Supplemental Materials for "**Estimation and Inference for Generalized Geoadditive Models**"

This supplementary document shows detailed proofs of the theoretical results in the main paper and the implementation of the bivariate spline smoothing over the triangulation. In Section B.1, we investigate the asymptotic properties of the oracle estimator, the first stage penalized spline pilot estimators, and the spline-backfitted local polynomial estimator. Section B.2 describes the implementation of bivariate spline. Section B.3 provides more simulation results from Examples 1 and 2 in the main paper.

Let $A(\Omega)$ be the area of the domain Ω, and without loss of generality, we assume $A(\Omega) = 1$ in the rest of the article. Note that the triangulation for different coefficient function can be different from each other. For notational convenience in the proof below, we consider a common triangulation for all the explanatory variables: $\mathbf{B}_0(\boldsymbol{z}) = \mathbf{B}_1(\boldsymbol{z}) = \cdots = \mathbf{B}_p(\boldsymbol{z}) = \mathbf{B}(\boldsymbol{z})$, and $\beta_\ell(\boldsymbol{z}_j) = \mathbf{B}^\top(\boldsymbol{z}_j)\boldsymbol{\gamma}_\ell$.

B.1. Theoretical Results with Details

B.1.1. Notations

First we introduce the general notations that we use in the following proof.

For a real value vector $\mathbf{a} \in \mathbb{R}^n$, we define its Euclidean norm as $\|\mathbf{a}\|^2 = \sum_{i=1}^n a_i^2$ and its supremum norm as $|\mathbf{a}| = \max_{1\le i\le n}|a_i|$. For any real symmetric matrix $\mathbf{A} = (a_{ij})_{i=1,j=1}^{m,n}$, denote by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ its smallest and largest eigenvalues, and its L_2 norm as $\|\mathbf{A}\|_2 = \max_{\mathbf{a}\in\mathbb{R}^n,\mathbf{a}\neq\mathbf{0}} \|\mathbf{A}\mathbf{a}\|_2 \|\mathbf{a}\|_2^{-1}$. For any Lebesgue measurable function $\psi(\mathbf{u})$ on a domain $\mathcal{D}$, $\mathcal{D} = [0,1]$, $[0,1]^p$, $\Omega \subseteq \mathbb{R}^2$ or $[0,1]^p \times \Omega$, let $\|\psi\|_\infty = \sup_{\mathbf{u}\in\mathcal{D}} |\psi(\mathbf{u})|$, and $\|\psi\|_{L_2}^2 = \int_{\mathcal{D}} \psi^2(\mathbf{u})d\mathbf{u}$. Define $\mathcal{J} = \{1,\ldots,J_n\}$ as the index set of univariate spline basis functions.

Define the model space $\mathcal{G}$ as

$$\mathcal{G} = \left\{ \psi = \sum_{k=1}^{p} \beta_k(x_k) + \alpha(\mathbf{s}) : \beta_k \in \mathcal{D}_k^0([0,1]), \alpha \in \mathcal{W}^{d+1,\infty}(\Omega) \right\},$$

where $\mathcal{D}_k^0$ and $\mathcal{W}^{d+1,\infty}$ are defined in (9) and (10) in the main paper. We define the norm on the space $\mathcal{G}$. For functions $\psi_1, \psi_2 \in \mathcal{G}$, define their theoretical inner product as $\langle \psi_1, \psi_2 \rangle = \mathrm{E}\psi_1(\mathbf{X}, \mathbf{S})\psi_2(\mathbf{X}, \mathbf{S})$. Define their empirical inner product as $\langle \psi_1, \psi_2 \rangle_n = \frac{1}{n}\sum_{i=1}^{n} \psi_1(\mathbf{X}_i, \mathbf{S}_i)\psi_2(\mathbf{X}_i, \mathbf{S}_i)$. Consequently, $\|\psi\| = \sqrt{\langle \psi, \psi \rangle}$ and $\|\psi\|_n = \sqrt{\langle \psi, \psi \rangle_n}$.

For the notation simplicity, let $\dot{g}^{-1}(x) = \{g^{-1}(x)\}'$. For the quasi-likelihood function $\ell\{g^{-1}(x), y\}$, let $q_1(x,y) = \frac{\partial}{\partial x}\ell\{g^{-1}(x), y\}$ and $q_2(x,y) = \frac{\partial^2}{\partial x^2}\ell\{g^{-1}(x), y\}$. It is clear that

$$q_1(x,y) = \{y - g^{-1}(x)\}\rho_1(x), \qquad q_2(x,y) = \{y - g^{-1}(x)\}\rho_1'(x) - \rho_2(x),$$

where $\rho_j(x) = \{\dot{g}^{-1}(x)\}^j / [\sigma^2 V\{g^{-1}(x)\}]$, $j = 1, 2$. Moreover, let

$$\eta(\mathbf{x}, \mathbf{s}) = \sum_{k=1}^{p} \beta_k(x_k) + \alpha(\mathbf{s}), \qquad \eta_i^0 = \eta(\mathbf{X}_i, \mathbf{S}_i), \qquad \eta_{i,-k}^0 = \sum_{k' \neq k}^{p} \beta_{k'}(X_{ik'}) + \alpha(\mathbf{S}_i)$$

and $\varepsilon_i = Y_i - g^{-1}(\eta_i^0)$ be the error term. Without loss of generality, for the bandwidth of the local polynomial, we consider $h = h_k$ to facilitate the development of the theoretical properties.

Denote $\nu_l = \int z^l K(z) dz$ and $\rho_{2,k}(x) = \mathrm{E}\rho_2\{\eta(\mathbf{X}, \mathbf{S}) | X_k = x\}$. Let $\mathbf{T}_2$, $\mathbf{N}_2$ and $\mathbf{M}_2$ be the 2×2 matrix with entry $\int z^{i+j-2} K(z)^2 dz$, ν_{i+j-2} and ν_{i+j-1}, respectively. Let

$$\mathbf{\Sigma}_k(x_k) = \rho_{2,k}(x_k) f_k(x_k) \mathbf{N}_2, \quad \mathbf{\Lambda}_k(x_k) = \{\rho_{2,k}(x_k) f_k(x_k)\}' \mathbf{M}_2, \quad \mathbf{\Xi}_k(x_k) = \rho_{2,k}(x_k) f_k(x_k) \mathbf{T}_2.$$

B.1.2. Properties of Penalized Quasi-likelihood Estimators

Recall that $u_{kj}(x_k)$, $j \in \mathcal{J}$, are the original B-spline basis functions for the kth covariate, where $\mathcal{J}$ is the index set of the basis functions. In the following we define their centered basis $u_{kj}^0(x_k)$ and the standardized basis $U_{kj}(x_k)$. Let $c_{kj} = \langle u_{kj}, 1\rangle$, we have

$$u_{kj}^0(x_k) = u_{kj}(x_k) - \frac{c_{kj}}{c_{k1}} u_{k1}(x_k), \qquad U_{kj}(x_k) = \frac{u_{kj}^0(x_k)}{\|u_{kj}^0\|},\ j \in \mathcal{J},$$

so that $\mathrm{E}U_{kj}(X_k) = 0$ and $\mathrm{E}U_{kj}^2(X_k) = 1$. Similarly, we define the standardized Bernstein basis polynomials as $B_m^*(\mathbf{s}) = B_m(\mathbf{s})/\|B_m\|$, $m \in \mathcal{M}$, where $\mathcal{M}$ is the index set of Bernstein basis functions. For example, for bivariate spline space $\mathbb{S}_d^r(\triangle)$ containing N triangles, $\mathcal{M} = \{1, 2, \cdots, \frac{(d+1)(d+2)N}{2}\}$. Define the approximate space as

$$\mathcal{A} = \left\{ \phi : \phi(\mathbf{x}, \mathbf{s}) = \sum_{k=1}^{p} \sum_{j \in \mathcal{J}} \theta_{kj} U_{kj}(x_k) + \sum_{m \in \mathcal{M}} \gamma_m B_m^*(\mathbf{s}), x_k \in [0,1], \mathbf{s} \in \Omega, \theta_{kj}, \gamma_m \in \mathbb{R} \right\}.$$

B.1.2.1. Preliminaries

Lemma B.1 *Under Assumptions (A2) and (A5), for $k = 1, \cdots, p$ and $j, j' \in \mathcal{J}$, $m, m' \in \mathcal{M}$ and $r \geq 1$,*

$$\mathrm{E}|U_{kj}(X_{ik})U_{kj'}(X_{ik})|^r \asymp H^{-r}\mathrm{E}|u_{kj}(X_{ik})u_{kj'}(X_{ik})|^r \asymp \begin{cases} 0, & |j - j'| > \varrho + 1, \\ H^{1-r}, & |j - j'| \leq \varrho + 1, \end{cases}$$

$$\mathrm{E}\left|B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i)\right|^r \asymp \begin{cases} 0, & \lceil m/d^* \rceil \neq \lceil m'/d^* \rceil, \\ |\triangle|^{2-2r}, & \lceil m/d^* \rceil = \lceil m'/d^* \rceil, \end{cases}$$

$$\mathrm{E}\left|U_{kj}(X_{ik})B_m^*(\mathbf{S}_i)\right|^r \asymp |\triangle|^{2-r} H^{(2-r)/2},$$

where $d^* = (d+1)(d+2)/2$.

Proof. By Assumptions (A2) and (A5), $\|u_{kj}^0\| \asymp \|u_{kj}\|_{L_2} \asymp H^{1/2}$ and $\|B_m\| \asymp \|B_m\|_{L_2} \asymp |\triangle|$, which imply that $U_{kj} \asymp H^{-1/2}u_{kj}$ and $B_m^* \asymp |\triangle|^{-1}B_m$. Then, we have

$$\mathrm{E}|U_{kj}(X_{ik})U_{kj'}(X_{ik})|^r \asymp H^{-r}\mathrm{E}|u_{kj}(X_{ik})u_{kj'}(X_{ik})|^r \asymp \begin{cases} 0, & |j-j'| > \varrho+1, \\ H^{1-r}, & |j-j'| \leq \varrho+1, \end{cases}$$

$$\mathrm{E}\left|B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i)\right|^r \asymp |\triangle|^{-2r}\mathrm{E}\left|B_m(\mathbf{S}_i)B_{m'}(\mathbf{S}_i)\right|^r \asymp \begin{cases} 0, & \lceil m/d^* \rceil \neq \lceil m'/d^* \rceil, \\ |\triangle|^{2-2r}, & \lceil m/d^* \rceil = \lceil m'/d^* \rceil, \end{cases}$$

$$\mathrm{E}\left|U_{kj}(X_{ik})B_m^*(\mathbf{S}_i)\right|^r \asymp |\triangle|^{-r}H^{-r/2}\mathrm{E}\left|u_{kj}(X_{ik})B_m(\mathbf{S}_i)\right|^r \asymp |\triangle|^{2-r}H^{(2-r)/2}.$$

Thus, the desired results are established. ∎

Lemma B.2 *Under Assumption (A5), there exist positive constants c_k, C_k, c_s, C_s, such that, for $k = 1, \cdots, p$,*

$$c_k \sum_{j\in\mathcal{J}} \theta_{kj}^2 \leq \left\|\sum_{j\in\mathcal{J}} \theta_{kj}U_{kj}\right\|_{L_2}^2 \leq C_k \sum_{j\in\mathcal{J}} \theta_{kj}^2, \; c_s \sum_{m\in\mathcal{M}} \gamma_m^2 \leq \left\|\sum_{m\in\mathcal{M}} \gamma_m B_m^*\right\|_{L_2}^2 \leq C_s \sum_{m\in\mathcal{M}} \gamma_m^2.$$

Lemma B.3 *Under Assumptions (A2) and (A5), there exist positive constants c, C such that*

$$c\left(\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}^2+\sum_{m\in\mathcal{M}}\gamma_m^2\right) \leq \left\|\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}U_{kj}+\sum_{m\in\mathcal{M}}\gamma_m B_m^*\right\|^2 \leq C\left(\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}^2+\sum_{m\in\mathcal{M}}\gamma_m^2\right).$$

Proof. By Assumption (A2) and Lemma B.2, we have

$$\begin{aligned}\left\|\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}U_{kj}+\sum_{m\in\mathcal{M}}\gamma_m B_m^*\right\|^2 &\leq C_f\left\|\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}U_{kj}+\sum_{m\in\mathcal{M}}\gamma_m B_m^*\right\|_{L_2}^2 \\ &\leq C\left(\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}^2+\sum_{m\in\mathcal{M}}\gamma_m^2\right).\end{aligned}$$

Recall that $\mathrm{E}U_{kj}(X_k) = 0$, then

$$\left\|\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}U_{kj}+\sum_{m\in\mathcal{M}}\gamma_m B_m^*\right\|^2 = \left\|\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj}U_{kj}+\sum_{m\in\mathcal{M}}(\gamma_m B_m^*-\mu_m)\right\|^2 + \left(\sum_{m\in\mathcal{M}}\gamma_m\mu_m\right)^2,$$

where $\mu_m = \mathrm{E}B_m^*(\mathbf{S})$. According to Lemma 1 by Stone (1985), we have

$$\left\|\sum_{k=1}^p \sum_{j\in\mathcal{J}} \theta_{kj}U_{kj} + \sum_{m\in\mathcal{M}}(\gamma_m B_m^* - \mu_m)\right\|^2 \geq c_0 \left\{\sum_{k=1}^p \left\|\sum_{j\in\mathcal{J}} \theta_{kj}U_{kj}\right\|^2 + \left\|\sum_{m\in\mathcal{M}}(\gamma_m B_m^* - \mu_m)\right\|^2\right\}.$$

Therefore, by Lemma B.2,

$$\begin{aligned}\left\|\sum_{k=1}^p \sum_{j\in\mathcal{J}} \theta_{kj}U_{kj} + \sum_{m\in\mathcal{M}} \gamma_m B_m^*\right\|^2 &\geq c_1 \left\{\sum_{k=1}^p \left\|\sum_{j\in\mathcal{J}} \theta_{kj}U_{kj}\right\|^2 + \left\|\sum_{m\in\mathcal{M}} \gamma_m B_m^*\right\|^2\right\} \\ &\geq c\left(\sum_{k=1}^p \sum_{j\in\mathcal{J}} \theta_{kj}^2 + \sum_{m\in\mathcal{M}} \gamma_m^2\right).\end{aligned}$$

The lemma follows. ■

Lemma B.4 *For any $k = 1, \cdots, p$, $\phi_k \in \mathcal{H}^{(\varrho)} \cap \mathcal{D}_k^0$, there exist a constant c and a function $\phi_k^* \in \mathcal{U}_k^0$ such that*

$$\|\phi_k - \phi_k^*\|_\infty \leq c\|\phi_k^{(\varrho+1)}\|_\infty H^{\varrho+1}.$$

Proof. By the result on page 149 of De Boor (2001), there exists $\phi_k^{**} = \sum_{j\in\mathcal{J}} \theta_{kj}u_{kj}$ such that $\|\phi_k - \phi_k^{**}\|_\infty \leq c_0\|\phi_k^{(\varrho+1)}\|_\infty H^{\varrho+1}$. Let $C_{kj} = \|u_{kj} - \mathrm{E}u_{kj}(X_k)\|$, $u_{kj}^* = C_{kj}^{-1}\{u_{kj} - \mathrm{E}u_{kj}(X_k)\}$ and $\theta_{kj}^* = C_{kj}\theta_{kj}$ then we have

$$\begin{aligned}\left\|\phi_k - \sum_{j\in\mathcal{J}} \theta_{kj}^* u_{kj}^*\right\|_\infty &= \|\phi_k - \sum_{j\in\mathcal{J}} \theta_{kj}\{u_{kj} - \mathrm{E}u_{kj}(X_k)\}\|_\infty \leq \|\phi_k - \sum_{j\in\mathcal{J}} \theta_{kj}u_{kj}\|_\infty + |\sum_{j\in\mathcal{J}} \theta_{kj}\mathrm{E}u_{kj}(X_k)| \\ &\leq c_0\|\phi_k^{(\varrho+1)}\|_\infty H^{\varrho+1} + |\mathrm{E}\{\sum_{j\in\mathcal{J}} \theta_{kj}u_{kj}(X_k) - \phi_k(X_k)\}| + |\mathrm{E}\phi_k(X_k)| \\ &\leq 2c_0\|\phi_k^{(\varrho+1)}\|_\infty H^{\varrho+1}.\end{aligned}$$

Therefore, $\|\phi_k - \sum_{j\in\mathcal{J}} \theta_{kj}^* u_{kj}^*\|_\infty \leq C\|\phi_k^{(\varrho+1)}\|_\infty H^{\varrho+1}$. Lemma B.4 is established. ■

Lemma B.5 (Theorem 10.2, Lai and Schumaker (2007)) *Suppose that $|\triangle|$ is a π-quasi-uniform triangulation of a polygonal domian Ω, and $\psi(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$.*

(i) For bi-integer (a_1, a_2) with $0 \leq a_1 + a_2 \leq d$, there exists a spline $\psi^*(\cdot) \in \mathbb{S}_d^0(\triangle)$ such that $\|\nabla_{z_1}^{a_1}\nabla_{z_2}^{a_2}(\psi - \psi^*)\|_\infty \leq C|\triangle|^{d+1-a_1-a_2}|\psi|_{d+1,\infty}$, where C is a constant depending on d, and the shape parameter π.

(ii) For bi-integer (a_1, a_2) with $0 \leq a_1 + a_2 \leq d$, there exists a spline function $\psi^{**}(\cdot) \in \mathbb{S}_d^r(\triangle)$ ($d \geq 3r+2$) such that $\|\nabla_{z_1}^{a_1}\nabla_{z_2}^{a_2}(\psi - \psi^{**})\|_\infty \leq C|\triangle|^{d+1-a_1-a_2}|\psi|_{d+1,\infty}$, where C is a constant depending on d, r, and the shape parameter π.

Lemma B.5 shows that $\mathbb{S}_d^0(\triangle)$ has full approximation power, and $\mathbb{S}_d^r(\triangle)$ also has full approximation power if $d \geq 3r+2$.

Lemma B.6 *Suppose that Assumptions (A2), (A5) and (A6) hold. Then*

$$\max_{\substack{1\leq k,k'\leq p\\ j,j'\in\mathcal{J}}}\left|\frac{1}{n}\sum_{i=1}^n U_{kj}(X_{ik})U_{k'j'}(X_{ik'}) - \mathrm{E}\left\{U_{kj}(X_{ik})U_{k'j'}(X_{ik'})\right\}\right| = O_{\mathrm{a.s.}}(n^{-1/2}H^{-1/2}\log^{1/2} n), \tag{B.1}$$

$$\max_{m,m'\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^n B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i)\right\}\right| = O_{\mathrm{a.s.}}(n^{-1/2}|\triangle|^{-1}\log^{1/2} n), \tag{B.2}$$

$$\max_{m\in\mathcal{M},1\leq k\leq p,j\in\mathcal{J}}\left|\frac{1}{n}\sum_{i=1}^n B_m^*(\mathbf{S}_i)U_{kj}(X_{ik}) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)U_{kj}(X_{ik})\right\}\right| = O_{\mathrm{a.s.}}(n^{-1/2}\log^{1/2} n). \tag{B.3}$$

Proof. For simplicity, we consider the case $k = k'$. Let

$$U_{i,kjj'} = n^{-1}U_{kj}(X_{ik})U_{kj'}(X_{ik}) - n^{-1}\mathrm{E}\{U_{kj}(X_{ik})U_{kj'}(X_{ik})\},$$

$$u_{i,kjj'} = n^{-1}u_{kj}(X_{ik})u_{kj'}(X_{ik}) - n^{-1}\mathrm{E}\{u_{kj}(X_{ik})u_{kj'}(X_{ik})\}.$$

Then $U_{i,kjj'} = \|u_{kj}\|^{-1}\|u_{kj'}\|^{-1}u_{i,kjj'}$ and $nu_{i,kjj'} \leq c$, as the B-spline basis function is bounded by a constant. Notice that $\mathrm{E}U_{i,kjj'} = \mathrm{E}u_{i,kjj'} = 0$, and

$$\begin{aligned}\mathrm{E}|U_{i,kjj'}|^r &= \|u_{kj}\|^{-r}\|u_{kj'}\|^{-r}\mathrm{E}|u_{i,kjj'}|^r \leq (cn^{-1}\|u_{kj}\|^{-1}\|u_{kj'}\|^{-1})^{r-2}\mathrm{E}|U_{i,kjj'}|^2\\ &\leq (Cn^{-1}H^{-1})^{r-2}\mathrm{E}|U_{i,kjj'}|^2.\end{aligned}$$

Thus, $U_{i,kjj'}$ satisfies the Cramér's condition with constant $Cn^{-1}H^{-1}$. Applying Bernstein inequality to $\sum_{i=1}^n U_{i,kjj'}$, for any $\delta > 0$, one has

$$P\left\{\left|\sum_{i=1}^n U_{i,kjj'}\right| \geq \delta n^{-1/2}H^{-1/2}\log^{1/2} n\right\} \leq 2\exp\left\{\frac{-\delta^2 \log n}{4 + CH^{-1/2}n^{-1/2}\log^{1/2} n}\right\}.$$

Assume $|\mathcal{J}| \asymp n^\tau$ for some $0 < \tau < \infty$. Under Assumption (A6), we have

$$\sum_{n=1}^\infty P\left\{\max_{k,j,j'}\left|\sum_{i=1}^n U_{i,kjj'}\right| \geq \delta n^{-1/2}H^{-1/2}\log^{1/2} n\right\} \leq \sum_{n=1}^\infty \sum_{k=1}^p |\mathcal{J}|n^{-2-\tau} < \infty.$$

Thus,

$$\max_{k,j,j'}\left|\sum_{i=1}^n U_{i,kjj'}\right| = O_{\text{a.s.}}\{n^{-1/2}H^{-1/2}\log^{1/2} n\}.$$

Similarly, under Assumptions (A5) and (A6), we have

$$\max_{m,m'\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^n B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i)\right\}\right| = O_{\text{a.s.}}\left\{n^{-1/2}|\triangle|^{-1}\log^{1/2} n\right\},$$

$$\max_{m\in\mathcal{M},1\leq k\leq p,j\in\mathcal{J}}\left|\frac{1}{n}\sum_{i=1}^n B_m^*(\mathbf{S}_i)U_{kj}(X_{ik}) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)U_{kj}(X_{ik})\right\}\right| = O_{\text{a.s.}}\left\{n^{-1/2}\log^{1/2} n\right\}.$$

The desired results are obtained. ■

Lemma B.7 *Suppose that Assumptions (A2), (A5) and (A6) hold. Then, we have*

$$R_n = \sup_{\psi_1,\psi_2\in\mathcal{A}}\left|\frac{\langle\psi_1,\psi_2\rangle_n - \langle\psi_1,\psi_2\rangle}{\|\psi_1\|\|\psi_2\|}\right| = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\}.$$

Proof. Without loss of generality, let

$$\psi_1 = \sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj,1}U_{kj} + \sum_{m\in\mathcal{M}}\gamma_{m,1}B_m^*, \quad \psi_2 = \sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj,2}U_{kj} + \sum_{m\in\mathcal{M}}\gamma_{m,2}B_m^*.$$

By Lemma B.3, we have

$$\|\psi_1\|\|\psi_2\| \asymp \left(\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj,1}^2 + \sum_{m\in\mathcal{M}}\gamma_{m,1}^2\right)^{1/2}\left(\sum_{k=1}^p\sum_{j\in\mathcal{J}}\theta_{kj,2}^2 + \sum_{m\in\mathcal{M}}\gamma_{m,2}^2\right)^{1/2}.$$

Also, notice that

$$
\begin{aligned}
&\langle\psi_1,\psi_2\rangle_n - \langle\psi_1,\psi_2\rangle \\
&\quad = \sum_{k,k'=1}^{p}\sum_{j\in\mathcal{J},j'\in\mathcal{J}}\theta_{kj,1}\theta_{k'j',2}\left[\frac{1}{n}\sum_{i=1}^{n}U_{kj}(X_{ik})U_{k'j'}(X_{ik'}) - \mathrm{E}\left\{U_{kj}(X_{ik})U_{k'j'}(X_{ik'})\right\}\right] \\
&\qquad + \sum_{m,m'\in\mathcal{M}}\gamma_{m,1}\gamma_{m',2}\left[\frac{1}{n}\sum_{i=1}^{n}B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i)\right\}\right] \\
&\qquad + \sum_{k=1}^{p}\sum_{j\in\mathcal{J},m\in\mathcal{M}}(\theta_{kj,1}\gamma_{m,2}+\theta_{kj,2}\gamma_{m,1})\left[\frac{1}{n}\sum_{i=1}^{n}B_m^*(\mathbf{S}_i)U_{kj}(X_{ik}) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)U_{kj}(X_{ik})\right\}\right] \\
&\quad \le \sum_{k,k'=1}^{p}\sum_{|j-j'|\le\varrho+1}|\theta_{kj,1}\theta_{k'j',2}|\max_{1\le k,k'\le p,j,j'\in\mathcal{J}}\left|\frac{1}{n}\sum_{i=1}^{n}U_{kj}(X_{ik})U_{k'j'}(X_{ik'}) - \mathrm{E}\left\{U_{kj}(X_{ik})U_{k'j'}(X_{ik'})\right\}\right| \\
&\qquad + \sum_{|m-m'|\le(d+1)(d+2)/2}|\gamma_{m,1}\gamma_{m',2}|\max_{m,m'\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^{n}B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)B_{m'}^*(\mathbf{S}_i)\right\}\right| \\
&\qquad + \sum_{k=1}^{p}\sum_{j\in\mathcal{J},m\in\mathcal{M}}(|\theta_{kj,1}\gamma_{m,2}|+|\theta_{kj,2}\gamma_{m,1}|)\max_{j\in\mathcal{J},m\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^{n}B_m^*(\mathbf{S}_i)U_{kj}(X_{ik}) - \mathrm{E}\left\{B_m^*(\mathbf{S}_i)U_{kj}(X_{ik})\right\}\right| \\
&\quad = I_1 + I_2 + I_3.
\end{aligned}
$$

By Lemma B.6, we have

$$
\begin{aligned}
I_1 &= \left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}\theta_{kj,1}^2\right)^{1/2}\left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}\theta_{kj,2}^2\right)^{1/2}\times O_{\mathrm{a.s.}}(n^{-1/2}H^{-1/2}\log^{1/2}n), \\
I_2 &= \left(\sum_{m\in\mathcal{M}}\gamma_{m,1}^2\right)^{1/2}\left(\sum_{m\in\mathcal{M}}\gamma_{m,2}^2\right)^{1/2}\times O_{\mathrm{a.s.}}(n^{-1/2}|\triangle|^{-1}\log^{1/2}n).
\end{aligned}
$$

By the Cauchy Schwarz inequality, we have

$$
\begin{aligned}
&\sum_{k=1}^{p}\sum_{j\in\mathcal{J},m\in\mathcal{M}}(|\theta_{kj,1}\gamma_{m,2}|+|\theta_{kj,2}\gamma_{m,1}|) \\
&\quad = \left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}|\theta_{kj,1}|\right)\left(\sum_{m\in\mathcal{M}}|\gamma_{m,2}|\right) + \left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}|\theta_{kj,2}|\right)\left(\sum_{m\in\mathcal{M}}|\gamma_{m,1}|\right) \\
&\quad \le CH^{-1/2}|\triangle|^{-1}\left\{\left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}\theta_{kj,1}^2\right)^{1/2}\left(\sum_{m\in\mathcal{M}}\gamma_{m,2}^2\right)^{1/2} + \left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}\theta_{kj,2}^2\right)^{1/2}\left(\sum_{m\in\mathcal{M}}\gamma_{m,1}^2\right)^{1/2}\right\},
\end{aligned}
$$

which implies that

$$I_3 = \left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}\theta_{kj,1}^2\right)^{1/2}\left(\sum_{m\in\mathcal{M}}\gamma_{m,2}^2\right)^{1/2} \times O_{\text{a.s.}}(H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n)$$
$$+\left(\sum_{k=1}^{p}\sum_{j\in\mathcal{J}}\theta_{kj,2}^2\right)^{1/2}\left(\sum_{m\in\mathcal{M}}\gamma_{m,1}^2\right)^{1/2} \times O_{\text{a.s.}}(H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n).$$

Combining I_1, I_2 and I_3, we obtain the desired result. ■

As a direct result of Lemma B.7, we obtain that

$$\sup_{\psi\in\mathcal{A}}\left|\|\psi\|_n^2/\|\psi\|^2 - 1\right| = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\}.$$

Denote

$$\mathbf{\Gamma}_{n,\lambda} = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top \\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top + \frac{\lambda}{n}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\end{pmatrix},$$

where $\widetilde{\mathbf{B}}(\mathbf{S}_i) = \mathbf{Q}_2^\top\mathbf{B}^*(\mathbf{S}_i)$, $\mathbf{B}^*(\mathbf{S}_i) = \{B_m^*(\mathbf{S}_i), m\in\mathcal{M}\}$.

Lemma B.8 *Under Assumptions (A2), (A5) and (A6), there exist constants* $0 < c_\Gamma < C_\Gamma < \infty$*, such that*

$$c_\Gamma \le \lambda_{\min}(\mathbf{\Gamma}_{n,\lambda}) \le \lambda_{\max}(\mathbf{\Gamma}_{n,\lambda}) \le C_\Gamma,$$

almost surely, for large enough n*.*

Proof. It is easy to see that for any vector $(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})^\top$,

$$(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\mathbf{\Gamma}_{n,0}(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})^\top = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top \\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\end{pmatrix}(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})^\top$$
$$= \|\boldsymbol{g}_{\boldsymbol{\gamma}^*,\boldsymbol{\theta}}\|_n^2,$$

where $\boldsymbol{g}_{\boldsymbol{\gamma}^*,\boldsymbol{\theta}}(\mathbf{x},\mathbf{s}) = \boldsymbol{\theta}^\top \mathbf{U}(\mathbf{x}) + \boldsymbol{\gamma}^{*\top}\widetilde{\mathbf{B}}(\mathbf{s})$. By Lemma B.3 and Lemma B.6, we have

$$c(1-R_n)\|(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\|^2 \le (1-R_n)\|\boldsymbol{g}_{\boldsymbol{\gamma}^*,\boldsymbol{\theta}}\|^2 \le \|\boldsymbol{g}_{\boldsymbol{\gamma}^*,\boldsymbol{\theta}}\|_n^2,$$

$$\|\boldsymbol{g}_{\boldsymbol{\gamma}^*,\boldsymbol{\theta}}\|_n^2 \le (1+R_n)\|\boldsymbol{g}_{\boldsymbol{\gamma}^*,\boldsymbol{\theta}}\|^2 \le C(1+R_n)\|(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\|^2.$$

By Assumption (A6), $R_n \to 0$, as $n \to \infty$, therefore,

$$c\|(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\|^2 \le (\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\boldsymbol{\Gamma}_{n,0}(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})^\top \le C\|(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\|^2, \tag{B.4}$$

almost surely, for large enough n.

The (m,m')th element of $\mathbf{P}$ is $P_{m,m'} = \int \nabla^2_{s_1} B^*_m(\mathbf{s})\nabla^2_{s_1} B^*_{m'}(\mathbf{s}) + \nabla^2_{s_2} B^*_m(\mathbf{s})\nabla^2_{s_2} B^*_{m'}(\mathbf{s}) + 2\nabla^2_{s_1 s_2} B^*_m(\mathbf{s})\nabla^2_{s_1 s_2} B^*_{m'}(\mathbf{s}) d\mathbf{s}$. By Theorem 2.19 in Lai and Schumaker (2007), we have

$$P_{m,m'} \asymp \begin{cases} |\triangle|^{-4}, & \lceil m/d^* \rceil = \lceil m'/d^* \rceil, \\ 0, & \lceil m/d^* \rceil \ne \lceil m'/d^* \rceil. \end{cases} \tag{B.5}$$

Then, by the Assumption (A6), we have

$$(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top}) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\lambda}{n}\mathbf{Q}_2^\top \mathbf{P}\mathbf{Q}_2 \end{pmatrix} (\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})^\top = O\{\lambda|\triangle|^{-4}n^{-1}\|(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\|^2\} = o\{\|(\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^{*\top})\|^2\}. \tag{B.6}$$

The desired result follows (B.4) and (B.6). ■

B.1.2.2. Consistency of Penalized Quasi-likelihood Estimators

By Lemmas B.4 and B.5, there exist $\widetilde{\beta}_k(x_k) = \mathbf{U}_k^\top(x_k)\widetilde{\boldsymbol{\theta}}_k$ and $\widetilde{\alpha}(\mathbf{s}) = \widetilde{\mathbf{B}}^\top(\mathbf{s})\widetilde{\gamma}^*$, which are the best approximation to β_k's and α with the approximation rate at $\|\beta_k - \widetilde{\beta}_k\|_\infty \le C_k\|\beta_k^{(\varrho+1)}\|_\infty H^{\varrho+1}$ and $\|\alpha - \widetilde{\alpha}\|_\infty \le C_\alpha|\alpha|_{d+1,\infty}|\triangle|^{d+1}$. Denote $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^\top, \cdots, \widetilde{\boldsymbol{\theta}}_p^\top)^\top$.

Denote that $\eta_i(\boldsymbol{\theta}, \boldsymbol{\gamma}^*) = \mathbf{U}(\mathbf{X}_i)^\top\boldsymbol{\theta} + \widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\boldsymbol{\gamma}^*$, which is a function with respect to subject i.

We have

$$\nabla L^P(\boldsymbol{\theta},\boldsymbol{\gamma}^*) = -\sum_{i=1}^n q_1\left\{\eta_i(\boldsymbol{\theta},\boldsymbol{\gamma}^*),Y_i\right\}\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\end{pmatrix} + \lambda\begin{pmatrix}\mathbf{0}\\ \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\boldsymbol{\gamma}^*\end{pmatrix},$$

$$\nabla^2 L^P(\boldsymbol{\theta},\boldsymbol{\gamma}^*) = -\sum_{i=1}^n q_2(\eta_i(\boldsymbol{\theta},\boldsymbol{\gamma}^*),Y_i)\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\end{pmatrix} + \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \lambda\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\end{pmatrix}.$$

For the notation simplicity, we denote $\eta_i^0 = \sum_{k=1}^p \beta_k(X_{ik}) + \alpha(\mathbf{S}_i)$, $\widehat{\eta}_i = \eta_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\gamma}}^*)$ and $\widetilde{\eta}_i = \eta_i(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^*)$.

Lemma B.9 *Under Assumptions (A1)–(A6), we have*

$$\left|\frac{1}{n}\nabla L^P(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^*)\right| = O_{a.s}\left\{\left(\frac{\log n}{n}\right)^{1/2} + H^{\varrho+3/2} + |\triangle|^{d+2} + H^{\varrho+1}|\triangle| + H^{1/2}|\triangle|^{d+1} + \frac{\lambda}{n|\triangle|^3}\right\},$$

$$\left\|\frac{1}{n}\nabla L^P(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^*)\right\| = O_{a.s}\left\{(H^{-1/2}+|\triangle|^{-1})\left(\frac{\log n}{n}\right)^{1/2} + H^{\varrho+1} + |\triangle|^{d+1} + \frac{\lambda}{n|\triangle|^4}\right\}.$$

Proof. Let $\widetilde{\eta}(\mathbf{x},\mathbf{s}) = \mathbf{U}(\mathbf{x})^\top\widetilde{\boldsymbol{\theta}} + \widetilde{\mathbf{B}}(\mathbf{s})^\top\widetilde{\boldsymbol{\gamma}}^*$, then $\|\widetilde{\eta}-\eta\|_\infty = O(H^{\varrho+1}+|\triangle|^{d+1})$. By Assumption (A2), we have

$$\begin{aligned}
n^{-1}\nabla L^P(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^*) &= -\frac{1}{n}\sum_{i=1}^n q_1\left\{\widetilde{\eta}_i,Y_i\right\}\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\end{pmatrix} + \frac{\lambda}{n}\begin{pmatrix}\mathbf{0}\\ \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\widetilde{\boldsymbol{\gamma}}^*\end{pmatrix}\\
&= -\frac{1}{n}\sum_{i=1}^n \frac{Y_i - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\widetilde{\eta}_i))}\dot{g}^{-1}(\widetilde{\eta}_i)\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\end{pmatrix} + \frac{\lambda}{n}\begin{pmatrix}\mathbf{0}\\ \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\widetilde{\boldsymbol{\gamma}}^*\end{pmatrix}\\
&= -\frac{1}{n}\sum_{i=1}^n \frac{Y_i - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)\left\{1+o(1)\right\}\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\end{pmatrix} + \frac{\lambda}{n}\begin{pmatrix}\mathbf{0}\\ \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\widetilde{\boldsymbol{\gamma}}^*\end{pmatrix}\\
&= \mathbf{V}_v\left\{1+o(1)\right\} + \mathbf{V}_b\left\{1+o(1)\right\} + \mathbf{V}_p,
\end{aligned}$$

where

$$\mathbf{V}_v = -\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\end{pmatrix},$$

$$\mathbf{V}_b = -\frac{1}{n}\sum_{i=1}^{n}\frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\end{pmatrix},\ \mathbf{V}_p = \frac{\lambda}{n}\begin{pmatrix}\mathbf{0}\\ \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\widetilde{\gamma}^*\end{pmatrix}. \quad \text{(B.7)}$$

For the vector $\mathbf{V}_v$, we have

$$\mathrm{E}\left[\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})\right] = 0, \qquad \mathrm{E}\left[\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})\right]^2 = O(1).$$

According to Assumption (A3) and applying the Bernstein inequality, for any $k = 1, \cdots, p$ and $j \in \mathcal{J}$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik}) = O_{\text{a.s.}}\left\{\left(\frac{\log n}{n}\right)^{1/2}\right\}. \quad \text{(B.8)}$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)B_m^*(\mathbf{S}_i) = O_{\text{a.s.}}\left\{\left(\frac{\log n}{n}\right)^{1/2}\right\}. \quad \text{(B.9)}$$

For the vector $\mathbf{V}_b$, we focus on the term $n^{-1}\sum_{i=1}^{n}\frac{g^{-1}(\eta_i^0)-g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})$. We write

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik}) = \sum_{i=1}^{n}\xi_i + \mathrm{E}\left[\frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})\right],$$

where

$$\xi_i = \frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{n\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik}) - \mathrm{E}\left[\frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{n\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})\right].$$

Note that,

$$\mathrm{E}\left[\frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})\right] = O\left\{\left(H^{\varrho+1} + |\triangle|^{d+1}\right)H^{1/2}\right\}. \quad \text{(B.10)}$$

As we have, for any $r \geq 3$,

$$\mathrm{E}|\xi_i|^r \leq \left\{Cn^{-1}|g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)|\|U_{kj}\|_\infty\right\}^{r-2} \mathrm{E}|\xi_i|^2 \leq \left\{Cn^{-1}(H^{\varrho+1} + |\triangle|^{d+1})H^{-1/2}\right\}^{r-2} \mathrm{E}|\xi_i|^2,$$

$\{\xi_i\}_{i=1}^n$ satisfy the Cramér's condition with constant $Cn^{-1}(H^{\varrho+1} + |\triangle|^{d+1})H^{-1/2}$. Also,

$$\begin{aligned}\mathrm{E}|\xi_i|^2 =& \mathrm{E}\left|\frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{n\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})\right|^2 - \left[\mathrm{E}\left\{\frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{n\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik})\right\}\right]^2 \\ =& O\left\{n^{-2}(H^{2\varrho+2} + |\triangle|^{2d+2})\right\}.\end{aligned}$$

Applying the Bernstein inequality, we have

$$P\left\{\left|\sum_{i=1}^n \xi_i\right| \geq \delta(H^{\varrho+1} + |\triangle|^{d+1})n^{-1/2}\log^{1/2} n\right\} \leq 2\exp\left\{\frac{-\delta^2 \log n}{4 + CH^{-1/2}n^{-1/2}\log^{1/2} n}\right\}.$$

Consequently,

$$\sum_{i=1}^n \xi_i = O_{\text{a.s.}}\left\{(H^{\varrho+1} + |\triangle|^{d+1})n^{-1/2}\log^{1/2} n\right\}. \tag{B.11}$$

Combining (B.10) and (B.11), we obtain

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)U_{kj}(X_{ik}) \\ &= O_{\text{a.s.}}\left\{(H^{\varrho+1} + |\triangle|^{d+1})n^{-1/2}\log^{1/2} n + \left(H^{\varrho+1} + |\triangle|^{d+1}\right)H^{1/2}\right\}.\end{aligned} \tag{B.12}$$

Similarly,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \frac{g^{-1}(\eta_i^0) - g^{-1}(\widetilde{\eta}_i)}{\sigma^2 V(g^{-1}(\eta_i^0))}\dot{g}^{-1}(\eta_i^0)B_m^*(\mathbf{S}_i) \\ &= O_{\text{a.s.}}\left\{(H^{\varrho+1} + |\triangle|^{d+1})n^{-1/2}\log^{1/2} n + \left(H^{\varrho+1} + |\triangle|^{d+1}\right)|\triangle|\right\}.\end{aligned} \tag{B.13}$$

For the vector $\mathbf{V}_p$, by (B.5) and $|\widetilde{\gamma}^*| \asymp |\triangle|$, then we have

$$|\mathbf{V}_p| = O(\lambda n^{-1}|\triangle|^{-3}), \qquad \|\mathbf{V}_p\| = O(\lambda n^{-1}|\triangle|^{-4}). \tag{B.14}$$

Combining (B.8), (B.9), (B.12), (B.13) and (B.14), we obtain

$$\left|\frac{1}{n}\nabla L^P(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^*)\right| = O_{a.s}\left\{\left(\frac{\log n}{n}\right)^{1/2} + H^{\varrho+3/2} + |\triangle|^{d+2} + H^{\varrho+1}|\triangle| + H^{1/2}|\triangle|^{d+1} + \frac{\lambda}{n|\triangle|^3}\right\},$$

$$\left\|\frac{1}{n}\nabla L^P(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^*)\right\| = O_{a.s}\left\{(H^{-1/2} + |\triangle|^{-1})\left(\frac{\log n}{n}\right)^{1/2} + H^{\varrho+1} + |\triangle|^{d+1} + \frac{\lambda}{n|\triangle|^4}\right\}.$$

Therefore, Lemma B.9 has been established. ■

Lemma B.10 *If $(\bar{\boldsymbol{\theta}}^\top, \bar{\boldsymbol{\gamma}}^{*\top})^\top$ is the vector that satisfies $\left\|(\bar{\boldsymbol{\theta}}^\top, \bar{\boldsymbol{\gamma}}^{*\top})^\top - (\widetilde{\boldsymbol{\theta}}^\top, \widetilde{\boldsymbol{\gamma}}^{*\top})^\top\right\| = O_{\mathrm{a.s.}}(H^{1/2} + |\triangle|)$, then, under Assumptions (A2), (A3), (A5) and (A6), there exists constants c and C, such that*

$$c\mathbf{I} \le -\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\gamma}}^*)\right\}^{-1} \le C\mathbf{I},$$

almost surely, for large enough n.

Proof. Let $\bar{\eta}_i = \eta_i(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\gamma}}^*)$, we have

$$\begin{aligned}
&n^{-1}\nabla^2 L^P(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\gamma}}^*)\\
&= -\frac{1}{n}\sum_{i=1}^n q_2(\bar{\eta}_i, Y_i)\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\end{pmatrix} + \frac{\lambda}{n}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\\
&= -\frac{1}{n}\sum_{i=1}^n \left[\{Y_i - g^{-1}(\eta_i^0)\}\rho_1'\{g^{-1}(\bar{\eta}_i)\} + \{g^{-1}(\bar{\eta}_i) - g^{-1}(\eta_i^0)\}\rho_1'\{g^{-1}(\bar{\eta}_i)\} - \rho_2\{g^{-1}(\bar{\eta}_i)\}\right]\\
&\quad\times\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\end{pmatrix} + \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \frac{\lambda}{n}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\end{pmatrix}\\
&= -\frac{1}{n}\sum_{i=1}^n\left[\rho_2\{g^{-1}(\bar{\eta}_i)\}\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\end{pmatrix}\right]\{1 + o_{a.s}(1)\} + \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \frac{\lambda}{n}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\end{pmatrix}.
\end{aligned}$$

By the boundedness of $\rho_2\{g^{-1}(\bar{\eta}_i)\}$ and Lemma B.8, we have

$$c\mathbf{I} \le \left[\frac{1}{n}\sum_{i=1}^n \rho_2\{g^{-1}(\bar{\eta}_i)\}\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\end{pmatrix} + \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \frac{\lambda}{n}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\end{pmatrix}\right] \le C\mathbf{I},$$

almost surely, for large enough n. $\blacksquare$

Proof of Theorem 1. We now prove that

$$\left\|(\widehat{\boldsymbol{\theta}}^{\top},\widehat{\boldsymbol{\gamma}}^{*\top})-(\widetilde{\boldsymbol{\theta}}^{\top},\widetilde{\boldsymbol{\gamma}}^{*\top})\right\|=O_{\text{a.s.}}\left\{(H^{-1/2}+|\triangle|^{-1})\left(\frac{\log n}{n}\right)^{1/2}+H^{\varrho+1}+|\triangle|^{d+1}+\frac{\lambda}{n|\triangle|^{4}}\right\}. \tag{B.15}$$

As $(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\gamma}}^{*})$ is the minimizer of $L^{P}(\boldsymbol{\theta},\boldsymbol{\gamma})$, we have $\nabla L^{P}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\gamma}}^{*})=\mathbf{0}$. Then by the mean value theorem, we obtain

$$(\widehat{\boldsymbol{\theta}}^{\top},\widehat{\boldsymbol{\gamma}}^{*\top})^{\top}-(\widetilde{\boldsymbol{\theta}}^{\top},\widetilde{\boldsymbol{\gamma}}^{*\top})^{\top}=-\left\{\nabla^{2}L^{P}(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\gamma}}^{*})\right\}^{-1}\nabla L^{P}(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^{*}),$$

where $\left(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\gamma}}^{*}\right)$ is some value between $(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\gamma}}^{*})$ and $(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}}^{*})$. Then (B.15) is obtained from Lemmas B.9 and B.10. Theorem 1 is obtained from Lemma B.2 and Assumption (A4). $\blacksquare$

For any $\phi\in\mathcal{U}_{k}$, $\psi\in\mathbb{S}_{d}^{r}$, , one has $\|\phi\|_{\infty}\leq CH^{-1/2}\|\phi\|$, $\|\psi\|_{\infty}\leq C|\triangle|^{-1}\|\psi\|$. Then

$$\sum_{k=1}^{p}\|\widehat{\beta}_{k}-\widetilde{\beta}_{k}\|_{\infty}+\|\widehat{\alpha}-\widetilde{\alpha}\|_{\infty}=O_{\text{a.s.}}\left\{(H^{-1}+|\triangle|^{-2})\left(\frac{\log n}{n}\right)^{1/2}+H^{\varrho+1/2}+|\triangle|^{d}+\frac{\lambda}{n}|\triangle|^{-5}\right\}.$$

Notice that $\|\widehat{\beta}_{k}-\beta_{k}\|_{\infty}\leq\|\widehat{\beta}_{k}-\widetilde{\beta}_{k}\|_{\infty}+\|\widetilde{\beta}_{k}-\beta_{k}\|_{\infty}$, for $k=1,\cdots,p$, and $\|\widehat{\alpha}-\alpha\|_{\infty}\leq\|\widehat{\alpha}-\widetilde{\alpha}\|_{\infty}+\|\widetilde{\alpha}-\alpha\|_{\infty}$. Consequently,

$$\sum_{k=1}^{p}\|\widehat{\beta}_{k}-\beta_{k}\|_{\infty}+\|\widehat{\alpha}-\alpha\|_{\infty}=O_{\text{a.s.}}\left\{(H^{-1}+|\triangle|^{-2})\left(\frac{\log n}{n}\right)^{1/2}+H^{\varrho+1/2}+|\triangle|^{d}+\frac{\lambda}{n|\triangle|^{5}}\right\}.$$

B.1.3. Properties of Local Polynomial Estimators

Denote that

$$\boldsymbol{\omega}(x_{k})=\left\{\boldsymbol{\omega}_{U}(x_{k})^{\top},\boldsymbol{\omega}_{B}(x_{k})^{\top}\right\}^{\top}=\left[\left\{\omega_{U,k'j}(x_{k})\right\}_{k'\neq k,j\in\mathcal{J}},\left\{\omega_{B,m}(x_{k})\right\}_{m\in\mathcal{M}}\right]^{\top},$$

where

$$\omega_{U,k'j}(x_{k})=n^{-1}\sum_{i=1}^{n}U_{k'j}(X_{ik'})K_{h_{k}}(X_{ik}-x_{k}),\quad \omega_{B,m}(x_{k})=n^{-1}\sum_{i=1}^{n}B_{m}^{*}(\mathbf{S}_{i})K_{h_{k}}(X_{ik}-x_{k}).$$

Similarly, we define

$$\boldsymbol{\omega}^{\varepsilon}(x_k)=\left\{\boldsymbol{\omega}_U^{\varepsilon}(x_k)^{\top},\boldsymbol{\omega}_B^{\varepsilon}(x_k)^{\top}\right\}^{\top}=\left[\left\{\omega_{U,k'j}^{\varepsilon}(x_k)\right\}_{k'\neq k,j\in\mathcal{J}},\left\{\omega_{B,m}^{\varepsilon}(x_k)\right\}_{m\in\mathcal{M}}\right]^{\top}.$$

Recall that $\varepsilon_i=Y_i-g^{-1}(\eta_i^0)$. Then

$$\omega_{U,k'j}^{\varepsilon}(x_k)=n^{-1}\sum_{i=1}^{n}\varepsilon_i U_{k'j}(X_{ik'})K_{h_k}(X_{ik}-x_k),\quad \omega_{B,m}^{\varepsilon}(x_k)=n^{-1}\sum_{i=1}^{n}\varepsilon_i B_m^*(\mathbf{S}_i)K_{h_k}(X_{ik}-x_k).$$

Lemma B.11 *Under Assumptions (A2), (A4), (A5), (A7) and (A8), as $n\to\infty$,*

$$\sup_{x_k\in[0,1],k'\neq k,j\in\mathcal{J}}|\omega_{U,k'j}(x_k)-\mathrm{E}\omega_{U,k'j}(x_k)|=O_{\mathrm{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2}n), \tag{B.16}$$

$$\sup_{x_k\in[0,1],m\in\mathcal{M}}|\omega_{B,m}(x_k)-\mathrm{E}\omega_{B,m}(x_k)|=O_{\mathrm{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2}n), \tag{B.17}$$

$$\sup_{x_k\in[0,1]}|\boldsymbol{\omega}_U(x_k)|=O_{\mathrm{a.s.}}(H^{1/2}), \tag{B.18}$$

$$\sup_{x_k\in[0,1]}|\boldsymbol{\omega}_B(x_k)|=O_{\mathrm{a.s.}}(|\triangle|), \tag{B.19}$$

$$\sup_{x_k\in[0,1]}|\boldsymbol{\omega}^{\varepsilon}(x_k)|=O_{\mathrm{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2}n). \tag{B.20}$$

Proof. First, we compute

$$\begin{aligned}\mathrm{E}\{\omega_{U,k'j}(x_k)\}&=\int\int U_{k'j}(u_{k'})K_{h_k}(u_k-x_k)f_{kk'}(u_k,u_{k'})du_kdu_{k'}\\&=\int\int U_{k'j}(u_{k'})K(v_k)f_{kk'}(x_k+h_kv_k,u_{k'})dv_kdu_{k'},\end{aligned}$$

where $f_{kk'}(u_k,u_{k'})$ is the joint density of $(X_k,X_{k'})$. By Assumption (A1), $f_{kk'}(u_k,u_{k'})$ is bounded below and above. Therefore, we have $\mathrm{E}\{\omega_{U,k'j}(x_k)\}=O(H^{1/2})$. Similarly, $\mathrm{E}|\omega_{U,k'j}(x_k)|^r=O(h_k^{1-r}H^{1-r/2})$. By the Bernstein inequality, we obtain (B.16). Note that

$$\|\omega_{U,k'j}(x_k)\|_\infty\leq\|\omega_{U,k'j}(x_k)-\mathrm{E}\{\omega_{U,k'j}(x_k)\}\|_\infty+\|\mathrm{E}\{\omega_{U,k'j}(x_k)\}\|_\infty,$$

therefore, (B.18) follows from (B.16). Similar arguments yield (B.17) and (B.19).

Next, we prove (B.20). Notice that $\mathrm{E}[\omega^{\varepsilon}_{U,k'j}(x_k)] = 0$ and $\mathrm{E}[\omega^{\varepsilon}_{U,k'j}(x_k)]^2 \asymp n^{-1}h_k^{-1}$.

Let $\varpi_i = \varepsilon_i U_{k'j}(X_{ik'})K_{h_k}(X_{ik} - x_k)$. We decompose the random variable ε_i into a tail part and a truncated part, $\varepsilon^{D_n}_{i,1} = \varepsilon_i I(|\varepsilon_i| > D_n)$, $\varepsilon^{D_n}_{i,2} = \varepsilon_i I(|\varepsilon_i| \le D_n) - \mu_i^{D_n}$ and $\mu_i^{D_n} = \mathrm{E}\{\varepsilon_i I(|\varepsilon_i| \le D_n)\}$, where $D_n = n^a$, $\frac{1}{2+\iota} < a < \frac{3-5\delta}{10}$. First, we show that tail part vanishes almost surely. Note that

$$\sum_{n=1}^{\infty} P\left\{|\varepsilon_i| > D_n\right\} \le \sum_{n=1}^{\infty} \frac{E\left|\varepsilon_i\right|^{2+\iota}}{D_n^{2+\iota}} \le C \sum_{n=1}^{\infty} D_n^{-(2+\iota)} < \infty.$$

By the Borel-Cantelli's lemma, we have $\left|\frac{1}{n}\sum_{i=1}^{n} U_{k'j}(X_{ik'})K_{h_k}(X_{ik} - x_k)\varepsilon^{D_n}_{i,1}\right| = O_{a.s.}\left(n^{-r}\right)$, for any $r > 0$. Also, it is straightforward to verify that $\mathrm{E}\{U_{k'j}(X_{ik'})K_{h_k}(X_{ik} - x_k)\varepsilon_i I(|\varepsilon_i| \le D_n)\} = \mathrm{E}\{U_{k'j}(X_{ik'})K_{h_k}(X_{ik} - x_k)\}\mu_i^{D_n} = o\{D_n^{-(1+\iota)}\}$. Note that $D_n^{-(1+\iota)}(nh_k \log n)^{1/2} \to 0$ by Assumption (A8), which implies that the truncated mean part is negligible compared to $(nh_k \log n)^{-1/2}$.

Next, notice that $\mathrm{Var}\left(\varpi^{D_n}_{i,2}\right) = E\left(\varpi_i^2\right) - E\left(\varpi^{D_n}_{i,1}\right)^2 - \left(\mu^{D_n}\right)^2 \asymp n^{-1}h_k^{-1}$. In addition, for any $r \ge 3$, $\mathrm{E}|\varpi^{D_n}_{i,2}|^r = (CD_nH^{-1/2}h_k^{-1})^{r-2}E|\varpi^{D_n}_{i,2}|^2$. By the Bernstein inequality,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\varpi^{D_n}_{i,2}\right| \ge \delta n^{-1/2}h_k^{-1/2}(\log n)^{1/2}\right\} \le 2\exp\left\{\frac{-\delta^2 \log n}{4c + 2\delta D_n H^{-1/2}n^{1/2}h_k^{1/2}\log^{1/2} n}\right\}.$$

Hence, for any large enough $\delta > 0$,

$$P\left\{\max_{k'=1,\ldots,p,j\in\mathcal{J}} \left|\omega^{\varepsilon}_{U,k'j}(x_k)\right| \ge \delta n^{-1/2}h_k^{-1/2}(\log n)^{1/2}\right\} \le CH^{-1}n^{-2/5} < \infty.$$

Similarly, we have

$$P\left\{\max_{m\in\mathcal{M}} \left|\omega^{\varepsilon}_{B,m}(x_k)\right| \ge \delta n^{-1/2}h_k^{-1/2}(\log n)^{1/2}\right\} \le C|\triangle|^{-2}n^{-2/5} < \infty.$$

Therefore, (B.20) is established. ■

Denote that

$$\boldsymbol{\tau}(x_k) = \left\{\boldsymbol{\tau}_U(x_k)^\top, \boldsymbol{\tau}_B(x_k)^\top\right\}^\top = \left[\left\{\tau_{U,k'j}(x_k)\right\}_{k'\neq k, j\in\mathcal{J}}, \left\{\tau_{B,m}(x_k)\right\}_{m\in\mathcal{M}}\right]^\top,$$

where

$$\tau_{U,k'j}(x_k) = n^{-1}h_k^{-1}\sum_{i=1}^{n} U_{k'j}(X_{ik'})(X_{ik}-x_k)K_{h_k}(X_{ik}-x_k),$$

$$\tau_{B,m}(x_k) = n^{-1}h_k^{-1}\sum_{i=1}^{n} B_m^*(\mathbf{S}_i)(X_{ik}-x_k)K_{h_k}(X_{ik}-x_k).$$

Similarly, we define

$$\boldsymbol{\tau}^{\varepsilon}(x_k) = \left\{\boldsymbol{\tau}_U^{\varepsilon}(x_k)^\top, \boldsymbol{\tau}_B^{\varepsilon}(x_k)^\top\right\}^\top = \left[\left\{\tau_{U,k'j}^{\varepsilon}(x_k)\right\}_{k'\neq k, j\in\mathcal{J}}, \left\{\tau_{B,m}^{\varepsilon}(x_k)\right\}_{m\in\mathcal{M}}\right]^\top,$$

where

$$\tau_{U,k'j}^{\varepsilon}(x_k) = n^{-1}h_k^{-1}\sum_{i=1}^{n} \varepsilon_i U_{k'j}(X_{ik'})(X_{ik}-x_k)K_{h_k}(X_{ik}-x_k),$$

$$\tau_{B,m}^{\varepsilon}(x_k) = n^{-1}h_k^{-1}\sum_{i=1}^{n} \varepsilon_i B_m^*(\mathbf{S}_i)(X_{ik}-x_k)K_{h_k}(X_{ik}-x_k).$$

Lemma B.12 *Under Assumptions (A2) – (A5), (A7) and (A8) as $n\to\infty$,*

$$\sup_{x_k\in[0,1],k'\neq k,j\in\mathcal{J}} |\tau_{U,k'j}(x_k) - \mathrm{E}\tau_{U,k'j}(x_k)| = O_{\text{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2} n),$$

$$\sup_{x_k\in[0,1],m\in\mathcal{M}} |\tau_{B,m}(x_k) - \mathrm{E}\tau_{B,m}(x_k)| = O_{\text{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2} n),$$

$$\sup_{x_k\in[0,1]} |\boldsymbol{\tau}_U(x_k)| = O_{\text{a.s.}}(H^{1/2}), \quad \sup_{x_k\in[0,1]} |\boldsymbol{\tau}_B(x_k)| = O_{\text{a.s.}}(|\triangle|),$$

$$\sup_{x_k\in[0,1]} |\boldsymbol{\tau}^{\varepsilon}(x_k)| = O_{\text{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2} n).$$

Proof. The proof follows from similar arguments as in the proof of Lemma B.11. ■

Lemma B.13 *Under Assumptions (A2)–(A6), for any $\boldsymbol{u}$, we have*

$$\left\|\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1}\boldsymbol{u} - \mathbf{D}^{-1}\boldsymbol{u}\right\| = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\}\|\boldsymbol{u}\|,$$

where

$$\mathbf{D} = \mathrm{E}\left[q_2(\bar{\eta}_i, Y_i)\begin{pmatrix}\mathbf{U}(\mathbf{X}_i)\mathbf{U}(\mathbf{X}_i)^\top & \mathbf{U}(\mathbf{X}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top \\ \widetilde{\mathbf{B}}(\mathbf{S}_i)\mathbf{U}(\mathbf{X}_i)^\top & \widetilde{\mathbf{B}}(\mathbf{S}_i)\widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\end{pmatrix} + \begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\lambda}{n}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\end{pmatrix}\right].$$

Proof. Let $\mathbf{A} = \frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*) - \mathbf{D}$. Observe that $\lambda_{\max}(\mathbf{A}^\top\mathbf{A}) = \max\{\lambda^2_{\max}(\mathbf{A}), \lambda^2_{\min}(\mathbf{A})\}$, then we have

$$\|\mathbf{A}\boldsymbol{u}\| = \sqrt{\boldsymbol{u}^\top\mathbf{A}^\top\mathbf{A}\boldsymbol{u}} \le \max\{|\lambda_{\max}(\mathbf{A})|, |\lambda_{\min}(\mathbf{A})|\}\|\boldsymbol{u}\|.$$

Similar to the proof of Lemma B.7, one can show that

$$\boldsymbol{u}^\top\mathbf{A}\boldsymbol{u} = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\}\|\boldsymbol{u}\|^2,$$

which implies

$$\max\{|\lambda_{\max}(\mathbf{A})|, |\lambda_{\min}(\mathbf{A})|\} = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\}.$$

Therefore, we have $\|\mathbf{A}\boldsymbol{u}\| = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\}\|\boldsymbol{u}\|$. Notice that

$$-\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\left[\{n^{-1}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\}^{-1} - \mathbf{D}^{-1}\right]\boldsymbol{u} = \mathbf{D}^{-1}\mathbf{A}\boldsymbol{u}.$$

Then, it is clear that

$$\left\|\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1}\boldsymbol{u} - \mathbf{D}^{-1}\boldsymbol{u}\right\| = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\}\|\boldsymbol{u}\|.$$

The desired result is obtained. ∎

B.1.3.1. Properties of Oracle Estimator

Proof of Theorem 2. First of all, we derive the pointwise asymptotically normality of the "oracle" estimator $\widehat{\beta}^o_k(x_k)$.

Let $\bar{\beta}_k(x_k, u) = \beta_k(x_k) + \beta_k'(x_k)(u - x_k)$,

$$\mathbf{a}^* = \sqrt{nh_k}\begin{pmatrix} a_0 - \beta_k(x_k) \\ \\ h_k\{a_1 - \beta_k'(x_k)\} \end{pmatrix}, \ \widehat{\mathbf{a}}^o = \sqrt{nh_k}\begin{pmatrix} \widehat{\beta}_k^o(x_k) - \beta_k(x_k) \\ \\ h_k\{\widehat{\beta}_k^{o\prime}(x_k) - \beta_k'(x_k)\} \end{pmatrix}, \ \mathbf{Z}_{ik} = \begin{pmatrix} 1 \\ \\ (X_{ik} - x_k)/h_k \end{pmatrix}.$$

Then, we have $a_0 + a_1(X_{ik} - x_k) = \bar{\beta}_k(x_k, X_{ik}) + nh_k\mathbf{a}^{*\top}\mathbf{Z}_{ik}$. It is easy to see that $\widehat{\mathbf{a}}^o$ is the maximizer of the following function

$$\begin{aligned}\ell_n(\mathbf{a}^*) =& \sum_{i=1}^n \ell\left[g^{-1}\{\bar{\beta}_k(x_k, X_{ik}) + \eta_{i,-k}^0 + (nh_k)^{-1/2}\mathbf{a}^{*\top}\mathbf{Z}_{ik}\}, Y_i\right] K\{(X_{ik} - x_k)/h_k\} \\ &- \sum_{i=1}^n \ell\left[g^{-1}\{\bar{\beta}_k(x_k, X_{ik}) + \eta_{i,-k}^0\}, Y_i\right] K\{(X_{ik} - x_k)/h_k\}.\end{aligned}$$

By Taylor expansion, we have

$$\begin{aligned}\ell_n(\mathbf{a}^*) =& (nh_k)^{-1/2}\sum_{i=1}^n q_1\{\bar{\beta}_k(x_k, X_{ik}) + \eta_{i,-k}^0, Y_i\}\mathbf{a}^{*\top}\mathbf{Z}_{ik}K\{(X_{ik} - x_k)/h_k\} \\ &+ \frac{1}{2nh_k}\sum_{i=1}^n q_2\{\bar{\beta}_k(x_k, X_{ik}) + \eta_{i,-k}^0, Y_i\}(\mathbf{a}^{*\top}\mathbf{Z}_{ik})^2K\{(X_{ik} - x_k)/h_k\} \\ &+ \frac{(nh_k)^{-3/2}}{6}\sum_{i=1}^n q_3\{\widetilde{\beta}_k(x_k, X_{ik}) + \eta_{i,-k}^0, Y_i\}(\mathbf{a}^{*\top}\mathbf{Z}_{ik})^3K\{(X_{ik} - x_k)/h_k\},\end{aligned}$$

where $\widetilde{\beta}_k(x_k, X_{ik})$ is between $\bar{\beta}_k(x_k, X_{ik})$ and $\bar{\beta}_k(x_k, X_{ik}) + (nh_k)^{-1/2}\mathbf{a}^{*\top}\mathbf{Z}_{ik}$ and the last term is $O_P(n^{-1/2}h_k^{-1/2})$. Moreover, if $nh_k^5 = O(1)$, we obtain

$$\ell_n(\mathbf{a}^*) = \mathbf{a}^{*\top}\mathbf{W}_n + \frac{1}{2}\mathbf{a}^{*\top}\mathbf{A}_n\mathbf{a}^* + o_P(h_k),$$

where

$$\mathbf{W}_n = (nh_k)^{-1/2}\sum_{i=1}^n q_1\left\{\bar{\beta}_k(x_k, X_{ik}) + \eta_{i,-k}^0, Y_i\right\} K\left\{(X_{ik} - x_k)/h_k\right\}\mathbf{Z}_{ik},$$

and

$$\mathbf{A}_n = \frac{1}{nh_k}\sum_{i=1}^n q_2\left\{\bar{\beta}_k(x_k, X_{ik}) + \eta_{i,-k}^0, Y_i\right\}\mathbf{Z}_{ik}^\top\mathbf{Z}_{ik}K\{(X_{ik} - x_k)/h_k\}.$$

Now we have $(\mathbf{A}_n)_{ij} = (\mathrm{E}\mathbf{A}_n)_{ij} + O_P\left[\{\mathrm{Var}(\mathbf{A}_n)_{ij}\}^{1/2}\right]$, and

$$\begin{aligned}\mathrm{E}(\mathbf{A}_n) &= h_k^{-1}\mathrm{E}\left[q_2\left\{\bar{\beta}_k(x_k, X_{ik}) + \eta^0_{i,-k}, Y_i\right\}\mathbf{Z}_{ik}^\top\mathbf{Z}_{ik}K\{(X_{ik}-x_k)/h_k\}\right] \\ &= h_k^{-1}\mathrm{E}\left[q_2\left\{\bar{\beta}_k(x_k, X_{ik}) + \eta^0_{i,-k}, \mathrm{E}(Y_i)\right\}\mathbf{Z}_{ik}^\top\mathbf{Z}_{ik}K\{(X_{ik}-x_k)/h_k\}\right].\end{aligned}$$

By Taylor expansion of q_2 around η_i^0 and $q_2\left\{\eta_i^0, \mathrm{E}(Y_i)\right\} = \rho_2(\eta_i^0)$, we obtain

$$(i-1)!(j-1)!(\mathrm{E}\mathbf{A}_n)_{ij} = -\rho_{2,k}(x_k)f_k(x_k)\nu_{i+j-2} - h(\rho_{2,k}f_k)'(x_k)\nu_{i+j-1} + o(h_k).$$

Therefore, $\ell_n(\mathbf{a}^*) = \mathbf{a}^{*\top}\mathbf{W}_n - \frac{1}{2}\mathbf{a}^{*\top}\{\mathbf{\Sigma}_k(x_k) + h_k\mathbf{\Lambda}_k(x_k)\}\mathbf{a}^* + o_P(h_k)$. By the Quadratic Approximation Lemma in Fan et al. (1995),

$$\widehat{\boldsymbol{a}}^o = \mathbf{\Sigma}_k(x_k)^{-1}\mathbf{W}_n - h_k\mathbf{\Sigma}_k(x_k)^{-1}\mathbf{\Lambda}_k(x_k)\mathbf{\Sigma}_k(x_k)^{-1}\mathbf{W}_n + o_P(h_k). \quad \text{(B.21)}$$

Next, we derive the asymptotic distribution of $\mathbf{W}_n$. Let $\mathbf{Y}_i^* = q_1\{\bar{\beta}_k(x_k, X_{ik}) + \eta^0_{i,-k}, Y_i\}K\{(X_{ik} - x_k)/h_k\}\mathbf{Z}_{ik}$, then $\mathbf{W}_n = (nh_k)^{-1/2}\sum_{i=1}^n \mathbf{Y}_i^*$. Therefore, $\mathrm{E}(\mathbf{W}_n) = n^{1/2}h_k^{-1/2}\mathrm{E}(\mathbf{Y}_1^*)$, $\mathrm{Cov}(\mathbf{W}_n) = h_k^{-1}\mathrm{Cov}(\mathbf{Y}_1^*)$. By Taylor expansion of q_1 around η_1^0, the jth component of $\mathbf{Y}_1^*$ satisfies

$$\begin{aligned}&(j-1)!\{\mathrm{E}(\mathbf{Y}_1^*)\}_j \\ =&\mathrm{E}\left[q_1\{\bar{\beta}_k(x_k, X_{1k}) + \eta^0_{1,-k}, Y_1\}\{(X_{1k}-x_k)/h_k\}^{j-1}K\{(X_{1k}-x_k)/h_k\}\right] \\ =&h_k\int\left\{\frac{\beta^{(2)}(x_k)}{2!}(uh_k)^2 + \frac{\beta^{(3)}(x_k)}{3!}(uh_k)^3 + o\{(uh_k)^4\}\right\}\rho_{2,k}(x_k+uh_k)f_k(x_k+uh_k)u^{j-1}K(u)du \\ =&h_k^3\frac{\beta^{(2)}(x_k)\rho_{2,k}(x_k)f_k(x_k)}{2!}\nu_{1+j} + h_k^4\frac{\beta^{(3)}(x_k)\rho_{2,k}(x_k)f_k(x_k)}{3!}\nu_{2+j} \\ &+ h_k^4\frac{\beta^{(2)}(x_k)(\rho_{2,k}f_k)'(x_k)}{2!}\nu_{2+j} + o(h_k^4). \quad \text{(B.22)}\end{aligned}$$

The covariance between the ith and jth component of $\mathbf{Y}_1^*$ is $\mathrm{E}\left\{(\mathbf{Y}_1^*)_i(\mathbf{Y}_1^*)_j\right\} + O(h_k^6)$. By Taylor expansion, we have

$$
\begin{aligned}
&(i-1)!(j-1)!\mathrm{E}\left\{(\mathbf{Y}_1^*)_i(\mathbf{Y}_1^*)_j\right\}\\
&=\mathrm{E}\left[q_1^2(\eta_1^0,Y_1)K^2\{(X_{1k}-x_k)/h_k\}\left\{(X_{1k}-x_k)/h_k\right\}^{i+j-2}\right]+O(h_k^2)\\
&=h_k\rho_{2,k}(x_k)f_k(x_k)\int u^{i+j-2}K(u)^2du.
\end{aligned}
$$

Therefore, $\mathrm{Cov}(\mathbf{W}_n)=\Xi_k(x_k)$. It is easy to verify that $\mathrm{Var}(\boldsymbol{u}^\top\mathbf{W}_n)=\boldsymbol{u}^\top\Xi_k(x_k)\boldsymbol{u}$. Then the Central Limit Theorem implies that $\boldsymbol{u}^\top\mathbf{W}_n-\mathrm{E}(\boldsymbol{u}^\top\mathbf{W}_n)\to N(0,\boldsymbol{u}^\top\Xi_k(x_k)\boldsymbol{u})$. By the Cramer-Wold device, we have $\Xi_k(x_k)^{-1/2}\{\mathbf{W}_n-\mathrm{E}(\mathbf{W}_n)\}\to N(0,\mathbf{I}_2)$. Then by (B.21), we obtain the first part of Theorem 2.

Secondly, we derive the uniform asymptotically normality of the "oracle" estimator $\widehat{\beta}_k^o(x_k)$ in the following. Define $M_{h_k}(t)=h_k^{-1/2}\int K\{(x-t)/h_k\}dW(x)$, where $W(x)$ is a Wiener process defined on $(0,\infty)$. By the Lemma 1 in Zheng et al. (2016), one has

$$
\lim_{n\to\infty}\mathbf{P}\left[(-2\log h_k)^{1/2}\left\{\sup_{t\in[h_k,1-h_k]}|M_{h_k}(t)|/\|K\|_{L_2}^2-d_{h_k}\right\}<x\right]=e^{-2e^{-x}}. \tag{B.23}
$$

where $d_{h_k}=(-2\log h_k)^{1/2}+(-2\log h_k)^{-1/2}\log\{\sqrt{C(K)}/(2\pi)\}$. Define the stochastic process $\widehat{\xi}_n(x_k)=n^{-1}\sum_{i=1}^n K_{h_k}(X_{ik}-x_k)\varepsilon_i,\ x_k\in[0,1]$, and $\widehat{\zeta}_n(x_k)=\sigma_{n,k}^{-1}(x_k)\widehat{\xi}_n(x_k)$. According to (B.21), we have

$$
\begin{aligned}
\widehat{\beta}_k^o(x_k)-\beta_k(x_k)=&\{\rho_{2,k}f_k(x_k)\}^{-1}\left[\frac{1}{n}\sum_{i=1}^n q_1\{\bar{\beta}_k(x_k,X_{ik})+\eta_{i,-k}^0,Y_i\}K_{h_k}(X_{ik}-x_k)\right.\\
&\left.+\frac{1}{n}\sum_{i=1}^n q_1\{\bar{\beta}_k(x_k,X_{ik})+\eta_{i,-k}^0,Y_i\}(X_{ik}-x_k)K_{h_k}(X_{ik}-x_k)\right].
\end{aligned}
$$

Notice that similar to the calculation in (B.22),

$$
\frac{1}{n}\sum_{i=1}^n q_1\{\bar{\beta}_k(x_k,X_{ik})+\eta_{i,-k}^0,Y_i\}K_{h_k}(X_{ik}-x_k)-\widehat{\xi}_n(x_k)=O_{\mathrm{a.s.}}(h_k^2+h_k^{3/2}n^{-1/2}\log^{1/2}n).
$$

Notice that

$$\frac{1}{n}\sum_{i=1}^{n} q_1\{\bar{\beta}_k(x_k, X_{ik}) + \eta^0_{i,-k}, Y_i\}(X_{ik} - x_k)K_{h_k}(X_{ik} - x_k) = O_{\text{a.s.}}(n^{-1/2}h_k^{1/2}\log^{1/2} n).$$

Therefore, $\sup_{x\in[h_k,1-h_k]} |\widehat{\beta}^o_k(x_k) - \beta_k(x_k) - \widehat{\xi}_n(x_k)| = O_{\text{a.s.}}(h_k^2 + n^{-1/2}h_k^{1/2}\log^{1/2} n)$. According to the proof of Theorem 1 in Zheng et al. (2016), we have

$$\sup_{x_k\in[h_k,1-h_k]} \left|\widehat{\zeta}_n(x_k) - M_{h_k}(x_k)/\|K\|^2_{L_2}\right| = o_P(\log^{-1/2} n). \tag{B.24}$$

Therefore, by (B.24), one has

$$\begin{aligned}
&\sup_{x_k\in[h_k,1-h_k]} \left|\sigma^{-1}_{n,k}(x_k)(\widehat{\beta}^{\text{o}}_k - \beta_k)(x_k) - M_{h_k}(x_k)/\|K\|_{L_2}\right| \\
&\le \sup_{x_k\in[h_k,1-h_k]} \left|\sigma^{-1}_{n,k}(x_k)(\widehat{\beta}^{\text{o}}_k - \beta_k)(x_k) - \widehat{\zeta}_n(x_k)\right| + \sup_{x_k\in[h_k,1-h_k]} \left|\widehat{\zeta}_n(x_k) - M_{h_k}(x_k)/\|K\|_{L_2}\right| \\
&= O_{\text{a.s.}}(n^{1/2}h_k^{5/2} + h_k^{1/2}\log^{1/2} n) + o_P(\log^{-1/2} n),
\end{aligned}$$

which implies that according to Assumption (A8) (ii), $\sup_{x_k\in[h_k,1-h_k]}(-2\log h_k)^{1/2}\left|\sigma^{-1}_{n,k}(x_k)(\widehat{\beta}^{\text{o}}_k - \beta_k)(x_k) - M_{h_k}(x_k)/\|K\|_{L_2}\right| = o_P(1)$. The uniform asymptotically normality of the "oracle" estimator $\widehat{\beta}^o_k(x_k)$ follows (B.23) and Slusky's Theorem. ■

B.1.3.2. Difference between SBL and Oracle Estimators

In this section, we investigate the difference between the SBL estimator and the oracle estimator.

Proof of Theorem 3. We denote

$$\eta^0_{i,-k} = \sum_{k'\neq k}^{p} \beta_{k'}(X_{ik'}) + \alpha(\mathbf{S}_i), \quad \widehat{\eta}_{i,-k} = \sum_{k'\neq k}^{p} \widehat{\beta}_{k'}(X_{ik'}) + \widehat{\alpha}(\mathbf{S}_i), \quad \widetilde{\eta}_{i,-k} = \sum_{k'\neq k}^{p} \widetilde{\beta}_{k'}(X_{ik'}) + \widetilde{\alpha}(\mathbf{S}_i).$$

Define

$$\mathbf{b}^*(x_k) = \sqrt{nh_k}\begin{pmatrix} b_0 - \widehat{\beta}^o_k(x_k) \\ h_k\left\{b_1 - \widehat{\beta}^{o\prime}_k(x_k)\right\}\end{pmatrix}, \ \widehat{\boldsymbol{b}}^*(x_k) = \sqrt{nh_k}\begin{pmatrix} \widehat{\beta}^{\text{SBL}}_k(x_k) - \widehat{\beta}^o_k(x_k) \\ h_k\left\{\widehat{\beta}^{\text{SBL}\prime}_k(x_k) - \widehat{\beta}^{o\prime}_k(x_k)\right\}\end{pmatrix}.$$

Then, $\widetilde{L}_n^{\mathrm{SBL}}\left\{\mathbf{b}^*(x_k)\right\}$ equals to

$$\sum_{i=1}^{n}\ell\left[g^{-1}\left\{\widehat{\beta}_k^{o}(x_k)+\widehat{\beta}_k^{o\prime}(x_k)(X_{ik}-x_k)+\widehat{\eta}_{i,-k}+(nh_k)^{-1/2}\mathbf{b}^{*\top}(x_k)\mathbf{Z}_{ik}(x_k)\right\},Y_i\right]K_{h_k}(X_{ik}-x_k).$$

It is clear to see that $\nabla\widetilde{L}_n^{\mathrm{SBL}}\left\{\widehat{\mathbf{b}}^*(x_k)\right\}=\mathbf{0}$, according to the property of the SBL estimator $\left\{\widehat{\beta}_k^{\mathrm{SBL}}(x_k),\widehat{\beta}_k^{\mathrm{SBL}\prime}(x_k)\right\}$. Applying the mean value theorem,

$$\widehat{\boldsymbol{b}}^*(x_k)=-\nabla^2\widetilde{L}_n^{\mathrm{SBL}}\left\{\bar{\mathbf{b}}^*(x_k)\right\}\nabla\widetilde{L}_n^{\mathrm{SBL}}(\mathbf{0}), \tag{B.25}$$

where $\bar{\mathbf{b}}^*(x_k)$ is some value between $\widehat{\boldsymbol{b}}^*(x_k)$ and $\mathbf{0}$. As

$$\frac{1}{\sqrt{nh_k}}\sum_{i=1}^{n}q_1\{\widehat{\beta}_k^{o}(x_k)+\widehat{\beta}_k^{o\prime}(x_k)(X_{ik}-x_k)+\eta_{i,-k}^{0},Y_i\}\mathbf{Z}_{ik}(x_k)K_{h_k}(X_{ik}-x_k)=0,$$

then

$$\begin{aligned}\nabla\widetilde{L}_n^{\mathrm{SBL}}(\mathbf{0})=&\frac{1}{\sqrt{nh_k}}\sum_{i=1}^{n}q_1\{\widehat{\beta}_k^{o}(x_k)+\widehat{\beta}_k^{o\prime}(x_k)(X_{ik}-x_k)+\widehat{\eta}_{i,-k},Y_i\}\mathbf{Z}_{ik}(x_k)K_{h_k}(X_{ik}-x_k)\\ =&\frac{1}{\sqrt{nh_k}}\sum_{i=1}^{n}q_1\{\widehat{\beta}_k^{o}(x_k)+\widehat{\beta}_k^{o\prime}(x_k)(X_{ik}-x_k)+\widehat{\eta}_{i,-k},Y_i\}\mathbf{Z}_{ik}(x_k)K_{h_k}(X_{ik}-x_k)\\ &-\frac{1}{\sqrt{nh_k}}\sum_{i=1}^{n}q_1\{\widehat{\beta}_k^{o}(x_k)+\widehat{\beta}_k^{o\prime}(x_k)(X_{ik}-x_k)+\eta_{i,-k}^{0},Y_i\}\mathbf{Z}_{ik}(x_k)K_{h_k}(X_{ik}-x_k)\\ =&(\iota_1,\iota_2)^{\top}.\end{aligned}$$

For the notation simplicity, we denote $\widehat{\eta}_i^{o}(x_k)=\widehat{\beta}_k^{o}(x_k)+\widehat{\beta}_k^{o\prime}(x_k)(X_{ik}-x_k)+\eta_{i,-k}^{0}$ in the following. At first, we focus on the first element of $\nabla\widetilde{L}_n^{\mathrm{SBL}}(\mathbf{0})$. By the Taylor expansion, we have

$$\iota_1=\frac{1}{\sqrt{nh_k}}\sum_{i=1}^{n}q_2\{\widehat{\eta}_i^{o}(x_k),Y_i\}(\widehat{\eta}_{i,-k}-\eta_{i,-k}^{0})K_{h_k}(X_{ik}-x_k)+O_P\left\{n^{1/2}h_k^{-1/2}\|\eta_{i,-k}^{0}-\widehat{\eta}_{i,-k}\|^2\right\}.$$

By Theorem 1 the second term in ι_1 has

$$\|\eta_{i,-k}^{0}-\widehat{\eta}_{i,-k}\|^2=O_{\mathrm{a.s.}}\left\{(H^{-1}+|\triangle|^{-2})n^{-1}\log n+H^{2\varrho+2}+|\triangle|^{2d+2}+\frac{\lambda^2}{n^2}|\triangle|^{-8}\right\}. \tag{B.26}$$

Also, the first term in ι_1 could be written as $n^{1/2}h_k^{-1/2}(I_1 + I_2)$, where

$$I_1 = \frac{1}{n}\sum_{i=1}^{n} q_2\{\widehat{\eta}_i^o(x_k), Y_i\}(\widetilde{\eta}_{i,-k} - \eta_{i,-k}^0)K_{h_k}(X_{ik} - x_k) = O_{\mathrm{a.s.}}(H^{\varrho+1} + |\triangle|^{d+1}), \tag{B.27}$$
$$I_2 = \frac{1}{n}\sum_{i=1}^{n} q_2\{\widehat{\eta}_i^o(x_k), Y_i\}(\widehat{\eta}_{i,-k} - \widetilde{\eta}_{i,-k})K_{h_k}(X_{ik} - x_k).$$

Recall that $\mathbf{V}_b$, $\mathbf{V}_v$ and $\mathbf{V}_p$ are the vectors defined in (B.7). Denote that

$$\mathbf{\Phi}_b = \left[\{\Phi_{U,kj,b}\}_{1\le k\le p, j\in\mathcal{J}}, \{\Phi_{B,m,b}(x_k)\}_{m\in\mathcal{M}}\right]^\top = -\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1}\mathbf{V}_b,$$
$$\mathbf{\Phi}_v = \left[\{\Phi_{U,kj,v}\}_{1\le k\le p, j\in\mathcal{J}}, \{\Phi_{B,m,v}(x_k)\}_{m\in\mathcal{M}}\right]^\top = -\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1}\mathbf{V}_v,$$
$$\mathbf{\Phi}_p = \left[\{\Phi_{U,kj,p}\}_{1\le k\le p, j\in\mathcal{J}}, \{\Phi_{B,m,p}(x_k)\}_{m\in\mathcal{M}}\right]^\top = -\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1}\mathbf{V}_p.$$

Following the proof of Lemma B.9, it is clear that

$$\|\mathbf{\Phi}_b\| = O_{\mathrm{a.s.}}(H^{\varrho+1}+|\triangle|^{d+1}),\ \|\mathbf{\Phi}_p\| = O(\lambda n^{-1}|\triangle|^{-4}),\ \|\mathbf{\Phi}_v\| = O_{\mathrm{a.s.}}\left\{(H^{-1/2} + |\triangle|^{-1})n^{-1/2}\log^{1/2} n\right\}.$$

Then, I_2 could be further written as $I_2 \asymp I_{2,b} + I_{2,v} + I_{2,p}$, where $I_{2,b}$, $I_{2,v}$ and $I_{2,p}$ are equal to

$$n^{-1}\sum_{i=1}^{n} q_2\{\widehat{\eta}_i^o(x_k), Y_i\}\left\{\sum_{k'\neq k}^{p}\sum_{j\in\mathcal{J}}\Phi_{U,k'j,b}U_{k'j}(X_{ik'}) + \sum_{m\in\mathcal{M}}\Phi_{B,m,b}\widetilde{B}_m^*(\mathbf{S}_i)\right\}K_{h_k}(X_{ik} - x_k),$$
$$n^{-1}\sum_{i=1}^{n} q_2\{\widehat{\eta}_i^o(x_k), Y_i\}\left\{\sum_{k'\neq k}^{p}\sum_{j\in\mathcal{J}}\Phi_{U,k'j,v}U_{k'j}(X_{ik'}) + \sum_{m\in\mathcal{M}}\Phi_{B,m,v}\widetilde{B}_m^*(\mathbf{S}_i)\right\}K_{h_k}(X_{ik} - x_k),$$
$$n^{-1}\sum_{i=1}^{n} q_2\{\widehat{\eta}_i^o(x_k), Y_i\}\left\{\sum_{k'\neq k}^{p}\sum_{j\in\mathcal{J}}\Phi_{U,k'j,p}U_{k'j}(X_{ik'}) + \sum_{m\in\mathcal{M}}\Phi_{B,m,p}\widetilde{B}_m^*(\mathbf{S}_i)\right\}K_{h_k}(X_{ik} - x_k),$$

respectively. By Assumption (A3) and the Cauchy Schwartz inequality, we have

$$|I_{2,b}| \le C_1\|\mathbf{\Phi}_b\|\|\boldsymbol{\omega}(x_k)\| + C_2\|\mathbf{\Phi}_b\|\|\boldsymbol{\omega}^\varepsilon(x_k)\|.$$

According to Lemma B.11, we have

$$I_{2,b} = O_{\mathrm{a.s.}}(H^{\varrho+1} + |\triangle|^{d+1}) \times O_{\mathrm{a.s.}}\{1 + (H^{-1/2} + |\triangle|^{-1})n^{-1/2}h_k^{-1/2}\log^{1/2} n\}. \tag{B.28}$$

Similarly,

$$
\begin{aligned}
I_{2,p} &\le C_1\|\boldsymbol{\Phi}_p\|\|\boldsymbol{\omega}(x_k)\| + C_2\|\boldsymbol{\Phi}_p\|\|\boldsymbol{\omega}^{\varepsilon}(x_k)\| \\
&= O_{\text{a.s.}}\left(\lambda n^{-1}|\triangle|^{-4}\right) \times O_{\text{a.s.}}\{1 + (H^{-1/2} + |\triangle|^{-1})n^{-1/2}h_k^{-1/2}\log^{1/2} n\}. \qquad \text{(B.29)}
\end{aligned}
$$

Now, we divide $I_{2,v}$ into two parts $I_{2,v} - \widetilde{I}_{2,v}$ and $\widetilde{I}_{2,v}$, where

$$
\begin{aligned}
\widetilde{I}_{2,v} = &\sum_{k' \neq k}^{p}\sum_{j \in \mathcal{J}} \Phi_{U,k'j,v}\mathrm{E}[U_{k'j}(X_{ik'})q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik} - x_k)] \\
&+ \sum_{m \in \mathcal{M}} \Phi_{B,m,v}\mathrm{E}[\widetilde{B}_m^*(\mathbf{S}_i)q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik} - x_k)].
\end{aligned}
$$

According to Lemma B.11 and Assumption (A3), we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} U_{k'j}(X_{ik'})q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik} - x_k) - \mathrm{E}[U_{k'j}(X_{ik'})q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik} - x_k)] \\
&\qquad = O_{\text{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2} n), \\
&\frac{1}{n}\sum_{i=1}^{n} B_m(\mathbf{S}_i)q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik} - x_k) - \mathrm{E}[B_m(\mathbf{S}_i)q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik} - x_k)] \\
&\qquad = O_{\text{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2} n).
\end{aligned}
$$

Applying Cauchy Schwarz inequality, we have

$$
\begin{aligned}
I_{2,v} - \widetilde{I}_{2,v} &\le (H^{-1/2} + |\triangle|^{-1}) \times O_{\text{a.s.}}(n^{-1/2}h_k^{-1/2}\log^{1/2} n) \times \|\boldsymbol{\Phi}_v\| \\
&= O_{\text{a.s.}}\left\{(H^{-1} + |\triangle|^{-2})n^{-1}h_k^{-1/2}\log n\right\}. \qquad \text{(B.30)}
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\widetilde{I}_{2,v} =& \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V\{g^{-1}(\eta_i^0)\}}\dot{g}^{-1}(\eta_i^0)\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1}\{\mathbf{U}(\mathbf{X}_i)^\top, \widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\}\boldsymbol{\mu}_{q_2} \\
=& \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V\{g^{-1}(\eta_i^0)\}}\dot{g}^{-1}(\eta_i^0)\left[\left\{\frac{1}{n}\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1} - \mathbf{D}^{-1}\right]\{\mathbf{U}(\mathbf{X}_i)^\top, \widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\}\boldsymbol{\mu}_{q_2} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V\{g^{-1}(\eta_i^0)\}}\dot{g}^{-1}(\eta_i^0)\mathbf{D}^{-1}\{\mathbf{U}(\mathbf{X}_i)^\top, \widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\}\boldsymbol{\mu}_{q_2},
\end{aligned}
$$

where $\boldsymbol{\mu}_{q_2}$ is a vector with elements $\mathrm{E}\left[U_{k'j}(X_{ik'})q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik}-x_k)\right]$, $1 \leq k \leq p$, $k \in \mathcal{J}$ for univariate spline, and $\mathrm{E}\left[\widetilde{B}_m(\mathbf{S}_i)q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik}-x_k)\right]$, $m \in \mathcal{M}$ for bivariate spline, and matrix $\mathbf{D}$ is defined in the Lemma B.13. Applying Lemma B.13, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V\{g^{-1}(\eta_i^0)\}}\dot{g}^{-1}(\eta_i^0)\left[\left\{\nabla^2 L^P(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\gamma}}^*)\right\}^{-1} - \mathbf{D}^{-1}\right]\{\mathbf{U}(\mathbf{X}_i)^\top, \widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\}\boldsymbol{\mu}_{q_2} \\
&\qquad = O_{\text{a.s.}}\left\{H^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n\right\} \times O_{\text{a.s.}}(n^{-1/2}\log^{1/2} n). \qquad \text{(B.31)}
\end{aligned}$$

By (B.18) and (B.19) in Lemma B.11, we have that $\mathrm{E}\left[U_{k'j}(X_{ik'})q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik}-x_k)\right] = O(H^{1/2})$ and $\mathrm{E}\left[\widetilde{B}_m(\mathbf{S}_i)q_2\{\widehat{\eta}_i^o(x_k), Y_i\}K_{h_k}(X_{ik}-x_k)\right] = O(|\triangle|)$. Therefore,

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - g^{-1}(\eta_i^0)}{\sigma^2 V\{g^{-1}(\eta_i^0)\}}\dot{g}^{-1}(\eta_i^0)\mathbf{D}^{-1}\{\mathbf{U}(\mathbf{X}_i)^\top, \widetilde{\mathbf{B}}(\mathbf{S}_i)^\top\}\boldsymbol{\mu}_{q_2}\right] = O(n^{-1}).$$

Then, it is easy to verify that

$$n^{-1}\sum_{i=1}^{n}\widetilde{I}_{2,v,i} = O_{\text{a.s.}}(n^{-1/2}\log^{1/2} n). \qquad \text{(B.32)}$$

Therefore, by (B.30), (B.31) and (B.32),

$$I_{2,v} = O_{\text{a.s.}}\left\{(H^{-1} + |\triangle|^{-2})n^{-1}h_k^{-1/2}\log n\right\} + O_{\text{a.s.}}(n^{-1/2}\log^{1/2} n). \qquad \text{(B.33)}$$

Then, by (B.28), (B.29) and (B.33), we have

$$\begin{aligned}
I_2 =& O_{\text{a.s.}}(H^{\varrho+1} + |\triangle|^{d+1} + \lambda n^{-1}|\triangle|^{-4}) \times O_{\text{a.s.}}\{1 + (H^{-1/2} + |\triangle|^{-1})n^{-1/2}h_k^{-1/2}\log^{1/2} n\} \\
&+ O_{\text{a.s.}}\left\{(H^{-1} + |\triangle|^{-2})n^{-1}h_k^{-1/2}\log n\right\} + O_{\text{a.s.}}(n^{-1/2}\log^{1/2} n). \qquad \text{(B.34)}
\end{aligned}$$

Combining (B.27) and (B.34), we obtain that $\iota_1 = n^{1/2}h_k^{-1/2}(I_1 + I_2) = O_{\text{a.s.}}(h_k^{-1/2}\log^{1/2} n)$. Following the above procedure and Lemma B.12, we obtain $\iota_2 = O_{\text{a.s.}}(h_k^{-1/2}\log^{1/2} n)$. Therefore,

$$|\nabla\widetilde{L}_n^{\mathrm{SBL}}(\mathbf{0})| = O_{\text{a.s.}}(h_k^{-1/2}\log^{1/2} n). \qquad \text{(B.35)}$$

Let $\eta_i^* = \widehat{\beta}_k^o(x_k) + (X_{ik} - x_k)\widehat{\beta}_k^{\prime o} + \widehat{\eta}_{i,-k} + (nh_k)^{-1/2}\bar{\mathbf{b}}^{*\top}\mathbf{Z}_{ik}$. Then we have

$$h_k^{-1}\left[\nabla^2 L_n^{\mathrm{SBL}}\left\{\bar{\mathbf{b}}^*(x_k)\right\}\right]^{-1} = \left\{\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1 & h_k^{-1}\sum_{i=1}^n (X_{ik}-x_k) \\ h_k^{-1}\sum_{i=1}^n (X_{ik}-x_k) & h_k^{-2}\sum_{i=1}^n (X_{ik}-x_k)^2 \end{pmatrix} \right.$$
$$\left. \times\, q_2\left(\eta_i^*, Y_i\right) K_{h_k}(X_{ik}-x_k)\right\}^{-1}.$$

For any vector $\mathbf{a} = (a_1, a_2)^\top \in \mathbb{R}^2$, $\|\mathbf{a}\| = 1$, we have

$$\mathbf{a}^\top \left\{\sum_{i=1}^n \begin{pmatrix} 1 & h_k^{-1}\sum_{i=1}^n (X_{ik}-x_k) \\ h_k^{-1}\sum_{i=1}^n (X_{ik}-x_k) & h_k^{-2}\sum_{i=1}^n (X_{ik}-x_k)^2 \end{pmatrix} q_2\left(\eta_i^*, Y_i\right) K_{h_k}(X_{ik}-x_k)\right\}\mathbf{a}$$
$$= \sum_{i=1}^n \left\{a_1 + a_2(X_{ik}-x_k)/h_k\right\}^2 q_2\left(\eta_i^*, Y_i\right) K_{h_k}(X_{ik}-x_k)$$
$$= \sum_{i=1}^n \left\{a_1 + a_2(X_{ik}-x_k)/h_k\right\}^2 \rho_2\left(\eta_i^*\right) K_{h_k}(X_{ik}-x_k)\{1 + O_{\mathrm{a.s.}}(1)\}.$$

As $\rho_2(\eta_i^*)$ is bounded from zero and infinity, therefore

$$c\left\{a_1^2 + a_2^2\int v^2 K(v)dv\right\} \le \int \left\{a_1 + a_2(X_{ik}-x_k)/h_k\right\}^2 \rho_2\left(\eta_i^*\right) K_{h_k}(X_{ik}-x_k),$$

and $\int \left\{a_1 + a_2(X_{ik}-x_k)/h_k\right\}^2 \rho_2\left(\eta_i^*\right) K_{h_k}(X_{ik}-x_k) \le C\left\{a_1^2 + a_2^2\int v^2 K(v)dv\right\}$. Consequently, the second order derivative matrix has the property

$$C^{-1}h_k\mathbf{I}_2 \le \left[\nabla^2 L_n^{\mathrm{SBL}}\left\{\bar{\mathbf{b}}^*(x_k)\right\}\right]^{-1} \le c^{-1}h_k\mathbf{I}_2 \tag{B.36}$$

almost surely, for large enough n.

Plugging (B.26), (B.35) and (B.36) into (B.25), by the Assumption (A6'), we have $|\widehat{\boldsymbol{b}}^*(x_k)| = O_{\mathrm{a.s.}}(h_k^{1/2}\log^{1/2} n)$. ■

B.2. Implementation for the Bivariate spline

B.2.1. An example of $\boldsymbol{\Psi}$

Consider the following simple triangulation. In the above figure, within this triangulation, one

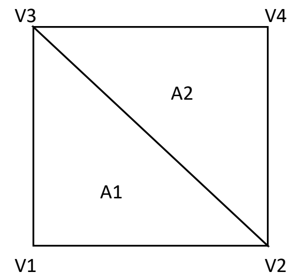


Figure B.1: A simple triangulation with two triangles.

sees that there are two triangles $A_1 :=< v_1, v_2, v_3 >$ and $A_2 :=< v_4, v_3, v_2 >$ sharing a common edge $e =< v_2, v_3 >$. Assume that there is a point v on the common edge e, then the barycentric coordinates (b_1, b_2, b_3) with respect to A_1 and A_2 are $(0, b_2, 1 - b_2)$ and $(0, 1 - b_2, b_2)$ respectively. As a result, the degree-d polynomials with respect to the two triangles can be written as

$$p_1(v) = \sum_{j+k=d} \gamma_{0jk}^{(1)} B_{d;0jk}^{(1)}(v) = \sum_{j+k=d} \gamma_{0jk}^{(1)} \frac{d!}{j!k!} b_2^j (1 - b_2)^k,$$

$$p_2(\widetilde{v}) = \sum_{j+k=d} \gamma_{0jk}^{(2)} B_{d;0jk}^{(2)}(v) = \sum_{j+k=d} \gamma_{0jk}^{(2)} \frac{d!}{j!k!} (1 - b_2)^j b_2^k.$$

Therefore, to have p_1 and p_2 join continuously on edge e, we require $\gamma_{0jk}^{(1)} = \gamma_{0kj}^{(2)}$ for $j, k \geq 0$ and $j + k = d$. For simplicity, we consider the case that $d = 2$ as an example. When $d = 2$, then the entire coefficients vector becomes

$$\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top)^\top = (\gamma_{2,0,0}^{(1)}, \gamma_{1,1,0}^{(1)}, \gamma_{1,0,1}^{(1)}, \gamma_{0,2,0}^{(1)}, \gamma_{0,1,1}^{(1)}, \gamma_{0,0,2}^{(1)}, \gamma_{2,0,0}^{(2)}, \gamma_{1,1,0}^{(2)}, \gamma_{1,0,1}^{(2)}, \gamma_{0,2,0}^{(2)}, \gamma_{0,1,1}^{(2)}, \gamma_{0,0,2}^{(2)})^\top.$$

and the constraints are

$$\gamma_{0,2,0}^{(1)} = \gamma_{0,0,2}^{(2)}, \quad \gamma_{0,1,1}^{(1)} = \gamma_{0,1,1}^{(2)}, \quad \gamma_{0,0,2}^{(1)} = \gamma_{0,2,0}^{(2)}.$$

Thus, in this case the constraint matrix $\mathbf{\Psi}$ is

$$\mathbf{\Psi} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \end{pmatrix}.$$

B.2.2. Construction of $\mathbf{P}$

This section is about how to construct $\mathbf{P}$. Given the triangulation Δ, the $\mathbf{P}$ is the block diagonal matrices with $\mathbf{P}_\tau$, $\tau \in \Delta$. Therefore, we focus on the construction of $\mathbf{P}_\tau$ instead.

Supposing the polynomial piece of spline is $\sum_{i+j+k=d} \gamma^\tau_{ijk} B^{\tau,d}_{ijk}(\mathbf{s})$ restricted on $\tau \in \triangle$, we have

$$\begin{aligned}
&\mathcal{E}\Big\{ \sum_{i+j+k=d} \gamma^\tau_{ijk} B^{\tau,d}_{ijk}(\mathbf{s}) \Big\} \\
&= \int_\tau \Big\{ \sum_{i+j+k=d} \gamma^\tau_{ijk} \nabla^2_{s_1} B^{\tau,d}_{ijk}(\mathbf{s}) \Big\}^2 ds_1 ds_2 + \int_\tau \Big\{ \sum_{i+j+k=d} \gamma^\tau_{ijk} \nabla_{s_1}\nabla_{s_2} B^{\tau,d}_{ijk}(\mathbf{s}) \Big\}^2 ds_1 ds_2 \\
&\quad + \int_\tau \Big\{ \sum_{i+j+k=d} \gamma^\tau_{ijk} \nabla^2_{s_2} B^{\tau,d}_{ijk}(\mathbf{s}) \Big\}^2 ds_1 ds_2 \\
&= \boldsymbol{\gamma}^{\tau\top} (\mathbf{P}_{1,\tau} + \mathbf{P}_{2,\tau} + \mathbf{P}_{3,\tau}) \boldsymbol{\gamma}^\tau = \boldsymbol{\gamma}^{\tau\top} \mathbf{P}_\tau \boldsymbol{\gamma}^\tau,
\end{aligned}$$

where $\boldsymbol{\gamma}^\tau = \left\{\gamma^\tau_{ijk}\right\}^\top_{i+j+k=d}$, and $\mathbf{P}_{1,\tau}$, $\mathbf{P}_{2,\tau}$ and $\mathbf{P}_{3,\tau}$ are the $\frac{(d+2)(d+1)}{2} \times \frac{(d+2)(d+1)}{2}$ matrices with entries being $\int_\tau \left(\nabla^2_{s_1} B^{\tau,d}_{ijk}\right)\left(\nabla^2_{s_1} B^{\tau,d}_{i'j'k'}\right)(\mathbf{s}) ds_1 ds_2$, $\int_\tau \left(\nabla_{s_1}\nabla_{s_2} B^{\tau,d}_{ijk}\right)\left(\nabla_{s_1}\nabla_{s_2} B^{\tau,d}_{i'j'k'}\right)(\mathbf{s}) ds_1 ds_2$ and $\int_\tau \left(\nabla^2_{s_2} B^{\tau,d}_{ijk}\right)\left(\nabla^2_{s_2} B^{\tau,d}_{i'j'k'}\right)(\mathbf{s}) ds_1 ds_2$, respectively.

Suppose $s^{(1)}$ and $s^{(2)}$ are two vectors in $\mathbb{R}^2$. According to the Theorem 2.13 in Lai and Schumaker (2007), the direction derivative of $B^{\tau,d}_{ijk}(\mathbf{s})$ with respect to $s^{(1)}$ has the expression

$$\begin{aligned}
\nabla_{s^{(1)}} B^{\tau,d}_{ijk}(\mathbf{s}) &= d\left\{ a^{(1)}_1 B^{d-1}_{(i-1)jk}(\mathbf{s}) + a^{(1)}_2 B^{d-1}_{i(j-1)k}(\mathbf{s}) + a^{(1)}_3 B^{d-1}_{ij(k-1)}(\mathbf{s}) \right\} \\
&= d \sum_{i+j+k=d-1} \gamma^{(1)}_{ijk}(s^{(1)}) B^{\tau,d-1}_{ijk}(\mathbf{s}),
\end{aligned}$$

where $(a_1^{(1)}, a_2^{(1)}, a_3^{(1)})$ is the barycentric coordinate of $s^{(1)}$. And the second direction derivative with respect to $s^{(1)}$ and $s^{(2)}$ is

$$\nabla_{s^{(2)}}\nabla_{s^{(1)}}B_{ijk}^{\tau,d}(\mathbf{s}) = d(d-1)\sum_{i+j+k=d-2}\gamma_{ijk}^{(2)}(s^{(1)}, s^{(2)})B_{ijk}^{\tau,d-2}(\mathbf{s}),$$

where $\gamma_{ijk}^{(2)}(s^{(1)}, s^{(2)}) = a_1^{(2)}\gamma_{(i+1)jk}^{(1)}(s^{(1)}) + a_2^{(2)}\gamma_{i(j+1)k}^{(1)}(s^{(1)}) + a_3^{(2)}\gamma_{ij(k+1)}^{(1)}(s^{(1)})$ and $\left(a_1^{(2)}, a_2^{(2)}, a_3^{(2)}\right)$ is the barycentric coordinate of $s^{(2)}$. Therefore, the terms $\nabla_{s_1}^2 B_{ijk}^{\tau,d}$, $\nabla_{s_1}\nabla_{s_2}B_{ijk}^{\tau,d}$ and $\nabla_{s_2}^2 B_{ijk}^{\tau,d}$ are the linear combinations of $\left\{B_{ijk}^{d-2}(\mathbf{s})\right\}_{i+j+k=d-2}$. Let $\boldsymbol{c}_{1,ijk}$, $\boldsymbol{c}_{2,ijk}$ and $\boldsymbol{c}_{3,ijk}$ be the corresponding coefficients. Then we have

$$\int_\tau \left(\nabla_{s_1}^2 B_{ijk}^{\tau,d}\right)\left(\nabla_{s_1}^2 B_{i'j'k'}^{\tau,d}\right)(\mathbf{s})ds_1ds_2 = \boldsymbol{c}_{1,ijk}^\top \mathbf{M}_\tau \boldsymbol{c}_{1,i'j'k'},$$

$$\int_\tau \left(\nabla_{s_1}\nabla_{s_2} B_{ijk}^{\tau,d}\right)\left(\nabla_{s_1}\nabla_{s_2} B_{i'j'k'}^{\tau,d}\right)(\mathbf{s})ds_1ds_2 = \boldsymbol{c}_{2,ijk}^\top \mathbf{M}_\tau \boldsymbol{c}_{2,i'j'k'},$$

$$\int_\tau \left(\nabla_{s_2}^2 B_{ijk}^{\tau,d}\right)\left(\nabla_{s_2}^2 B_{i'j'k'}^{\tau,d}\right)(\mathbf{s})ds_1ds_2 = \boldsymbol{c}_{3,ijk}^\top \mathbf{M}_\tau \boldsymbol{c}_{3,i'j'k'},$$

where $\mathbf{M}_\tau$ is the $\frac{(d+2)(d+1)}{2} \times \frac{(d+2)(d+1)}{2}$ matrix with entries $\int_\tau B_{ijk}^{\tau,d-2}B_{i'j'k'}^{\tau,d-2}ds_1ds_2$. By the Theorem 2.34 in Lai and Schumaker (2007), we have

$$\int_\tau B_{ijk}^{\tau,d-2}B_{i'j'k'}^{\tau,d-2}ds_1ds_2 = \frac{\binom{i+i'}{i}\binom{j+j'}{j}\binom{k+k'}{k}}{\binom{2d-4}{d-2}\binom{2d-2}{2}}A_\tau,$$

where A_τ is the area of triangle τ. Denote $\mathbf{C}_1 = \{\boldsymbol{c}_{1,ijk}\}_{i+j+k=d}$, $\mathbf{C}_2 = \{\boldsymbol{c}_{2,ijk}\}_{i+j+k=d}$ and $\mathbf{C}_3 = \{\boldsymbol{c}_{3,ijk}\}_{i+j+k=d}$. We finally obtain $\mathbf{P}_{1,\tau} = \mathbf{C}_1^\top\mathbf{M}_\tau\mathbf{C}_1$, $\mathbf{P}_{2,\tau} = \mathbf{C}_2^\top\mathbf{M}_\tau\mathbf{C}_2$, $\mathbf{P}_{3,\tau} = \mathbf{C}_3^\top\mathbf{M}_\tau\mathbf{C}_3$ and $\mathbf{P}_\tau = \mathbf{P}_{1,\tau} + \mathbf{P}_{2,\tau} + \mathbf{P}_{3,\tau}$.

Next we provide a simple example for $\mathbf{P}_\tau$. Consider the triangle A_1 in the Figure B.1 and $d = 2$. The barycentric coordinate of s_1 is $(-1, 1, 0)$. The barycentric coordinate of s_2 is $(-1, 0, 1)$. The set of degree-d bivariate Bernstein basis polynomials within τ is

$$\left\{B_{2,0,0}^{\tau,2}(\mathbf{s}), B_{1,1,0}^{\tau,2}(\mathbf{s}), B_{1,0,1}^{\tau,2}(\mathbf{s}), B_{0,2,0}^{\tau,2}(\mathbf{s}), B_{0,1,1}^{\tau,2}(\mathbf{s}), B_{0,0,2}^{\tau,2}(\mathbf{s})\right\}.$$

The second directional derivative of these bivariate Bernstein basis polynomials are the linear combination of $B^{\tau,0}_{0,0,0}(\mathbf{s})$. For example,

$$\nabla_{s_1} B^{\tau,2}_{1,1,0}(\mathbf{s}) = -2B^{\tau,1}_{0,1,0}(\mathbf{s}) + 2B^{\tau,1}_{1,0,0}(\mathbf{s}),\ \nabla^2_{s_1} B^{\tau,1}_{1,1,0}(\mathbf{s}) = -4B^{\tau,0}_{0,0,0}(\mathbf{s}).$$

In this case, we have $\mathbf{C}_1 = (2, -4, 2, 0, 0, 0)$, $\mathbf{C}_2 = (2, -2, 0, -2, 2, 0)$, $\mathbf{C}_3 = (2, 0, 0, -4, 0, 2)$ and $\mathbf{M}_\tau = 1/2$. Therefore, the penalty matrix $\mathbf{P}_\tau$ is

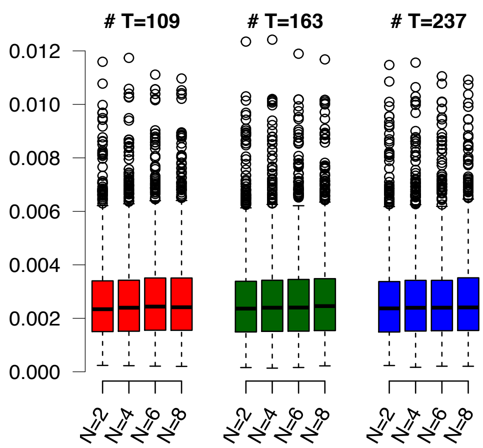
$$\begin{pmatrix} 8 & -8 & 2 & -8 & 4 & 2 \\ -8 & 12 & -4 & 4 & -4 & 0 \\ 2 & -4 & 2 & 0 & 0 & 0 \\ -8 & 4 & 0 & 12 & -4 & -4 \\ 4 & -4 & 0 & -4 & 4 & 0 \\ 2 & 0 & 0 & -4 & 0 & 2 \end{pmatrix}.$$

B.3. More Simulation Results

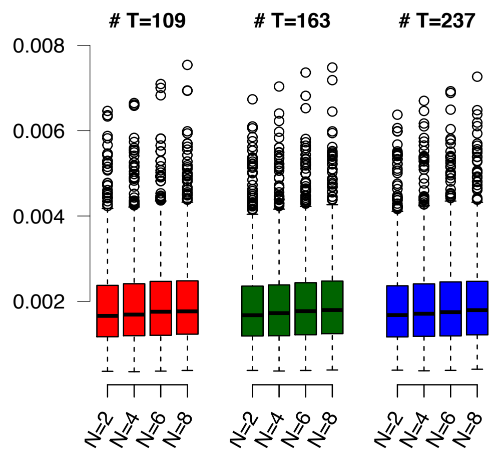
In this section, we present more simulation results from Examples 1 and 2 in the main paper.

For Example 1, Figures B.2 and B.3 present the boxplots of the MISEs of the estimators based on different combinations of knots and triangulations for Case I and Case II, where the number of knots for the cubic B-splines is $J_n = 2, 4, 6, 8$, and the triangulation for the bivariate splines are taken to be $\triangle_1$, $\triangle_2$ and $\triangle_3$ with 109 triangles, 163 and 237 triangles, respectively. The results are generally similar, although some of the MISEs of $\widehat{\alpha}$ are larger for $\triangle_1$ and those based on $\triangle_2$ and $\triangle_3$.

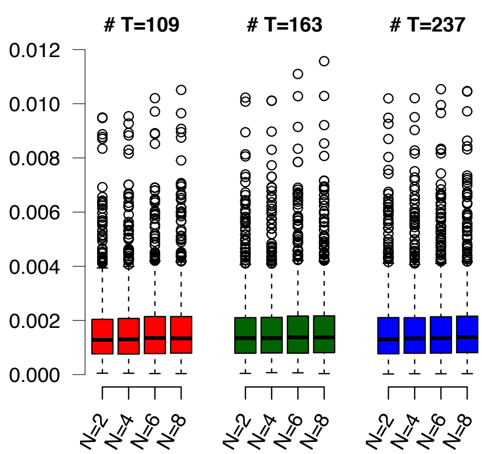
For Example 2, Figure B.4 and Figure B.5 present the estimators and the 95% SCBs from a typical simulation trial.



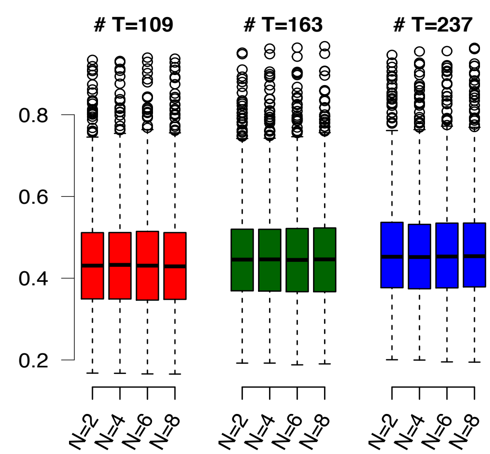
(a) β_1



(b) β_2

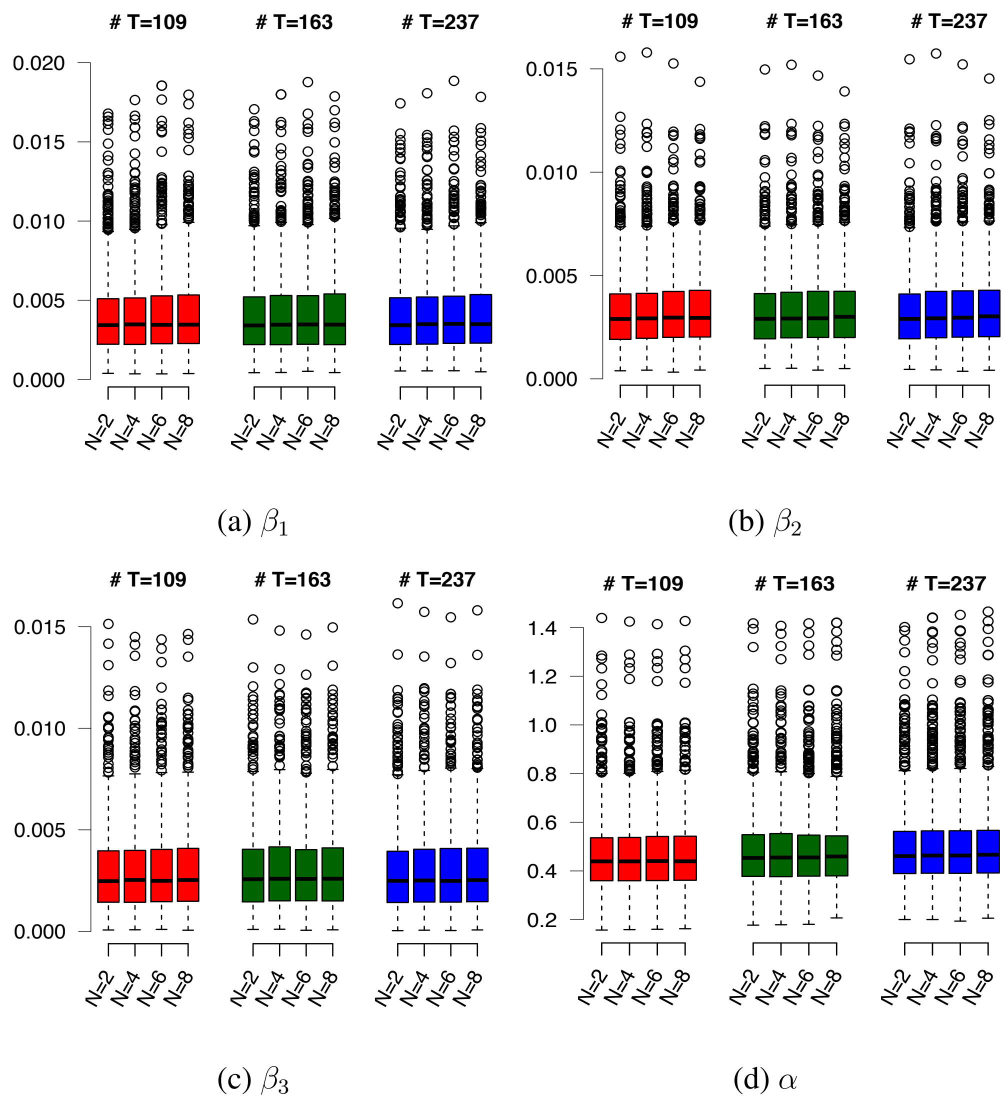


(c) β_3



(d) α

Figure B.2: Boxplots of the MISEs of the estimators of functional components in Case I using different combinations of knots and triangulations.



(a) β_1 (b) β_2

(c) β_3 (d) α

Figure B.3: Boxplots of the MISEs of the estimators of functional components in Case II using different combinations of knots and triangulations.

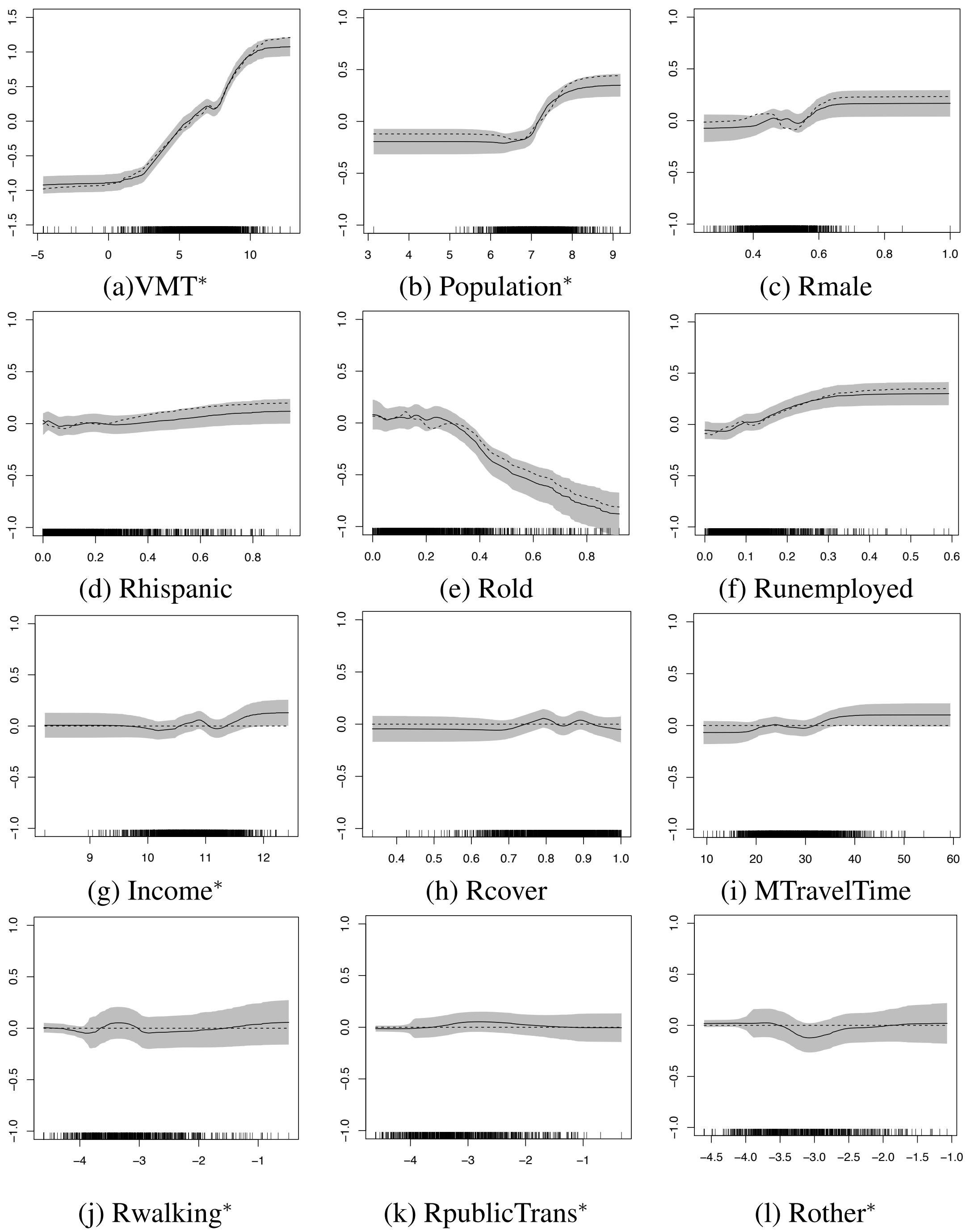


Figure B.4: Plots of the true function (dashed line), its SBL estimator (black curve) and the 95% SCB (grey band).

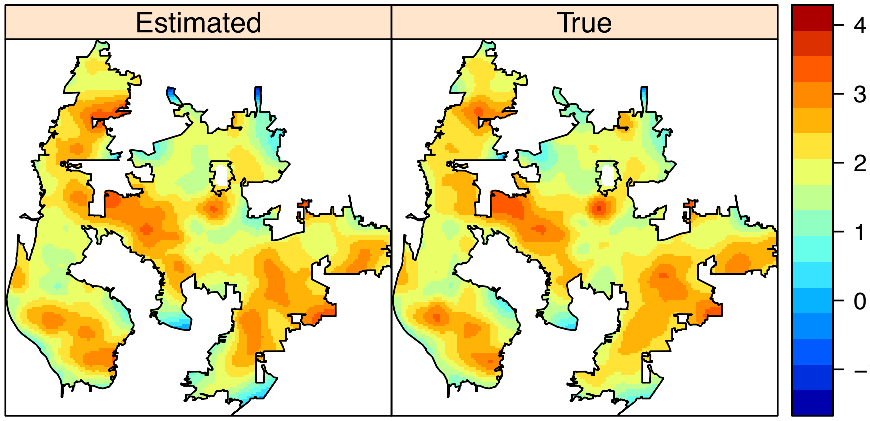


Figure B.5: Plots of the estimated $\alpha(\cdot)$ (left plot) and the true $\alpha(\cdot)$ (right plot).

References

De Boor, C. (2001). A practical guide to splines (applied mathematical sciences). Sciences, 55:75309.

Fan, J., Heckman, N. E., and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. Journal of the American Statistical Association, 90(429):141–150.

Lai, M. J. and Schumaker, L. L. (2007). Spline functions on triangulations. Cambridge University Press.

Stone, C. J. (1985). Additive regression and other nonparametric models. The Annals of Statistics, pages 689–705.

Zheng, S., Liu, R., Yang, L., and Härdle, W. K. (2016). Statistical inference for generalized additive models: simultaneous confidence corridors and variable selection. Test, 25(4):607–626.