Supplementary Material to "Estimation for Functional Single Index Models with Unknown Link Functions"

Yunxiang Huang and Qihua Wang

Academy of Mathematics and Systems Science, CAS,Beijing 100080, China

Supplementary Material

This supplementary material gives proofs of Lemmas 1 and 2 and Theorems 1–3 in the main paper.

S1 Proofs of Lemmas 1 and 2

We first introduce some useful results depending on the FPCA. In what follows, we use c and δ to denote positive constants that may change from line to line. Define $|||\Sigma|||^2 = \iint_{\mathcal{I}^2}[\Sigma(s,t)]^2\,ds\,dt$. By using the result in Bhatia et al. (1983), we can obtain

$$\sup_{j\geq 1}|\lambda_j - \hat{\lambda}_j| \leq \left|\left|\left|\Sigma - \hat{\Sigma}\right|\right|\right|, \text{ and } \sup_{j\geq 1}\delta_j\|\phi_j - \hat{\phi}_j\|_2 \leq 8^{1/2}\left|\left|\left|\Sigma - \hat{\Sigma}\right|\right|\right|,$$

where $\delta_j = \min_{1\leq k\leq j}(\lambda_k - \lambda_{k+1})$. Such results can also be found in Bosq (1991) and Chapter 4 of Bosq (2000). Provided $\int_{\mathcal{I}} \mathrm{E}[X(t)]^4\,dt < \infty$, it is not hard to prove $\int_{\mathcal{I}}|\mu(t) - \hat{\mu}(t)|\,dt = O_p(n^{-1/2})$ and $\mathrm{E}\left|\left|\left|\Sigma - \hat{\Sigma}\right|\right|\right|^2 = O(n^{-1})$.

Therefore, we have

$$\sup_{j\geq 1} |\lambda_j - \hat{\lambda}_j| = O_p(n^{-1/2}).$$

Lemma 1 is a direct consequence of Lemma 6.1 in Hall and Horowitz (2007).

Lemma 1. *If* $\inf_{k\neq j} |\hat{\lambda}_j - \lambda_k| > 0$*, then*

$$\hat{\phi}_j(t) - \phi_j(t) = \sum_{k:\, k\neq j} \left(\hat{\lambda}_j - \lambda_k\right)^{-1} \phi_k(t) \iint_{\mathcal{I}^2} \left[\hat{\Sigma}(s,\, t) - \Sigma(s,\, t)\right] \hat{\phi}_j(s)\phi_k(t)\, ds\, dt \\ + \phi_j(t) \int_{\mathcal{I}} \left[\hat{\phi}_j(s) - \phi_j(s)\right] \phi_j(s)\, ds\,. \tag{S1.1}$$

Define the event $\mathcal{E}_p$ by

$$\mathcal{E}_p = \mathcal{E}_p(n) = \left\{ \left|\left|\left| \Sigma - \hat{\Sigma} \right|\right|\right| \leq n^{-1/2+\delta} \right\},$$

for some small $\delta > 0$ and the event $\mathcal{F}_p$ by

$$\mathcal{F}_p = \mathcal{F}_p(n) = \left\{ (\hat{\lambda}_j - \lambda_k)^{-2} \leq 2(\lambda_j - \lambda_k)^{-2},\ \text{for all } 1 \leq j \neq k \leq p \right\}.$$

Then we have $\mathrm{P}(\mathcal{E}_p) \to 1$. By using $\lambda_j - \lambda_{j+1} \geq j^{-\alpha_0 - 1}$ in condition (A1)(i), we have $\mathcal{E}_p \subset \mathcal{F}_p$ for some small $\delta > 0$ when n and p are large enough. So it suffices to derive asymptomatic results when $\mathcal{F}_p$ holds. Under conditions (A1) and (A2), it follows that

$$\mathrm{E}\left\{ \sup_{1\leq j\leq p} [j^{-2} \|\phi_j - \hat{\phi}_j\|_2^2\, \mathrm{I}(\mathcal{F}_p)] \right\} = O(n^{-1}), \tag{S1.2}$$

whose proof can be found in Hall and Horowitz (2007), from page 83 to the end.

The following lemma gives a useful uniform consistency result of the FPC score estimates and is also of independent interest in FPCA. Also see Proposition 1 of Wong et al. (2019).

Lemma 2. *Under conditions (A1) and (A2), we have, uniformly over* $1 \leq j \leq p$,

$$\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\hat{x}_{ij})^2 = O_p(\max\{n^{-1}j^{2-\alpha_0},\, n^{-1}\}), \tag{S1.3}$$

and

$$\mathrm{E}(x_{\cdot j}-\hat{x}_{\cdot j})^2 = O(\max\{n^{-1}j^{2-\alpha_0},\, n^{-1}\}). \tag{S1.4}$$

Proof. We first prove (S1.4). Note that $(\hat{\lambda}_j,\, \hat{\phi}_j)$ is determined by $\hat{\Sigma}$ for all j. By using Lemma 1 and the orthonormalities of eigenfunctions and FPC scores, we have, uniformly over $1 \leq j \leq p$,

$$\begin{aligned}
&\lambda_j^{-1}\,\mathrm{E}[(x_{\cdot j}-\hat{x}_{\cdot j})^2 \mid \hat{\Sigma}] \\
&\leq 2\lambda_j^{-1}\,\mathrm{E}\left\{\left[\int_{\mathcal{I}}(X(t)-\mu(t))(\phi_j(t)-\hat{\phi}_j(t))\,dt\right]^2 \,\middle|\, \hat{\Sigma}\right\} \\
&\qquad + 2\lambda_j^{-1}\,\mathrm{E}\left\{\left[\int_{\mathcal{I}}(\mu(t)-\hat{\mu}(t))\hat{\phi}_j(t)\,dt\right]^2 \,\middle|\, \hat{\Sigma}\right\} \\
&\leq 4\lambda_j^{-1}\sum_{k:\,k\neq j}\lambda_k\Big(\hat{\lambda}_j-\lambda_k\Big)^{-2}\left\{\iint_{\mathcal{I}^2}\Big[\hat{\Sigma}(s,\,t)-\Sigma(s,\,t)\Big]\hat{\phi}_j(s)\phi_k(t)\,ds\,dt\right\}^2 \\
&\qquad + 4\left[\int_{\mathcal{I}}\Big(\hat{\phi}_j(s)-\phi_j(s)\Big)\phi_j(s)\,ds\right]^2 + 2\lambda_j^{-1}\,\mathrm{E}[\|\mu-\hat{\mu}\|_2^2 \mid \hat{\Sigma}] \\
&=: S_1(j)+S_{2,1}(j)+S_{2,2}(j).
\end{aligned} \tag{S1.5}$$

For $S_{2,1}(j)$, by using (S1.2), we have

$$\mathrm{E}\left\{\sup_{1\le j\le p}[j^{-2}S_{2,1}(j)\,\mathrm{I}(\mathcal{F}_p)]\right\} = O(n^{-1}). \tag{S1.6}$$

For $S_{2,2}(j)$, given $\int_{\mathcal{I}} \mathrm{E}[X(t)]^4\,dt < \infty$, we can obtain

$$\mathrm{E}[\lambda_j S_{2,2}(j)] = O(n^{-1}). \tag{S1.7}$$

We now consider $S_1(j)$. Write

$$\begin{aligned}
&\left\{\iint_{\mathcal{I}^2}\left[\hat\Sigma(s,\,t)-\Sigma(s,\,t)\right]\hat\phi_j(s)\phi_k(s)\,dt\right\}^2\\
&\le 2\left\{\iint_{\mathcal{I}^2}\left[\hat\Sigma(s,\,t)-\Sigma(s,\,t)\right]\phi_j(s)\phi_k(t)\,ds\,dt\right\}^2\\
&\quad + 2\left\{\iint_{\mathcal{I}^2}\left[\hat\Sigma(s,\,t)-\Sigma(s,\,t)\right][\hat\phi_j(s)-\phi_j(s)]\phi_k(t)\,ds\,dt\right\}^2 \qquad (\text{S1.8})\\
&=: S_{1,1}(jk)+S_{1,2}(jk).
\end{aligned}$$

By using $\mathrm{E}\left|\!\left|\!\left|\Sigma-\hat\Sigma\right|\!\right|\!\right|^2 = O(n^{-1})$ and (S1.2), we have

$$\mathrm{E}\left\{\sup_{1\le j,\,k\le p}[j^{-2}S_{1,2}(jk)\,\mathrm{I}(\mathcal{F}_p)]\right\} = O(n^{-2}). \tag{S1.9}$$

It can be shown that

$$\mathrm{E}\left\{\sup_{1\le j,\,k\le p}[(\lambda_j\lambda_k)^{-1}S_{1,1}(jk)\,\mathrm{I}(\mathcal{F}_p)]\right\} = O(n^{-1}). \tag{S1.10}$$

See Section 5.3 of Hall and Horowitz (2007). Given conditions (A1) and (A2), a direct calculation yields

$$n^{-1}\sup_{1\le j,\,k\le p}(\lambda_j^{-1}\lambda_k^{-1}j^2) \le cn^{-1}p^{2+2\alpha_0} = o(1). \tag{S1.11}$$

Combining this with (S1.8)–(S1.10) yields

$$\mathrm{E}\sup_{1\le j\le p}\left\{\left[\sum_{k:\,k\neq j}(\lambda_j-\lambda_k)^{-2}\lambda_k^2\right]^{-1}S_1(j)\,\mathrm{I}(\mathcal{F}_p)\right\}=O\big(n^{-1}\big). \tag{S1.12}$$

It remains to bound the summation on the left side of (S1.12), which we now discuss.

Let $p_1=[j/2]$ and $p_2=2j$, where $[a]$ denotes the integer part of a. From condition (A1), we have, for all $j=1,\ldots,p$,

$$\frac{\lambda_k^2}{(\lambda_j-\lambda_k)^2}\le\frac{2(\lambda_j-\lambda_k)^2+2\lambda_j^2}{(\lambda_j-\lambda_k)^2}\le c,\ \text{for } 1\le k\le p_1, \tag{S1.13}$$

$$\frac{\lambda_j^2}{(\lambda_j-\lambda_k)^2}\le c,\ \text{for } k\ge p_2, \tag{S1.14}$$

and

$$\frac{\lambda_k^2}{(\lambda_j-\lambda_k)^2}\le\frac{cj^{-2\alpha_0}}{(j-k)^2j^{-2(\alpha_0+1)}}\le\frac{cj^2}{(j-k)^2},\ \text{for } p_1<k<p_2,\ k\neq j, \tag{S1.15}$$

where $c>0$ is a constant not depending on j and k. Combing (S1.13)–

(S1.15), we have, uniformly over $1 \le j \le p$,

$$\begin{aligned}
&\sum_{k:\, k\neq j} (\lambda_j - \lambda_k)^{-2}\lambda_k^2 \\
&= \sum_{k=1}^{p_1} \frac{\lambda_k^2}{(\lambda_j - \lambda_k)^2} + \sum_{k=p_2}^{\infty} \frac{\lambda_k^2}{(\lambda_j - \lambda_k)^2} + \sum_{k:\, p_1<k<p_2,\, k\neq j} \frac{\lambda_k^2}{(\lambda_j - \lambda_k)^2} \\
&\le cp_1 + \sum_{k=p_2}^{\infty} \frac{\lambda_k^2}{\lambda_j^2} \frac{\lambda_j^2}{(\lambda_j - \lambda_k)^2} + \sum_{k:\, p_1<k<p_2,\, k\neq j} \frac{cj^2}{(j-k)^2} \qquad \text{(S1.16)} \\
&\le cj + c\lambda_j^{-2} \sum_{k=p_2}^{\infty} \lambda_k^2 + cj^2 \\
&\le cj^2,
\end{aligned}$$

where c is a constant not depending on j. Combining this with (S1.12) gives

$$\begin{aligned}
&\mathrm{E}\left\{\sup_{1\le j\le p} j^{-2}[S_1(j)\,\mathrm{I}(\mathcal{F}_p)]\right\} \\
&\le \sup_{1\le j\le p}\left[j^{-2}\sum_{k:\, k\neq j}\frac{\lambda_k^2}{(\lambda_j-\lambda_k)^2}\right]\mathrm{E}\sup_{1\le j\le p}\left\{\left[\sum_{k:\, k\neq j}\frac{\lambda_k^2}{(\lambda_j-\lambda_k)^2}\right]^{-1} S_1(j)\,\mathrm{I}(\mathcal{F}_p)\right\} \\
&= O(n^{-1}).
\end{aligned} \qquad \text{(S1.17)}$$

Combining this with (S1.5)– (S1.7) yields, uniformly over $1 \le j \le p$,

$$\mathrm{E}[(x_{\cdot j} - \hat{x}_{\cdot j})^2\,\mathrm{I}(\mathcal{F}_p)] = O(\max\{n^{-1}\lambda_j j^2,\, n^{-1}\}). \qquad \text{(S1.18)}$$

Condition (A1)(iii) assumes $\lambda_j \le cj^{-\alpha_0}$ for all $j > 0$. Therefore, the proof of Lemma 2 completes if the factor $\mathrm{I}(\mathcal{F}_p)$ can be removed from the left side. Since we assumed that all moments of the principal component scores are finite, by using Markov's inequality, it can be shown that $\mathrm{P}(\mathcal{E}_p) = 1 - O(n^{-c})$ and hence $\mathrm{P}(\mathcal{F}_p) = 1 - O(n^{-c})$ for any $c > 0$, which completes the proof of (S1.4).

We now consider (S1.3). By an argument similar to (S1.5), we can prove, uniformly over $1 \le j \le p$,

$$\iint_{\mathcal{I}^2} \Sigma(s,\, t)(\phi_j(s) - \hat{\phi}_j(s))(\phi_j(t) - \hat{\phi}_j(t))\, ds\, dt = O_p(n^{-1} j^{2-\alpha_0}). \quad \text{(S1.19)}$$

By using this, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \hat{x}_{ij})^2 \\
&\le \frac{2}{n}\sum_{i=1}^{n}\left[\int_{\mathcal{I}}(X_i(t) - \hat{\mu}(t))(\phi_j(t) - \hat{\phi}_j(t))\, dt\right]^2 + 2\left[\int_{\mathcal{I}}(\mu(t) - \hat{\mu}(t))\phi_j(t)\, dt\right]^2 \\
&\le 4\iint_{\mathcal{I}^2} \Sigma(s,\, t)(\phi_j(s) - \hat{\phi}_j(s))(\phi_j(t) - \hat{\phi}_j(t))\, ds\, dt \\
&\qquad + 4\left|\!\left|\!\left|\Sigma - \hat{\Sigma}\right|\!\right|\!\right| \|\phi_j(t) - \hat{\phi}_j(t)\|_2^2 + 2\,\mathrm{E}\,\|\mu - \hat{\mu}\|_2^2 \\
&= O_p(\max\{n^{-1} j^{2-\alpha_0},\, n^{-1}\}),
\end{aligned} \quad \text{(S1.20)}$$

which completes the proof of (S1.3). □

Proof of Lemma 1. We first prove (3.1). By the construction of $\tilde{\beta}_p$, we have

$$\begin{aligned}
&\|\tilde{\beta}_p - \beta_0\|_2^2 \\
&= \left\| \sum_{j=1}^{p} b_{0j}(\hat{\phi}_j - \phi_j) + \Big(\tilde{b}_p - b_{0p}\Big)\hat{\phi}_j - \sum_{j=p+1}^{\infty} b_{0j}\phi_j \right\|_2^2 \qquad \text{(S1.21)} \\
&\le 3p \sum_{j=1}^{p} b_{0j}^2 \|\hat{\phi}_j - \phi_j\|_2^2 + 3\Big(\tilde{b}_p - b_{0p}\Big)^2 + 3 \sum_{j=p+1}^{\infty} b_{0j}^2.
\end{aligned}$$

When $\mathcal{F}_p$ holds, by using (S1.2), we have

$$p \sum_{j=1}^{p} b_{0j}^2 \, \mathrm{E}[\|\hat{\phi}_j - \phi_j\|_2^2 \, \mathrm{I}(\mathcal{F}_p)] = O\left(n^{-1} p \sum_{j=1}^{p} j^{-2\alpha_1+2} \right) = O\big(n^{-1}p\big), \quad \text{(S1.22)}$$

where the last equality is due to $-2\alpha_1 + 2 < -\alpha_0 < -1$. Simple calculations yield

$$\sum_{j=p+1}^{\infty} b_{0j}^2 = O\big(n^{-(2\alpha_1-1)/(\alpha_0+2\alpha_1)}\big), \qquad \text{(S1.23)}$$

and

$$(\tilde{b}_p - b_{0p})^2 \le \sum_{j=p+1}^{\infty} b_{0j}^2 = O\big(n^{-(2\alpha_1-1)/(\alpha_0+2\alpha_1)}\big). \qquad \text{(S1.24)}$$

Combing (S1.21)–(S1.24), we have

$$\mathrm{E}[\|\tilde{\beta}_p - \beta_0\|_2^2 \, \mathrm{I}(\mathcal{F}_p)] = O\big(n^{-(2\alpha_1-1)/(\alpha_0+2\alpha_1)}\big). \qquad \text{(S1.25)}$$

Therefore, the proof of (3.1) completes if the factor $\mathrm{I}(\mathcal{F}_p)$ can be removed from the left side. Since we assumed that all moments of the principal component scores are finite, by using Markov's inequality, it can be shown that $\mathrm{P}(\mathcal{E}_p) = 1 - O(n^{-c})$ and hence $\mathrm{P}(\mathcal{F}_p) = 1 - O(n^{-c})$ for any $c > 0$, which complete the proof of (3.1).

We now turn to proving (3.2). Direct calculations yield

$$\begin{aligned}
&\mathrm{E}\left[(\hat{\boldsymbol{x}}_{\cdot p})^T\tilde{\boldsymbol{b}}_p - \sum_{j=1}^{\infty} x_{\cdot j} b_{0j}\right]^2 \\
&\le c\,\mathrm{E}\left[\sum_{j=1}^{p}(\hat{x}_{\cdot j} - x_{\cdot j})b_{0j}\right]^2 + c\,\mathrm{E}[\hat{x}_{\cdot p}(\tilde{b}_p - b_{0p})]^2 + c\,\mathrm{E}\left(\sum_{j=p+1}^{\infty} x_{\cdot j} b_{0j}\right)^2 \\
&\le cp\sum_{j=1}^{p}\mathrm{E}(\hat{x}_{\cdot j} - x_{\cdot j})^2 b_{0j}^2 + c(\tilde{b}_p - b_{0p})^2\,\mathrm{E}\,\hat{x}_{\cdot p}^2 + c\sum_{j=p+1}^{\infty} b_{0j}^2\,\mathrm{E}\,x_{\cdot j}^2 \\
&\le cp\sum_{j=1}^{p} n^{-1} j^{-2\alpha_1}\max\{j^{2-\alpha_0},\, 1\} + cp^{-2\alpha_1+1}p^{-\alpha_0} + c\sum_{j=p+1}^{\infty} j^{-\alpha_0-2\alpha_1} \\
&\le cn^{-1}(p^{-\alpha_0-2\alpha_1+4} + p^{-2\alpha_1+2}) + cp^{-\alpha_0-2\alpha_1+1} + cp^{-\alpha_0-2\alpha_1+1} \\
&\le c(n^{-1}p),
\end{aligned}$$

where c is a positive constant and the third inequality results from (S1.4) in Lemma 2. From this, (3.2) follows immediately. □

Proof of Lemma 2. Observing that

$$\|\beta_p - \tilde{\beta}\|_2^2 = \sum_{j=1}^{p}(b_j - \tilde{b}_j)^2 \le \lambda_p^{-1}\sum_{j=1}^{p}\lambda_j(\hat{b}_j - \tilde{b}_j)^2 \le \frac{C_1 p}{n\lambda_p},$$

(3.4) follows immediately.

We now consider (3.5). Given $\{\hat{\phi}_j\}_{j=1}^p$, let Γ_p be a $p\times p$ random matrix with the $(j,\, k)$ entry

$$(\Gamma_p)_{jk} = \lambda_j^{-1/2}\lambda_k^{-1/2}\hat{x}_{\cdot j}\hat{x}_{\cdot k}.$$

Let $|||\Gamma_p|||$ be the spectral norm of Γ_p. We have

$$\mathrm{E}\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}[\hat{\boldsymbol{x}}_{\cdot p}^T(\boldsymbol{b}_p-\tilde{\boldsymbol{b}}_p)]^2 \le \sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sum_{j=1}^{p}\lambda_j(b_j-\tilde{b}_j)^2 \times |||\mathrm{E}\,\Gamma_p||| \le \frac{C_1 p}{n}|||\mathrm{E}\,\Gamma_p|||. \tag{S1.26}$$

By using Lemma 2, we can obtain, uniformly over $1 \le j,\ k \le p$,

$$\mathrm{E}\,|(\Gamma_p)_{jk} - \lambda_j^{-1/2}\lambda_k^{-1/2}x_{\cdot j}x_{\cdot k}| = O\left(\frac{\max\{p,\ p^{\alpha_0/2}\}}{\sqrt{n}}\right). \tag{S1.27}$$

It is well known that the largest eigenvalue of a semi-positive definite matrix is not larger than its maximum row sum of absolute values. Therefore, by using (S1.27) and the orthonormality of FPC scores, it follows that

$$|||\mathrm{E}\,\Gamma_p||| \le 1 + p \times O\left(\frac{\max\{p,\ p^{\alpha_0/2}\}}{\sqrt{n}}\right) = 1 + o(1).$$

Substituting this into (S1.26) completes the proof of (3.5). □

S2 Auxiliary Lemmas

We first give some auxiliary lemmas about kernel estimators with functional predictors. Lemma 3 provides the uniform convergence for $s_l(u \mid \boldsymbol{b}_p,\ h)$ and $s_l(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\ h)$, which is very useful for kernel estimators. By using Lemma 3, we can obtain Lemmas 4–7, which provide the rates of convergence for the estimators of the nonparametric parts. Lemma 8 is a direct generalization of Lemma A.1 in Wang et al. (2010) with diverging

p. Lemma 9 gives the rates of convergence for the variance parts of the estimators $\hat{\eta}$ and $\hat{\eta}'$.

Before proceeding further, we introduce some useful notations. For two set of random variables Z_n and Z'_n and a constant $r > 0$, we write $Z_n = O_r(Z'_n)$ if $\mathrm{E}(|Z_n/Z'_n|^r) = O(1)$ as $n \to \infty$. The notation o_r is defined in a similar way. Recall

$$U_i(\beta) = \int_{\mathcal{I}} [X_i(t) - \mu(t)]\beta(t)\, dt\,,\ \hat{U}_i(\boldsymbol{b}_p) = \int_{\mathcal{I}} [X_i(t) - \hat{\mu}(t)][\boldsymbol{b}_p^T \hat{\boldsymbol{\phi}}_p(t)]\, dt\,,$$

and

$$s_l(u \mid \boldsymbol{b}_p,\, h) = \frac{1}{n}\sum_{i=1}^{n}(\hat{U}_i(\boldsymbol{b}_p) - u)^l K_h\Big[\hat{U}_i(\boldsymbol{b}_p) - u\Big],\ l = 0,\, 1, \ldots.$$

Given a new observation X, the notations $U(\beta)$ and $\hat{U}(\boldsymbol{b}_p)$ are defined in a similar way. We also write

$$s_l(u \mid \beta_0,\, h) = \frac{1}{n}\sum_{i=1}^{n}(U_i(\beta_0) - u)^l K_h[U_i(\beta_0) - u],\ l = 0,\, 1, \ldots,$$

and

$$\mathcal{U}(\beta) = \left\{ \int_{\mathcal{I}} [X(t) - \mu(t)]\beta(t)\, dt \,\middle|\, X \in \mathcal{A}_x \right\}.$$

Recall that $f_{\beta_0}(\cdot)$ is the probability density function of $U(\beta_0)$. When the the argument $u \notin \mathcal{U}(\beta_0)$, the value for $\eta(u)$ (or $\eta'(u)$) should be interpreted as that of an extended version of η (or η') such that condition (A6)(i) holds on the whole real line.

Lemma 3. *Suppose that conditions (A1)–(A5) hold except for condition (A4)(iii). Then we have, for any integers $l \geq 0$ and $r \geq 1$, there exists a constant $c > 0$ such that*

$$\sup_{(u,\boldsymbol{b}_p)\in(\mathbb{R}\times\mathcal{B}_p)} \frac{|s_l(u \mid \boldsymbol{b}_p, h) - h^l \nu_l f_{\beta_0}(u)|}{\max\{f_{\beta_0}(u), d_n\} h^l} = O_r(n^{-c}). \tag{S5.1}$$

Given a sequence $\{v_n\}_{n=1}^{\infty}$, define the sequence of sets $\mathcal{A}_x(v_n) = \{X \in \mathcal{A}_x : \|X - \mu\|_2 \leq v_n\}$. If $v_n = O(h^* n^{(\alpha_1 - 1/2)/(\alpha_0 + 2\alpha_1)})$, then we also have, for any integers $l \geq 0$ and $r \geq 1$, there exists a constant $c > 0$ such that

$$\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \frac{|s_l(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p, h) - h^l \nu_l f_{\beta_0}(U(\beta_0))|}{\max\{f_{\beta_0}(U(\beta_0)), d_n\} h^l} = O_r(n^{-c}). \tag{S5.2}$$

Proof. We first prove (S5.1). As we assumed $n^{1-c_2} h d_n \to \infty$ in condition (A4)(ii), equation (S5.1) follows if we can prove that, for any integers $l \geq 0$ and $r \geq 1$, there exists a constant $c > 0$ not depending on n such that

$$\sup_{(u,\boldsymbol{b}_p)\in(\mathbb{R}\times\mathcal{B}_p)} \frac{|s_l(u \mid \boldsymbol{b}_p, h) - s_l(u \mid \beta_0, h)|}{\max\{f_{\beta_0}(u), d_n\} h^l} = O_r(n^{-c}), \tag{S5.3}$$

$$\sup_{u\in\mathbb{R}} \frac{|s_l(u \mid \beta_0, h) - \mathrm{E}\, s_l(u \mid \beta_0, h)|}{(\max\{f_{\beta_0}(u), d_n\})^{1/2} h^l} = O_r\left(\sqrt{\frac{\log n}{nh}}\right), \tag{S5.4}$$

and

$$\sup_{u\in\mathbb{R}} \frac{|\mathrm{E}\, s_l(u \mid \beta_0, h) - h^l \nu_l f_{\beta_0}(u)|}{\max\{f_{\beta_0}(u), d_n\} h^l} = O_r(n^{-c}). \tag{S5.5}$$

Note that (S5.5) is a direct consequence of condition (A5)(iii). So we only need to prove (S5.3) and (S5.4).

We begin the proof of (S5.4) by truncation. Define

$$\zeta_i(u \mid \beta_0, l) = \left(\frac{U_i(\beta_0) - u}{h}\right)^l K_h(U_i(\beta_0) - u), \, l = 0, 1, \ldots, \tag{S5.6}$$

and

$$\tilde{\zeta}_i(u \mid \beta_0, l) = \zeta_i(u \mid \beta_0, l) \, \mathrm{I}[|U_i(\beta_0)| \leq a_n], \tag{S5.7}$$

where $\{a_n\}$ is a positive constant sequence to be determined. By using Markov's inequality, we can obtain that, for any integer $s \geq 1$,

$$\begin{aligned} & \mathrm{E}\left\{\sup_{u\in\mathbb{R}} |\zeta_i(u \mid \beta_0, l) - \tilde{\zeta}_i(u \mid \beta_0, l)|\right\}^r \\ \leq \; & \mathrm{E}\{\mathrm{I}[|U_i(\beta_0)| > a_n]\} \times \sup_{u\in\mathbb{R}} |\zeta_i(u \mid \beta_0, l)|^r \\ \leq \; & a_n^{-s} \, \mathrm{E}\, |U_i(\beta_0)|^s \times h^{-r} \\ \leq \; & c a_n^{-s} h^{-r}, \end{aligned} \tag{S5.8}$$

where c is a constant only depending on s. Write

$$\tilde{s}_l(u \mid \beta_0, h) = h^l n^{-1} \sum_{i=1}^{n} \tilde{\zeta}_i(u \mid \beta_0, l). \tag{S5.9}$$

Then we have

$$\begin{aligned} & \mathrm{E}\left\{\sup_{u\in\mathbb{R}} h^{-l} |s_l(u \mid \beta_0, h) - \tilde{s}_l(u \mid \beta_0, h)|\right\}^r \\ = \; & \frac{1}{n^r} \, \mathrm{E}\left\{\sup_{u\in\mathbb{R}} \sum_{i=1}^{n} |\zeta_i(u \mid \beta_0, l) - \tilde{\zeta}_i(u \mid \beta_0, l)|\right\}^r \\ \leq \; & \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left\{\sup_{u\in\mathbb{R}} |\zeta_i(u \mid \beta_0, l) - \tilde{\zeta}_i(u \mid \beta_0, l)|^r\right\} \\ = \; & O(a_n^{-s} h^{-r}). \end{aligned} \tag{S5.10}$$

Therefore, by taking

$$a_n = \sqrt{\frac{n}{hd_n}}, \tag{S5.11}$$

and $s = r + 1$, it follows that

$$\sup_{u\in\mathbb{R}} \frac{|s_l(u \mid \beta_0,\, h) - \tilde{s}_l(u \mid \beta_0,\, h)|}{(\max\{f_{\beta_0}(u),\, d_n\})^{1/2}h^l} = o_r\left(\sqrt{\frac{\log n}{nh}}\right). \tag{S5.12}$$

Write

$$\Theta_0 = [-a_n - h,\, a_n + h] \cap \mathcal{U}(\beta_0). \tag{S5.13}$$

Then we have $\tilde{s}_l(u \mid \beta_0,\, h) = 0$ on $u \in \mathbb{R} \setminus \Theta_0$. Therefore, in view of (S5.12), to prove (S5.4), it suffices to prove

$$\sup_{u\in\Theta_0} \frac{|s_l(u \mid \beta_0,\, h) - \mathrm{E}\, s_l(u \mid \beta_0,\, h)|}{(\max\{f_{\beta_0}(u),\, d_n\})^{1/2}h^l} = O_r\left(\sqrt{\frac{\log n}{nh}}\right). \tag{S5.14}$$

Given some $\delta > 0$, define

$$\Theta_1 = \{u \mid f_{\beta_0}(u) \geq (1+\delta)d_n\}, \text{ and } \Theta_2 = \Theta_0 \setminus \Theta_1. \tag{S5.15}$$

We shall prove (S5.14) by showing that

$$\sup_{u\in\Theta} \frac{|s_l(u \mid \beta_0,\, h) - \mathrm{E}\, s_l(u \mid \beta_0,\, h)|}{(\max\{f_{\beta_0}(u),\, d_n\})^{1/2}h^l} = O_r\left(\sqrt{\frac{\log n}{nh}}\right),\ \Theta = \Theta_1 \text{ and } \Theta_2. \tag{S5.16}$$

We first prove (S5.16) with $\Theta = \Theta_1$. By using condition (A5)(iii), there is a constant c not depending on k such that

$$\sup_{u\in\Theta_1} \frac{h\, \mathrm{E}[\zeta_i^2(u \mid \beta_0,\, l)]}{f_{\beta_0}(u)} \leq c. \tag{S5.17}$$

From this, equation (S5.16) holds with $\Theta = \Theta_1$ if we can prove

$$\sup_{u\in\Theta_1}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\zeta_i(u\mid\beta_0,\, l)-\mathrm{E}\,\zeta_i(u\mid\beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u\mid\beta_0,\, l)]^{1/2}}\right| = O_r\left(\sqrt{\frac{\log n}{n}}\right). \tag{S5.18}$$

The main idea is to cover the set Θ_1 with intervals $[u_k - r_n,\, u_k + r_n]$, $k = 1, 2, \ldots, \mathcal{N}_n$ centered at $u_k \in \Theta_1$ with lengths $2r_n$. From condition (A5)(ii), we have

$$\sup_{\{u: f_{\beta_0}(u)\geq(1+\delta)d_n\}} |u| = O(\log n). \tag{S5.19}$$

Therefore, the cover number $\mathcal{N}_n = O(r_n^{-1}\log n)$. Observe that

$$\begin{aligned}
&\sup_{u\in\Theta_1}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\zeta_i(u\mid\beta_0,\, l)-\mathrm{E}\,\zeta_i(u\mid\beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u\mid\beta_0,\, l)]^{1/2}}\right| \\
&\leq \max_{1\leq k\leq\mathcal{N}_n}\sup_{u\in[u_k-r_n,\, u_k+r_n]\cap\Theta_1}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\zeta_i(u\mid\beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u\mid\beta_0,\, l)]^{1/2}}-\frac{\zeta_i(u_k\mid\beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u_k\mid\beta_0,\, l)]^{1/2}}\right]\right| \\
&\quad + \max_{1\leq k\leq\mathcal{N}_n}\sup_{u\in[u_k-r_n,\, u_k+r_n]\cap\Theta_1}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathrm{E}\,\zeta_i(u\mid\beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u\mid\beta_0,\, l)]^{1/2}}-\frac{\mathrm{E}\,\zeta_i(u_k\mid\beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u_k\mid\beta_0,\, l)]^{1/2}}\right]\right| \\
&\quad + \max_{1\leq k\leq\mathcal{N}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\zeta_i(u_k\mid\beta_0,\, l)-\mathrm{E}\,\zeta_i(u_k\mid\beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u_k\mid\beta_0,\, l)]^{1/2}}\right| \\
&=: S_3 + S_4 + S_5.
\end{aligned} \tag{S5.20}$$

Since K is defined on a compact support and satisfies the Lipschitz condition of order 1, we have

$$\max_{1\leq k\leq\mathcal{N}_n}\sup_{u\in[u_k-r_n,\, u_k+r_n]\cap\Theta_1}|\zeta_i(u\mid\beta_0,\, l)-\zeta_i(u_k\mid\beta_0,\, l)| = O\left(\frac{r_n}{h^2}\right), \tag{S5.21}$$

and

$$\max_{1\le k\le \mathcal{N}_n} \sup_{u\in[u_k-r_n,\, u_k+r_n]\cap\Theta_1} \left|\zeta_i^2(u \mid \beta_0,\, l) - \zeta_i^2(u_k \mid \beta_0,\, l)\right| = O\Big(\frac{r_n}{h^3}\Big). \quad \text{(S5.22)}$$

Note that

$$\begin{aligned}
&\frac{\zeta_i(u \mid \beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u \mid \beta_0,\, l)]^{1/2}} - \frac{\zeta_i(u_k \mid \beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u_k \mid \beta_0,\, l)]^{1/2}} \\
&= \frac{[\mathrm{E}\,\zeta_i^2(u_k \mid \beta_0,\, l) - \mathrm{E}\,\zeta_i^2(u \mid \beta_0,\, l)]\zeta_i(u \mid \beta_0,\, l)}{[\mathrm{E}\,\zeta_i^2(u \mid \beta_0,\, l)]^{1/2}[\mathrm{E}\,\zeta_i^2(u_k \mid \beta_0,\, l)]^{1/2}\{[\mathrm{E}\,\zeta_i^2(u \mid \beta_0,\, l)]^{1/2} + [\mathrm{E}\,\zeta_i^2(u_k \mid \beta_0,\, l)]^{1/2}\}} \\
&\qquad + \frac{1}{[\mathrm{E}\,\zeta_i^2(u_k \mid \beta_0,\, l)]^{1/2}}[\zeta_i(u \mid \beta_0,\, l) - \zeta_i(u_k \mid \beta_0,\, l)].
\end{aligned} \quad \text{(S5.23)}$$

So if we take

$$r_n = d_n^{3/2} h^{5/2} \varepsilon_n, \quad \text{(S5.24)}$$

where $\varepsilon_n = \sqrt{(\log n)/n}$, then we have

$$S_3 = O\left(\sqrt{\frac{\log n}{n}}\right), \text{ and } S_4 = O\left(\sqrt{\frac{\log n}{n}}\right). \quad \text{(S5.25)}$$

It remains to bound S_5. By using condition (A5)(iii), there is a constant c not depending on k such that

$$\frac{|\zeta_i(u_k \mid \beta_0,\, l) - \mathrm{E}\,\zeta_i(u_k \mid \beta_0,\, l)|}{[\mathrm{E}\,\zeta_i^2(u_k \mid \beta_0,\, l)]^{1/2}} \le \frac{c}{h^{1/2} d_n^{1/2}}. \quad \text{(S5.26)}$$

Obviously,

$$\mathrm{Var}[\zeta_i(u_k \mid \beta_0,\, l)] \le \mathrm{E}\,\zeta_i^2(u_k \mid \beta_0,\, l). \quad \text{(S5.27)}$$

Since $\mathcal{N}_n$ grows at a rate of at most n^c for some $c > 0$ not depending on n, by applying the Bernstein inequality for independent variables (see, for example, Chapter 2 of Wainwright (2019)), there exists a sufficiently large C such that

$$\begin{aligned}
&\mathrm{P}[S_5 \geq C\varepsilon_n] \\
&\leq 2\mathcal{N}_n \exp\left\{\frac{-C^2 n\varepsilon_n^2}{2+(2/3)(c/h^{1/2}d_n^{1/2})C\varepsilon_n}\right\} \\
&\leq 2\mathcal{N}_n \exp\left\{-C^2(\log n)\left(2+\frac{2Cc}{3}\sqrt{\frac{\log n}{nhd_n}}\right)^{-1}\right\} \\
&\leq 2\mathcal{N}_n n^{-C/3} = o(1),
\end{aligned} \tag{S5.28}$$

which implies that $S_5 = O_p(\varepsilon_n)$. In addition, from (S5.28), the sequence $|S_5/\varepsilon_n|^r$, $n = 1, 2, \ldots$ is uniformly integrable for any integer $r \geq 1$. Therefore,

$$S_5 = O_r\left(\sqrt{\frac{p\log n}{n}}\right) \tag{S5.29}$$

holds for any integer $r \geq 1$. This, together with (S5.20) and (S5.25), proves (S5.18) with $\Theta = \Theta_1$.

Recall that $f_{\beta_0}(u) < (1+\delta)d_n$ on Θ_2. Therefore, to prove (S5.16) with $\Theta = \Theta_2$, it suffices to show

$$\sup_{u\in\Theta_2}\left|\frac{1}{n}\sum_{i=1}^{n} d_n^{-1/2}[\zeta_i(u \mid \beta_0, l) - \mathrm{E}\,\zeta_i(u \mid \beta_0, l)]\right| = O_r\left(\sqrt{\frac{\log n}{nh}}\right). \tag{S5.30}$$

This part of the proof is quite similar to that with $\Theta = \Theta_1$ and so is omitted.

In summary, we have proved (S5.16). From this, (S5.14) and therefore (S5.4) follow.

We now turn to (S5.3). Define

$$\zeta_i(u \mid \boldsymbol{b}_p,\, l) = \left(\frac{\hat{U}_i(\boldsymbol{b}_p) - u}{h}\right)^l K_h(\hat{U}_i(\boldsymbol{b}_p) - u),\; l = 0,\, 1, \ldots. \tag{S5.31}$$

Since K is defined on a compact support and satisfies the Lipschitz condition of order 1, we can obtain

$$\begin{aligned}
&|\zeta_i(u \mid \boldsymbol{b}_p,\, l) - \zeta_i(\beta_0 \mid \boldsymbol{b}_p,\, l)| \\
\le\, &\frac{c}{h}\{\mathrm{I}[|\hat{U}_i(\boldsymbol{b}_p) - u| < h] + \mathrm{I}[|U_i(\beta_0) - u| < h]\} \times \max\{h^{-1}|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)|,\, 2\} \\
\le\, &\frac{c}{h}\{\mathrm{I}[|U_i(\beta_0) - u| < 2h] + \mathrm{I}[|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)| > h]\} \times \max\{h^{-1}|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)|,\, 2\},
\end{aligned} \tag{S5.32}$$

where $c > 0$ is a constant not depending on n. From this, we have

$$\begin{aligned}
&\frac{|s_l(u \mid \boldsymbol{b}_p,\, h) - s_l(u \mid \beta_0,\, h)|}{\max\{f_{\beta_0}(u),\, d_n\}h^l} \\
=\, &\frac{1}{n \max\{f_{\beta_0}(u),\, d_n\}} \sum_{i=1}^{n} |\zeta_i(u \mid \boldsymbol{b}_p,\, l) - \zeta_i(\beta_0 \mid \boldsymbol{b}_p,\, l)| \\
\le\, &\frac{c}{nh \max\{f_{\beta_0}(u),\, d_n\}} \sum_{i=1}^{n} \mathrm{I}[|U_i(\beta_0) - u| < 2h] \max\{h^{-1}|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)|,\, 2\} \\
&\quad + \frac{c}{nhd_n} \sum_{i=1}^{n} \mathrm{I}[|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)| > h] \\
=:\, &S_6 + S_7.
\end{aligned} \tag{S5.33}$$

We first consider the asymptotic properties of $\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)$. As we assumed that all moments of the FPC scores are finite, it follows that

$$\mathrm{E}[\hat{U}_i(\tilde{\boldsymbol{b}}_p) - U_i(\beta_0)]^{2r} = O[(p/n)^r] \tag{S5.34}$$

for any integer $r \geq 1$ from (3.2) in Lemma 1. By using the Cauchy–Schwarz inequality, the Minkowski inequality and (S1.4) in Lemma 2, we also have, uniformly over $\boldsymbol{b} \in \mathcal{B}_p$, for any integer $r \geq 1$,

$$\begin{aligned} \mathrm{E}\left\{\sup_{\boldsymbol{b}\in\mathcal{B}_p}[\hat{U}_i(\tilde{\boldsymbol{b}}_p) - \hat{U}_i(\boldsymbol{b}_p)]^{2r}\right\} &= \mathrm{E}\left\{\sup_{\boldsymbol{b}\in\mathcal{B}_p}\sum_{j=1}^{p}[\lambda_j^{-1/2}\hat{x}_{ij}\lambda_j^{1/2}(b_j - \tilde{b}_j)]\right\}^{2r} \\ &\leq \mathrm{E}\left[\sum_{j=1}^{p}\lambda_j^{-1}\hat{x}_{ij}^2\right]^r \times \left[\sup_{\boldsymbol{b}\in\mathcal{B}_p}\sum_{j=1}^{p}\lambda_j(b_j - \tilde{b}_j)^2\right]^r \\ &\leq \left\{\sum_{j=1}^{p}[\mathrm{E}(\lambda_j^{-1}\hat{x}_{ij}^2)^r]^{1/r}\right\}^r \times \left(\frac{p}{n}\right)^r = O(p^{2r}/n^r). \end{aligned} \tag{S5.35}$$

Combining the above two results yields

$$\mathrm{E}\left\{\sup_{\boldsymbol{b}\in\mathcal{B}_p}|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)|^r\right\} = O(p^r/n^{r/2}) \tag{S5.36}$$

for any integers $r \geq 1$.

Note that $p/(n^{1/2}h) = o(n^{-c_1})$ by condition (A4)(i). By using the Minkowski inequality and Markov's inequality, we can obtain

$$\begin{aligned} &\mathrm{E}\left\{\sup_{\boldsymbol{b}\in\mathcal{B}_p}\sum_{i=1}^{n}\mathrm{I}[|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)| > h]\right\}^r \\ &\leq \left\{\sum_{i=1}^{n}\left[\mathrm{E}\,\mathrm{I}\left[\sup_{\boldsymbol{b}\in\mathcal{B}_p}|\hat{U}_i(\boldsymbol{b}_p) - U_i(\beta_0)| > h\right]\right]^{1/r}\right\}^r = O\left[n\left(\frac{p}{n^{1/2}h}\right)^{s/r}\right]^r = o(n^{-c_1 r}) \end{aligned} \tag{S5.37}$$

by taking s large enough. From this, by using $(nhd_n)^{-1} = o(n^{-c_2})$ in condition (A4)(ii), it follows that, for any integer $r \geq 1$,

$$\mathrm{E}\left[\sup_{\boldsymbol{b}\in\mathcal{B}_p} S_7\right]^r = o(n^{-cr}) \tag{S5.38}$$

for some $c > 0$.

It remains to bound S_6. Let $\tilde{K}$ be a kernel function satisfying condition (A3) such that $\tilde{K}(u/2) > 1$ for $u \in [-1, 1]$. By using (S5.4) and (S5.5), we can obtain

$$\begin{aligned}\hat{m}(u) :=& \sup_{u\in\mathbb{R}}\left\{\frac{1}{nh\max\{f_{\beta_0}(u),\, d_n\}}\sum_{i=1}^{n}\mathrm{I}[|U_i(\beta_0)-u| < 2h]\right\}\\ \leq& \sup_{u\in\mathbb{R}}\left\{\frac{1}{nh\max\{f_{\beta_0}(u),\, d_n\}}\sum_{i=1}^{n}\tilde{K}\left[\frac{|U_i(\beta_0)-u|}{4h}\right]\right\} = O_r(1)\end{aligned} \tag{S5.39}$$

for any integer $r \geq 1$. By using (S5.36) and condition (A4)(i), we have

$$\begin{aligned}&\mathrm{E}\left[\sup_{(u,\,\boldsymbol{b}_p)\in(\mathbb{R}\times\mathcal{B}_p)}|S_6|\right]^r\\ &\leq c^r\,\mathrm{E}\left\{\hat{m}(u)\left[\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sup_{1\leq i\leq n}\max\{h^{-1}|\hat{U}_i(\boldsymbol{b}_p)-U_i(\beta_0)|,\,2\}\right]\right\}^r\\ &\leq c^r\,\mathrm{E}\left\{\hat{m}(u)\times n^{-c_1/2}\right\}^r\\ &\qquad + (2c)^r\,\mathrm{E}\left\{[\hat{m}(u)]^r\times\mathrm{I}\left[\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sup_{1\leq i\leq n}|\hat{U}_i(\boldsymbol{b}_p)-U_i(\beta_0)|\geq hn^{-c_1/2}\right]\right\}\\ &= O(n^{-c_1 r/2}) + O(1)\times\left\{n\,\mathrm{P}\left[\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}|\hat{U}_1(\boldsymbol{b}_p)-U_1(\beta_0)|\geq hn^{-c_1/2}\right]\right\}^{1/2}\\ &= O(n^{-c_1 r/2}) + O\left[n^{1/2}\Big(\frac{p}{n^{1/2-c_1/2}h}\Big)^{s/2}\right],\end{aligned} \tag{S5.40}$$

for any integer $s \geq 1$, where the last equality results from Markov's inequality. By taking s large enough, we have, for any integer $r \geq 1$,

$$\mathrm{E}\left[\sup_{(u,\,\boldsymbol{b}_p)\in(\mathbb{R}\times\mathcal{B}_p)} |S_6|\right]^r = O(n^{-c_1 r/2}). \tag{S5.41}$$

Now (S5.3) follows by plugging this and (S5.38) into (S5.33), which completes the proof of (S5.1).

It remains to prove (S5.2). In view of (S5.1), equation (S5.2) follows if we can prove, for any integer $r \geq 1$,

$$\sup_{(X,\,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \left|\frac{\max\{f_{\beta_0}(\hat{U}(\boldsymbol{b}_p)),\, d_n\}}{\max\{f_{\beta_0}(U(\beta_0)),\, d_n\}} - 1\right| = o_r(1). \tag{S5.42}$$

By using condition (A5)(iii) and (S1.25), it is easy to verify

$$\sup_{(X,\,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \left|\frac{\max\{f_{\beta_0}(\hat{U}(\boldsymbol{b}_p)),\, d_n\}}{\max\{f_{\beta_0}(U(\beta_0)),\, d_n\}} - 1\right| \mathrm{I}[\mathcal{F}_p] = o(1). \tag{S5.43}$$

Now (S5.42) follows from $\mathrm{P}(\mathcal{F}_p) = 1 - O(n^{-c})$ for any $c > 0$, which completes the proof of (S5.2). □

Lemma 4. *Under the assumptions in Lemma 3, we have, for any integers* $r,\, s \geq 1$,

$$\sup_{(X,\,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \Bigg[\sum_{k=1}^{n} |W_{nk}(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p)|^r \times [\max\{f_{\beta_0}(U(\beta_0)),\, d_n\}]^{r-1} \times \left(\frac{g_n(u \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)}\right)^r\Bigg] = O_s\big(n^{-r+1}h^{-r+1}\big), \tag{S5.44}$$

and

$$\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \Bigg[\sum_{k=1}^{n} |\tilde{W}_{nk}(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p)|^r \times [\max\{f_{\beta_0}(U(\beta_0)),\, d_n\}]^{r-1} \times \left(\frac{g_n(u \mid \boldsymbol{b}_p,\, h^*)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h^*)}\right)^r\Bigg] = O_s\big(n^{-r+1}(h^*)^{-2r+1}\big). \tag{S5.45}$$

Proof. By using Lemma 3, this lemma can be proved by direct calculations. The manipulation is not particularly difficult. Part of this proof is similar to that of Lemma 5, so we shall not reproduce here. □

Lemma 5. *Under the assumptions in Lemma 4 and condition (A6)(i), we have, for any $\delta > 0$ and integer $r \geq 1$,*

$$\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \left|\eta(\hat{U}(\boldsymbol{b}_p)) - \frac{g_n(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h)} \sum_{k=1}^{n} W_{nk}[\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p]\eta(\hat{U}_k(\boldsymbol{b}_p))\right| \times \mathrm{I}[f_{\beta_0}(U(\beta_0)) \geq (1+\delta)d_n] = O_r\big(h^2\big), \tag{S5.46}$$

$$\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \left|\eta'(\hat{U}(\boldsymbol{b}_p)) - \frac{g_n(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h^*)}{g_{d_n}(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h^*)} \sum_{k=1}^{n} \tilde{W}_{nk}[\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p]\eta(\hat{U}_k(\boldsymbol{b}_p))\right| \times \mathrm{I}[f_{\beta_0}(U(\beta_0)) \geq (1+\delta)d_n] = O_r(h^*), \tag{S5.47}$$

and

$$\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)}\left|\eta'(\hat{U}(\boldsymbol{b}_p))-\frac{g_n(\hat{U}(\boldsymbol{b}_p)\mid \boldsymbol{b}_p,\,h^*)}{g_{d_n}(\hat{U}(\boldsymbol{b}_p)\mid \boldsymbol{b}_p,\,h^*)}\sum_{k=1}^n\tilde{W}_{nk}[\hat{U}(\boldsymbol{b}_p)\mid \boldsymbol{b}_p]\eta(\hat{U}_k(\boldsymbol{b}_p))\right|$$
$$\times\,\mathrm{I}[f_{\beta_0}(U(\beta_0))<(1+\delta)d_n]=O_r(1). \tag{S5.48}$$

Proof. We first prove (S5.46). To simplify notations, write $u=\hat{U}(\boldsymbol{b}_p)$ and $u_k=\hat{U}_k(\boldsymbol{b}_p)$. Define

$$H(u,\,u_k)=\eta(u)-\eta(u_k)-\eta'(u)(u-u_k). \tag{S5.49}$$

Note that

$$\sum_{k=1}^n W_{nk}(u\mid \boldsymbol{b}_p)=1 \text{ and } \sum_{k=1}^n W_{nk}(u\mid \boldsymbol{b}_p)(u-u_k)=0. \tag{S5.50}$$

By a simple calculation, we have,

$$\begin{aligned}
&\eta(u)-\frac{g_n(u\mid \boldsymbol{b}_p,\,h)}{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)}\sum_{k=1}^n W_{nk}(u\mid \boldsymbol{b}_p)\eta(u_k)\\
&=\frac{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)-g_n(u\mid \boldsymbol{b}_p,\,h)}{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)}\eta(u)+\frac{g_n(u\mid \boldsymbol{b}_p,\,h)}{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)}\eta(u)\\
&\qquad-\frac{(nh^2)^{-1}\sum_{k=1}^n\xi_k(u\mid \boldsymbol{b}_p)\eta(u_k)}{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)}-\frac{\eta'(u)(nh^2)^{-1}\sum_{k=1}^n\xi_k(u\mid \boldsymbol{b}_p)(u-u_k)}{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)}\\
&=\frac{(nh^2)^{-1}\sum_{k=1}^n\xi_k(u\mid \boldsymbol{b}_p)H(u,\,u_k)}{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)}+\frac{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)-g_n(u\mid \boldsymbol{b}_p,\,h)}{g_{d_n}(u\mid \boldsymbol{b}_p,\,h)}\eta(u)\\
&=:S_8+S_9.
\end{aligned} \tag{S5.51}$$

We first consider S_9. Noting that η has a bounded derivative, by using

(3.6), we have,

$$\mathrm{E}\left\{\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}|S_9\,\mathrm{I}[f_{\beta_0}(u)\geq(1+\delta)d_n]|^r\right\}$$
$$=O(\log^r n)\,\mathrm{E}\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\left\{\mathrm{I}[g_n(u\mid\boldsymbol{b}_p,\,h)<\nu_2 d_n^2]\,\mathrm{I}[f_{\beta_0}(u)\geq(1+\delta)d_n]\right\}. \tag{S5.52}$$

By using Lemma 3, we can obtain that, there is a $c>0$ such that for any integer $s\geq 2$,

$$\sup_{(X,\,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)}\left\{\frac{|g_n(u\mid\boldsymbol{b}_p,\,h)-\nu_2 f_{\beta_0}^2(u)|}{\nu_2 f_{\beta_0}^2(u)}\times\mathrm{I}[f_{\beta_0}(u)\geq(1+\delta)d_n]\right\}=O_s(n^{-c}). \tag{S5.53}$$

From this, we can obtain

$$\mathrm{E}\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\left\{\mathrm{I}[g_n(u\mid\boldsymbol{b}_p,\,h)<\nu_2 d_n^2]\,\mathrm{I}[f_{\beta_0}(u)\geq(1+\delta)d_n]\right\}=O(n^{-C}). \tag{S5.54}$$

for any $C>0$ by taking s large enough. Therefore, it follows that, for any integer $r\geq 1$,

$$\mathrm{E}\left\{\sup_{(X,\,\boldsymbol{b}_p)\in(\mathcal{A}_x\times\mathcal{B}_p)}\{|S_9|^r\times\mathrm{I}[f_{\beta_0}(u)\geq(1+\delta)d_n]\}\right\}=o(h^{2r}). \tag{S5.55}$$

We now turn to S_8. An elementary calculation yields

$$\sum_{k=1}^n \xi_k(u\mid\beta_0)H(u_k,\,u)$$
$$=s_2(u\mid\beta_0,\,h)\sum_{k=1}^n H(u_k,\,u)K_h(u_k-u)-s_1(u\mid\beta_0,\,h)\sum_{k=1}^n H(u_k,\,u)(u_k-u)K_h(u_k-u). \tag{S5.56}$$

From condition (A6)(i), we have

$$|H(u_k,\,u)|\leq(u_k-u)^2\sup_{u\in\hat{\mathcal{U}}(\boldsymbol{b}_p)}|\eta''(u)|. \tag{S5.57}$$

From this, by an argument similar to the proof of (S5.2) in Lemma 3, we have, for $l = 0$ and 1 and any integer $s \geq 1$,

$$\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \frac{1}{nh^{2+l}\max\{f_{\beta_0}(u),\, d_n\}} \sum_{k=1}^{n} |H(u_k,\, u)|(u_k - u)^l K_h(u_k - u) = O_s(1). \tag{S5.58}$$

Combining this with (S5.1) of Lemma 3, we have

$$\mathrm{E}\left\{\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} |S_8|^r\right\} = O(h^{2r}). \tag{S5.59}$$

It is easy to show, for any positive integer r,

$$\begin{aligned} &\mathrm{E}\left[\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x\times\mathcal{B}_p)} \{|S_8+S_9|^r \times \mathrm{I}[f_{\beta_0}(u) \geq (1+\delta)d_n]\}\right] \\ &\leq 2^{r-1}\,\mathrm{E}\left[\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x\times\mathcal{B}_p)} \{|S_8|^r \times \mathrm{I}[f_{\beta_0}(u) \geq (1+\delta)d_n]\}\right] \\ &\quad + 2^{r-1}\,\mathrm{E}\left[\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x\times\mathcal{B}_p)} \{|S_9|^r \times \mathrm{I}[f_{\beta_0}(u) \geq (1+\delta)d_n]\}\right]. \end{aligned} \tag{S5.60}$$

This, together with (S5.51), (S5.55) and (S5.59), completes the proof of (S5.46).

We now consider (S5.47). Again, write $u = \hat{U}(\boldsymbol{b}_p)$ and $u_k = \hat{U}_k(\boldsymbol{b}_p)$. Note that

$$\sum_{k=1}^{n} \tilde{\xi}_{nk}(u \mid \boldsymbol{b}_p) = 0 \text{ and } \sum_{k=1}^{n} \tilde{\xi}_{nk}(u \mid \boldsymbol{b}_p)(u_k - u) = \sum_{k=1}^{n} \xi_{nk}(u \mid \boldsymbol{b}_p). \tag{S5.61}$$

By direct calculations, we have

$$\begin{aligned}
&\eta'(u) - \frac{g_n(u \mid \boldsymbol{b}_p,\, h^*)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h^*)} \sum_{k=1}^{n} \tilde{W}_{nk}(u \mid \boldsymbol{b}_p)\eta(u_k) \\
&= \frac{g_{d_n}(u \mid \boldsymbol{b}_p,\, h) - g_n(u \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)}\eta'(u) + \frac{g_n(u \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)}\eta'(u) \times \frac{\sum_{k=1}^{n} \tilde{\xi}_{nk}(u \mid \boldsymbol{b}_p)(u_k - u)}{\sum_{k=1}^{n} \xi_{nk}(u \mid \boldsymbol{b}_p)} \\
&\quad - \frac{(nh^2)^{-1}\sum_{k=1}^{n} \tilde{\xi}_k(u \mid \boldsymbol{b}_p)\eta(u_k)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)} + \frac{\eta(u)(nh^2)^{-1}\sum_{k=1}^{n} \tilde{\xi}_k(u \mid \boldsymbol{b}_p)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)} \\
&= \frac{(nh^2)^{-1}\sum_{k=1}^{n} \tilde{\xi}_k(u \mid \beta_0)H(u_k,\, u)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)} + \frac{g_{d_n}(u \mid \boldsymbol{b}_p,\, h) - g_n(u \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)}\eta'(u).
\end{aligned} \tag{S5.62}$$

The reminder of the proof is quite similar to that of (S5.46) and is omitted.

The proof (S5.48) is simple by using (S5.62) and is also omitted. □

Lemma 6. *Under the assumptions in Lemma 5, we have,*

$$\begin{aligned}
&\sup_{\boldsymbol{b}_p \in \mathcal{B}_p} \sum_{i=1}^{n} \left[\sum_{k=1}^{n} |W_{nk}(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p)| \times \left[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big) - \eta(U_k(\beta_0)) \right] \right. \\
&\quad \left. \times \left(\frac{g_n(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h)} \right) \right]^2 = O_p(p),
\end{aligned} \tag{S5.63}$$

$$\begin{aligned}
&\sup_{\boldsymbol{b}_p \in \mathcal{B}_p} \sum_{i=1}^{n} \left[\sum_{k=1}^{n} |W_{nk}(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p)| \times \left[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big) - \eta(U_k(\beta_0)) \right] \right. \\
&\quad \left. \times \left(\frac{g_n(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h)} \right) \right]^2 = O_1(pn^c),
\end{aligned} \tag{S5.64}$$

$$\begin{aligned}
&\sup_{\boldsymbol{b}_p \in \mathcal{B}_p} \sum_{i=1}^{n} \left[\sum_{k=1}^{n} |\tilde{W}_{nk}(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p)| \times \left[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big) - \eta(U_k(\beta_0)) \right] \right. \\
&\quad \left. \times \left(\frac{g_n(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h^*)} \right) \right]^2 = O_p\left(\frac{p}{(h^*)^2} \right),
\end{aligned} \tag{S5.65}$$

and

$$\begin{aligned}\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sum_{i=1}^{n}\Bigg[\sum_{k=1}^{n}|\tilde{W}_{nk}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p)|\times\Big[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big)-\eta(U_k(\beta_0))\Big]\\ \times\left(\frac{g_n(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h^*)}\right)\Bigg]^2=O_1\bigg(\frac{pn^c}{(h^*)^2}\bigg).\end{aligned}\tag{S5.66}$$

for any $c>0$ not depending on n.

Proof. We only prove (S5.63) and (S5.64), the proof of (S5.65) and (S5.66) is similar. From (S5.36), by using Markov's inequality, we have

$$\mathrm{P}(\|X\|_2\geq n^c)\leq n^{-D}\tag{S5.67}$$

for any c, $D > 0$. Direct calculations yield

$$
\begin{aligned}
&\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sum_{i=1}^{n}\left[\sum_{k=1}^{n}|W_{nk}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p)|\times\left[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big)-\eta(U_k(\beta_0))\right]\right.\\
&\qquad\left.\times\left(\frac{g_n(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}\right)\right]^2\\
&\le\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sum_{i=1}^{n}\sum_{k=1}^{n}\left\{\frac{W_{nk}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p)}{\sqrt{K[h^{-1}(\hat{U}_i(\boldsymbol{b}_p)-\hat{U}_k(\boldsymbol{b}_p))]}}\times\left[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big)-\eta(U_k(\beta_0))\right]\right\}^2\\
&\qquad\times\left(\frac{g_n(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}\right)^2\sum_{k=1}^{n}K[h^{-1}(\hat{U}_i(\boldsymbol{b}_p)-\hat{U}_k(\boldsymbol{b}_p))]\\
&\le\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sum_{k=1}^{n}\left[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big)-\eta(U_k(\beta_0))\right]^2\sum_{i=1}^{n}\left\{\frac{W_{nk}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p)}{\sqrt{K[h^{-1}(\hat{U}_i(\boldsymbol{b}_p)-\hat{U}_k(\boldsymbol{b}_p))]}}\right\}^2\\
&\qquad\times\left(\frac{g_n(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}\right)^2\times nh\max\{f_{\beta_0}(\hat{U}_i(\boldsymbol{b}_p)),\,d_n\}\times O_r(1)\\
&=\sup_{\boldsymbol{b}_p\in\mathcal{B}_p}\sum_{k=1}^{n}\left[\eta\Big(\hat{U}_k(\boldsymbol{b}_p)\Big)-\eta(U_k(\beta_0))\right]^2\mathrm{I}[\|X_k\|_2\le n^c]\times O_{r/2}(1)+o_1(1)\\
&=O_p(p)
\end{aligned}
\tag{S5.68}
$$

where the first inequality can be obtained by using the Cauchy–Schwarz inequality, the second inequality results from (S5.1) in Lemma 3, and the third equality can be shown by an similar argument to that of (S5.44) in Lemma 4 and using (S5.67). Now (S5.63) follows by using Lemma 2.

By using Hölder's inequality, the proof of (S5.64) is straightforward, so we omit here. □

Lemma 7. *Under the assumptions in Lemma 5 and condition (A6)(ii), we have, uniformly over* $1 \le j \le p$,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\rho_j[U_i(\beta_0)] - \sum_{k=1}^{n}\frac{g_n(U_k(\beta_0)\mid \beta_0,\, h)}{g_{d_n}(U_k(\beta_0)\mid \beta_0,\, h)}W_{ni}[U_k(\beta_0)\mid \beta_0]\rho_j[U_k(\beta_0)]\right\}^2 = o_1(1), \tag{S5.69}$$

where $\rho_j(u) = \eta'(u)\eta_{2j}(u)$.

Proof. From condition (A5)(iii), we have, uniformly over $1 \le i < k \le n$, if $|U_i(\beta_0) - U_k(\beta_0)| \le h$, then

$$\frac{1}{\max\{f_{\beta_0}(U_k(\beta_0)),\, d_n\}} = \frac{1}{\max\{f_{\beta_0}(U_i(\beta_0)),\, d_n\}}(1 + o(1)). \tag{S5.70}$$

Note that $\mathrm{P}[f_{\beta_0}(u) \le (1+\delta)d_n] = O(d_n \log n)$ holds for any $\delta > 0$ not depending on n. Similar to (S5.1) of Lemma 3, we can obtain that there exists a $c > 0$ such that

$$\sup_{u\in\mathbb{R}}\frac{|s_l(u\mid \beta_0,\, h) - h^l\nu_l f_{\beta_0}(u)|}{\max\{f_{\beta_0}(u),\, d_n\}h^l} = O_r(n^{-c}), \tag{S5.71}$$

for any integer $r \ge 1$. By using condition (A6), it is not hard to verify

$$\sup_{1\le i\le n}\frac{1}{n}\sum_{i=1}^{n}\left\{\rho_j[U_i(\beta_0)] - \sum_{k=1}^{n}\frac{g_n(U_k(\beta_0)\mid \beta,\, h)}{g_{d_n}(U_k(\beta_0)\mid \beta,\, h)}W_{ni}[U_k(\beta_0)\mid \beta_0]\rho_j[U_k(\beta_0)]\right\}^2$$
$$\times\, \mathrm{I}[f_{\beta_0}(U_i(\beta_0)) \le (1+\delta)d_n] = o_1(1). \tag{S5.72}$$

Combining (S5.71) and (S5.72), we have, to prove (S5.69), it suffices to show, uniformly over $1 \le j \le p$,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\rho_j[U_i(\beta_0)] - \sum_{k=1}^{n}\frac{K_h(U_i(\beta_0) - U_k(\beta_0))}{nf_{\beta_0}(U_k(\beta_0))}\rho_j[U_k(\beta_0)]\right\}^2 \times \mathrm{I}[f_{\beta_0}(U_i(\beta_0)) > (1+\delta)d_n] = o_1(1). \tag{S5.73}$$

By using condition (A6), it follows that, for all $1 \le j \le p$ and $1 \le i < k \le n$, if $|U_k(\beta_0) - U_i(\beta_0)| \le h$, then

$$|\rho_j[U_k(\beta_0)] - \rho_j[U_i(\beta_0)]| \le ch \tag{S5.74}$$

for some $c > 0$ not depending on i, j and k. By using (S5.70), (S5.71) and (S5.74), equation (S5.73) can be proved by direct calculations which are not particularly difficult. Further details are omitted. □

Lemma 8. *Let $V_1, \ldots, V_n$ be a sequence of random variables. For a given V_i, let $\varphi_{u,\boldsymbol{b}_p}(V_i)$ be a function on $\mathcal{C}_n = \{(u, \boldsymbol{b}_p) \in \mathbb{R}^{p+1} : |u-u_0| \le n^{a_1},\ \|\boldsymbol{b}_p - \boldsymbol{b}_{p,0}\|_2 \le n^{a_2}\}$ for constants $a_1, a_2 < \infty$ not depending on n. Assume that $\varphi_{u,\boldsymbol{b}_p}(\cdot)$ satisfies*

$$\frac{1}{n}\sum_{i=1}^{n}|\varphi_{u,\boldsymbol{b}_p}(V_i) - \varphi_{u^*,\boldsymbol{b}_p^*}(V_i)| \le n^{a_3}\|(u, \boldsymbol{b}_p) - (u^*, \boldsymbol{b}_p^*)\|_2 \tag{S5.75}$$

for some constants u^*, $\boldsymbol{b}_p^*$, and $a_3 > 0$. Let $\varepsilon_n > 0$ depend only on n. If

$$\mathrm{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\varphi_{u,\boldsymbol{b}_p}(V_i)\right| > \frac{1}{2}\varepsilon_n\right\} \le \frac{1}{2} \tag{S5.76}$$

for all $(u,\,\boldsymbol{b}_p) \in \mathcal{C}_n$*, then we have*

$$\begin{aligned} &\mathrm{P}\left\{\sup_{(u,\,\boldsymbol{b}_p)\in\mathcal{C}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\varphi_{u,\,\boldsymbol{b}_p}(V_i)\right| > \frac{1}{2}\varepsilon\right\} \\ &\le (n/\varepsilon_n)^{cp}\,\mathrm{E}\left\{\sup_{(u,\,\boldsymbol{b}_p)\in\mathcal{C}_n} 2\exp\left(\frac{-n^2\varepsilon_n^2/128}{\sum_{i=1}^{n}\varphi_{u,\,\boldsymbol{b}_p}^2(V_i)}\right)\right\} \end{aligned} \tag{S5.77}$$

for some $c > 0$ *not depending on* n*.*

Proof. Lemma 8 can be proved by using a symmetrization method. See Lemma A.1 of Wang et al. (2010). □

Lemma 9. *Under the assumptions in Lemma 5 and condition (A7), we have, for any integer* $r \geq 1$*,*

$$\sup_{(u,\,\boldsymbol{b}_p)\in(\mathbb{R}\times\mathcal{B}_p)}\sum_{k=1}^{n}\frac{g_n(u\mid\boldsymbol{b}_p,\,h)}{g_{d_n}(u\mid\boldsymbol{b}_p,\,h)}\sqrt{\max\{f_{\beta_0}(u),\,d_n\}}W_{nk}[u\mid\boldsymbol{b}_p]e_k = O_r\left(\sqrt{\frac{p\log n}{nh}}\right), \tag{S5.78}$$

$$\begin{aligned} &\sup_{(X,\,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)}\sum_{k=1}^{n}\frac{g_n(\hat{U}(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}{g_{d_n}(\hat{U}(\boldsymbol{b}_p)\mid\boldsymbol{b}_p,\,h)}W_{nk}[\hat{U}(\boldsymbol{b}_p)\mid\boldsymbol{b}_p]e_k \\ &\times\sqrt{\max\{f_{\beta_0}(U(\beta_0)),\,d_n\}} = O_r\left(\sqrt{\frac{p\log n}{nh}}\right), \end{aligned} \tag{S5.79}$$

$$\sup_{(u,\,\boldsymbol{b}_p)\in(\mathbb{R}\times\mathcal{B}_p)}\sum_{k=1}^{n}\frac{g_n(u)\mid\boldsymbol{b}_p,\,h^*)}{g_{d_n}(u)\mid\boldsymbol{b}_p,\,h^*)}\sqrt{\max\{f_{\beta_0}(u),\,d_n\}}\tilde{W}_{nk}[u\mid\boldsymbol{b}_p]e_k = O_r\left(\sqrt{\frac{p\log n}{n(h^*)^3}}\right), \tag{S5.80}$$

and

$$\sup_{(X,\boldsymbol{b}_p)\in(\mathcal{A}_x(v_n)\times\mathcal{B}_p)} \sum_{k=1}^{n} \frac{g_n(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h^*)}{g_{d_n}(\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p,\, h^*)} \tilde{W}_{nk}[\hat{U}(\boldsymbol{b}_p) \mid \boldsymbol{b}_p] e_k \times \sqrt{\max\{f_{\beta_0}(U(\beta_0)),\, d_n\}} = O_r\left(\sqrt{\frac{p\log n}{n(h^*)^3}}\right). \tag{S5.81}$$

Proof. We only prove (S5.78) and (S5.79) here. The proof of (S5.80) and (S5.81) are similar.

Note that $W_{nk}[u \mid \boldsymbol{b}_p] g_n(u \mid \boldsymbol{b}_p,\, h)/g_{d_n}(u \mid \boldsymbol{b}_p,\, h) = 0$ if $u - \hat{U}_k(\boldsymbol{b}_p) > h$. As we assume the density function $f_{\beta_0}(u)$ is sub-exponential, by an argument similar to that in the proof of Lemma 3, it is not hard to verify, for any integer $r \geq 1$, there exists a $\delta > 0$ such that

$$\sup_{(u,\boldsymbol{b}_p)\in((\mathbb{R}\setminus[-n^\delta,n^\delta])\times\mathcal{B}_p)} \sum_{k=1}^{n} \frac{g_n(u \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)} \times \sqrt{\max\{f_{\beta_0}(u),\, d_n\}} W_{nk}[u \mid \boldsymbol{b}_p] e_k = O_r\left(\sqrt{\frac{p\log n}{nh}}\right). \tag{S5.82}$$

Therefore, it suffices to prove

$$\sup_{(u,\boldsymbol{b}_p)\in([-n^\delta,n^\delta]\times\mathcal{B}_p)} \sum_{k=1}^{n} \frac{g_n(u \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)} \sqrt{\max\{f_{\beta_0}(u),\, d_n\}} W_{nk}[u \mid \boldsymbol{b}_p] e_k = O_r\left(\sqrt{\frac{p\log n}{nh}}\right) \tag{S5.83}$$

for any $\delta > 0$. Let $V_i = (e_i,\, \hat{\boldsymbol{x}}_p^T)^T$ and

$$\varphi_{u,\boldsymbol{b}_p}(V_k) = n\sqrt{\frac{nh}{p\log n}} \frac{g_n(u \mid \boldsymbol{b}_p,\, h)}{g_{d_n}(u \mid \boldsymbol{b}_p,\, h)} \sqrt{\max\{f_{\beta_0}(u),\, d_n\}} W_{nk}[u \mid \boldsymbol{b}_p] e_k. \tag{S5.84}$$

The reminder of the proof of (S5.78) is straightforward by using Lemma 4 and Lemma 8.

By using (S5.43), equation (S5.79) holds on $\mathcal{F}_p$. Since $\mathrm{P}(\mathcal{F}_p) = 1 - O(n^{-c})$ for any $c > 0$, equation (S5.79) follows. □

S6 Proof of Theorem 1

We prove Theorem 1 by showing that under conditions (A1)–(A8), the estimating equations (2.13) have a root $\check{\boldsymbol{b}}_p$ such that

$$\mathrm{P}(\boldsymbol{b}_p(\check{\boldsymbol{b}}_p) \in \mathcal{B}_p(n,\, C_1)) \to 1 \tag{S6.1}$$

for some $C_1 < \infty$, where $\mathcal{B}_p(n,\, C_1)$ is defined in (3.3). Recall

$$\check{\boldsymbol{b}}_{0p} = (\check{b}_{02}, \ldots,\, \check{b}_{0p})^T = (\lambda_2^{1/2}\tilde{b}_2, \ldots, \ldots,\, \lambda_p^{1/2}\tilde{b}_p)^T, \tag{S6.2}$$

$$\mathcal{B}_{2p} = \mathcal{B}_{2p}(n,\, C_2) = \left\{ \check{\boldsymbol{b}}_p \in \mathbb{R}^{p-1} : \sum_{j=2}^{p} (\check{b}_j - \check{b}_{0j})^2 = C_2 p/n \right\}, \tag{S6.3}$$

and

$$\mathcal{B}'_{2p} = \mathcal{B}'_{2p}(n,\, C_2) = \bigcup_{0 \le C \le C_2} \mathcal{B}_{2p}(n,\, C). \tag{S6.4}$$

Our first goal is to show that provided $\mathcal{E}_p$, if $\check{\boldsymbol{b}}_p \in \mathcal{B}'_{2p}$ for some finite C_2, then the corresponding $\boldsymbol{b}_p = \boldsymbol{b}_p(\check{\boldsymbol{b}}_p)$ lies in $\mathcal{B}_p(n,\, C_1)$ for some C_1 only depending on C_2. With this, in order to prove (S6.1), it suffices to prove

$$\mathrm{P}(\check{\boldsymbol{b}}_p \in \mathcal{B}'_{2p}(n,\, C_2)) \to 1 \tag{S6.5}$$

for some $C_2 < \infty$. In rest of this section we shall drop the superscript p for simplicity in notations when there is no ambiguity.

For a given $\check{\boldsymbol{b}} \in \mathcal{B}_2$ with some $C_2 > 0$, direct calculation yields

$$\begin{aligned}
|b_1(\check{b}) - b_1(\check{b}_0)| &= \left| \left(1 - \sum_{j=2}^{p} \check{b}_j^2 \hat{\lambda}_j^{-1}\right)^{1/2} - \left(1 - \sum_{j=2}^{p} \check{b}_{0j}^2 \lambda_j^{-1}\right)^{1/2} \right| \\
&= \frac{\left|\sum_{j=2}^{p} \check{b}_j^2 \hat{\lambda}_j^{-1} - \sum_{j=2}^{p} \check{b}_{0j}^2 \lambda_j^{-1}\right|}{\left(1 - \sum_{j=2}^{p} \check{b}_j^2 \hat{\lambda}_j^{-1}\right)^{1/2} + \left(1 - \sum_{j=2}^{p} \check{b}_{0j}^2 \lambda_j^{-1}\right)^{1/2}} \\
&\le \frac{1}{b_{01}^{1/2}} \left| \sum_{j=2}^{p} \hat{\lambda}_j^{-1} (\check{b}_j^2 - \check{b}_{0j}^2) \right| + \frac{1}{b_{01}^{1/2}} \left| \sum_{j=2}^{p} \check{b}_{0j}^2 (\hat{\lambda}_j^{-1} - \lambda_j^{-1}) \right| \\
&=: S_{10}(\check{\boldsymbol{b}}) + S_{11}(\check{\boldsymbol{b}}).
\end{aligned} \tag{S6.6}$$

We first consider $S_{11}(\check{\boldsymbol{b}})$. Provided $\mathcal{E}_p$, simple calculation yields, uniformly over $1 \le j \le p$,

$$j^{-2\alpha_0} |\hat{\lambda}_j^{-1} - \lambda_j^{-1}| = O(n^{-1/2+\delta}), \tag{S6.7}$$

for any $\delta > 0$. Therefore, by taking δ small enough, we have

$$\begin{aligned}
\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} S_{11}(\check{\boldsymbol{b}}) &\le \frac{1}{b_{01}^{1/2}} \sum_{j=2}^{p-1} \check{b}_{0j}^2 (\hat{\lambda}_j^{-1} - \lambda_j^{-1}) + \frac{1}{b_{01}^{1/2}} \check{b}_{0p}^2 (\hat{\lambda}_p^{-1} - \lambda_p^{-1}) \\
&= O(n^{-1/2+\delta}) \sum_{j=2}^{p-1} j^{2\alpha_0} j^{-\alpha_0 - 2\alpha_1} + O(p^{-\alpha_0 - 2\alpha_1 + 1} n^{-1/2+\delta} p^{2\alpha_0}) = o\left(\sqrt{\frac{p}{n}}\right).
\end{aligned} \tag{S6.8}$$

For $S_{10}(\check{\boldsymbol{b}})$, we first observe that, by using (S6.7), it follows that, uniformly over $1 \le j \le p$,

$$\hat{\lambda}_j^{-1} = \lambda_j^{-1}(1 + o(1)). \tag{S6.9}$$

Now applying the Cauchy–Schwarz inequality yields

$$\begin{aligned} S_{10}(\check{\boldsymbol{b}}) &\le \frac{1}{b_{01}^{1/2}} \left[\sum_{j=2}^{p} (\check{b}_j - \check{b}_{0j})^2\right]^{1/2} \times \left[\sum_{j=2}^{p} \lambda_j^{-2} (\check{b}_j + \check{b}_{0j})^2\right]^{1/2} \times (1 + o(1)) \\ &\le c\sqrt{\frac{C_2 p}{b_{01} n}} \times \left\{\sum_{j=2}^{p} \lambda_j^{-2} [2(\check{b}_j - \check{b}_{0j})^2 + 8\check{b}_{0j}^2]\right\}^{1/2} \times (1 + o(1)) \\ &=: c\sqrt{\frac{C_2 p}{b_{01} n}} \sqrt{S_{10}'(\check{\boldsymbol{b}})} \times (1 + o(1)), \end{aligned} \tag{S6.10}$$

for some constant $c > 0$ not depending on n. Note that for $\check{\boldsymbol{b}} \in \mathcal{B}_{2p}$, $\sup_{1 \le j \le p} (\check{b}_j - \check{b}_{0j})^2 \le C_2 p/n$. Therefore, we have, for any $\check{\boldsymbol{b}} \in \mathcal{B}_2$,

$$\begin{aligned} S_{10}'(\check{\boldsymbol{b}}) &\le \frac{2C_2 p}{n} \sum_{j=2}^{p} \lambda_j^{-2} + 8 \sum_{j=2}^{p} \lambda_j^{-2} \check{b}_{0j}^2 \\ &\le \frac{2C_2 p}{n} \times c p^{2\alpha_0 + 1} + c \sum_{j=2}^{p-1} j^{\alpha_0 - 2\alpha_1} + c p^{\alpha_0 - 2\alpha_1 + 1} \end{aligned} \tag{S6.11}$$

for some $c > 0$ not depending on C_2. By condition (A1) and (A2), we have $p^{2\alpha_0+2}/n = o(1)$ and $\alpha_0 - 2\alpha_1 < -2$. So we have $S_{10}'(\check{\boldsymbol{b}}) = O(1)$, and

$$\sup_{\check{\boldsymbol{b}}_p \in \mathcal{B}_2} S_{10}(\check{\boldsymbol{b}}) = O\left(\sqrt{\frac{p}{n}}\right). \tag{S6.12}$$

Substituting this and (S6.8) into (S6.6) yields

$$\sup_{\check{\boldsymbol{b}}_p \in \mathcal{B}_2} |b_1(\check{b}) - b_1(\check{b}_0)| = O\left(\sqrt{\frac{p}{n}}\right). \tag{S6.13}$$

From this, we find that, provided $\mathcal{E}_p$, if $\check{\boldsymbol{b}} \in \mathcal{B}_2$, then $\boldsymbol{b} \in \mathcal{B}(n, C_1)$ holds for some $C_1 < \infty$ when n is large enough.

Note that provided $\mathcal{E}_p$, all the results for $\boldsymbol{b} \in \mathcal{B}$ in Section S2 remain valid for $\boldsymbol{b} = \boldsymbol{b}(\check{\boldsymbol{b}})$ with $\check{\boldsymbol{b}} \in \mathcal{B}_2'$. Before proving (S6.5), we first give some useful results. As we assumed the density function $f_{\beta_0}(\cdot)$ is sub-exponential, it is easy to show

$$\mathrm{E}\{[\max\{f_{\beta_0}(u),\, d_n\}]^{-1}\} = O(\log n). \tag{S6.14}$$

Define

$$\tilde{\boldsymbol{J}}_p = \boldsymbol{\Lambda}_p^{-1/2} \begin{pmatrix} -\lambda_1^{1/2}\tilde{b}_1^{-1}(\lambda_2^{-1}\check{b}_{02}, \ldots, \lambda_p^{-1}\check{b}_{0p}) \\ \boldsymbol{I}_{p-1} \end{pmatrix}. \tag{S6.15}$$

By direct calculations, we can obtain, for any integer $r \geq 1$,

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2'} \left|\left|\left| \hat{\boldsymbol{\Lambda}}^{1/2}\hat{\boldsymbol{J}}(\check{\boldsymbol{b}}) \right|\right|\right| = O_r(1), \tag{S6.16}$$

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2'} \left|\left|\left| \hat{\boldsymbol{\Lambda}}^{1/2}\hat{\boldsymbol{J}}(\check{\boldsymbol{b}}) - \boldsymbol{\Lambda}^{1/2}\tilde{\boldsymbol{J}} \right|\right|\right| = O_r\left(n^{(-\alpha_1+\alpha_0/2+1/2)/(\alpha_0+2\alpha_1)}\right) = o_r(n^{-c}p^{-1/2}) \tag{S6.17}$$

for some small $c > 0$, and

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2'} \sum_{i=1}^{n} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i|^2 = C_2 O_1(p). \tag{S6.18}$$

By using Lemma 2, we can obtain

$$\lambda_j|\hat{\lambda}_j^{-1/2}\hat{x}_{\cdot j} - \lambda_j^{-1/2}x_{\cdot j}| = O_r(n^{-1/2}) \tag{S6.19}$$

for any integer $r \geq 1$. By using the Cauchy–Schwarz inequality and the Minkowski inequality, we also have, for $r \geq 2$,

$$\begin{aligned} & \mathrm{E} \sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2`} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}|^{2r} \\ \leq & \left| \sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \sum_{j=2}^{p} (\check{b}_j - \check{b}_{0j})^2 \right|^r \times \mathrm{E} \left[c x_{\cdot 1}^2 + \sum_{j=2}^{p} \lambda_j^{-1} x_{\cdot j}^2 + p \times o_r(1) \right]^r \\ \leq & C_2^r \times O(p^r/n^r) \times \left[c(\mathrm{E}\, x_{\cdot 1}^{2r})^{1/2} + \sum_{j=2}^{p} \lambda_j^{-1} (\mathrm{E}\, x_{\cdot j}^{2r})^{1/r} + o(p) \right]^r \\ = & O(p^{2r}/n^r), \end{aligned} \tag{S6.20}$$

where the constant on the right side of the second inequality

$$c = \sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \left[\frac{1}{b_1} \sum_{j=2}^{p} \lambda_j^{-1} \check{b}_j (\check{b}_j - \check{b}_{j0}) \right]^2 = o(1). \tag{S6.21}$$

Therefore, by using Markov's inequality, it follows that

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2'} \sup_{1 \leq i \leq n} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}(\check{\boldsymbol{b}})^T \hat{\boldsymbol{x}}_i| = o_p(n^c p/\sqrt{n}) \tag{S6.22}$$

for any $c > 0$.

Our task now is to prove (S6.5). Define

$$R(\check{\boldsymbol{b}}) = \sum_{i=1}^{n} \left[Y_i - \hat{\eta}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) \right] \hat{\eta}'_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}) \hat{\boldsymbol{x}}_i. \tag{S6.23}$$

Separating $R(\check{\boldsymbol{b}})$, we have

$$\begin{aligned}
R(\check{\boldsymbol{b}}) =& \sum_{i=1}^{n} e_i \eta'(U_i(\beta_0)) \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})[\hat{\boldsymbol{x}}_i - \mathrm{E}(\boldsymbol{x}_i \mid U_i(\beta_0))] \\
&+ \sum_{i=1}^{n} e_i \Big[\hat{\eta'}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) - \eta'(U_i(\beta_0))\Big] \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i \\
&- \sum_{i=1}^{n} \eta'(U_i(\beta_0)) \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i \Big[\hat{\eta}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) - \hat{\eta}_{d_n}\Big(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}}\Big)\Big] \\
&- \sum_{i=1}^{n} \eta'(U_i(\beta_0)) \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}) \Big\{\hat{\boldsymbol{x}}_i \Big[\hat{\eta}_{d_n}\Big(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}}\Big) - \eta(U_i(\beta_0))\Big] - e_i \, \mathrm{E}[\boldsymbol{x}_i \mid U_i(\beta_0)]\Big\} \\
&- \sum_{i=1}^{n} \Big[\hat{\eta}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) - \eta(U_i(\beta_0))\Big] \Big[\hat{\eta'}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) - \eta'(U_i(\beta_0))\Big] \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i \\
=:& R_1(\check{\boldsymbol{b}}) + R_2(\check{\boldsymbol{b}}) - R_3(\check{\boldsymbol{b}}) - R_4(\check{\boldsymbol{b}}) - R_5(\check{\boldsymbol{b}}).
\end{aligned} \tag{S6.24}$$

For $R_1(\check{\boldsymbol{b}})$, by using (S6.22) and that η has a bounded derivatives in condition (A6)(i), it is not hard to prove

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_1(\check{\boldsymbol{b}})| = o_p(p) \tag{S6.25}$$

by utilizing Lemma 8.

Next, for $R_2(\check{\boldsymbol{b}})$, we observe that

$$\begin{aligned}
R_2(\check{\boldsymbol{b}}) =& \sum_{i=1}^{n} e_i \left| \eta'(U_i(\beta_0)) - \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)} \sum_{k=1}^{n} \tilde{W}_{nk}[\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}]\eta(U_k(\beta_0)) \right| \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i \\
&+ \sum_{i=1}^{n} e_i \left[\sum_{k=1}^{n} \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)} \tilde{W}_{nk}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}) e_k \right] \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i \\
=:& R_{21}(\check{\boldsymbol{b}}) + R_{22}(\check{\boldsymbol{b}})
\end{aligned} \tag{S6.26}$$

For $R_{21}(\check{\boldsymbol{b}})$, we shall prove

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} |(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T R_{21}(\check{\boldsymbol{b}})| = o_p(p) \tag{S6.27}$$

by utilizing Lemma 8. Let $V_i = (e_i,\, \hat{\boldsymbol{x}}_p^T)^T$ and

$$\varphi_{\boldsymbol{b}}(V_i) = \frac{n}{p} e_i \left| \eta'(U_i(\beta_0)) - \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)} \sum_{k=1}^{n} \tilde{W}_{nk}[\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}]\eta(U_k(\beta_0)) \right| \times (\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i. \tag{S6.28}$$

As we assumed that the kernel K satisfies the Lipschitz condition of order 1, it is easy to verify (S5.75). Recall that $\mathrm{P}(\|X\|_2 \geq n^c) \leq n^{-D}$ for any $c,\, D > 0$. By using (S6.22), Lemma 5, Lemma 6 and (S6.14), we have

$$\begin{aligned}
&\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\varphi_{\boldsymbol{b}}(V_i)\right]^2 \leq \mathrm{E}\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\left[\frac{1}{n^2}\sum_{i=1}^{n}\varphi_{\boldsymbol{b}}^2(V_i)\right] \\
&= \frac{1}{p^2}\sum_{i=1}^{n}\mathrm{E}\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\left\{\left[\eta'(U_i(\beta_0)) - \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}\sum_{k=1}^{n}\tilde{W}_{nk}[\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}]\eta(U_k(\beta_0))\right]^2 \right. \\
&\quad \left. \times \mathrm{I}[\|X_i\|_2 \leq n^c]\right\} \times O\left(\frac{n^c p^2}{n}\right) + o(p^{-1}n^{-c}) \\
&\leq O\left(n(h^*)^2 + nd_n\log n + \frac{pn^c}{(h^*)^2}\right) \times O(n^{-1+c}) + o(p^{-1}n^{-c}) = o(p^{-1}n^{-c})
\end{aligned} \tag{S6.29}$$

for some $c > 0$ by taking c small enough. By using Markov's inequality, equation (S5.75) holds. Now (S6.27) follows by using (S6.29) and applying

Lemma 8. For $R_{22}(\check{\boldsymbol{b}})$, we shall prove

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_{22}(\check{\boldsymbol{b}})| = o_p(p) \tag{S6.30}$$

in Section S6.1 This, together with (S6.26) and (S6.27) proves

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_2(\check{\boldsymbol{b}})| = o_p(p). \tag{S6.31}$$

For $R_3(\check{\boldsymbol{b}})$, we shall show in Section S6.2 that

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_3(\check{\boldsymbol{b}}) + n(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \boldsymbol{V}_p(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)| = o_p(p). \tag{S6.32}$$

For $R_4(\check{\boldsymbol{b}})$, we shall show in Section S6.3 that

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_4(\check{\boldsymbol{b}})| = o_p(p). \tag{S6.33}$$

For $R_5(\check{\boldsymbol{b}})$, we shall show in Section S6.4 that

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_5(\check{\boldsymbol{b}})| = o_p(p). \tag{S6.34}$$

Now, combining (S6.24), (S6.25) and (S6.31)–(S6.34) yields

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} (\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R(\check{\boldsymbol{b}}) = -n(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \boldsymbol{V}_p(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0) + o_p(p). \tag{S6.35}$$

By condition (A8), the minimum eigenvalue of $\boldsymbol{V}_p$ is bounded away from zero for all $p \geq 2$. Therefore, for an arbitrary $\delta > 0$, there exists a C_2 not depending on n such that $\mathrm{P}\{\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}[(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R(\check{\boldsymbol{b}})] < 0\} \geq 1 - \delta$ when n is large enough. With this, (S6.5) follows from Theorem 6.3.4 of Ortega and Rheinboldt (2000), which completes the proof.

S6.1 Proof of (S6.30)

We shall prove (S6.30) by utilizing Lemma 8. Let $V_i = (e_i,\, \hat{\boldsymbol{x}}_p^T)^T$ and

$$\varphi_{\boldsymbol{b}}(V_i) = \frac{n}{p} e_i \left[\sum_{k=1}^{n} \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)} \tilde{W}_{nk}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}) e_k \right] (\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}) \hat{\boldsymbol{x}}_i. \tag{S6.36}$$

To simplify notations, in this section, we write

$$\tilde{W}_{nk}(i) = \tilde{W}_{nk}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}), \text{ and } \frac{g_n^*(i)}{g_{d_n}^*(i)} = \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}. \tag{S6.37}$$

As we assumed that the kernel K satisfies the Lipschitz condition of order 1, it is not hard to verify (S5.75). For (S5.76), we have, uniformly over $\check{\boldsymbol{b}} \in \mathcal{B}_2$,

$$\begin{aligned}
&\mathrm{E}\left\{ \sum_{i=1}^{n} e_i \left[\sum_{k=1}^{n} \frac{g_n^*(i)}{g_{d_n}^*(i)} \tilde{W}_{nk}(i) e_k \right] \right\}^2 \\
&= \mathrm{E}\left[\sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{m=1}^{n} \frac{g_n^*(i)}{g_{d_n}^*(i)} \tilde{W}_{nk}(i) \frac{g_n^*(l)}{g_{d_n}^*(l)} \tilde{W}_{nm}(l) e_i e_k e_l e_m \right] \\
&\leq c\,\mathrm{E} \sum_{i=1}^{n} \left[\frac{g_n^*(i)}{g_{d_n}^*(i)} \tilde{W}_{ni}(i) \right]^2 + c\,\mathrm{E}\left[\sum_{i=1}^{n} \sum_{k:k\neq i,\, 1\leq k\leq n} \left| \frac{g_n^*(i)}{g_{d_n}^*(i)} \tilde{W}_{nk}(i) \right|^2 \right] \\
&= \mathrm{E}\left\{ \sum_{i=1}^{n} \left[\frac{\mathrm{I}[\|X_i\|_2 \leq n^c]}{n(h^*)^2 \max\{f_{\beta_0}(U_i(\beta_0)),\, d_n\}} \right]^2 \times O_r(1) \right\} \\
&\qquad + \mathrm{E}\left[\sum_{i=1}^{n} \frac{\mathrm{I}[\|X_i\|_2 \leq n^c]}{n(h^*)^3 \max\{f_{\beta_0}(U_i(\beta_0)),\, d_n\}} \times O_r(1) \right] + o(1) \\
&= o\left(\frac{n^c}{n(h^*)^4 d_n} \right) + o\left(\frac{n^c}{(h^*)^3} \right) = o\left(\frac{n^c}{(h^*)^3} \right),
\end{aligned} \tag{S6.38}$$

by taking $c > 0$ small enough and r large enough, where the third equality can be shown by using Lemmas 3 and 4 and (S5.67), and the second last equality results from Hölder's inequality and (S6.14). Therefore, by taking $c > 0$ small enough, we have

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} \varphi_{\boldsymbol{b}}(V_i)\right]^2 = \left(\frac{n^c}{(h^*)^3}\right) \times O(n^{-1+c}) = o(n^{-c}). \tag{S6.39}$$

So (S5.76) follows for any fixed $\varepsilon_n = \varepsilon > 0$. By using Lemma 9, (S6.22) and (S6.14), it is not hard to verify

$$\begin{aligned}\mathrm{E}\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2}\left[\frac{1}{n^2}\sum_{i=1}^{n} \varphi_{\boldsymbol{b}}^2(V_i)\right] &= \mathrm{E}\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2}\left[\frac{1}{n^2}\sum_{i=1}^{n} \varphi_{\boldsymbol{b}}^2(V_i)\,\mathrm{I}[\|X_i\|_2 \le n^c]\right] + o(n^{-1}) \\ &= o\left(\frac{n^c p}{n(h^*)^3}\right) = o(p^{-1}n^{-c})\end{aligned} \tag{S6.40}$$

for some $c > 0$ small enough. Now (S6.30) follows immediately by applying Lemma 8.

S6.2 Proof of (S6.32)

By using (S5.67), it is not hard to verify

$$\mathrm{E}\left\{[R_3(\check{\boldsymbol{b}})]^2\,\mathrm{I}\left[\sup_{1\le i\le n}\|X_i\|_2 > n^c\right]\right\} = o(n) \tag{S6.41}$$

for any $c > 0$. From this, without loss of generality, we assume $\sup_{1\le i\le n}\|X_i\| \le n^c$ where $c > 0$ does not depend on n and can be arbitrarily small in this section. Define the event $\mathcal{G}_i = \{f_{\beta_0}(U_i(\beta_0) \ge (1+\delta)d_n\}$ and the event $\mathcal{G}_i' = \{\inf_{\check{\boldsymbol{b}} \in \mathcal{B}_2'} g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h) \ge \nu_2 d_n^2\}$. By using Lemma 3, we can obtain

$$\sum_{i=1}^{n} \mathrm{P}(\mathcal{G}_i \setminus \mathcal{G}_i') \to 0. \tag{S6.42}$$

When $g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}_p,\, h) \geq \nu_2 d_n^2$ holds for all $\check{\boldsymbol{b}} \in \mathcal{B}_2'$, by a first-order Taylor expansion of $\hat{\eta}_{d_n}(\cdot)$ with respect to $\check{\boldsymbol{b}}$, we have

$$\hat{\eta}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) - \hat{\eta}_{d_n}\Big(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}}\Big) = \left.\frac{\partial \hat{\eta}_{d_n}(\hat{U}_i(\boldsymbol{b}(\check{\boldsymbol{b}})) \mid \boldsymbol{b}(\check{\boldsymbol{b}}))}{\partial \check{\boldsymbol{b}}}\right|_{\check{\boldsymbol{b}}=\check{\boldsymbol{b}}^*} (\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0), \tag{S6.43}$$

where $\check{\boldsymbol{b}}^* = \theta\check{\boldsymbol{b}} + (1-\theta)\check{\boldsymbol{b}}_0$ for some $\theta \in [0,\, 1]$. From this, it follows that

$$\begin{aligned} R_3(\check{\boldsymbol{b}}) = &\sum_{i=1}^n \eta'\Big(\hat{U}_i(\tilde{\boldsymbol{b}})\Big)\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i\Big[\hat{\eta}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) - \hat{\eta}_{d_n}\Big(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}}\Big)\Big]\,\mathrm{I}[\mathcal{G}_i^c] \\ &+ \sum_{i=1}^n \eta'\Big(\hat{U}_i(\tilde{\boldsymbol{b}})\Big)\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \left.\frac{\partial \hat{\eta}(\hat{U}_i(\boldsymbol{b}(\check{\boldsymbol{b}})) \mid \boldsymbol{b}(\check{\boldsymbol{b}}))}{\partial \check{\boldsymbol{b}}}\right|_{\check{\boldsymbol{b}}=\check{\boldsymbol{b}}^*} \mathrm{I}[\mathcal{G}_i \cap \mathcal{G}_i'] \\ =:&\, R_{31} + R_{32} \end{aligned} \tag{S6.44}$$

holds in probability.

For R_{31}, by using Lemmas 5 and 9, we have, uniformly over $1 \leq i \leq n$,

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \left|\hat{\eta}_{d_n}\Big(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}\Big) - \hat{\eta}_{d_n}\Big(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}}\Big)\right| \leq 2\eta(U_i(\beta_0)) + O_p\left(\sqrt{\frac{p\log n}{nhd_n}}\right). \tag{S6.45}$$

Therefore, by using (S6.22), condition (A6)(i) and $\sum_{i=1}^n \mathrm{I}[\mathcal{G}_i^c] = O_p(nd_n)$, it is easy to verify

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_{31}(\check{\boldsymbol{b}})| = o_p(p). \tag{S6.46}$$

For R_{32}, to simplify notations, in rest of this section, we write

$$u_i(\check{\boldsymbol{b}}) = \hat{U}_i(\boldsymbol{b}(\check{\boldsymbol{b}})),\ s_l(i \mid \check{\boldsymbol{b}}) = s_l(u_i(\check{\boldsymbol{b}}) \mid \boldsymbol{b}(\check{\boldsymbol{b}}),\, h), \tag{S6.47}$$

$$\hat{\eta}(i \mid \check{\boldsymbol{b}}) = \hat{\eta}(u_i(\check{\boldsymbol{b}}) \mid \boldsymbol{b}(\check{\boldsymbol{b}})),\ \hat{\eta'}(i \mid \check{\boldsymbol{b}}) = \hat{\eta'}(u_i(\check{\boldsymbol{b}}) \mid \boldsymbol{b}(\check{\boldsymbol{b}}), h), \tag{S6.48}$$

and

$$W_{nk}(i \mid \check{\boldsymbol{b}}) = W_{nk}(u_i(\check{\boldsymbol{b}}) \mid \boldsymbol{b}(\check{\boldsymbol{b}})),\ \tilde{W}_{nk}(i \mid \check{\boldsymbol{b}}) = \tilde{W}_{nk}(u_i(\check{\boldsymbol{b}}) \mid \boldsymbol{b}(\check{\boldsymbol{b}}), h). \tag{S6.49}$$

Here we use the notation $\hat{\eta'}(\cdot \mid \cdot, h)$ and $\tilde{W}_{nk}(\cdot \mid \cdot, h)$ to emphasize that the bandwidth is h. Recall that $(\hat{\eta}(\cdot \mid \check{\boldsymbol{b}}), \hat{\eta'}(\cdot \mid \check{\boldsymbol{b}}))$ is the minimizer of (2.9) for a given $\check{\boldsymbol{b}}$. Therefore, for a given $u_i(\check{\boldsymbol{b}})$, the corresponding minimizer $(\hat{\eta}(u_i(\check{\boldsymbol{b}}) \mid \check{\boldsymbol{b}}), \hat{\eta'}(u_i(\check{\boldsymbol{b}}) \mid \check{\boldsymbol{b}}))$ satisfies

$$\sum_{k=1}^{n} \Big[Y_k - \hat{\eta}(i \mid \check{\boldsymbol{b}}) - \hat{\eta'}(i \mid \check{\boldsymbol{b}})(u_k(\check{\boldsymbol{b}}) - u_i(\check{\boldsymbol{b}}))\Big] K_h(u_k(\check{\boldsymbol{b}}) - u_i(\check{\boldsymbol{b}})) = 0. \tag{S6.50}$$

Taking derivatives with respect to $\check{\boldsymbol{b}}$ on both sides yields

$$\begin{aligned}
\frac{\partial \hat{\eta}(i \mid \check{\boldsymbol{b}})}{\partial \check{\boldsymbol{b}}} = & -\frac{1}{s_0(i \mid \check{\boldsymbol{b}})} \hat{\eta'}(i \mid \check{\boldsymbol{b}}) \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}) \sum_{k=1}^{n} (\hat{\boldsymbol{x}}_k - \hat{\boldsymbol{x}}_i) K_h(u_k(\check{\boldsymbol{b}}) - u_i(\check{\boldsymbol{b}})) \\
& - \frac{1}{s_0(i \mid \check{\boldsymbol{b}})} \frac{\partial \hat{\eta'}(i \mid \check{\boldsymbol{b}})}{\partial \check{\boldsymbol{b}}} \sum_{k=1}^{n} (u_k(\check{\boldsymbol{b}}) - u_i(\check{\boldsymbol{b}})) K_h(u_k(\check{\boldsymbol{b}}) - u_i(\check{\boldsymbol{b}})) \\
& + \frac{1}{s_0(i \mid \check{\boldsymbol{b}})} \sum_{k=1}^{n} \Big[Y_k - \hat{\eta}(i \mid \check{\boldsymbol{b}}) - \hat{\eta'}(i \mid \check{\boldsymbol{b}})(u_k(\check{\boldsymbol{b}}) - u_i(\check{\boldsymbol{b}}))\Big] \\
& \times K_h'(u_k(\check{\boldsymbol{b}}) - u_i(\check{\boldsymbol{b}})) \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}) \frac{(\hat{\boldsymbol{x}}_k - \hat{\boldsymbol{x}}_i)}{h} \\
= & R_{32,1} + R_{32,2} + R_{32,3}.
\end{aligned} \tag{S6.51}$$

Before proceeding further, we introduce a useful lemma.

Lemma 10. *Under the assumptions in Lemma 3 and condition (A6)(ii), we have that, uniformly over* $1 \le i \le n$, $1 \le j \le p$, *and* $\check{\boldsymbol{b}} \in \mathcal{B}_2'$, *if* $\mathcal{G}_i$ *holds and* $\|X_i\|_2 \in \mathcal{A}_x(v_n)$ *where* $\mathcal{A}_x(v_n)$ *is defined in Lemma 3, then for any integer* $r \ge 1$, *there exists a* $c > 0$ *such that*

$$\begin{aligned} &\frac{1}{n f_{\beta_0}(U_i(\beta_0))} \sum_{k=1}^{n} \hat{\lambda}_{\cdot j}^{-1/2} \hat{x}_{kj} K_h(\hat{U}_i(\boldsymbol{b}) - \hat{U}_k(\boldsymbol{b})) \\ &= \mathrm{E}[\lambda_j^{-1/2} x_{\cdot j} \mid U(\beta_0) = U_i(\beta_0)] + o_r(n^{-c}), \end{aligned} \tag{S6.52}$$

and additionally, if condition (A6)(i) also holds, then for any integer $r \ge 1$, *there exists a* $c > 0$ *such that*

$$\begin{aligned} &\frac{1}{n f_{\beta_0}(U_i(\beta_0))} \sum_{k=1}^{n} \hat{\lambda}_{\cdot j}^{-1/2} \hat{x}_{kj} K_h(\hat{U}_i(\boldsymbol{b}) - \hat{U}_k(\boldsymbol{b})) \eta(U_k(\beta_0)) \\ &= \mathrm{E}[\lambda_j^{-1/2} x_{\cdot j} \mid U(\beta_0) = U_i(\beta_0)] \eta(U_i(\beta_0)) + o_r(n^{-c}). \end{aligned} \tag{S6.53}$$

Proof. By using (S6.19), this lemma can be proved by an argument similar to that of Lemma 3. The proof is not particularly difficult. So we omit it. □

For $R_{32,1}$, by using Lemmas 2 and 10 and (S6.17), we can obtain, uniformly over $1 \le i \le n$ and $\check{\boldsymbol{b}}$, $\check{\boldsymbol{b}}^* \in \mathcal{B}_2'$,

$$\begin{aligned} &-(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}^*) \frac{1}{s_0(i \mid \check{\boldsymbol{b}}^*)} \sum_{k=1}^{n} (\hat{\boldsymbol{x}}_k - \hat{\boldsymbol{x}}_i) K_h(u_k(\check{\boldsymbol{b}}^*) - u_i(\check{\boldsymbol{b}}^*)) \, \mathrm{I}[\mathcal{G}_i \cap \mathcal{G}_i'] \\ &= (\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \tilde{\boldsymbol{J}}^T (\boldsymbol{x}_i - \mathrm{E}[\boldsymbol{x} \mid U(\beta_0) = U_i(\beta_0)]) \, \mathrm{I}[\mathcal{G}_i \cap \mathcal{G}_i'] + o_2(\sqrt{p/n}). \end{aligned} \tag{S6.54}$$

By using this, we have

$$\sup_{\check{\boldsymbol{b}},\,\check{\boldsymbol{b}}^*\in\mathcal{B}_2'}\Bigg|-(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\sum_{i=1}^{n}\eta'(U_i(\tilde{\boldsymbol{b}}))\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i\left[\sum_{l=1}^{n}\tilde{W}_{nl}(i\mid\check{\boldsymbol{b}}^*)\eta(U_l(\beta_0))\right]$$
$$\times(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}^*)\left[\frac{1}{s_0(i\mid\check{\boldsymbol{b}}^*)}\sum_{k=1}^{n}(\hat{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}_i)K_h(u_k(\check{\boldsymbol{b}})-u_i(\check{\boldsymbol{b}}))\right]\mathrm{I}[\mathcal{G}_i\cap\mathcal{G}_i']$$
$$-\,n(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\boldsymbol{V}_p(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)\Bigg|=o_p(p)\tag{S6.55}$$

For the term containing e_k, $1\le k\le n$, letting $V_i=(e_i,\,\hat{\boldsymbol{x}}_i^T)^T$, we claim

$$\sup_{\check{\boldsymbol{b}},\,\check{\boldsymbol{b}}^*\in\mathcal{B}_2'}\left|\frac{1}{n}\sum_{i=1}^{n}\varphi_{\boldsymbol{b},\,\boldsymbol{b}^*}(V_i)\right|$$
$$=:\sup_{\check{\boldsymbol{b}},\,\check{\boldsymbol{b}}^*\in\mathcal{B}_2'}\Bigg|\frac{1}{n}\sum_{i=1}^{n}\frac{n}{p}(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\eta'(\hat{U}_i(\tilde{\boldsymbol{b}}))\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i\sum_{l=1}^{n}\left[\tilde{W}_{nk}(i\mid\check{\boldsymbol{b}})e_l\right]$$
$$\times(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}^*)\left[\frac{1}{s_0(i\mid\check{\boldsymbol{b}}^*)}\sum_{k=1}^{n}(\hat{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}_i)K_h(u_k(\check{\boldsymbol{b}})-u_i(\check{\boldsymbol{b}}))\right]\mathrm{I}[\mathcal{G}_i\cap\mathcal{G}_i']\Bigg|.$$
$$=o_p(1).\tag{S6.56}$$

The proof is straightforward by using Lemma 8 so we omit it here. Combining (S6.55) and (S6.56) yields

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}|(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^TR_{32,\,1}(\check{\boldsymbol{b}})-n(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\boldsymbol{V}_p(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)|=o_p(p).\tag{S6.57}$$

For $R_{32,\,2}$, we observe that the asymptotic behavior of $\lambda_j^{-1/2}\,\partial\tilde{W}_{nl}(i\mid\check{\boldsymbol{b}})\Big/\partial b_j$ is similar to that of $h^{-1}\tilde{W}_{nl}(i\mid\check{\boldsymbol{b}})\lambda_j^{-1/2}(\hat{x}_{lj}-\hat{x}_{ij})$ for $1\le i,\,l\le n$ and $1\le j\le p$, which can be verified by direct calculations. Therefore, by using

Lemma 3 and arguments similar the proofs of Lemmas 5, 6 and 9, we can obtain, uniformly over $\check{\boldsymbol{b}}$, $\check{\boldsymbol{b}}^* \in \mathcal{B}_2'$,

$$\begin{aligned}\sum_{i=1}^{n}[(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T R_{32,2}]^2 = \sum_{i=1}^{n}\Bigg\{&(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T \boldsymbol{J}^T(\check{\boldsymbol{b}}^*)\hat{\boldsymbol{\Lambda}}_p^{1/2}\,\mathrm{I}[\mathcal{G}_i\cap\mathcal{G}_i'] \times O_r(n^{-c}h)\\ &\times\left[\hat{\boldsymbol{\Lambda}}_p^{-1/2}\sum_{l=1}^{n}\frac{\partial \tilde{W}_{nl}(i \mid \check{\boldsymbol{b}})}{\partial \boldsymbol{b}}\bigg|_{\boldsymbol{b}=\boldsymbol{b}(\check{\boldsymbol{b}}^*)}(\eta(U_l(\beta_0))+e_l)\right]\Bigg\}^2\\ = o_p(p).\end{aligned} \tag{S6.58}$$

Therefore, by using Cauchy–Schwarz inequality, it follows that

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}|(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T R_{32,2}(\check{\boldsymbol{b}})| = o_p(p) \tag{S6.59}$$

For $R_{32,3}$, by using a second-order Taylor expansion of $\eta(\cdot)$, we have, uniformly over $\check{\boldsymbol{b}} \in \mathcal{B}_2'$ and $1 \le i,\, k \le n$, if $|(u_i(\check{\boldsymbol{b}}) - u_k(\check{\boldsymbol{b}})| \le h$, then

$$\eta(U_i(\beta_0)) = \eta(u_k(\check{\boldsymbol{b}})) + \eta'(u_k(\check{\boldsymbol{b}}))(u_i(\check{\boldsymbol{b}}) - u_k(\check{\boldsymbol{b}})) + O(h^2) + O_r(pn^{-1/2+c}), \tag{S6.60}$$

for any $c > 0$ and integer $r \ge 1$ not depending on n. By substituting this into the expression of $R_{32,3}$, we can obtain

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}|(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T R_{32,3}(\check{\boldsymbol{b}})| = o_p(p) \tag{S6.61}$$

The detailed proof is not particularly difficult by using Lemma 10 and the results in Section S2. We omit it here.

Now (S6.2) follows from (S6.44), (S6.46), (S6.51), (S6.57), (S6.59) and (S6.61).

S6.3 Proof of (S6.33)

To simplify notations, in this section, we write

$$\frac{g_n(i)}{g_{d_n}(i)} = \frac{g_n(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}},\, h)}{g_{d_n}(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}},\, h)}, \text{ and } W_{nk}(i) = W_{nk}(\hat{U}_i(\tilde{\boldsymbol{b}}) \mid \tilde{\boldsymbol{b}}). \tag{S6.62}$$

Separating (S6.33) yields

$$\begin{aligned} R_4(\check{\boldsymbol{b}}) = & \sum_{k=1}^{n} \left\{ \sum_{i=1}^{n} \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}) \frac{g_n(i)}{g_{d_n}(i)} W_{nk}(i)\eta'(U_i(\beta_0))\hat{\boldsymbol{x}}_i - \eta'(U_k(\beta_0)) \, \mathrm{E}(\boldsymbol{x}_k \mid U(\beta_0)) \right\} e_k \\ & + \sum_{i=1}^{n} \eta'(U_i(\beta_0)) \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}}) \hat{\boldsymbol{x}}_i \left[\frac{g_n(i)}{g_{d_n}(i)} \sum_{k=1}^{n} W_{nk}(i)\eta(U_k(\beta_0)) - \eta(U_i(\beta_0)) \right] \\ & =: R_{41}(\check{\boldsymbol{b}}) + R_{42}(\check{\boldsymbol{b}}). \end{aligned} \tag{S6.63}$$

By using (S5.67), it is not hard to verify

$$\mathrm{E}\left\{ [R_4(\check{\boldsymbol{b}})]^2 \, \mathrm{I}\left[\sup_{1\le i\le n} \|X_i\|_2 > n^c \right] \right\} = o(n) \tag{S6.64}$$

for any $c > 0$. From this, without loss of generality, we assume $\sup_{1\le i\le n} \|X_i\| \le n^c$ where $c > 0$ is an arbitrarily small constant not depending on n in this section.

Write $R_{41}(\check{\boldsymbol{b}}) = \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\Lambda^{1/2} R_{41}^*(\check{\boldsymbol{b}})$. Let $R_{41,j}^*$ denote the jth component of $R_{41}(\check{\boldsymbol{b}})$. Similar to (S6.62), in this section, we write

$$\frac{\bar{g}_n(i)}{\bar{g}_{d_n}(i)} = \frac{g_n(U_i(\beta_0) \mid \beta_0,\, h)}{g_{d_n}(U_i(\beta_0) \mid \beta_0,\, h)}, \text{ and } \bar{W}_{nk}(i) = W_{nk}(\hat{U}_i(\beta_0) \mid \beta_0). \tag{S6.65}$$

Then we have

$$\begin{aligned}
&\mathrm{E}(R^*_{41,\,j})^2 \\
&= \lambda_j^{-1} \sum_{k=1}^{n} \mathrm{E}\left[\sum_{i=1}^{n} \frac{g_n(i)}{g_{d_n}(i)} W_{nk}(i)\eta'(U_i(\beta_0))\hat{x}_{ij} - \eta'(U_k(\beta_0))\,\mathrm{E}(x_{kj} \mid U_k(\beta_0))\right]^2 \\
&\le c\lambda_j^{-1} \sum_{k=1}^{n} \mathrm{E}\left[\sum_{i=1}^{n} \frac{g_n(i)}{g_{d_n}(i)} W_{nk}(i)\eta'(U_i(\beta_0))\hat{x}_{ij} - \frac{\bar{g}_n(i)}{\bar{g}_{d_n}(i)} \bar{W}_{nk}(i)\eta'(U_i(\beta_0))x_{ij}\right]^2 \\
&\qquad + c\lambda_j^{-1} \sum_{k=1}^{n} \mathrm{E}\left[\sum_{i=1}^{n} \frac{\bar{g}_n(i)}{\bar{g}_{d_n}(i)} \bar{W}_{nk}(i)\eta'(U_i(\beta_0))x_{ij} - \eta'(U_k(\beta_0))\,\mathrm{E}(x_{kj} \mid U_k(\beta_0))\right]^2 \\
&=: R^*_{41,\,j1} + R^*_{41,\,j2}.
\end{aligned} \tag{S6.66}$$

For $R^*_{41,\,j1}$, by using (S5.3)–(S5.5) and (S5.43) in the proof of Lemma 4, (S5.70) and (S6.19), it is not hard to prove

$$R^*_{41,\,j1} = o(n). \tag{S6.67}$$

By some standard computations, it is easy to verify

$$\begin{aligned}
&\mathrm{E}\left\{\frac{1}{n}\sum_{k=1}^{n}\left[\sum_{i=1}^{n} \frac{\bar{g}_n(i)}{\bar{g}_{d_n}(i)} \bar{W}_{nk}(i)\eta'(U_i(\beta_0))\lambda_j^{-1/2}[x_{ij} - \mathrm{E}(x_{ij} \mid U_i(\beta_0))]\right]^2\right\} \\
&= \mathrm{E}\left\{\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\left[\frac{\bar{g}_n(i)}{\bar{g}_{d_n}(i)} \bar{W}_{nk}(i)\eta'(U_i(\beta_0))\lambda_j^{-1/2}[x_{ij} - \mathrm{E}(x_{ij} \mid U_i(\beta_0))]\right]^2\right\} = o(1)
\end{aligned} \tag{S6.68}$$

From this, by using Lemma 7, we can obtain

$$R^{*}_{41,\,j2} = o(n). \tag{S6.69}$$

Substituting (S6.67) and (S6.69) into (S6.72) yields

$$\mathrm{E}(R^{*}_{41,\,j})^2 = o(n), 1 \le j \le p. \tag{S6.70}$$

By using (S6.16), we can also obtain

$$\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \left|\left|\left| \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\Lambda^{1/2} \right|\right|\right| = O_p(1). \tag{S6.71}$$

Therefore, we have

$$\begin{aligned} &\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T R_{41}(\check{\boldsymbol{b}})| \\ &\le \sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \|\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0\| \times \sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \left|\left|\left| \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\Lambda^{1/2} \right|\right|\right| \times \sqrt{p} \times o_p(\sqrt{n}) = o_p(p). \end{aligned} \tag{S6.72}$$

We now turn to $R_{42}(\check{\boldsymbol{b}})$. By using (S6.22) and Lemmas 4 and 5, we have

$$\begin{aligned} &\sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \sum_{i=1}^{n} \eta'(U_i(\beta_0))(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i \\ &\quad \times \left[\frac{g_n(i)}{g_{d_n}(i)} \sum_{k=1}^{n} W_{nk}(i)\eta(U_k(\beta_0)) - \eta(U_i(\beta_0)) \right] \mathrm{I}[f_{\beta_0}(U_i(\beta_0)) \le (1+\delta)d_n] \\ &\le \sup_{\check{\boldsymbol{b}} \in \mathcal{B}_2} \sup_{1 \le i \le n} |(\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i| \times \sup_{1 \le i \le n} |\eta'(U_i(\beta_0))| \times \sup_{1 \le i \le n} |\eta(U_i(\beta_0))| \\ &\quad \times \sum_{i=1}^{n} \mathrm{I}[f_{\beta_0}(U_i(\beta_0)) \le (1+\delta)d_n] \times O_p(1) \\ &= O_p(n^{-1/2+c}p) \times O_p(n^c) \times O_p(nd_n \log n) = o_p(p) \end{aligned} \tag{S6.73}$$

by taking c small enough. By an argument similar to the proof of (S5.46) in Lemma 5, we can obtain

$$\sup_{1\le i\le n}\left|\frac{g_n(i)}{g_{d_n}(i)}\left[\sum_{k=1}^{n}W_{nk}(i)\eta(U_k(\beta_0))-\eta(U_i(\beta_0))\right]\mathrm{I}[f_{\beta_0}(U_i(\beta_0))\ge(1+\delta)d_n]\right|=O_r(h^2) \tag{S6.74}$$

for any integer $r\ge 1$. From this, by using the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&\mathrm{E}\left\{\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\sum_{i=1}^{n}\eta'(U_i(\beta_0))(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i\right.\\
&\qquad\left.\times\left[\frac{g_n(i)}{g_{d_n}(i)}\sum_{k=1}^{n}W_{nk}(i)\eta(U_k(\beta_0))-\eta(U_i(\beta_0))\right]\mathrm{I}[f_{\beta_0}(U_i(\beta_0))>(1+\delta)d_n]\right\}\\
\le\ &\mathrm{E}\left\{\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\sum_{i=1}^{n}\left[\eta'\left(\hat{U}_i(\tilde{\boldsymbol{b}})\right)(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^T\hat{\boldsymbol{J}}^T(\check{\boldsymbol{b}})\hat{\boldsymbol{x}}_i\right]^2\right\}^{1/2}\times O(n^{1/2}h^2)\\
=\ &O((np)^{1/2}h^2)=o(p),
\end{aligned} \tag{S6.75}$$

since we assumed $nh^4/p=o(1)$ in condition (A4)(iii). Combining (S6.75) and (S6.73) yields

$$\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}|(\check{\boldsymbol{b}}-\check{\boldsymbol{b}}_0)^TR_{42}(\check{\boldsymbol{b}})|=o_p(p) \tag{S6.76}$$

This, together with (S6.63) and (S6.72), proves (S6.33).

S6.4 Proof of (S6.34)

To simplify notations, in this section, we write

$$W_{nk}(i) = W_{nk}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}),\ \frac{g_n(i)}{g_{d_n}(i)} = \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h)}; \tag{S6.77}$$

and

$$\tilde{W}_{nk}(i) = \tilde{W}_{nk}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b}),\ \frac{g_n^*(i)}{g_{d_n}^*(i)} = \frac{g_n(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}{g_{d_n}(\hat{U}_i(\boldsymbol{b}) \mid \boldsymbol{b},\, h^*)}. \tag{S6.78}$$

Define

$$A_1(i) = \sum_{k=1}^{n} \frac{g_n(i)}{g_{d_n}(i)} W_{nk}(i)\eta(\hat{U}_k(\boldsymbol{b}_p)) - \eta(\hat{U}_i(\boldsymbol{b}_p)), \tag{S6.79}$$

$$A_2(i) = \sum_{k=1}^{n} \frac{g_n(i)}{g_{d_n}(i)} W_{nk}(i)[\eta(U_k(\beta_0)) - \eta(\hat{U}_k(\boldsymbol{b}_p))], \tag{S6.80}$$

$$A_3(i) = \eta(\hat{U}_i(\boldsymbol{b}_p)) - \eta(U_i(\beta_0)), \tag{S6.81}$$

$$A_4(i) = \sum_{k=1}^{n} \frac{g_n(i)}{g_{d_n}(i)} W_{nk}(i)e_k, \tag{S6.82}$$

$$B_1(i) = \sum_{k=1}^{n} \frac{g_n^*(i)}{g_{d_n}^*(i)} \tilde{W}_{nk}(i)\eta(\hat{U}_k(\boldsymbol{b}_p)) - \eta'(\hat{U}_i(\boldsymbol{b}_p)), \tag{S6.83}$$

$$B_2(i) = \sum_{k=1}^{n} \frac{g_n^*(i)}{g_{d_n}^*(i)} \tilde{W}_{nk}(i)[\eta(U_k(\beta_0)) - \eta(\hat{U}_k(\boldsymbol{b}_p))], \tag{S6.84}$$

$$B_3(i) = \eta'(\hat{U}_i(\boldsymbol{b}_p)) - \eta'(U_i(\beta_0)), \tag{S6.85}$$

$$B_4(i) = \sum_{k=1}^{n} \frac{g_n^*(i)}{g_{d_n}^*(i)} \tilde{W}_{nk}(i) e_k, \tag{S6.86}$$

$$C(i) = (\check{\boldsymbol{b}} - \check{\boldsymbol{b}}_0)^T \hat{\boldsymbol{J}}^T (\check{\boldsymbol{b}}) \hat{\boldsymbol{x}}_i, \tag{S6.87}$$

and

$$D(i) = \mathrm{I}[\|X_i\|_2 \leq n^c], \tag{S6.88}$$

where c is an arbitrarily small constant not depending on n. It is easy to verify

$$\sum_{i=1}^{n} \sum_{l=1}^{4} \sum_{m=1}^{4} A_l(i) B_m(i) C(i)[1 - D(i)] = o_p(p). \tag{S6.89}$$

Therefore, to prove (S6.34), it suffices to prove

$$\sum_{i=1}^{n} A_l(i) B_m(i) C(i) D(i) = o_p(p);\ l,\, m \in \{1,\, 2,\, 3,\, 4\}. \tag{S6.90}$$

Here we only prove

$$\sum_{i=1}^{n} A_l(i) B_m(i) C(i) D(i) = o_p(p),\ (l,\, m) = (4,\, 2) \text{ and } (4,\, 4). \tag{S6.91}$$

The reminder of the proof of (S6.90) is not particularly difficult by using the results in Section S2.

We first consider the case $(l,\, m) = (4,\, 2)$. Let

$$\varphi_{\boldsymbol{b}}(V_k) = \frac{n}{p}\sum_{i=1}^{n} B_2(i)C(i)D(i)\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)e_k. \tag{S6.92}$$

We have

$$\begin{aligned}
&\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2} \mathrm{E}\left[\frac{1}{n}\sum_{k=1}^{n}\varphi_{\boldsymbol{b}}(V_k)\right]^2 \leq \mathrm{E}\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\left[\frac{1}{n^2}\sum_{k=1}^{n}\varphi^2_{\check{\boldsymbol{b}}}(V_k)\right]\\
&= \frac{1}{p^2}\,\mathrm{E}\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\sum_{k=1}^{n}\left\{\sum_{i=1}^{n} B_2(i)C(i)D(i)\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)\right\}^2\\
&\leq \frac{1}{p^2}\times O\left(\frac{n^c p^2}{n^2 h d_n}\right)\times \mathrm{E}\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\sum_{i=1}^{n}[B_2(i)]^2\\
&= O\left(\frac{n^c p}{n^2 h (h^*)^2 d_n}\right) = o(p^{-1}n^{-c})
\end{aligned} \tag{S6.93}$$

for some $c > 0$. Then (S6.91) with $(l,\, m) = (4,\, 2)$ follows by applying Lemma 8.

For the case $(l,\, m) = (4,\, 4)$, let

$$\varphi_{\boldsymbol{b}}(V_i) = \frac{n}{p}\sum_{i=1}^{n}\left[\sum_{k=1}^{n}\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)e_k\right]\left[\sum_{l=1}^{n}\frac{g^*_n(i)}{g^*_{d_n}(i)}\tilde{W}_{nl}(i)e_l\right]C(i)D(i). \tag{S6.94}$$

It is not hard to verify (S5.75). For (S5.76), we observe that, uniformly over $\check{\boldsymbol{b}} \in \mathcal{B}_2$,

$$
\begin{aligned}
& \mathrm{E}\left\{\sum_{i=1}^{n}\left[\sum_{k=1}^{n}\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)e_k\right]\left[\sum_{l=1}^{n}\frac{g_n^*(i)}{g_{d_n}^*(i)}\tilde{W}_{nl}(i)e_l\right]D(i)\right\}^2 \\
= \; & \mathrm{E}\left\{\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{n}\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)\frac{g_n^*(i)}{g_{d_n}^*(i)}\tilde{W}_{nl}(i)e_k e_l D(i)\right\}^2 \\
\leq \; & c\,\mathrm{E}\left[\sum_{k=1}^{n}\sum_{i=1}^{n}\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)\frac{g_n^*(i)}{g_{d_n}^*(i)}\tilde{W}_{nk}(i)D(i)\right]^2 \\
& \quad + c\,\mathrm{E}\left\{\sum_{k=1}^{n}\sum_{l:l\neq k,\,1\leq l\leq n}\sum_{i=1}^{n}\left[\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)\frac{g_n^*(i)}{g_{d_n}^*(i)}\tilde{W}_{nl}(i)D(i)\right]^2\right\} \\
\leq \; & \mathrm{E}\left[\sum_{i=1}^{n}\frac{1}{n(h^*)^2\max\{f_{\beta_0}(U_i(\beta_0)),\, d_n\}}\times O_r(1)\right]^2 \\
& \quad + \mathrm{E}\left\{\sum_{i=1}^{n}\frac{1}{nh(h^*)^3\max\{f_{\beta_0}(U_i(\beta_0)),\, d_n\}}\times O_r(1)\right\} \\
= \; & O\left(\frac{n^c}{h(h^*)^3}\right),
\end{aligned}
\tag{S6.95}
$$

where the second last inequality can be shown by an argument similar to (S5.68) and the last equality results from Hölder's inequality and (S6.14). By using (S6.18) and taking c small enough, it follows that

$$
\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\varphi_{\boldsymbol{b}}(V_i)\right]^2 \leq \frac{1}{p^2}\times O\left(\frac{n^c}{h(h^*)^3}\right)\times O\left(\frac{n^c p^2}{n}\right) = o(n^{-c}). \tag{S6.96}
$$

Therefore, (S5.76) holds for any fixed $\varepsilon_n = \varepsilon > 0$. By using Lemma 9, we can obtain

$$\begin{aligned}
&\mathrm{E}\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\left\{\sum_{i=1}^{n}\left[\sum_{k=1}^{n}\frac{g_n(i)}{g_{d_n}(i)}W_{nk}(i)e_k\sum_{l=1}^{n}\frac{g_n^*(i)}{g_{d_n}^*(i)}\tilde{W}_{nl}(i)e_l\right]^2 D(i)\right\}\\
&=\mathrm{E}\sup_{\check{\boldsymbol{b}}\in\mathcal{B}_2}\left\{\sum_{i=1}^{n}\left[\frac{p^2\log^2 n}{n^2h(h^*)^3\max\{f_{\beta_0}^2(U_i(\beta_0)),\, d_n^2\}}\times O_r(1)\right]\right\} \qquad \text{(S6.97)}\\
&=o\left(\frac{n^c p^2}{nh(h^*)^3 d_n}\right),
\end{aligned}$$

for any $c>0$. So we have

$$\mathrm{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\varphi_{\boldsymbol{b}}^2(V_i)\right]=\frac{1}{p^2}\times o\left(\frac{n^c p^2}{nh(h^*)^3 d_n}\right)\times O\left(\frac{n^c p^2}{n}\right)=o(p^{-1}n^{-c}) \qquad \text{(S6.98)}$$

by using condition (A4)(iii) and taking $c>0$ small enough. Now given $(l,\, m)=(4,\, 4)$, (S6.91) holds by applying Lemma 8.

S7 Proof of Theorem 2

Separating $[\hat{\eta}_{d_n}(\hat{U}(\beta))-\eta(U(\beta_0))]^2$, we have

$$\begin{aligned}
&[\hat{\eta}_{d_n}(\hat{U}(\beta))-\eta(U(\beta_0))]^2\\
&\le c\left\{\eta(\hat{U}(\beta))-\frac{g_n(\hat{U}(\beta)\mid\beta,\, h)}{g_{d_n}(\hat{U}(\beta)\mid\beta,\, h)}\sum_{k=1}^{n}W_{nk}[\hat{U}(\beta)\mid\beta]\eta(\hat{U}_k(\beta))\right\}^2\\
&\quad+c\left\{\frac{g_n(\hat{U}(\beta)\mid\beta,\, h)}{g_{d_n}(\hat{U}(\beta)\mid\beta,\, h)}\sum_{k=1}^{n}W_{nk}[\hat{U}(\beta)\mid\beta]e_k\right\}^2+c[\eta(\hat{U}(\beta))-\eta(U(\beta_0))]^2\\
&\quad+c\left\{\frac{g_n(\hat{U}(\beta)\mid\beta,\, h)}{g_{d_n}(\hat{U}(\beta)\mid\beta,\, h)}\sum_{k=1}^{n}W_{nk}[\hat{U}(\beta)\mid\beta][\eta(\hat{U}_k(\beta))-\eta(U_k(\beta_0))]\right\}^2\\
&=: R_6(\beta)+R_7(\beta)+R_8(\beta)+R_9(\beta).
\end{aligned}$$

(S7.1)

By using Lemma 5, we have

$$\sup_{(X,\beta)\in\mathcal{A}_x(u_n)\times\mathcal{B}_p} R_6(\beta) = O_p(h^4). \tag{S7.2}$$

By using Lemma 9, we have

$$\sup_{(X,\beta)\in\mathcal{A}_x(u_n)\times\mathcal{B}_p} R_7(\beta) = O_p\bigg(\frac{p\log n}{nhd_n'}\bigg). \tag{S7.3}$$

As we assumed that η has a bounded derivative, it follows that

$$\sup_{(X,\beta)\in\mathcal{A}_x(u_n)\times\mathcal{B}_p} R_8(\beta) \le c[\hat{U}(\beta) - U(\beta_0)]^2 = O_p\big(n^{(-2\alpha_1+1)/(\alpha_0+2\alpha_1)}u_n^2\big). \tag{S7.4}$$

For $R_9(\beta)$, by using the Cauchy–Schwarz inequality and Lemma 4, we have, uniformly over $(X,\,\beta)\in\mathcal{A}_x(u_n)\times\mathcal{B}_p$,

$$\begin{aligned} R_9(\beta) &\le c\bigg[\frac{g_n(\hat{U}(\beta)\mid\beta,\,h)}{g_{d_n}(\hat{U}(\beta)\mid\beta,\,h)}\bigg]^2 \sum_{k=1}^n W_{nk}[\hat{U}(\beta)\mid\beta]^2 \times \sum_{k=1}^n [\eta(\hat{U}_k(\beta)) - \eta(U_k(\beta_0))]^2 \\ &= O_p\bigg(\frac{1}{nhd_n'}\bigg)O_p(p) = O_p\bigg(\frac{p}{nhd_n'}\bigg). \end{aligned} \tag{S7.5}$$

Combining (S7.1)–(S7.5) yields

$$\begin{aligned} &\sup_{(X,\beta)\in\mathcal{A}_x(u_n)\times\mathcal{B}_p} [\hat{\eta}_{d_n}(\hat{U}(\beta)) - \eta(U(\beta_0))]^2 \\ &= O_p(h^4) + O_p\bigg(\frac{p\log n}{nhd_n'}\bigg) + O_p\big(n^{(-2\alpha_1+1)/(\alpha_0+2\alpha_1)}u_n^2\big), \end{aligned} \tag{S7.6}$$

which completes the proof of (3.9).

Observe that $\mathrm{E}\{[\max\{f_{\beta_0}(U(\beta_0)),\,d_n\}]^{-1}\,\mathrm{I}[X\in\mathcal{A}_x(u_n)]\} = O(u_n)$. From this, (3.10) can be proved by a similar argument using (S7.1). Further details are not reproduced here.

S8 Proof of Theorem 3

Decomposing $\hat{\sigma}_e^2$, we have

$$\begin{aligned}\hat{\sigma}_e^2 &= \frac{1}{n}\sum_{i=1}^{n} e_i^2 + \frac{1}{n}\sum_{i=1}^{n}[\eta(U_i(\beta_0)) - \hat{\eta}_{d_n}(\hat{U}_i(\hat{\beta}))]^2 \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} 2e_i[\eta(U_i(\beta_0)) - \hat{\eta}_{d_n}(\hat{U}_i(\hat{\beta}))] \qquad \text{(S8.1)} \\ &=: T_1 + T_2 + T_3.\end{aligned}$$

From the classical central limit theorem, it follows that

$$\sqrt{n}(T_1 - \sigma_e^2)[\mathrm{Var}(e_1^2)]^{-1/2} \xrightarrow{d} N(0,\, 1). \qquad \text{(S8.2)}$$

By an argument similar to the proof of Theorem 2, it is not hard to show

$$\sqrt{n}T_2 = o_p(1). \qquad \text{(S8.3)}$$

Similarly to the proof of (S6.31), we can also obtain

$$\sqrt{n}T_3 = o_p(1). \qquad \text{(S8.4)}$$

By Slutsky's Theorem, it follows that

$$\sqrt{n}(\hat{\sigma}_e^2 - \sigma_e^2)(\mathrm{Var}(e_1^2))^{-1/2} \xrightarrow{d} N(0,\, 1). \qquad \text{(S8.5)}$$

From this, to prove (3.11), it suffices to prove

$$\frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^4 - \left[\frac{1}{n}\sum_{i=1}^{n}(\hat{e}_i^2)\right]^2 - \mathrm{Var}(e_1^2) \xrightarrow{p} 0. \tag{S8.6}$$

By using the law of large numbers, we can obtain

$$\frac{1}{n}\sum_{i=1}^{n}e_i^4 - \left[\frac{1}{n}\sum_{i=1}^{n}(e_i^2)\right]^2 \xrightarrow{p} \mathrm{Var}(e_1^2).$$

Therefore, it suffices to show

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{e}_i^4 - e_i^4) \xrightarrow{p} 0, \text{ and} \tag{S8.7}$$

$$\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{e}_i^2)\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}(e_i^2)\right]^2 \xrightarrow{p} 0. \tag{S8.8}$$

Note that (S8.8) can be proved by using (S8.5). So we only need to prove (S8.7). Writing $\hat{e}_i = Y_i - \hat{\eta}_{d_n}(\hat{U}(\hat{\beta})) = e_i + \eta(U_i(\beta_0)) - \hat{\eta}_{d_n}(\hat{U}_i(\hat{\beta}))$, equation (S8.7) can be proved by direct calculations. The details are omitted.

Bibliography

Bhatia, R., C. Davis, and A. McIntosh (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra Appl. 52/53*, 45–67.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In *Nonparametric functional estimation* *and related topics (Spetses, 1990)*, Volume 335 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pp. 509–529. Kluwer Acad. Publ., Dordrecht.

Bosq, D. (2000). *Linear processes in function spaces*, Volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Theory and applications.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist. 35*(1), 70–91.

Ortega, J. M. and W. C. Rheinboldt (2000). *Iterative solution of nonlinear equations in several variables*, Volume 30 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1970 original.

Wainwright, M. J. (2019). *High-dimensional statistics*, Volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. A non-asymptotic viewpoint.

Wang, J.-L., L. Xue, L. Zhu, and Y. S. Chong (2010). Estimation for a partial-linear single-index model. *Ann. Statist. 38*(1), 246–274.

Wong, R. K. W., Y. Li, and Z. Zhu (2019). Partially linear functional additive models for multivariate functional data. *J. Amer. Statist. Assoc. 114*(525), 406–418.