathbf{g}}(\cdot,\boldsymbol{\alpha}_0)-\mathbf{g}_0(\cdot,\boldsymbol{\alpha}_0)\right]\right. \\
&\qquad\left.-(1+o_P(1))h^2\mu_{0,K}^{-1}\mu_{2,K}E\left(\ddot{\mathbf{g}}_0(Z_t^o))^{\mathsf{T}}\mathbb{X}_t\mathbf{U}_t\right)\right\}\xrightarrow{D}N(\mathbf{0},\mathcal{V}), \qquad (5.44)
\end{aligned}$$

where

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

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

□

6 Appendix: Uniform convergence.

We collect and prove some uniform convergence results which were used in Section 5. All limits are taken as $n \to \infty$ unless stated otherwise.

6.1 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 6.1 (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 $q > 2$,*

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

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

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

and that the mixing coefficients satisfy

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

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

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

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

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

Lemma 6.2 (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 $f_s(x_1, \cdots, x_s)$ of $(X_{t_1}, \cdots, X_{t_s})$ exists and is bounded uniformly for $s = 1, \cdots, 2r - 1$, where r is some positive integer such that $E[|\boldsymbol{\theta}_t|^{qr}] < \infty$ 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 $|K(x) - K(y)| \leq M|x - y|$ for $x,\ y \in C_K$ and $0 < M < \infty$. The bandwidth $h = h_n$ satisfies that

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

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

$$E\left[\left(\sum_{i=1}^{n} \xi_i\right)^{2r}\right] \leq C\,(nh)^r. \tag{6.1}$$

Proof. It is a special case of Theorem 1.1 of Gao, Lu and Tjøstheim (2003) with $N = 1$ there.

□

6.2 uniform convergence

Lemma 6.3 *Under the conditions of Theorem 3.2, assume that $|K(x) - K(y)| \leq C\|x - y\|$ for any $x, y \in \mathbb{R}^1$, and $|w(x) - w(y)| \leq C\|x - y\|$ for any $x, y \in S_w$, and that $|a(z, \boldsymbol{\alpha}) - a(z', \boldsymbol{\alpha}')| \leq C(|z - z'| + \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|)$ and $\|\mathbf{b}(z, \boldsymbol{\alpha}) - \mathbf{b}(z', \boldsymbol{\alpha}')\| \leq C(|z - z'| + \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|)$ for $z, z' \in \mathbb{R}^1$ and $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathbb{B}$, and that $\sup_{\boldsymbol{\alpha}\in\mathbb{B}} E|\varepsilon_t(\boldsymbol{\alpha})|^2 < \infty$. 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 $G_{ij}(z, \boldsymbol{\alpha}) = \bar{g}_{i-1,j-1}(z, \boldsymbol{\alpha})$ defined in Subsection 5.1,*

$$\sup_{z\in S_w, \boldsymbol{\alpha}\in\mathbb{B}} |\widehat{\Phi}_{i,j}(z, \boldsymbol{\alpha}) - \mu_{0,K} G_{ij}(z, \boldsymbol{\alpha}) w(z) f_Z(z, \boldsymbol{\alpha})| = O_P\left[\left(nh^{1+2d/r}\right)^{-r/(2r+d)} + h^2\right] \tag{6.2}$$

$$\sup_{z\in S_w,\boldsymbol{\alpha}\in\mathbb{B}} |\widehat{\Phi}_{i+d,j+d}(z,\boldsymbol{\alpha}) - \mu_{2,K} G_{ij}(z,\boldsymbol{\alpha}) w(z) f_Z(z,\boldsymbol{\alpha})| = O_P\left[h^{-1}\left(nh^{1+2d/r}\right)^{-r/(2r+d)} + h\right] \quad (6.3)$$

for $i, j = 1, \cdots, d,$

$$\sup_{z\in S_w,\boldsymbol{\alpha}\in\mathbb{B}} \|\widehat{\mathbf{g}}(z,\boldsymbol{\alpha}) - \mathbf{g}_0(z,\boldsymbol{\alpha})\| = O_P\left[\left(nh^{1+2d/r}\right)^{-r/(2r+d)} + h^2\right], \quad (6.4)$$

$$\sup_{z\in S_w,\boldsymbol{\alpha}\in\mathbb{B}} \|\widehat{\mathbf{g}}_1(z,\boldsymbol{\alpha}) - \mathbf{g}_{01}(z,\boldsymbol{\alpha})\| = O_P\left[h^{-1}\left(nh^{1+2d/r}\right)^{-r/(2r+d)} + h\right], \quad (6.5)$$

$$\sup_{z\in S_w,\boldsymbol{\alpha}\in\mathbb{B}} \|\widehat{\mathbf{g}}_2(z,\boldsymbol{\alpha}) - \mathbf{g}_{02}(z,\boldsymbol{\alpha})\| = O_P\left[h^{-1}\left(nh^{1+2d/r}\right)^{-r/(2r+d)} + h\right]. \quad (6.6)$$

Proof of Lemma 6.3. As the proofs of (6.2)–(6.6) are similar, so we only sketch the proof of (6.4) below. It follows from conditions (C2) 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 $E\left(\mathbb{X}_t\, \mathbb{X}_t^\top \middle| \boldsymbol{\alpha}^\top \mathbf{X}_t = z\right)$, are bounded away from zero over $z \in S_w, \boldsymbol{\alpha} \in \mathbb{B}$. Therefore, it is derived from (5.30) 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 (5.31)) 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 (5.32). 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

$$\sup_{\|\boldsymbol{\alpha}\|=1} \sup_{z\in S_w} \|\widehat{W}_2(z,\boldsymbol{\alpha})\| = \left(nh^{1+2d/r}\right)^{-r/(2r+d)}, \quad (6.7)$$

owing to $\mathbb{B} \subset \{\boldsymbol{\alpha} \in \mathbb{R}^d : \ \|\boldsymbol{\alpha}\| = 1\}$.

For notation convenience, we denote $\mathbb{B} = \{\boldsymbol{\alpha} \in \mathbb{R}^d : \ \|\boldsymbol{\alpha}\| = 1\}$ below. 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 $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \le \text{const.}/M^{1/(d-1)}$ 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 $|z - z_\ell| \le \text{const.}/N$ for $z \in J_\ell$, respectively, where M and N are to be specified later. Therefore

$$\begin{aligned}\sup_{\|\boldsymbol{\alpha}\|=1} \sup_{z\in S_w} \|\widehat{W}_2(z,\boldsymbol{\alpha})\| &\le \max_{1\le k\le M} \sup_{z\in S_w} \|\widehat{W}_2(z,\boldsymbol{\alpha}_k)\| + \max_{1\le k\le M} \sup_{\boldsymbol{\alpha}\in I_k} \sup_{z\in S_w} \|\widehat{W}_2(z,\boldsymbol{\alpha}) - \widehat{W}_2(z,\boldsymbol{\alpha}_k)\| \\ &\equiv W_{21} + W_{22}. \qquad (6.8)\end{aligned}$$

We first consider W_{22}. Note that

$$\begin{aligned}
&\widehat{W}_2(z,\boldsymbol{\alpha}) - \widehat{W}_2(z,\boldsymbol{\alpha}_k) \\
&\quad = \frac{1}{n}\sum_{t=1}^{n} \mathbb{X}_t\{\varepsilon_t(\boldsymbol{\alpha})K_h(\boldsymbol{\alpha}^\top \mathbf{X}_t - z)w(\boldsymbol{\alpha}^\top \mathbf{X}_t) - \varepsilon_t(\boldsymbol{\alpha}_k)K_h(\boldsymbol{\alpha}_k^\top \mathbf{X}_t - z)w(\boldsymbol{\alpha}_k^\top \mathbf{X}_t)\}, \qquad (6.9)
\end{aligned}$$

and that

$$\begin{aligned}
|\varepsilon_t(\boldsymbol{\alpha}) - \varepsilon_t(\boldsymbol{\alpha}_k)| &\leq |a(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - a(\boldsymbol{\alpha}_k^\top \mathbf{X}_t, \boldsymbol{\alpha}_k)| + \|\mathbf{b}(\boldsymbol{\alpha}^\top \mathbf{X}_t, \boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}_k^\top \mathbf{X}_t, \boldsymbol{\alpha}_k)\| \; \|\mathbf{X}_{t,-d}\| \\
&\leq C\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|(1 + \|\mathbf{X}_t\|)^2,
\end{aligned}$$

and

$$|K_h(\boldsymbol{\alpha}^\top \mathbf{X}_t - z)w(\boldsymbol{\alpha}^\top \mathbf{X}_t) - K_h(\boldsymbol{\alpha}_k^\top \mathbf{X}_t - z)w(\boldsymbol{\alpha}_k^\top \mathbf{X}_t)| \leq Ch^{-2}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \; \|\mathbf{X}_t\|.$$

Thus

$$\begin{aligned}
&\widehat{W}_2(z,\boldsymbol{\alpha}) - \widehat{W}_2(z,\boldsymbol{\alpha}_k) \\
&\quad \leq \frac{1}{n}\sum_{t=1}^{n} \|\mathbb{X}_t\|\{h^{-1}(1 + \|\mathbf{X}_t\|)^2 + |\varepsilon_t(\boldsymbol{\alpha}_k)|h^{-2} \; \|\mathbf{X}_t\|\}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \\
&\quad \leq \frac{1}{n}\sum_{t=1}^{n} \left\{h^{-1}(1 + \|\mathbf{X}_t\|)^3 + |\varepsilon_t(\boldsymbol{\alpha}_k)|h^{-2} \; \|\mathbf{X}_t\|^2\right\} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|, \qquad (6.10)
\end{aligned}$$

and it follows from (6.8) and (6.10) that

$$\begin{aligned}
W_{22} &\leq \max_{1\leq k\leq M} \sup_{\boldsymbol{\alpha}\in I_k} \frac{1}{n}\sum_{t=1}^{n} \left\{h^{-1}(1 + \|\mathbf{X}_t\|)^3 + |\varepsilon_t(\boldsymbol{\alpha}_k)|h^{-2} \; \|\mathbf{X}_t\|^2\right\} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \\
&\leq CM^{-1/(d-1)} \max_{1\leq k\leq M} \frac{1}{n}\sum_{t=1}^{n} \left\{h^{-1}(1 + \|\mathbf{X}_t\|)^3 + |\varepsilon_t(\boldsymbol{\alpha}_k)|h^{-2} \; \|\mathbf{X}_t\|^2\right\} \\
&\leq CM^{-1/(d-1)} \left\{h^{-1}\frac{1}{n}\sum_{t=1}^{n}(1 + \|\mathbf{X}_t\|)^3 + \max_{1\leq k\leq M} \frac{1}{n}\sum_{t=1}^{n} |\varepsilon_t(\boldsymbol{\alpha}_k)|h^{-2} \; \|\mathbf{X}_t\|^2\right\} \\
&\leq CM^{-1/(d-1)} \left\{h^{-1}O_P(1) + h^{-2}W_{222} + h^{-2}W_{223}\right\}, \qquad (6.11)
\end{aligned}$$

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

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

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

$$W_{223} \le \max_{1\le k\le M} \left\{E|\varepsilon_t(\boldsymbol{\alpha}_k)|^2\right\}^{1/2} \left\{E\|\mathbf{X}_t\|^4\right\}^{1/2} = O(1), \tag{6.12}$$

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 6.1 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

$$\begin{aligned} P\{W_{222} > 2A\} = P\left\{\max_{1\le k\le M} |\Delta_k| > 2A\right\} &\le \sum_{k=1}^M P\{|\Delta_k| > A\} \\ &= C_r M A^{-r} n^{-r} = CA^{-r} \to 0 \end{aligned} \tag{6.13}$$

as $A \to \infty$, which leads to $W_{222} = O_P(1)$. This together with (6.11) and (6.12) implies

$$W_{22} = O_P(M^{-1/(d-1)} h^{-2}) = O_P\left\{\delta_n\right\}, \tag{6.14}$$

where we take $M = (h^2\delta_n)^{-(d-1)}$ with δ_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 (6.8). 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:

$$W_{21} \le \max_{1\le k\le M}\max_{1\le \ell\le N} \widehat{W}_2(z_\ell, \boldsymbol{\alpha}_k) + \max_{1\le k\le M}\max_{1\le \ell\le N}\sup_{z\in J_\ell} \|\widehat{W}_2(z, \boldsymbol{\alpha}_k) - \widehat{W}_2(z_\ell, \boldsymbol{\alpha}_k)\| \equiv W_{211} + W_{212}. \tag{6.15}$$

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

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

therefore, noting $\|\mathbf{X}_{t,-d}\| \le \|\mathbb{X}_t\|$,

$$
\begin{aligned}
W_{212} &= \max_{1\le k\le M}\max_{1\le \ell\le N}\sup_{z\in J_\ell}\|\widehat{W}_2(z,\boldsymbol{\alpha}_k)-\widehat{W}_2(z_\ell,\boldsymbol{\alpha}_k)\| \\
&= C\max_{1\le k\le M}\max_{1\le \ell\le N}\sup_{z\in J_\ell} h^{-2}|z-z_\ell|\ \frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t(\boldsymbol{\alpha}_k)|\ \|\mathbb{X}_t\| w(\boldsymbol{\alpha}_k^\top \mathbf{X}_t) \\
&\le Ch^{-2}N^{-1}\max_{1\le k\le M}\frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t(\boldsymbol{\alpha}_k)|\ \|\mathbb{X}_t\| w(\boldsymbol{\alpha}_k^\top \mathbf{X}_t) \\
&= O_P\left(h^{-2}N^{-1}\right) = O_P(\delta_n), \qquad (6.16)
\end{aligned}
$$

where in the final equality of (6.16), we take $N=(h^2\delta_n)^{-1}$, 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 (6.11) in the above. Now we consider W_{211} in (6.15). With $\xi_t=\varepsilon_t(\boldsymbol{\alpha}_k)\ \mathbb{X}_t K((\boldsymbol{\alpha}_k^\top \mathbf{X}_t - z_\ell)/h)w(\boldsymbol{\alpha}_k^\top \mathbf{X}_t)$ and $\boldsymbol{\theta}_t=\varepsilon_t(\boldsymbol{\alpha}_k)\ \mathbb{X}_t w(\boldsymbol{\alpha}_k^\top \mathbf{X}_t)$ in Lemma 6.2, it follows from Lemma 6.2 that

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

Therefore

$$
W_{211} = O_P\left((nh)^{-1/2}(h^2\delta_n)^{-d/(2r)}\right), \qquad (6.17)
$$

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 $\delta_n=(nh^{1+2d/r})^{-r/(2r+d)}$, then $N=(h^2\delta_n)^{-1}=(nh^{-3})^{r/(2r+d)}$, and $M=(h^2\delta_n)^{-(d-1)}=(nh^{-3})^{(d-1)r/(2r+d)}=O(n^r)$ as $\lim_{n\to\infty}n^{2r+1}h^{3(d-1)}>0$. For such a δ_n, (6.13), (6.14) and (6.16) hold simultaneously. Thus the result of (6.7) follows from (6.8), (6.14), (6.15), (6.16) and (6.17). □

Acknowledgement:

References

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series models. *Journal of the American Statistical Association*, **95**, 888-902.

Carroll, R.J., Fan. J., Gijbels, I. and Wand, M.P. (1997). Generalized partially linear single–index models. *Journal of the American Statistical Association*, **92**, 477–489.

Chan, R. and Tsay, R. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, **88**, 298-308.

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

Cline, D.B.H. and Pu, H.M.H. (1999). Geometric ergodicity of nonlinear time series. *Statistica Sinica*, **9**, 1103-1118.

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

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*. London: Chapman and Hall.

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. *Journal of the Royal Statistical Society*, **B**, **65**, 57-80.

Gao, J., Tong, H. and Wolff, R. (2002). Adaptive series estimation in additive stochastic regression models. *Statistica Sinica* **12** 409-428.

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.

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

Hallin, M., Lu, Z. and Tran, L.T. (2004). Local linear spatial regression. *The Annals of Statistics*, **32**, to appear.

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.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

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

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

Lu, Z. and Cheng, P. (1997). Distribution-free strong consistency for nonparametric kernel regression involving nonlinear time series. *Journal of Statistical Planning and 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 .

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

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 coecient regression models. *J. Amer. Statist. Ass.*, **94**, 1275-1285.