SPLINE ESTIMATION OF SINGLE-INDEX MODELS

Li Wang and Lijian Yang

University of Georgia and Michigan State University

Supplementary Material

This note contains proofs for the main results. The following two propositions play an important role in the proof. Proposition A.1 establishes the uniform convergence rate of the derivatives of $\hat{\gamma}_\theta$ up to order 2 to those of γ_θ in θ. Proposition A.2 shows that the derivatives of the risk function up to order 2 are uniformly almost surely approximated by their empirical versions.

Proposition A.1. *Under Assumptions A2-A6, with probability* 1

$$\sup_{\theta \in S_c^{d-1}} \sup_{u \in [0,1]} \left| \hat{\gamma}_\theta(u) - \gamma_\theta(u) \right| = O\left\{ (nh)^{-1/2} \log n + h^4 \right\}, \tag{A.1}$$

$$\sup_{1 \le p \le d} \sup_{\theta \in S_c^{d-1}} \max_{1 \le i \le n} \left| \frac{\partial}{\partial \theta_p} \left\{ \hat{\gamma}_\theta(U_{\theta,i}) - \gamma_\theta(U_{\theta,i}) \right\} \right| = O\left(\frac{\log n}{\sqrt{nh^3}} + h^3 \right), \tag{A.2}$$

$$\sup_{1 \le p,q \le d} \sup_{\theta \in S_c^{d-1}} \max_{1 \le i \le n} \left| \frac{\partial^2}{\partial \theta_p \partial \theta_q} \left\{ \hat{\gamma}_\theta(U_{\theta,i}) - \gamma_\theta(U_{\theta,i}) \right\} \right| = O\left(\frac{\log n}{\sqrt{nh^5}} + h^2 \right). \tag{A.3}$$

Proposition A.2. *Under Assumptions A2-A6, one has for* $k = 0, 1, 2$

$$\sup_{\|\theta_{-d}\| \le \sqrt{1-c^2}} \left| \frac{\partial^k}{\partial^k \theta_{-d}} \left\{ \hat{R}^*(\theta_{-d}) - R^*(\theta_{-d}) \right\} \right| = o(1), a.s..$$

Proofs of Theorem 2, Propositions A.1 and A.2 are given in the following. Wherever proofs are incomplete, it is referred to Wang and Yang (2007b).

A1. Preliminaries

In this section, we introduce some properties of the B-spline.

Lemma A.1. *There exist constants* $c > 0$ *such that for* $\sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k}$ *up to order* $k = 4$

$$\begin{cases} ch^{1/r} \|\alpha\|_r \le \left\| \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k} \right\|_r \le \left(3^{r-1} h\right)^{1/r} \|\alpha\|_r, & 1 \le r \le \infty \\ ch^{1/r} \|\alpha\|_r \le \left\| \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k} \right\|_r \le (3h)^{1/r} \|\alpha\|_r, & 0 < r < 1 \end{cases},$$

where $\alpha := (\alpha_{-1,2}, \alpha_{0,2}, ..., \alpha_{N,2}, ..., \alpha_{N,4})$. *In particular, under Assumption A2, for any fixed* θ

$$ch^{1/2} \|\alpha\|_2 \leq \left\| \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k} \right\|_{2,\theta} \leq Ch^{1/2} \|\alpha\|_2 .$$

Proof. It follows from the B-spline property on page 96 of de Boor (2001), $\sum_{k=2}^{4} \sum_{j=-k+1}^{N} B_{j,k} \equiv 3$ on $[0,1]$. So the right inequality follows immediate for $r = \infty$. When $1 \leq r < \infty$, Hölder's inequality implies that

$$\left| \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k} \right| \leq 3^{1-1/r} \left(\sum_{k=2}^{4} \sum_{j=-k+1}^{N} |\alpha_{j,k}|^r B_{j,k} \right)^{1/r} .$$

Since all the knots are equally spaced, $\int_{-\infty}^{\infty} B_{j,k}(u)\, du \leq h$, the right inequality follows from

$$\int_0^1 \left| \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k}(u) \right|^r du \leq 3^{r-1} h \|\alpha\|_r^r .$$

When $r < 1$, we have

$$\left| \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k} \right|^r \leq \sum_{k=2}^{4} \sum_{j=-k+1}^{N} |\alpha_{j,k}|^r B_{j,k}^r .$$

Since $\int_{-\infty}^{\infty} B_{j,k}^r(u)\, du \leq t_{j+k} - t_j = kh$ and

$$\int_0^1 \left| \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k}(u) \right|^r du \leq \|\alpha\|_r^r \int_{-\infty}^{\infty} B_{j,k}^r(u)\, du \leq 3h \|\alpha\|_r^r ,$$

the right inequality follows in this case as well. For the left inequalities, we derive from Theorem 5.4.2, DeVore and Lorentz (1993), for any $0 < r \leq \infty$

$$|\alpha_{j,k}|^r \leq C_1^r h^{-1} \int_{t_j}^{t_{j+1}} \left| \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k}(u) \right|^r du.$$

Since each $u \in [0,1]$ appears in at most k intervals (t_j, t_{j+k}), adding up these inequalities, we obtain that

$$\|\alpha\|_r^r \leq C_1 h^{-1} \sum_{k=1}^{4} \int_{t_j}^{t_{j+k}} \left| \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k}(u) \right|^r du \leq 3Ch^{-1} \left\| \sum_{j=-k+1}^{N} \alpha_{j,k} B_{j,k} \right\|_r^r .$$

The left inequality follows.

For any functions ϕ and φ, define the empirical inner product and the empirical norm as

$$\langle \phi, \varphi \rangle_\theta = \int_0^1 \phi(u) \varphi(u) f_\theta(u)\, du, \;\; \|\phi\|_{2,n,\theta}^2 = n^{-1} \sum_{i=1}^{n} \phi^2(U_{\theta,i}) .$$

In addition, if functions ϕ, φ are $L_2[0,1]$-integrable, define the theoretical inner product and its corresponding theoretical L_2 norm as

$$\|\phi\|_{2,\theta}^2 = \int_0^1 \phi^2(u) f_\theta(u)\, du, \;\; \langle \phi, \varphi \rangle_{n,\theta} = n^{-1} \sum_{i=1}^{n} \phi(U_{\theta,i}) \varphi(U_{\theta,i}) .$$

Denote by $\Gamma = \Gamma^{(0)} \cup \Gamma^{(1)} \cup \Gamma^{(2)}$ the space of all linear, quadratic and cubic spline functions on $[0,1]$. We establish the uniform rate at which the empirical inner product approximates the theoretical inner product for all B-splines $B_{j,k}$ with $k = 2, 3, 4$.

Lemma A.2. *Under Assumptions A2, A5 and A6, with probability* 1

$$A_n = \sup_{\theta \in S_c^{d-1}} \sup_{\gamma_1, \gamma_2 \in \Gamma} \left| \frac{\langle \gamma_1, \gamma_2 \rangle_{n,\theta} - \langle \gamma_1, \gamma_2 \rangle_\theta}{\|\gamma_1\|_{2,\theta} \|\gamma_2\|_{2,\theta}} \right| = O\left\{ (nh)^{-1/2} \log n \right\} . \tag{A.4}$$

Proof. Denote $\gamma_1 = \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \alpha_{jk} B_{j,k}$, $\gamma_2 = \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \beta_{jk} B_{j,k}$ without loss of generality. Then for fixed θ

$$\begin{aligned}
\langle \gamma_1, \gamma_2 \rangle_{n,\theta} &= \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \sum_{k'=2}^{4} \sum_{j'=-k+1}^{N} \alpha_{j,k} \beta_{j',k'} \langle B_{j,k}, B_{j',k'} \rangle_{n,\theta} , \\
\|\gamma_1\|_{2,\theta}^2 &= \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \sum_{k'=2}^{4} \sum_{j'=-k+1}^{N} \alpha_{j,k} \alpha_{j',k'} \langle B_{j,k}, B_{j',k'} \rangle_\theta , \\
\|\gamma_2\|_{2,\theta}^2 &= \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \sum_{k'=2}^{4} \sum_{j'=-k+1}^{N} \beta_{j,k} \beta_{j',k'} \langle B_{j,k}, B_{j',k'} \rangle_\theta .
\end{aligned}$$

Let $\alpha = (\alpha_{-1,2}, \alpha_{0,2}, ..., \alpha_{N,2}, ..., \alpha_{N,4})$ and $\beta = (\beta_{-1,2}, \beta_{0,2}, ..., \beta_{N,2}, ..., \beta_{N,4})$. According to Lemma A.1, one has for any $\theta \in S_c^{d-1}$,

$$c_1 h \|\alpha\|_2^2 \le \|\gamma_1\|_{2,\theta}^2 \le c_2 h \|\alpha\|_2^2, c_1 h \|\beta\|_2^2 \le \|\gamma_2\|_{2,\theta}^2 \le c_2 h \|\beta\|_2^2,$$

$$c_1 h \|\alpha\|_2 \|\beta\|_2 \le \|\gamma_1\|_{2,\theta} \|\gamma_2\|_{2,\theta} \le c_2 h \|\alpha\|_2 \|\beta\|_2 .$$

Hence

$$\begin{aligned} A_n &= \sup_{\theta \in S_c^{d-1}} \sup_{\gamma_1 \in \gamma, \gamma_2 \in \Gamma} \left| \frac{\langle \gamma_1, \gamma_2 \rangle_{n,\theta} - \langle \gamma_1, \gamma_2 \rangle_\theta}{\|\gamma_1\|_{2,\theta} \|\gamma_2\|_{2,\theta}} \right| \le \frac{\|\alpha\|_\infty \|\beta\|_\infty}{c_1 h \|\alpha\|_2 \|\beta\|_2} \\ &\times \sup_{\theta \in S_c^{d-1}} \max_{\substack{k,k'=2,3,4 \\ 1 \le j,j' \le N}} \left| \frac{1}{n} \sum_{i=1}^n \left\{ \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\theta} - \left\langle B_{j,k}, B_{j',k'} \right\rangle_\theta \right\} \right|, \end{aligned}$$

$$A_n \le c_0 h^{-1} \sup_{\theta \in S_c^{d-1}} \max_{\substack{k,k'=2,3,4 \\ 1 \le j,j' \le N}} \left| \frac{1}{n} \sum_{i=1}^n \left\{ \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\theta} - \left\langle B_{j,k}, B_{j',k'} \right\rangle_\theta \right\} \right|,$$

which, together with Lemma A.2 in Wang and Yang (2007b), imply (A.4).

A2. Proof of Proposition A.1

For any fixed θ, we write the response $\mathbf{Y}^T = (Y_1, ..., Y_n)$ as the sum of a signal vector γ_θ, a parametric noise vector $\mathbf{E}_\theta$ and a systematic noise vector $\mathbf{E}$, i.e., $\mathbf{Y} = \gamma_\theta + \mathbf{E}_\theta + \mathbf{E}$, in which the vectors $\gamma_\theta^T = \{\gamma_\theta(U_{\theta,1}), ..., \gamma_\theta(U_{\theta,n})\}$, $\mathbf{E}_\theta^T = \{m(\mathbf{X}_1) - \gamma_\theta(U_{\theta,1}), ..., m(\mathbf{X}_n) - \gamma_\theta(U_{\theta,n})\}$, $\mathbf{E}^T = \{\sigma(\mathbf{X}_1)\varepsilon_1, ..., \sigma(\mathbf{X}_n)\varepsilon_n\}$.

Remark A.1. If m is a genuine single-index function, then $\mathbf{E}_{\theta_0} \equiv 0$, thus the proposed model given by (1.1) and (1.2) is exactly the single-index model.

We break the cubic spline estimation error $\hat{\gamma}_\theta(u_\theta) - \gamma_\theta(u_\theta)$ into a bias term $\tilde{\gamma}_\theta(u_\theta) - \gamma_\theta(u_\theta)$ and two noise terms $\tilde{\varepsilon}_\theta(u_\theta)$ and $\hat{\varepsilon}_\theta(u_\theta)$

$$\hat{\gamma}_\theta(u_\theta) - \gamma_\theta(u_\theta) = \{\tilde{\gamma}_\theta(u_\theta) - \gamma_\theta(u_\theta)\} + \tilde{\varepsilon}_\theta(u_\theta) + \hat{\varepsilon}_\theta(u_\theta), \tag{A.5}$$

where

$$\tilde{\gamma}_\theta(u) = \{B_{j,4}(u)\}_{-3 \le j \le N}^T \mathbf{V}_{n,\theta}^{-1} \left\{ \langle \gamma_\theta, B_{j,4} \rangle_{n,\theta} \right\}_{j=-3}^N, \tag{A.6}$$

$$\tilde{\varepsilon}_\theta(u) = \{B_{j,4}(u)\}_{-3 \le j \le N}^T \mathbf{V}_{n,\theta}^{-1} \left\{ \langle \mathbf{E}_\theta, B_{j,4} \rangle_{n,\theta} \right\}_{j=-3}^N, \tag{A.7}$$

$$\hat{\varepsilon}_\theta(u) = \{B_{j,4}(u)\}_{-3 \le j \le N}^T \mathbf{V}_{n,\theta}^{-1} \left\{ \langle \mathbf{E}, B_{j,4} \rangle_{n,\theta} \right\}_{j=-3}^N. \tag{A.8}$$

In the above, we denote by $\mathbf{V}_{n,\theta}$ the empirical inner product matrix of the cubic B-spline basis and similarly, the theoretical inner product matrix as $\mathbf{V}_\theta$

$$\mathbf{V}_{n,\theta} = \frac{1}{n}\mathbf{B}_\theta^T\mathbf{B}_\theta = \left\{\langle B_{j',4}, B_{j,4}\rangle_{n,\theta}\right\}_{j,j'=-3}^{N}, \mathbf{V}_\theta = \left\{\langle B_{j',4}, B_{j,4}\rangle_\theta\right\}_{j,j'=-3}^{N}. \tag{A.9}$$

The next lemma provides the uniform upper bound of $\left\|\mathbf{V}_{n,\theta}^{-1}\right\|_\infty$ and $\left\|\mathbf{V}_\theta^{-1}\right\|_\infty$.

Lemma A.3. *Under Assumptions A2, A5 and A6, there exist constants* $0 < c_V < C_V$ *such that* $c_V N^{-1}\left\|\mathbf{w}\right\|_2^2 \le \mathbf{w}^T\mathbf{V}_\theta\mathbf{w} \le C_V N^{-1}\left\|\mathbf{w}\right\|_2^2$ *and*

$$c_V N^{-1}\left\|\mathbf{w}\right\|_2^2 \le \mathbf{w}^T\mathbf{V}_{n,\theta}\mathbf{w} \le C_V N^{-1}\left\|\mathbf{w}\right\|_2^2, a.s.,$$

with matrices $\mathbf{V}_\theta$ *and* $\mathbf{V}_{n,\theta}$ *in (A.9). Consequently, there exists a constant* $C > 0$ *such that*

$$\sup_{\theta\in S_c^{d-1}}\left\|\mathbf{V}_{n,\theta}^{-1}\right\|_\infty \le CN, a.s., \sup_{\theta\in S_c^{d-1}}\left\|\mathbf{V}_\theta^{-1}\right\|_\infty \le CN. \tag{A.10}$$

In the following, we denote by $Q_T(m)$ the 4-th order quasi-interpolant of m corresponding to the knots T, see equation (4.12), page 146 of DeVore and Lorentz (1993).

Lemma A.4. *Under Assumptions A2, A3, A5 and A6, there exists an absolute constant* $C > 0$*, such that for function* $\tilde{\gamma}_\theta(u)$ *in (A.6)*

$$\sup_{\theta\in S_c^{d-1}}\left\|\frac{d^k}{du^k}\left(\tilde{\gamma}_\theta - \gamma_\theta\right)\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^{4-k}, a.s., 0 \le k \le 2, \tag{A.11}$$

Proof. According to Theorem A.1 of Huang (2003), there exists a constant $C > 0$, such that

$$\sup_{\theta\in S_c^{d-1}}\left\|\tilde{\gamma}_\theta - \gamma_\theta\right\|_\infty \le C\sup_{\theta\in S_c^{d-1}}\inf_{\gamma\in\Gamma^{(2)}}\left\|\gamma - \gamma_\theta\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^4, a.s., \tag{A.12}$$

which proves (A.11) for the case $k = 0$. Applying Theorem 7.7.4 in DeVore and Lorentz (1993), one has for $0 \le k \le 2$

$$\sup_{\theta\in S_c^{d-1}}\left\|\frac{d^k}{du^k}\left\{Q_T(\gamma_\theta) - \gamma_\theta\right\}\right\|_\infty \le C\sup_{\theta\in S_c^{d-1}}\left\|\gamma_\theta^{(4)}\right\|_\infty h^{4-k} \le C\left\|m^{(4)}\right\|_\infty h^{4-k}, \tag{A.13}$$

As a consequence of (A.12) and (A.13) for the case $k = 0$, one has

$$\sup_{\theta\in S_c^{d-1}}\left\|Q_T(\gamma_\theta) - \tilde{\gamma}_\theta\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^4, a.s.,$$

which, according to the differentiation of B-spline given in de Boor (2001), entails that

$$\sup_{\theta\in S_c^{d-1}}\left\|\frac{d^k}{du^k}\left\{Q_T\left(\gamma_\theta\right)-\tilde{\gamma}_\theta\right\}\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^{4-k}, a.s.,\ 0\le k\le 2. \tag{A.14}$$

Combining (A.13) and (A.14) proves (A.11) for $k=1,2$.

Lemma A.5. *Under Assumptions A2, A4 and A5, there exists a constant* $C>0$ *such that with probability* 1

$$\sup_{1\le p\le d}\sup_{\theta\in S_c^{d-1}}\left\|\frac{\partial}{\partial\theta_p}\left\{\tilde{\gamma}_\theta\left(U_{\theta,i}\right)-\gamma_\theta\left(U_{\theta,i}\right)\right\}_{i=1}^n\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^3, \tag{A.15}$$

$$\sup_{1\le p,q\le d}\sup_{\theta\in S_c^{d-1}}\left\|\frac{\partial^2}{\partial\theta_p\partial\theta_q}\left\{\tilde{\gamma}_\theta\left(U_{\theta,i}\right)-\gamma_\theta\left(U_{\theta,i}\right)\right\}_{i=1}^n\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^2. \tag{A.16}$$

Proof. According to the definition of $\tilde{\gamma}_\theta$ in (A.6), and the fact that $Q_T\left(\gamma_\theta\right)$ is a cubic spline on the knots T, one has

$$\begin{aligned}&\frac{\partial}{\partial\theta_p}\left\{\left\{Q_T\left(\gamma_\theta\right)-\tilde{\gamma}_\theta\right\}\left(U_{\theta,i}\right)\right\}_{i=1}^n = \frac{\partial}{\partial\theta_p}\mathbf{P}_\theta\left\{\left\{Q_T\left(\gamma_\theta\right)-\gamma_\theta\right\}\left(U_{\theta,i}\right)\right\}_{i=1}^n\\ = &\ \dot{\mathbf{P}}_p\left\{\left\{Q_T\left(\gamma_\theta\right)-\gamma_\theta\right\}\left(U_{\theta,i}\right)\right\}_{i=1}^n + \mathbf{P}_\theta\frac{\partial}{\partial\theta_p}\left\{\left\{Q_T\left(\gamma_\theta\right)-\gamma_\theta\right\}\left(U_{\theta,i}\right)\right\}_{i=1}^n.\end{aligned}$$

Applying (A.14) to the following decomposition

$$\begin{aligned}&\frac{\partial}{\partial\theta_p}\left\{\left\{Q_T\left(\gamma_\theta\right)-\gamma_\theta\right\}\left(U_{\theta,i}\right)\right\}_{i=1}^n = \left\{\left\{Q_T\left(\frac{\partial}{\partial\theta_p}\gamma_\theta\right)-\frac{\partial}{\partial\theta_p}\gamma_\theta\right\}\left(U_{\theta,i}\right)\right\}_{i=1}^n\\ &+\left\{\frac{d}{du}\left\{Q_T\left(\gamma_\theta\right)-\gamma_\theta\right\}\left(U_{\theta,i}\right)X_{ip}\right\}_{i=1}^n\end{aligned}$$

yields (A.15). The proof of (A.16) is similar.

Lemma A.6. *Under Assumptions A2, A5 and A6, there exists a constant* $C>0$ *such that with probability 1*

$$\sup_{\theta\in S_c^{d-1}}\left\|n^{-1}\mathbf{B}_\theta^T\right\|_\infty \le Ch,\ \sup_{1\le p\le d}\sup_{\theta\in S_c^{d-1}}\left\|n^{-1}\dot{\mathbf{B}}_p^T\right\|_\infty \le C, \tag{A.17}$$

$$\sup_{\theta\in S_c^{d-1}}\left\|\mathbf{P}_\theta\right\|_\infty \le C,\ \sup_{1\le p\le d}\sup_{\theta\in S_c^{d-1}}\left\|\dot{\mathbf{P}}_p\right\|_\infty \le Ch^{-1}. \tag{A.18}$$

Proof. To prove (A.17), note that for any vector $\mathbf{a} \in R^n$, with probability 1

$$\left\| n^{-1}\mathbf{B}_\theta^T \mathbf{a} \right\|_\infty \leq \|\mathbf{a}\|_\infty \max_{-3\leq j\leq N} \left| n^{-1} \sum_{i=1}^n B_{j,4}\left(U_{\theta,i}\right) \right| \leq Ch \|\mathbf{a}\|_\infty ,$$

$$\begin{aligned} \left\| n^{-1}\dot{\mathbf{B}}_p^T \mathbf{a} \right\|_\infty &\leq \|\mathbf{a}\|_\infty \max_{-3\leq j\leq N} \left| \frac{1}{nh} \sum_{i=1}^n \left\{ \left(B_{j,3} - B_{j+1,3}\right)\left(U_{\theta,i}\right) \right\} \dot{F}_d\left(\mathbf{X}_{\theta,i}\right) X_{i,p} \right| \\ &\leq C \|\mathbf{a}\|_\infty . \end{aligned}$$

To prove (A.18), one only needs to use (A.10), (A.17).

Lemma A.7. *Under Assumptions A2 and A4-A6, with probability* 1

$$\sup_{\theta\in S_c^{d-1}} \left\| \frac{\mathbf{B}_\theta^T \mathbf{E}}{n} \right\|_\infty = O\left(\frac{\log n}{\sqrt{nN}}\right), \sup_{1\leq p\leq d} \sup_{\theta\in S_c^{d-1}} \left\| \frac{\partial}{\partial\theta_p} \left(\frac{\mathbf{B}_\theta^T \mathbf{E}}{n}\right) \right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}}\right), \tag{A.19}$$

$$\sup_{\theta\in S_c^{d-1}} \left\| \frac{\mathbf{B}_\theta^T \mathbf{E}_\theta}{n} \right\|_\infty = O\left(\frac{\log n}{\sqrt{nN}}\right), \sup_{1\leq p\leq d} \sup_{\theta\in S_c^{d-1}} \left\| \frac{\partial}{\partial\theta_p} \left(\frac{\mathbf{B}_\theta^T \mathbf{E}_\theta}{n}\right) \right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}}\right). \tag{A.20}$$

Proof. We decompose the noise variable ε_i into a truncated part and a tail part $\varepsilon_i = \varepsilon_{i,1}^{D_n} + \varepsilon_{i,2}^{D_n} + m_i^{D_n}$, where $D_n = n^\eta \left(1/3 < \eta < 2/5\right)$, $\varepsilon_{i,1}^{D_n} = \varepsilon_i I\left\{|\varepsilon_i| > D_n\right\}$,

$$\varepsilon_{i,2}^{D_n} = \varepsilon_i I\left\{|\varepsilon_i| \leq D_n\right\} - m_i^{D_n}, m_i^{D_n} = E\left[\varepsilon_i I\left\{|\varepsilon_i| \leq D_n\right\} |\mathbf{X}_i\right].$$

Note that the B-spline basis and the conditional variance function σ^2 are bounded, so it is straightforward to verify that

$$\sup_{\theta\in S_c^{d-1}} \left| \frac{1}{n} \sum_{i=1}^n B_{j,4}\left(U_{\theta,i}\right) \sigma\left(\mathbf{X}_i\right) m_i^{D_n} \right| = O\left(D_n^{-2}\right) = o\left(n^{-2/3}\right).$$

The tail part vanishes almost surely

$$\sum_{n=1}^\infty P\left\{|\varepsilon_n| > D_n\right\} \leq \sum_{n=1}^\infty D_n^{-3} < \infty.$$

So Borel-Cantelli Lemma implies that $\left| \frac{1}{n} \sum_{i=1}^n B_{j,4}\left(U_{\theta,i}\right) \sigma\left(\mathbf{X}_i\right) \varepsilon_{i,1}^{D_n} \right| = O\left(n^{-k}\right)$ for any $k > 0$. For the truncated part, using Bernstein's inequality and discretization method

$$\sup_{\theta\in S_c^{d-1}} \sup_{1\leq j\leq N} \left| n^{-1} \sum_{i=1}^n B_{j,4}\left(U_{\theta,i}\right) \sigma\left(\mathbf{X}_i\right) \varepsilon_{i,2}^{D_n} \right| = O\left(\log n/\sqrt{nN}\right), a.s..$$

Therefore with probability 1

$$\sup_{\theta\in S_c^{d-1}}\left\|\frac{1}{n}\mathbf{B}_\theta^{\mathbf{T}}\mathbf{E}\right\|_\infty = o\left(n^{-2/3}\right)+O\left(n^{-k}\right)+O\left(\log n/\sqrt{nN}\right)=O\left(\frac{\log n}{\sqrt{nN}}\right).$$

The proofs of the second part of (A.19) and the first part of (A.20) are similar since the conditional expectation of $m(\mathbf{X}_i)-\gamma_\theta(U_{\theta,i})$ given $U_{\theta,i}$ is 0, but no truncation is needed for the first part of (A.20) as

$$\sup_{\theta\in S_c^{d-1}}\max_{1\le i\le n}\left|m(\mathbf{X}_i)-\gamma_\theta(U_{\theta,i})\right|\le C<\infty.$$

Meanwhile, to prove the second part of (A.20), note that for any $p=1,...,d$

$$\frac{\partial}{\partial\theta_p}\left(\mathbf{B}_\theta^T\mathbf{E}_\theta\right)=\left\{\sum_{i=1}^n\frac{\partial}{\partial\theta_p}\left[B_{j,4}(U_{\theta,i})\{m(\mathbf{X}_i)-\gamma_\theta(U_{\theta,i})\}\right]\right\}_{j=-3}^N.$$

By (2.6), $\gamma_\theta(U_\theta)\equiv E\{m(\mathbf{X})|U_\theta\}$, hence $E\left[B_{j,4}(U_\theta)\{m(\mathbf{X})-\gamma_\theta(U_\theta)\}\right]\equiv 0$, for any $\theta\in S_c^{d-1}$, $-3\le j\le N$. Applying Assumptions A2 and A3, one can differentiate through the expectation, thus

$$E\left\{\frac{\partial}{\partial\theta_p}\left[B_{j,4}(U_\theta)\{m(\mathbf{X})-\gamma_\theta(U_\theta)\}\right]\right\}\equiv 0,$$

for any $\theta\in S_c^{d-1}$, $1\le p\le d$, $-3\le j\le N$. Applying the Bernstein's inequality, with probability 1

$$\left\|\left\{\frac{1}{n}\sum_{i=1}^n\frac{\partial}{\partial\theta_p}\left[B_{j,4}(U_{\theta,i})\{m(\mathbf{X}_i)-\gamma_\theta(U_{\theta,i})\}\right]\right\}_{j=-3}^N\right\|_\infty = O\left\{(nh)^{-1/2}\log n\right\},$$

thus the desired result follows.

Lemma A.8. *Under Assumptions A2 and A4-A6, for $\hat{\varepsilon}_\theta(u)$ and $\tilde{\varepsilon}_\theta(u)$ in (A.8) and (A.7)*

$$\sup_{\theta\in S_c^{d-1}}\sup_{u\in[0,1]}\left|\hat{\varepsilon}_\theta(u)\right|=O\left\{(nh)^{-1/2}\log n\right\},a.s.,\tag{A.21}$$

$$\sup_{\theta\in S_c^{d-1}}\sup_{u\in[0,1]}\left|\tilde{\varepsilon}_\theta(u)\right|=O\left\{(nh)^{-1/2}\log n\right\},a.s..\tag{A.22}$$

Proof. Denote

$$\hat{\mathbf{a}} \equiv (\hat{a}_{-3}, \cdots, \hat{a}_N)^T = \left(\mathbf{B}_\theta^T \mathbf{B}_\theta\right)^{-1} \mathbf{B}_\theta^T \mathbf{E} = \mathbf{V}_{n,\theta}^{-1} \left(n^{-1} \mathbf{B}_\theta^{\mathbf{T}} \mathbf{E}\right).$$

By Theorem 5.4.2 in DeVore and Lorentz (1993)

$$\sup_{\theta \in S_c^{d-1}} \sup_{u \in [0,1]} |\hat{\varepsilon}_\theta(u)| \leq \sup_{\theta \in S_c^{d-1}} \|\hat{\mathbf{a}}\|_\infty \leq CN \sup_{\theta \in S_c^{d-1}} \left\|n^{-1} \mathbf{B}_\theta^{\mathbf{T}} \mathbf{E}\right\|_\infty, a.s.,$$

where the last inequality follows from (A.10) of Lemma A.3. Applying (A.19) of Lemma A.7, we have established (A.21). Similarly, (A.22) can be proved.

The next result evaluates the uniform size of the noise derivatives.

Lemma A.9. *Under Assumptions A2-A6, one has with probability* 1

$$\sup_{1 \leq p \leq d} \sup_{\theta \in S_c^{d-1}} \max_{1 \leq i \leq n} \left| \frac{\partial}{\partial \theta_p} \hat{\varepsilon}_\theta(U_{\theta,i}) \right| = O\left\{(nh^3)^{-1/2} \log n\right\}, \quad \text{(A.23)}$$

$$\sup_{1 \leq p \leq d} \sup_{\theta \in S_c^{d-1}} \max_{1 \leq i \leq n} \left| \frac{\partial}{\partial \theta_p} \tilde{\varepsilon}_\theta(U_{\theta,i}) \right| = O\left\{(nh^3)^{-1/2} \log n\right\}, \quad \text{(A.24)}$$

$$\sup_{1 \leq p,q \leq d} \sup_{\theta \in S_c^{d-1}} \max_{1 \leq i \leq n} \left| \frac{\partial^2}{\partial \theta_p \partial \theta_q} \hat{\varepsilon}_\theta(U_{\theta,i}) \right| = O\left\{(nh^5)^{-1/2} \log n\right\}, \quad \text{(A.25)}$$

$$\sup_{1 \leq p,q \leq d} \sup_{\theta \in S_c^{d-1}} \max_{1 \leq i \leq n} \left| \frac{\partial^2}{\partial \theta_p \partial \theta_q} \tilde{\varepsilon}_\theta(U_{\theta,i}) \right| = O\left\{(nh^5)^{-1/2} \log n\right\}. \quad \text{(A.26)}$$

Proof. Note that $\left\{\frac{\partial}{\partial \theta_p} \hat{\varepsilon}_\theta(U_{\theta,i})\right\}_{i=1}^n$ is equal to

$$(\mathbf{I} - \mathbf{P}_\theta) \dot{\mathbf{B}}_p \left(\mathbf{B}_\theta^T \mathbf{B}_\theta\right)^{-1} \mathbf{B}_\theta^T \mathbf{E} + \mathbf{B}_\theta \left(\mathbf{B}_\theta^T \mathbf{B}_\theta\right)^{-1} \dot{\mathbf{B}}_p^T (\mathbf{I} - \mathbf{P}_\theta) \mathbf{E}.$$

Applying Lemmas A.3, A.6 and A.7, one derives (A.23). To prove (A.24), note that

$$\left\{\frac{\partial}{\partial \theta_p} \tilde{\varepsilon}_\theta(U_{\theta,i})\right\}_{i=1}^n = \frac{\partial}{\partial \theta_p} \{\mathbf{P}_\theta \mathbf{E}_\theta\} = \dot{\mathbf{P}}_p \mathbf{E}_\theta + \mathbf{P}_\theta \frac{\partial}{\partial \theta_p} \mathbf{E}_\theta = T_1 + T_2,$$

in which

$$T_1 = \left\{(\mathbf{I} - \mathbf{P}_\theta) \dot{\mathbf{B}}_p - \mathbf{B}_\theta \left(\mathbf{B}_\theta^T \mathbf{B}_\theta\right)^{-1} \dot{\mathbf{B}}_p^T \mathbf{B}_\theta\right\} \left(\mathbf{B}_\theta^T \mathbf{B}_\theta\right)^{-1} \mathbf{B}_\theta^T \mathbf{E}_\theta,$$

$$T_2 = \mathbf{B}_\theta \left(\frac{\mathbf{B}_\theta^T \mathbf{B}_\theta}{n}\right)^{-1} \frac{\partial}{\partial \theta_p} \left(\frac{\mathbf{B}_\theta^T \mathbf{E}_\theta}{n}\right).$$

By (A.10), (A.17), (A.18) and (A.19), one derives

$$\sup_{\theta\in S_c^{d-1}} \|T_1\|_\infty = O\left(n^{-1/2}N^{3/2}\log n\right), a.s.,$$

while (A.20) of Lemma A.7, (A.10) of Lemma A.3

$$\sup_{\theta\in S_c^{d-1}} \|T_2\|_\infty = N \times O\left(n^{-1/2}h^{-1/2}\log n\right) = O\left(n^{-1/2}h^{-3/2}\log n\right), a.s..$$

Thus (A.24) has been established. The proof for (A.25) and (A.26) are similar.
Proof of Proposition A.1. According to the decomposition (A.5), one has

$$\frac{\partial}{\partial\theta_p}\left\{\left(\hat\gamma_\theta-\gamma_\theta\right)\left(U_{\theta,i}\right)\right\} = \frac{\partial}{\partial\theta_p}\left(\tilde\gamma_\theta-\gamma_\theta\right)\left(U_{\theta,i}\right) + \frac{\partial}{\partial\theta_p}\tilde\gamma_\theta\left(U_{\theta,i}\right) + \frac{\partial}{\partial\theta_p}\hat\varepsilon_\theta\left(U_{\theta,i}\right).$$

It is clear from (A.15), (A.23) and (A.24) that with probability 1

$$\sup_{1\le p\le d}\sup_{\theta\in S_c^{d-1}}\max_{1\le i\le n}\left|\frac{\partial}{\partial\theta_p}\left(\tilde\gamma_\theta-\gamma_\theta\right)\left(U_{\theta,i}\right)\right| = O\left(h^3\right),$$

$$\sup_{1\le p\le d}\sup_{\theta\in S_c^{d-1}}\max_{1\le i\le n}\left\{\left|\frac{\partial}{\partial\theta_p}\tilde\varepsilon_\theta\left(U_{\theta,i}\right)\right| + \left|\frac{\partial}{\partial\theta_p}\hat\varepsilon_\theta\left(U_{\theta,i}\right)\right|\right\} = O\left\{\left(nh^3\right)^{-1/2}\log n\right\}.$$

Putting together all the above yields (A.2). The proofs of (A.3) are similar.

A.3. Proof of Proposition A.2

Lemma A.10. *Under Assumptions A2-A6,* $\sup_{\theta\in S_c^{d-1}}\left|\hat R(\theta) - R(\theta)\right| = o(1), a.s.$.

Proof. Let

$$I_1 = \sup_{\theta\in S_c^{d-1}}\left|n^{-1}\sum_{i=1}^n\left\{\hat\gamma_\theta\left(U_{\theta,i}\right)-\gamma_\theta\left(U_{\theta,i}\right)\right\}^2\right|,$$

$$I_2 = \sup_{\theta\in S_c^{d-1}}\left|2n^{-1}\sum_{i=1}^n\left\{\hat\gamma_\theta\left(U_{\theta,i}\right)-\gamma_\theta\left(U_{\theta,i}\right)\right\}\left\{\gamma_\theta\left(U_{\theta,i}\right)-m\left(\mathbf{X}_i\right)-\sigma\left(\mathbf{X}_i\right)\varepsilon_i\right\}\right|,$$

$$I_3 = \sup_{\theta\in S_c^{d-1}}\left|n^{-1}\sum_{i=1}^n\left\{\gamma_\theta\left(U_{\theta,i}\right)-m\left(\mathbf{X}_i\right)\right\}^2 - E\left\{\gamma_\theta\left(U_\theta\right)-m\left(\mathbf{X}\right)\right\}^2\right|,$$

$$\begin{aligned} I_4 &= \sup_{\theta \in S_c^{d-1}} \left\{ \left| \frac{1}{n} \sum_{i=1}^{n} \sigma^2 \left(\mathbf{X}_i \right) \varepsilon_i^2 - E \left[\sigma^2 \left(\mathbf{X} \right) \right] \right| \right. \\ & \left. + \left| \frac{2}{n} \sum_{i=1}^{n} \left\{ \gamma_\theta \left(U_{\theta,i} \right) - m \left(\mathbf{X}_i \right) \right\} \sigma \left(\mathbf{X}_i \right) \varepsilon_i \right| \right\}, \end{aligned}$$

then

$$\sup_{\theta \in S_c^{d-1}} \left| \hat{R}(\theta) - R(\theta) \right| \leq I_1 + I_2 + I_3 + I_4.$$

Bernstein inequality and strong law of large number for α-mixing sequence imply that with probability 1, $I_3 + I_4 = o(1)$. Now (A.1) of Proposition A.1 provides that

$$\sup_{\theta \in S_c^{d-1}} \sup_{u \in [0,1]} \left| \hat{\gamma}_\theta (u) - \gamma_\theta (u) \right| = O \left(n^{-1/2} h^{-1/2} \log n + h^4 \right), a.s.,$$

which entail that $I_1 = O\left\{ \left(n^{-1/2} h^{-1/2} \log n \right)^2 + \left(h^4 \right)^2 \right\}$ almost surely. On the other hand

$$I_2 \leq O \left\{ (nh)^{-1/2} \log n + h^4 \right\} \times \sup_{\theta \in S_c^{d-1}} 2n^{-1} \sum_{i=1}^{n} \left| \gamma_\theta \left(U_{\theta,i} \right) - m \left(\mathbf{X}_i \right) - \sigma \left(\mathbf{X}_i \right) \varepsilon_i \right|.$$

Hence $I_2 \leq O\left(n^{-1/2} h^{-1/2} \log n + h^4 \right)$ almost surely. The lemma follows from Assumption A6.

Lemma A.11. *Under Assumptions A2 - A6, for $k = 1, 2$, with probability 1*

$$\sup_{\theta \in S_c^{d-1}} \left| \frac{\partial^k}{\partial \theta^k} \left\{ \hat{R}(\theta) - R(\theta) \right\} \right| = O \left(n^{-1/2} h^{-1/2-k} \log n + h^{4-k} \right). \qquad \text{(A.27)}$$

Proof. Note that for any $p = 1, 2, ..., d$

$$\begin{aligned} \frac{1}{2} \frac{\partial}{\partial \theta_p} \hat{R}(\theta) &= n^{-1} \sum_{i=1}^{n} \left\{ \hat{\gamma}_\theta \left(U_{\theta,i} \right) - Y_i \right\} \frac{\partial}{\partial \theta_p} \hat{\gamma}_\theta \left(U_{\theta,i} \right), \\ \frac{1}{2} \frac{\partial}{\partial \theta_p} R(\theta) &= E \left[\left\{ \gamma_\theta \left(U_\theta \right) - m \left(\mathbf{X} \right) - \sigma \left(\mathbf{X} \right) \varepsilon \right\} \frac{\partial}{\partial \theta_p} \gamma_\theta \left(U_\theta \right) \right]. \end{aligned}$$

Denote

$$\xi_{\theta,i,p} = 2 \left\{ \gamma_\theta \left(U_{\theta,i} \right) - Y_i \right\} \frac{\partial}{\partial \theta_p} \gamma_\theta \left(U_{\theta,i} \right) - \frac{\partial}{\partial \theta_p} R(\theta),$$

then $E\left(\xi_{\theta,i,p}\right) = 2E\left[\left\{\gamma_\theta\left(U_{\theta,i}\right) - Y_i\right\} \frac{\partial}{\partial\theta_p}\gamma_\theta\left(U_{\theta,i}\right)\right] - \frac{\partial}{\partial\theta_p}R\left(\theta\right) = 0$ and

$$\frac{1}{2}\frac{\partial}{\partial\theta_p}\left\{\hat{R}\left(\theta\right) - R\left(\theta\right)\right\} = \left(2n\right)^{-1}\sum_{i=1}^{n}\xi_{\theta,i,p} + J_{1,\theta,p} + J_{2,\theta,p} + J_{3,\theta,p}, \qquad \text{(A.28)}$$

with

$$\begin{aligned}
J_{1,\theta,p} &= n^{-1}\sum_{i=1}^{n}\left\{\hat{\gamma}_\theta\left(U_{\theta,i}\right) - \gamma_\theta\left(U_{\theta,i}\right)\right\}\frac{\partial}{\partial\theta_p}\left(\hat{\gamma}_\theta - \gamma_\theta\right)\left(U_{\theta,i}\right), \\
J_{2,\theta,p} &= n^{-1}\sum_{i=1}^{n}\left\{\gamma_\theta\left(U_{\theta,i}\right) - m\left(\mathbf{X}_i\right) - \sigma\left(\mathbf{X}_i\right)\varepsilon_i\right\}\frac{\partial}{\partial\theta_p}\left(\hat{\gamma}_\theta - \gamma_\theta\right)\left(U_{\theta,i}\right), \\
J_{3,\theta,p} &= n^{-1}\sum_{i=1}^{n}\left\{\hat{\gamma}_\theta\left(U_{\theta,i}\right) - \gamma_\theta\left(U_{\theta,i}\right)\right\}\frac{\partial}{\partial\theta_p}\gamma_\theta\left(U_{\theta,i}\right).
\end{aligned}$$

Bernstein inequality implies that

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|n^{-1}\sum_{i=1}^{n}\xi_{\theta,i,p}\right| = O\left(n^{-1/2}\log n\right), a.s..$$

Meanwhile, applying (A.1) and (A.2) of Proposition A.1, one obtains that

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|J_{1,\theta,p}\right| = O\left(n^{-1}h^{-2}\log^2 n + h^7\right), a.s.. \qquad \text{(A.29)}$$

Note that

$$\begin{aligned}
J_{2,\theta,p} &= n^{-1}\sum_{i=1}^{n}\left\{\gamma_\theta\left(U_{\theta,i}\right) - m\left(\mathbf{X}_i\right) - \sigma\left(\mathbf{X}_i\right)\varepsilon_i\right\}\frac{\partial}{\partial\theta_p}\left(\tilde{\gamma}_\theta - \gamma_\theta\right)\left(U_{\theta,i}\right) \\
&\quad - n^{-1}\left(\mathbf{E} + \mathbf{E}_\theta\right)^T\frac{\partial}{\partial\theta_p}\left\{\mathbf{P}_\theta\left(\mathbf{E} + \mathbf{E}_\theta\right)\right\}.
\end{aligned}$$

Applying (A.1), one gets

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|J_{2,\theta,p} + n^{-1}\left(\mathbf{E} + \mathbf{E}_\theta\right)^T\frac{\partial}{\partial\theta_p}\left\{\mathbf{P}_\theta\left(\mathbf{E} + \mathbf{E}_\theta\right)\right\}\right| = O\left(h^3\right), a.s.,$$

while Lemmas A.3 and A.7 entail that with probability 1

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|n^{-1}\left(\mathbf{E} + \mathbf{E}_\theta\right)^T\frac{\partial}{\partial\theta_p}\left\{\mathbf{P}_\theta\left(\mathbf{E} + \mathbf{E}_\theta\right)\right\}\right| = O\left\{n^{-1}N\log^2 n\right\},$$

thus

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|J_{2,\theta,p}\right| = O\left(h^3+n^{-1}N\log^2 n\right), a.s.. \tag{A.30}$$

Lastly

$$J_{3,\theta,p}-\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\gamma}_\theta-\gamma_\theta\right)\frac{\partial}{\partial\theta_p}\gamma_\theta\left(U_{\theta,i}\right)=\frac{1}{n}\left(\mathbf{E}+\mathbf{E}_\theta\right)^T\mathbf{B}_\theta\left(\frac{\mathbf{B}_\theta^T\mathbf{B}_\theta}{n}\right)^{-1}\frac{\mathbf{B}_\theta^T}{n}\frac{\partial}{\partial\theta_p}\gamma_\theta.$$

Applying Lemmas A.3 and A.7 again, one has with probability 1

$$\begin{aligned}&\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|\left(n^{-1}\mathbf{B}_\theta^T\mathbf{E}+n^{-1}\mathbf{B}_\theta^{\mathbf{T}}\mathbf{E}_\theta\right)^T\left(\frac{\mathbf{B}_\theta^T\mathbf{B}_\theta}{n}\right)^{-1}\frac{\mathbf{B}_\theta^T}{n}\frac{\partial}{\partial\theta_p}\gamma_\theta\right|\\ =\;& O\left\{n^{-1}\log^2 n+(nN)^{-1/2}\log n\right\},\end{aligned}$$

while by applying (A.11) of Lemma A.4, one has

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|n^{-1}\sum_{i=1}^{n}\left(\tilde{\gamma}_\theta-\gamma_\theta\right)\frac{\partial}{\partial\theta_p}\gamma_\theta\left(U_{\theta,i}\right)\right|=O\left(h^4\right), a.s.,$$

together, the above entail that

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|J_{3,\theta,p}\right|=O\left\{h^4+n^{-1}\log^2 n+(nN)^{-1/2}\log n\right\}, a.s.. \tag{A.31}$$

Therefore, (A.28), (A.29), (A.30), (A.31) and Assumption A6 lead to

$$\sup_{\theta\in S_c^{d-1}}\sup_{1\le p\le d}\left|\frac{\partial}{\partial\theta_p}\left\{\hat{R}\left(\theta\right)-R\left(\theta\right)\right\}-n^{-1}\sum_{i=1}^{n}\xi_{\theta,i,p}\right|=o\left(n^{-1/2}\right),\ a.s.,$$

which establishes (A.27) for $k=1$. Note that the second order derivative of $\hat{R}(\theta)$ and $R(\theta)$ with respect to θ_p, θ_q are

$$2n^{-1}\left[\sum_{i=1}^{n}\left\{\hat{\gamma}_\theta\left(U_{\theta,i}\right)-Y_i\right\}\frac{\partial^2}{\partial\theta_p\partial\theta_q}\hat{\gamma}_\theta\left(U_{\theta,i}\right)+\sum_{i=1}^{n}\frac{\partial}{\partial\theta_q}\hat{\gamma}_\theta\left(U_{\theta,i}\right)\frac{\partial}{\partial\theta_p}\hat{\gamma}_\theta\left(U_{\theta,i}\right)\right],$$

$$2\left[E\left\{\gamma_\theta\left(U_\theta\right)-m\left(\mathbf{X}\right)\right\}\frac{\partial^2}{\partial\theta_p\partial\theta_q}\gamma_\theta\left(U_\theta\right)+E\left\{\frac{\partial}{\partial\theta_q}\gamma_\theta\left(U_\theta\right)\frac{\partial}{\partial\theta_p}\gamma_\theta\left(U_\theta\right)\right\}\right].$$

The proof of (A.27) for $k=2$ follows from (A.1), (A.2) and (A.3).

Proof of Proposition A.2. The result follows from Lemma A.10, Lemma A.11.

A.4. Proof of Theorem 2

For any $p = 1, 2, ..., d-1$, let $\hat{S}_p^* (\theta_{-d})$ be the p-th element of $\hat{S}^* (\theta_{-d})$, and for any $t \in [0,1]$, let $f_p(t) = \hat{S}_p^* \left(t\hat{\theta}_{-d} + (1-t)\,\theta_{0,-d} \right)$, then

$$\frac{d}{dt} f_p(t) = \sum_{q=1}^{d-1} \frac{\partial}{\partial \theta_q} \hat{S}_p^* \left(t\hat{\theta}_{-d} + (1-t)\,\theta_{0,-d} \right) \left(\hat{\theta}_q - \theta_{0,q} \right).$$

Note that $\hat{S}^* (\theta_{-d})$ attains its minimum at $\hat{\theta}_{-d}$, i.e., $\hat{S}_p^* \left(\hat{\theta}_{-d} \right) \equiv 0$. Thus, for any $p = 1, 2, ..., d-1$, $t_p \in [0,1]$, one has

$$\begin{aligned} -\hat{S}_p^* (\theta_{0,-d}) &= f_p(1) - f_p(0) \\ &= \left\{ \frac{\partial^2}{\partial \theta_q \theta_p} \hat{R}^* \left(t_p \hat{\theta}_{-d} + (1-t_p)\,\theta_{0,-d} \right) \right\}_{q=1,...,d-1}^T \left(\hat{\theta}_{-d} - \theta_{0,-d} \right), \end{aligned}$$

then

$$-\hat{S}^* (\theta_{0,-d}) = \left\{ \frac{\partial^2}{\partial \theta_q \partial \theta_p} \hat{R}^* \left(t_p \hat{\theta}_{-d} + (1-t_p)\,\theta_{0,-d} \right) \right\}_{p,q=1,...,d-1} \left(\hat{\theta}_{-d} - \theta_{0,-d} \right).$$

Now Theorem 1 and Proposition A.2 with $k = 2$ imply that

$$\frac{\partial^2}{\partial \theta_q \partial \theta_p} \hat{R}^* \left(t_p \hat{\theta}_{-d} + (1-t_p)\,\theta_{0,-d} \right) \longrightarrow l_{q,p}, a.s., \; p, q = 1, 2, ..., d-1 \qquad \text{(A.32)}$$

where $l_{p,q}$ is given in Theorem 2. Noting that $\sqrt{n} \left(\hat{\theta}_{-d} - \theta_{0,-d} \right)$ is represented as

$$-\left[\left\{ \frac{\partial^2}{\partial \theta_q \partial \theta_p} \hat{R}^* \left(t_p \hat{\theta}_{-d} + (1-t_p)\,\theta_{0,-d} \right) \right\}_{p,q=1,...,d-1} \right]^{-1} \sqrt{n} \hat{S}^* (\theta_{0,-d}),$$

where $\hat{S}^* (\theta_{0,-d}) = \left\{ \hat{S}_p^* (\theta_{0,-d}) \right\}_{p=1}^{d-1}$. For γ_θ in (2.6), denote

$$\eta_{i,p} := 2 \left\{ \dot{\gamma}_p - \theta_{0,p} \theta_{0,d}^{-1} \dot{\gamma}_d \right\} (U_{\theta_0,i}) \left\{ \gamma_{\theta_0} (U_{\theta_0,i}) - Y_i \right\},$$

where $\dot{\gamma}_p$ is value of $\frac{\partial}{\partial\theta_p}\gamma_\theta$ taking at $\theta = \theta_0$, for any $p, q = 1, 2, ..., d-1$. According to Lemma A.16 in Wang and Yang (2007b), with probability 1, one has

$$\hat{S}_p^* \left(\theta_{0,-d}\right) = n^{-1} \sum_{i=1}^{n} \eta_{p,i} + o\left(n^{-1/2}\right), \ E\left(\eta_{p,i}\right) = 0.$$

Let $\Psi\left(\theta_0\right) = \left(\psi_{pq}\right)_{p,q=1}^{d-1}$ be the covariance matrix of $\sqrt{n}\left\{\hat{S}_p^*\left(\theta_{0,-d}\right)\right\}_{p=1}^{d-1}$ with ψ_{pq} given in Theorem 2. Cramér-Wold device and central limit theorem for α mixing sequences entail that $\sqrt{n}\hat{S}^*\left(\theta_{0,-d}\right) \xrightarrow{d} N\left\{\mathbf{0}, \Psi\left(\theta_0\right)\right\}$. Let

$$\Sigma\left(\theta_0\right) = \left\{H^*\left(\theta_{0,-d}\right)\right\}^{-1} \Psi\left(\theta_0\right) \left[\left\{H^*\left(\theta_{0,-d}\right)\right\}^T\right]^{-1},$$

with $H^*\left(\theta_{0,-d}\right)$ being the Hessian matrix defined in (2.3). The above limiting distribution of $\sqrt{n}\hat{S}^*\left(\theta_{0,-d}\right)$, (A.32) and Slutsky's theorem imply the desired result.