Supplementary materials of the article 'Nonparametric Estimation and Inference for Spatiotemporal Epidemic Models'

In this document, we provide proofs of the main asymptotic results and some technical details used in the proofs.

B.1. *Notations*

To facilitate the proof, we first introduce some notations to make the model more general. Suppose there are n locations, where $\mathbf{U}_i \in \Omega$ are the spatial coordinate of the i-th location. Let Y_{it} be the response variable observed at time t, and in the STEM models, Y_{it} is the number of new infected cases or new fatal cases on day t for county i. To simplify our proof, here we abuse our notations from the main article and let $\mathbf{Z}_{it} = (Z_{it1}, \ldots, Z_{itp_1})^\top$ be a p_1-dimensional vector of time-varying explanatory variables which are linearly associated with the response variable. In the STEM-infection model, $Z_{it1} = \log(S_{it-1}/N_i)$ and Z_{itj}'s, $j = 2, \ldots, p_1$, are the mobility information and dummy variables of interventions or control measures observed on day $t-7$ for county i. In addition, let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip_2})^\top$ be a p_2-dimensional vector of explanatory variables, which are not varying with time, and have nonlinear effect on the response, and in the STEM models, X_{ik}'s are the time invariant county-level features. Finally, let $\mathbf{W}_{it} = (1, W_{it1}, \ldots, W_{itp_3})^\top$ be a (p_3+1)-dimensional vector of explanatory variables, which have a varying relationship with the response across different locations. For example, in the STEM models proposed in the main paper, W_{it1} can be the logarithm of the first lagged value of the cumulative number of active cases at time t, i.e., $W_{it1} = \log(I_{i,t-1}+1)$. We assume $\{(Y_{it}, \mathbf{Z}_{it}^\top, \mathbf{X}_i^\top, \mathbf{W}_{it}^\top)\}_{i=1}^n$ are independent and identically distributed across all the locations, but correlation may exist for the time series observed at the same location. The following proof studies the theoretical properties of the discrete-time spatial epidemic model at any fixed time point t. The results hold for all the time points within the studying period. For notation simplicity, we drop the time index t for $\boldsymbol{\alpha}_t$, γ_{1t}, ..., γ_{p_2t}, β_{0t}, ..., β_{p_3t} and spline coefficients $\boldsymbol{\xi}_{1t}, \ldots, \boldsymbol{\xi}_{p_2t}$, $\boldsymbol{\theta}_{1t}, \ldots, \boldsymbol{\theta}_{p_3t}$, and $\boldsymbol{\theta}^*_{1t}, \ldots, \boldsymbol{\theta}^*_{p_3t}$.

Furthermore, in the proof, we assume that the conditional mean of Y_{it} depends only on the covariate vector for the tth observation. We consider the following spatiotemporal generalized partially linear model, and we assume that μ_{it}^0 can be modeled via a link function g as follows:

$$g(\mu_{it}^0) = \boldsymbol{\alpha}^{0\top}\mathbf{Z}_{it} + \sum_{k=1}^{p_2} \gamma_k^0(X_{ik}) + \sum_{\ell=0}^{p_3} \beta_\ell^0(\mathbf{U}_i) W_{it\ell}$$

and rewrite the quasilikelihood as the following:

$$L_n(\boldsymbol{\alpha}; \boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{i=1}^n \sum_{s=t_1}^{t_2} L\left[g^{-1}\left\{\boldsymbol{\alpha}^\top \mathbf{Z}_{is} + \sum_{k=1}^{p_2} \gamma_k(X_{ik}) + \sum_{\ell=0}^{p_3} \beta_\ell(\mathbf{U}_i) W_{is\ell}\right\}, Y_{is}\right].$$

First we introduce the general notations that we use in the following proof. Without loss of generality, let $x_k \in [a_k, b_k] = [0,1]$, for $k = 1, \cdots, p_2$, and the area of the domain Ω be 1, $A(\Omega) = 1$, in the rest of the article. For the univariate splines, we consider equally-spaced knots in our theoretical derivation, and denote h as the length of the equally-spaced subintervals, then it is clear $h \asymp |\mathcal{J}|^{-1}$. For a real value vector $\mathbf{a} \in \mathbb{R}^n$, we define its Euclidean norm as $\|\mathbf{a}\|_2^2 = \sum_{i=1}^n a_i^2$ and its supremum norm as $|\mathbf{a}| = \max_{1\le i\le n} |a_i|$. For any real symmetric matrix $\mathbf{A} = (a_{ij})_{i=1,j=1}^{n,n}$, 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 $\psi(\boldsymbol{u})$ on a domain Ω, let $\|\psi\|_\infty = \sup_{\boldsymbol{u}\in\Omega} |\psi(\boldsymbol{u})|$, and $\|\psi\|_{L_2}^2 = \int_\Omega \psi^2(\boldsymbol{u})d\boldsymbol{u}$. For any bivariate function $g : \Omega \to R$, denote $|g|_{\upsilon,\infty,\Omega} = \max_{i+j=\upsilon} \|\nabla_{s_1}^i \nabla_{s_2}^j g_j(\boldsymbol{u})\|_{\infty,\Omega}$. Let v be a nonnegative integer, and $\delta \in (0,1]$ such that $\varrho = \delta + v \ge 1$. Let $\mathcal{H}^{(\varrho)}([0,1])$ be the class of functions ψ on $[0,1]$ whose vth derivative exists and satisfies a Lipschitz condition of order δ: $|\psi^{(v)}(x) - \psi^{(v)}(x')| \le C_v|x - x'|^\delta$, for $x, x' \in [0,1]$. Let

$$\mathcal{D}_k^0([0,1]) = \{g : \mathrm{E}g(X_k) = 0, \mathrm{E}g^2(X_k) < \infty\}$$

be the functional space defined on $[0,1]$ and

$$\mathcal{S}^{d+1,\infty}(\Omega) = \{g : |g|_{k,\infty,\Omega} < \infty, 0 \le k \le d+1\}$$

be the standard Sobolev space.

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

$$\mathcal{G} = \left\{ \eta = \sum_{j=1}^{p_1} \alpha_j z_j + \sum_{k=1}^{p_2} \gamma_k(x_k) + \sum_{\ell=0}^{p_3} \beta_\ell(\boldsymbol{u}) w_\ell : \alpha_j \in \mathbb{R}, \gamma_k \in \mathcal{H}^{(\varrho)} \cap \mathcal{D}_k^0, \beta_\ell \in \mathcal{S}^{d+1,\infty}(\Omega) \right\}.$$

We define the norm on the space $\mathcal{G}$. For functions $\eta_1, \eta_2 \in \mathcal{G}$, define their theoretical inner product as

$$\langle \eta_1, \eta_2 \rangle = \mathrm{E}\{\eta_1(\mathbf{Z}, \mathbf{X}, \mathbf{W}, \mathbf{U})\eta_2(\mathbf{Z}, \mathbf{X}, \mathbf{W}, \mathbf{U})\},$$

and define their empirical inner product as

$$\langle \eta_1, \eta_2 \rangle_n = \frac{1}{n_T} \sum_{i=1}^n \sum_{s=t_1}^{t_2} \eta_1(\mathbf{Z}_{is}, \mathbf{X}_i, \mathbf{W}_{is}, \mathbf{U}_i)\eta_2(\mathbf{Z}_{is}, \mathbf{X}_i, \mathbf{W}_{is}, \mathbf{U}_i),$$

where $n_T = n(t_2 - t_1 + 1)$.

Consequently, $\|\eta\|^2 = \langle \eta, \eta \rangle$ and $\|\eta\|_n^2 = \langle \eta, \eta \rangle_n$.

Denote the corresponding empirical and theoretical norms $\|\cdot\|_n$ and $\|\cdot\|$. Furthermore, let $\|\cdot\|_{\mathcal{E}}$ be the norm introduced by the inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$, where, for $\boldsymbol{\beta}^{(1)}(\boldsymbol{u})$ and $\boldsymbol{\beta}^{(2)}(\boldsymbol{u})$,

$$\left\langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \right\rangle_{\mathcal{E}} = \sum_{\ell,\ell'=0}^{p_3} \int_\Omega \left\{ \sum_{i+j=2} \binom{2}{i} (\nabla_{u_1}^i \nabla_{u_2}^j \beta_\ell^{(1)}) \right\} \left\{ \sum_{i+j=2} \binom{2}{i} (\nabla_{u_1}^i \nabla_{u_2}^j \beta_{\ell'}^{(2)}) \right\} du_1 du_2.$$

For the notation simplicity, let $\dot{g}^{-1}(x) = \{g^{-1}(x)\}'$. For the quasi-likelihood function $L\{g^{-1}(x), y\}$, let $q_1(x,y) = \frac{\partial}{\partial x} L\{g^{-1}(x), y\}$ and $q_2(x,y) = \frac{\partial^2}{\partial x^2} L\{g^{-1}(x), y\}$. 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 $\rho_j(x) = \{\dot{g}^{-1}(x)\}^j/[\sigma^2 V\{g^{-1}(x)\}]$, and σ^2 is the dispersion parameter, $j = 1, 2$. Moreover, let

$$\eta(\mathbf{z}, \mathbf{x}, \boldsymbol{w}, \boldsymbol{u}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{j=1}^{p_1} z_j \alpha_j + \sum_{k=1}^{p_2} \gamma_k(x_k) + \sum_{\ell=0}^{p_3} \beta_\ell(\boldsymbol{u}) w_\ell,$$

and denote $\eta_{is}^0 = \eta(\mathbf{Z}_{is}, \mathbf{X}_i, \mathbf{W}_{is}, \mathbf{U}_i; \boldsymbol{\alpha}^0, \boldsymbol{\gamma}^0, \boldsymbol{\beta}^0)$, and $\varepsilon_i = Y_{is} - g^{-1}(\eta_{is}^0)$ be the error term.

Recall that $\Phi_{kJ}(x_k)$, $J \in \mathcal{J}$, are the standardized B-spline basis functions for the kth covariate, where $\mathcal{J}$ is the index set of the basis functions. Thus, $\mathrm{E}\Phi_{kJ}(X_k) = 0$ and $\mathrm{E}\Phi_{kJ}^2(X_k) = 1$. Similarly, we define the standardized Bernstein basis polynomials as $B_M^*(\boldsymbol{u}) = B_M(\boldsymbol{u})/\|B_M\|$, $M \in \mathcal{M}$, where $\mathcal{M}$ is the index set of Bernstein basis functions. Define the approximate space as

$$\mathcal{A} = \Bigg\{ \eta^* : \eta^*(\mathbf{x}, \mathbf{z}, \mathbf{w}, \boldsymbol{u}) = \sum_{j=1}^{p_1} z_j \alpha_j + \sum_{k=1}^{p_2} \sum_{J \in \mathcal{J}} \xi_{kJ} \Phi_{kJ}(x_k) + \sum_{\ell=0}^{p_3} \sum_{M \in \mathcal{M}} \theta_{\ell M} B_M^*(\boldsymbol{u}) w_\ell,$$
$$x_k \in [0,1], \boldsymbol{u} \in \Omega, \xi_{kJ}, \theta_{\ell M}, \alpha_j \in \mathbb{R} \Bigg\}.$$

Let $\varepsilon_{is} = Y_{is} - g^{-1}\{\boldsymbol{\alpha}^{0\top} \mathbf{Z}_{is} + \sum_{k=1}^{p_2} \gamma_k^0(X_{ik}) + \sum_{\ell=0}^{p_3} \beta_\ell^0(\mathbf{U}_i) W_{is\ell}\}$ be the error term.

B.2. *Assumptions*

The following are the technical assumptions needed to facilitate the technical details, though they may not be the weakest conditions.

(A1) For $k = 1, \ldots, p_2$, $\gamma_k^0 \cap \mathcal{D}_k^0$. For $\ell = 0, \ldots, p_3$, $\beta_\ell^0 \in \mathcal{S}^{d+1,\infty}(\Omega)$.

(A2) The density function $f_U(\boldsymbol{u})$ of $\mathbf{U}$ is bounded away from zero and infinity on Ω. The density function $f_X(\boldsymbol{x})$ of $\mathbf{X}$ is absolutely continuously and bounded away from zero and infinity on $[0,1]^{p_2}$.

(A3) The function $q_2(x, y) < 0$, $c_1 < |q_2(x, y)| < C_1$ and $c_2 < |\frac{\partial}{\partial x} q_2(x, y)| < C_2$ for $x \in \mathbb{R}$ and y in the range of the response variable. The functions $V(\cdot)$, $g^{-1}(\cdot)$, the first order derivative of $g^{-1}(\cdot)$ are continuous, and there exist positive constants c_ρ and C_ρ such that $c_\rho \le \rho_2(\cdot) \le C_\rho$. For each $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}, \boldsymbol{u})$, $\mathrm{Var}(Y|\mathbf{X} = \boldsymbol{x}, \mathbf{Z} = \boldsymbol{z}, \mathbf{W} = \boldsymbol{w}, \mathbf{U} = \boldsymbol{u})$ and $g'(\mu(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}, \boldsymbol{u}))$ are nonzero.

(A4) The errors satisfy $\mathrm{E}\{\varepsilon_{is}|\mathbf{X}_i = \boldsymbol{x}, \mathbf{Z}_{is} = \boldsymbol{z}, \mathbf{W}_{is} = \boldsymbol{w}, \mathbf{U}_i = \boldsymbol{u}\} = 0$ and $\mathrm{E}(|\varepsilon_{is}|^{2+\iota}|\mathbf{X}_i = \boldsymbol{x}, \mathbf{Z}_{is} = \boldsymbol{z}, \mathbf{W}_{is} = \boldsymbol{w}, \mathbf{U}_i = \boldsymbol{u}) < \infty$ for some $\iota \in (1/2, \infty)$.

(A5) For any $\ell = 0, \ldots, p_3$, there exists a positive constant C_ℓ such that $|W_\ell| \le C_\ell$. Denote $\mathbf{Q}(\boldsymbol{x}, \boldsymbol{u}) = \mathrm{E}\left\{(1, \mathbf{Z}^\top, \mathbf{W}^\top)^\top (1, \mathbf{Z}^\top, \mathbf{W}^\top) \middle| \mathbf{X} = \boldsymbol{x}, \mathbf{U} = \boldsymbol{u}\right\}$. The eigenvalues $\psi_0(\boldsymbol{x}, \boldsymbol{u}) \le \psi_1(\boldsymbol{x}, \boldsymbol{u}) \le \cdots \le \psi_{p_1+p_3+1}(\boldsymbol{x}, \boldsymbol{u})$ of $\mathbf{Q}(\boldsymbol{x}, \boldsymbol{u})$ are bounded away from 0 and infinity uniformly for all $\boldsymbol{x} \in [0,1]^{p_2}$, $\boldsymbol{u} \in \Omega$; that is, there are positive constants C_1 and C_2 such that $C_1 \le \psi_0(\boldsymbol{x}, \boldsymbol{u}) \le \psi_1(\boldsymbol{x}, \boldsymbol{u}) \le \cdots \le \psi_{p_1+p_3+1}(\boldsymbol{x}, \boldsymbol{u}) \le C_2$ for all $\boldsymbol{x} \in [0,1]^{p_2}$, $\boldsymbol{u} \in \Omega$.

(A6) The triangulation $\triangle$ is π-quasi-uniform, that is, there exists a positive constant π such that $|\triangle|/\rho_\triangle \le \pi$, where $|\triangle| = \max\{|T|, T \in \triangle\}$ and $\rho_\triangle = \min\{\rho_T\}$ with ρ_T being the radius of the largest circle inscribed in T.

(A7) The length of subintervals h for the univariate spline and the triangulation size $|\triangle|$ for bivariate spline satisfy that $(\log n)^{2/5} n^{-1/5} \ll h \ll n^{-1/(4\varrho+1)}$, $\varrho > 1$, $(\log n)^{1/5} n^{-1/10} \ll |\triangle| \ll n^{-1/(4d-2)}, d > 3$; and the smoothness penalty parameter satisfies $\lambda \ll \min\{n^{3/4} h^{3/4} |\triangle|^4, n^{3/4} |\triangle|^{11/2}\}$.

(A8) Let $\eta^0(\mathbf{X},\mathbf{Z},\mathbf{U},\mathbf{W}) = \mathbf{Z}^\top\boldsymbol{\alpha}_t^0 + \sum_{k=1}^{p_2}\gamma_k^0(X_k) + \sum_{\ell=0}^{p_3}\beta_\ell^0(\mathbf{U})W_\ell$. Define

$$\Gamma(\mathbf{x},\mathbf{u},\mathbf{w}) = \frac{\mathrm{E}\left[\rho_2\{\eta^0(\mathbf{Z},\mathbf{X},\mathbf{U},\mathbf{W})\}\mathbf{Z}|\mathbf{X}=\mathbf{x},\mathbf{U}=\mathbf{u},\mathbf{W}=\mathbf{w}\right]}{\mathrm{E}[\rho_2\{\eta^0(\mathbf{Z},\mathbf{X},\mathbf{U},\mathbf{W})\}|\mathbf{X}=\mathbf{x},\mathbf{U}=\mathbf{u},\mathbf{W}=\mathbf{w}]},\ \widetilde{\mathbf{Z}} = \mathbf{Z} - \boldsymbol{\Gamma}(\mathbf{X},\mathbf{U},\mathbf{W}), \tag{B.1}$$

and $\boldsymbol{\Sigma} = \mathrm{E}\{\rho_2(\eta^0)\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\}$. The matrix $\boldsymbol{\Sigma}$ is positive definite.

The above assumptions are mild conditions that can be satisfied in many practical situations. Assumption (A1) is frequently used in the literature of nonparametric estimation. It illustrates the smoothness requirement on the true functional components. The first part of Assumption (A2) requires that the locations of the observations are randomly scattered over the domain Ω. The second part of Assumption (A2) is a mild condition on the joint density of $X_1,\ldots,X_{p_2}$. Assumption (A3) is a standard condition for the quasi-likelihood method; see, for example, Carroll et al. (1997), Wang et al. (2011) and Wang and Cao (2018). Assumptions (A2) and (A4) are similar to Assumptions (A3) and (A4) in Liu et al. (2013). Under Assumption (A5), the vector $(X_{i1},\ldots,X_{ip_2})^\top$ is not multicolinear. Assumption (A6) imposes basic regularity conditions for the triangulation in BPST; see Lai and Wang (2013), Wang et al. (2020) and Mu et al. (2020). The class of triangulations whose smallest angles are larger than some positive constant satisfy this condition. The requirements of smoothing parameters and triangulation size in the BPST estimation are given in the Assumption (A7).

B.3. *Preliminaries*

LEMMA B.1 *Under Assumptions (A2) and (A6), for $k=1,\cdots,p_2$, $J,J'\in\mathcal{J}$, $M,M'\in\mathcal{M}$ and $r\geq 1$,*

$$\mathrm{E}|\Phi_{kJ}(X_{ik})\Phi_{kJ'}(X_{ik})|^r \asymp h^{-r}\mathrm{E}|\varphi_{kJ}(X_{ik})\varphi_{kJ'}(X_{ik})|^r \asymp \begin{cases}0, & |J-J'|>\varrho+1,\\ h^{1-r}, & |J-J'|\leq\varrho+1,\end{cases}$$

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

$$\mathrm{E}\,|\Phi_{kJ}(X_{ik})B_M^*(\mathbf{U}_i)|^r \asymp |\triangle|^{2-r}h^{(2-r)/2},$$

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

Proof. By Assumptions (A2) and (A6), $\|\varphi_{kJ}^0\| \asymp \|\varphi_{kJ}\|_{L_2} \asymp h^{1/2}$ and $\|B_M\| \asymp \|B_M\|_{L_2} \asymp |\triangle|$, which imply that $\Phi_{kJ}\asymp h^{-1/2}\phi_{kJ}$ and $B_M^*\asymp|\triangle|^{-1}B_M$. Then, we have

$$\mathrm{E}|\Phi_{kJ}(X_{ik})\Phi_{kJ'}(X_{ik})|^r \asymp h^{-r}\mathrm{E}|\varphi_{kJ}(X_{ik})\varphi_{kJ'}(X_{ik})|^r \asymp \begin{cases}0, & |J-J'|>\varrho+1,\\ h^{1-r}, & |J-J'|\leq\varrho+1,\end{cases}$$

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

$$\mathrm{E}\,|\Phi_{kJ}(X_{ik})B_M^*(\mathbf{U}_i)|^r \asymp |\triangle|^{-r}h^{-r/2}\mathrm{E}\,|\varphi_{kJ}(X_{ik})B_M(\mathbf{U}_i)|^r \asymp |\triangle|^{2-r}h^{(2-r)/2}.$$

Thus, the desired results are established. ■

LEMMA B.2 *Under Assumption (A6), there exist positive constants c_k, C_k, c_s, C_s, such that, for $k = 1, \cdots, p$,*

$$c_k \sum_{J\in\mathcal{J}} \xi_{kJ}^2 \le \left\| \sum_{J\in\mathcal{J}} \xi_{kJ}\Phi_{kJ} \right\|_{L_2}^2 \le C_k \sum_{J\in\mathcal{J}} \xi_{kJ}^2,\ c_s \sum_{M\in\mathcal{M}} \theta_{\ell M}^2 \le \left\| \sum_{M\in\mathcal{M}} \theta_{\ell M} B_M^* \right\|_{L_2}^2 \le C_s \sum_{M\in\mathcal{M}} \theta_{\ell M}^2.$$

LEMMA B.3 *Let $\eta = \sum_{j=1}^{p_1}\alpha_j z_j + \sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}\Phi_{kJ}(x_k) + \sum_{\ell=0}^{p_3}\sum_{M\in\mathcal{M}}\theta_{\ell M}B_M^*(\boldsymbol{u})w_\ell$. Under Assumptions (A2), (A5) and (A6), there exist positive constants c, C such that*

$$c\left(\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\xi}\|_2^2 + \|\boldsymbol{\theta}\|_2^2\right) \le \|\eta\|^2 \le C\left(\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\xi}\|_2^2 + \|\boldsymbol{\theta}\|_2^2\right),$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p_1})^\top$, $\boldsymbol{\xi} = \{\xi_{kJ}, k = 1, \ldots, p_2, J \in \mathcal{J}\}^\top$, and $\boldsymbol{\theta} = \{\theta_{\ell M}, \ell = 0, \ldots, p_3, M \in \mathcal{M}\}^\top$

Proof. Denote

$$\boldsymbol{a}(\mathbf{X},\mathbf{U}) = \left\{ \sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}\Phi_{kJ}(X_k), \boldsymbol{\alpha}, \sum_{M\in\mathcal{M}}\theta_{1M}B_M^*(\mathbf{U}), \ldots, \sum_{M\in\mathcal{M}}\theta_{p_3 M}B_M^*(\mathbf{U}) \right\},$$

$$\mathbf{Q}(\boldsymbol{x},\boldsymbol{u}) = \mathrm{E}\left\{ (1, \mathbf{Z}^\top, \mathbf{W}^\top)^\top (1, \mathbf{Z}^\top, \mathbf{W}^\top) \middle| \mathbf{X} = \boldsymbol{x}, \mathbf{U} = \boldsymbol{u} \right\}.$$

By Assumptions (A2), (A5) and (A6) and Lemma B.2, we have

$$\begin{aligned}
\|\eta\|^2 &= \mathrm{E}\left[\mathrm{E}\left\{ \sum_{j=1}^{p_1}\alpha_j Z_j + \sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}\Phi_{kJ}(X_k) + \sum_{\ell=0}^{p_3} W_\ell \sum_{M\in\mathcal{M}}\theta_{\ell M}B_M^*(\mathbf{U}) \right\}^2 \middle| (\mathbf{X},\mathbf{U}) \right] \\
&= \mathrm{E}\left\{ \boldsymbol{a}(\mathbf{X},\mathbf{U})\mathbf{Q}(\mathbf{X},\mathbf{U})\boldsymbol{a}(\mathbf{X},\mathbf{U})^\top \right\} \\
&\le C\sum_{j=1}^{p_1}\alpha_j^2 + C\left\| \sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}\Phi_{kJ} \right\|_{L_2}^2 + C\sum_{\ell=0}^{p_3}\left\| \sum_{M\in\mathcal{M}}\theta_{\ell M}B_M^* \right\|_{L_2}^2 \\
&\le C\left(\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\xi}\|_2^2 + \|\boldsymbol{\theta}\|_2^2\right).
\end{aligned}$$

Similarly, we have $\|\eta\|^2 \ge c\sum_{j=1}^{p_1}\alpha_j^2 + c\left\|\sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}\Phi_{kJ}\right\|_{L_2}^2 + c\sum_{\ell=0}^{p_3}\left\|\sum_{M\in\mathcal{M}}\theta_{\ell M}B_M^*\right\|_{L_2}^2$. According to Lemma 1 by Stone (1985), we have

$$\left\| \sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}\Phi_{kJ} \right\|^2 \ge c_0 \sum_{k=1}^{p_2}\left\| \sum_{J\in\mathcal{J}}\xi_{kJ}\Phi_{kJ} \right\|^2.$$

Then, $\|\eta\|^2 \ge c\left(\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\xi}\|_2^2 + \|\boldsymbol{\theta}\|_2^2\right)$ holds. The lemma follows. ∎

LEMMA B.4 (Lemma B.4, Yu et al. (2020)) *For any $k = 1, \ldots, p_2$, $\gamma_k \in \mathcal{H}^{(\varrho)} \cap \mathcal{D}_k^0$, there exist a constant c and a function $\gamma_k^* \in \mathcal{U}_k^0$ such that $\|\gamma_k - \gamma_k^*\|_\infty \le c\|\gamma_k^{(\varrho+1)}\|_\infty h^{\varrho+1}$.*

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

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

(ii) For bi-integer (a_1, a_2) with $0 \le a_1 + a_2 \le d$, there exists a spline function $\beta^{**}(\cdot) \in \mathbb{S}_d^r(\triangle)$ ($d \ge 3r + 2$) such that $\|\nabla_{z_1}^{a_1}\nabla_{z_2}^{a_2}(\beta - \beta^{**})\|_\infty \le C|\triangle|^{d+1-a_1-a_2}|\beta|_{d+1,\infty}$, 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 $d \ge 3r + 2$.

Define basis functions of the approximation space $\mathcal{A}$ as $\eta_j^Z(\mathbf{z}, \mathbf{x}, \boldsymbol{w}, \boldsymbol{u}) = z_j, j = 1, \ldots, p_1$, $\eta_{kJ}^X(\mathbf{z}, \mathbf{x}, \boldsymbol{w}, \boldsymbol{u}) = \Phi_{kJ}(x_j), k = 1, \ldots, p_2, J \in \mathcal{J}$, and $\eta_{\ell M}^W(\mathbf{z}, \mathbf{x}, \boldsymbol{w}, \boldsymbol{u}) = B_M^*(\boldsymbol{u})w_\ell, \ell = 0, \ldots, p_3, M \in \mathcal{M}$. Notice that for any $\eta \in \mathcal{A}$ can be expressed as the linear combination of the above functions.

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

$$\max_{\substack{0\le\ell,\ell'\le p_3\\ M,M'\in\mathcal{M}}} |\langle\eta_{\ell M}^W, \eta_{\ell' M'}^W\rangle_n - \langle\eta_{\ell M}^W, \eta_{\ell' M'}^W\rangle| = O_{\text{a.s.}}\{n^{-1/2}|\triangle|^{-1}(\log n)^{1/2}\}, \tag{B.2}$$

$$\max_{\substack{1\le k,k'\le p_2\\ J,J'\in\mathcal{J}}} |\langle\eta_{kJ}^X, \eta_{k'J'}^X\rangle_n - \langle\eta_{kJ}^X, \eta_{k'J'}^X\rangle| = O_{\text{a.s.}}\{n^{-1/2}h^{-1/2}(\log n)^{1/2}\}, \tag{B.3}$$

$$\max_{1\le j,j'\le p_1} |\langle\eta_j^Z, \eta_{j'}^Z\rangle_n - \langle\eta_j^Z, \eta_{j'}^Z\rangle| = O_{\text{a.s.}}\left\{n^{-1/2}(\log n)^{1/2}\right\}, \tag{B.4}$$

$$\max_{\substack{1\le k\le p_2, 1\le\ell\le p_3\\ M\in\mathcal{M}, J\in\mathcal{J}}} |\langle\eta_{\ell M}^W, \eta_{kJ}^X\rangle_n - \langle\eta_{\ell M}^W, \eta_{kJ}^X\rangle| = O_{\text{a.s.}}\left\{n^{-1/2}(\log n)^{1/2}\right\}, \tag{B.5}$$

$$\max_{\substack{1\le j\le p_1, 0\le\ell\le p_3\\ M\in\mathcal{M}}} |\langle\eta_{\ell M}^W, \eta_j^Z\rangle_n - \langle\eta_{\ell M}^W, \eta_j^Z\rangle| = O_{\text{a.s.}}\left\{n^{-1/2}(\log n)^{1/2}\right\}, \tag{B.6}$$

$$\max_{\substack{1\le j\le p_1, 1\le k\le p_2\\ J'\in\mathcal{J}}} |\langle\eta_j^Z, \eta_{kJ'}^X\rangle_n - \langle\eta_j^Z, \eta_{kJ'}^X\rangle| = O_{\text{a.s.}}\left\{n^{-1/2}(\log n)^{1/2}\right\}. \tag{B.7}$$

Proof. Notice that

$$\begin{aligned}
&\max_{\substack{0\le\ell,\ell'\le p_3\\ M,M'\in\mathcal{M}}} |\langle\eta_{\ell M}^W, \eta_{\ell' M'}^W\rangle_n - \langle\eta_{\ell M}^W, \eta_{\ell' M'}^W\rangle| \\
&= \max_{\substack{0\le\ell,\ell'\le p_3\\ M,M'\in\mathcal{M}}} \left| \frac{1}{nT}\sum_{i=1}^n\sum_{s=t_1}^{t_2} B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}W_{is\ell'} - \mathrm{E}\left\{B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}W_{is\ell'}\right\}\right| \\
&\le \max_{t_1\le s\le t_2}\max_{\substack{0\le\ell,\ell'\le p_3\\ M,M'\in\mathcal{M}}} \left| \frac{1}{n}\sum_{i=1}^n B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}W_{is\ell'} - \mathrm{E}\left\{B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}W_{is\ell'}\right\}\right|.
\end{aligned}$$

Thus, in the following, we derive the order of terms

$$\max_{\substack{0\le\ell,\ell'\le p_3\\ M,M'\in\mathcal{M}}} \left| \frac{1}{n}\sum_{i=1}^n B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}W_{is\ell'} - \mathrm{E}\left\{B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}W_{is\ell'}\right\}\right|.$$

For simplicity, we consider the case $\ell = \ell'$. Let

$$\mathcal{U}_{is\ell MM'} = n^{-1}B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}^2 - n^{-1}\mathrm{E}\left\{B_M^*(\mathbf{U}_i)B_{M'}^*(\mathbf{U}_i)W_{is\ell}^2\right\},$$
$$u_{is\ell MM'} = n^{-1}B_M(\mathbf{U}_i)B_{M'}(\mathbf{U}_i)W_{is\ell}^2 - n^{-1}\mathrm{E}\left\{B_M(\mathbf{U}_i)B_{M'}(\mathbf{U}_i)W_{is\ell}^2\right\}.$$

Then $\mathcal{U}_{is\ell MM'} = \|B_M\|^{-1}\|B_{M'}\|^{-1}u_{is\ell MM'}$, as the basis function is bounded by a constant. Notice that $\mathrm{E}\mathcal{U}_{is,\ell MM'} = \mathrm{E}u_{is\ell MM'} = 0$, and

$$\begin{aligned}\mathrm{E}|\mathcal{U}_{is\ell MM'}|^r &= \|B_M\|^{-r}\|B_{M'}\|^{-r}\mathrm{E}|u_{is\ell MM'}|^r \le (cn^{-1}\|B_M\|^{-1}\|B_{M'}\|^{-1})^{r-2}\mathrm{E}|\mathcal{U}_{is,\ell MM'}|^2\\ &\le (Cn^{-1}|\triangle|^{-2})^{r-2}\mathrm{E}|\mathcal{U}_{is,\ell MM'}|^2.\end{aligned}$$

Thus, $\mathcal{U}_{is\ell MM'}$ satisfies the Cramér's condition with constant $Cn^{-1}|\triangle|^{-2}$. Applying Bernstein inequality to $\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\mathcal{U}_{is,\ell MM'}$, for any $\delta > 0$, one has

$$P\left\{\left|\sum_{i=1}^{n}\mathcal{U}_{is\ell MM'}\right| \ge \delta n^{-1/2}|\triangle|^{-1}\log^{1/2}n\right\} \le 2\exp\left\{\frac{-\delta^2\log n}{4 + C|\triangle|^{-1}n^{-1/2}\log^{1/2}n}\right\}.$$

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

$$\sum_{n=1}^{\infty}P\left\{\max_{\ell,M,M'}\left|\sum_{i=1}^{n}\mathcal{U}_{is\ell MM'}\right| \ge \delta n^{-1/2}|\triangle|^{-1}\log^{1/2}n\right\} \le \sum_{n=1}^{\infty}\sum_{\ell=0}^{p_3}|\mathcal{M}|n^{-2-\tau} < \infty.$$

Thus,

$$\max_{\ell,M,M'}\left|\sum_{i=1}^{n}\mathcal{U}_{is\ell MM'}\right| = O_{\mathrm{a.s.}}\{n^{-1/2}|\triangle|^{-1}\log^{1/2}n\}.$$

Similarly, under Assumptions (A5) and (A6), we have (B.3) and (B.5). The desired results are obtained. ■

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

$$R_n = \sup_{\eta_1,\eta_2\in\mathcal{A}}\left|\frac{\langle\eta_1,\eta_2\rangle_n - \langle\eta_1,\eta_2\rangle}{\|\eta_1\|\|\eta_2\|}\right| = O_{\mathrm{a.s.}}\left\{h^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2}n\right\}, \tag{B.8}$$

and consequently,

$$\sup_{\eta\in\mathcal{A}}\left|\|\eta\|_n^2/\|\eta\|^2 - 1\right| = O_{\mathrm{a.s.}}\left\{h^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2}n\right\}. \tag{B.9}$$

Proof. Without loss of generality, let

$$\eta_1 = \sum_{j=1}^{p_1}\alpha_j^{(1)}z_j + \sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}^{(1)}\Phi_{kJ} + \sum_{\ell=0}^{p_3}\sum_{M\in\mathcal{M}}\theta_{\ell M}^{(1)}B_M^*,\ \ \eta_2 = \sum_{j=1}^{p_1}\alpha_j^{(2)}z_j\sum_{k=1}^{p_2}\sum_{J\in\mathcal{J}}\xi_{kJ}^{(2)}\Phi_{kJ} + \sum_{\ell=0}^{p_3}\sum_{M\in\mathcal{M}}\theta_{\ell M}^{(2)}B_M^*.$$

Denote $\boldsymbol{\alpha}^{(b)} = (\alpha_1^{(b)}, \ldots, \alpha_{p_1}^{(b)})^\top$, $\boldsymbol{\xi}^{(b)} = \left\{\xi_{kJ}^{(b)}, k = 1, \ldots, p_2, J \in \mathcal{J}\right\}^\top$ and $\boldsymbol{\theta}^{(b)} = \left\{\theta_{\ell M}^{(b)}, \ell = 0, \ldots, p_3, M \in \mathcal{M}\right\}^\top$, for $b = 1, 2$. By Lemma B.3, we have

$$\|\eta_1\|\|\eta_2\| \asymp \left(\|\boldsymbol{\alpha}^{(1)}\|_2^2 + \|\boldsymbol{\xi}^{(1)}\|_2^2 + \|\boldsymbol{\theta}^{(1)}\|_2^2\right)^{1/2} \left(\|\boldsymbol{\alpha}^{(2)}\|_2^2 + \|\boldsymbol{\xi}^{(2)}\|_2^2 + \|\boldsymbol{\theta}^{(2)}\|_2^2\right)^{1/2}.$$

Also, notice that

$$\begin{aligned}
\langle \eta_1, \eta_2 \rangle_n - \langle \eta_1, \eta_2 \rangle = & \sum_{k,k'=1}^{p_2} \sum_{J,J' \in \mathcal{J}} \xi_{kJ}^{(1)} \xi_{k'J'}^{(2)} (\langle \eta_{kJ}^X, \eta_{k'J'}^X \rangle_n - \langle \eta_{kJ}^X, \eta_{k'J'}^X \rangle) \\
& + \sum_{\ell,\ell'=0}^{p_3} \sum_{M,M' \in \mathcal{M}} \theta_{\ell M}^{(1)} \theta_{\ell',M'}^{(2)} (\langle \eta_{\ell M}^W, \eta_{\ell' M'}^W \rangle_n - \langle \eta_{\ell M}^W, \eta_{\ell' M'}^W \rangle) + \sum_{j,j'=1}^{p_1} \alpha_j^{(1)} \alpha_{j'}^{(2)} (\langle \eta_j^Z, \eta_{j'}^Z \rangle_n - \langle \eta_j^Z, \eta_{j'}^Z \rangle) \\
& + \sum_{k=1}^{p_2} \sum_{\ell=0}^{p_3} \sum_{J \in \mathcal{J}, M \in \mathcal{M}} (\xi_{kJ}^{(1)} \theta_{\ell M}^{(2)} + \xi_{kJ}^{(2)} \theta_{\ell M}^{(1)}) (\langle \eta_{\ell M}^W, \eta_{kJ}^X \rangle_n - \langle \eta_{\ell M}^W, \eta_{kJ}^X \rangle) \\
& + \sum_{j=1}^{p_1} \sum_{\ell=0}^{p_3} \sum_{M \in \mathcal{M}} (\theta_{\ell M}^{(1)} \alpha_j^{(2)} + \theta_{\ell M}^{(2)} \alpha_j^{(1)}) (\langle \eta_{\ell M}^W, \eta_j^Z \rangle_n - \langle \eta_{\ell M}^W, \eta_j^Z \rangle) \\
& + \sum_{j=1}^{p_1} \sum_{k=1}^{p_2} \sum_{J \in \mathcal{J}} (\xi_{kJ}^{(1)} \alpha_j^{(2)} + \xi_{kJ}^{(2)} \alpha_j^{(1)}) (\langle \eta_j^Z, \eta_{kJ}^X \rangle_n - \langle \eta_j^Z, \eta_{kJ}^X \rangle) \\
= & I_1 + I_2 + I_3 + I_4 + I_5 + I_6.
\end{aligned}$$

By Lemma B.6, we have

$$\begin{aligned}
I_1 & \leq \sum_{k,k'=1}^{p_2} \sum_{|J-J'| \leq \varrho+1} |\xi_{kJ}^{(1)} \xi_{k'J'}^{(2)}| \max_{1 \leq \ell, \ell' \leq p, J, J' \in \mathcal{J}} |\langle \Phi_{kJ}, \Phi_{k'J'} \rangle_n - \langle \Phi_{kJ}, \Phi_{k'J'} \rangle| \\
& = \|\boldsymbol{\xi}^{(1)}\|_2 \|\boldsymbol{\xi}^{(2)}\|_2 \times O_{\text{a.s.}}(n^{-1/2} h^{-1/2} \log^{1/2} n).
\end{aligned}$$

Similarly, we have

$$I_2 = \|\boldsymbol{\theta}^{(1)}\|_2 \|\boldsymbol{\theta}^{(2)}\|_2 \times O_{\text{a.s.}}(n^{-1/2} |\triangle|^{-1} \log^{1/2} n), \;\; I_3 = \|\boldsymbol{\alpha}^{(1)}\|_2 \|\boldsymbol{\alpha}^{(2)}\|_2 \times O_{\text{a.s.}}(n^{-1/2} \log^{1/2} n).$$

By the Cauchy Schwarz inequality, we have

$$\begin{aligned}
& \sum_{k=1}^{p_2} \sum_{\ell=0}^{p_3} \sum_{J \in \mathcal{J}, M \in \mathcal{M}} (|\xi_{kJ}^{(1)} \theta_{\ell M}^{(2)}| + |\xi_{kJ}^{(2)} \theta_{\ell M}^{(1)}|) \\
& \quad = \left(\sum_{k=1}^{p_2} \sum_{J \in \mathcal{J}} |\xi_{kJ}^{(1)}|\right) \left(\sum_{\ell=0}^{p_3} \sum_{M \in \mathcal{M}} |\theta_{\ell M}^{(2)}|\right) + \left(\sum_{k=1}^{p_2} \sum_{J \in \mathcal{J}} |\xi_{kJ}^{(2)}|\right) \left(\sum_{\ell=0}^{p_3} \sum_{M \in \mathcal{M}} |\theta_{\ell M}^{(1)}|\right) \\
& \quad \leq C h^{-1/2} |\triangle|^{-1} \left(\|\boldsymbol{\xi}^{(1)}\|_2 \|\boldsymbol{\theta}^{(2)}\|_2 + \|\boldsymbol{\xi}^{(2)}\|_2 \|\boldsymbol{\theta}^{(1)}\|_2\right),
\end{aligned}$$

which implies that

$$I_4 = \left(\|\boldsymbol{\xi}^{(1)}\|_2\|\boldsymbol{\theta}^{(2)}\|_2 + \|\boldsymbol{\xi}^{(2)}\|_2\|\boldsymbol{\theta}^{(1)}\|_2\right) \times O_{\text{a.s.}}(h^{-1/2}|\triangle|^{-1}n^{-1/2}\log^{1/2} n).$$

Similarly, we obtain

$$I_5 = \left(\|\boldsymbol{\alpha}^{(1)}\|_2\|\boldsymbol{\theta}^{(2)}\|_2 + \|\boldsymbol{\alpha}^{(2)}\|_2\|\boldsymbol{\theta}^{(1)}\|_2\right) \times O_{\text{a.s.}}(h^{-1/2}n^{-1/2}\log^{1/2} n),$$
$$I_6 = \left(\|\boldsymbol{\alpha}^{(1)}\|_2\|\boldsymbol{\xi}^{(2)}\|_2 + \|\boldsymbol{\alpha}^{(2)}\|_2\|\boldsymbol{\xi}^{(1)}\|_2\right) \times O_{\text{a.s.}}(|\triangle|^{-1}n^{-1/2}\log^{1/2} n).$$

Combining $I_1 - I_6$, we obtain (B.8). As a direct result of (B.8), we obtain that (B.9) ■

Next, denote

$$\mathbf{B}^*(\mathbf{U}_i) = \{B_M^*(\mathbf{U}_i), M \in \mathcal{M}\},\ \widetilde{\mathbf{B}}(\mathbf{W}_{is}, \mathbf{U}_i) = \mathbf{W}_{is} \otimes \mathbf{Q}_2^\top \mathbf{B}^*(\mathbf{U}_i), \tag{B.10}$$

and

$$\mathbf{F}_{is} = \begin{pmatrix} \mathbf{A}_{is} \\ \boldsymbol{\Phi}(\mathbf{X}_i) \\ \widetilde{\mathbf{B}}(\mathbf{W}_{is}, \mathbf{U}_i) \end{pmatrix}, \quad \mathbb{D}_\lambda = \lambda \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_2^\top \mathbf{P}\mathbf{Q}_2 \end{pmatrix}. \tag{B.11}$$

Denote

$$\mathbf{H}_{n,0}^* = \frac{1}{n_T}\sum_{i=1}^n \sum_{s=t_1}^{t_2} \mathbf{F}_{is}\mathbf{F}_{is},\ \mathbf{H}_n^* \equiv \mathbf{H}_{n,\lambda}^* = \frac{1}{n_T}\sum_{i=1}^n \sum_{s=t_1}^{t_2} \mathbf{F}_{is}\mathbf{F}_{is} + \mathbb{D}_\lambda. \tag{B.12}$$

LEMMA B.8 *Under Assumptions (A2) and (A5)–(A7), there exist constants* $0 < c_H < C_H < \infty$*, such that* $c_H \le \lambda_{\min}(\mathbf{H}_{n,\lambda}^*) \le \lambda_{\max}(\mathbf{H}_{n,\lambda}^*) \le C_H$*, almost surely, for large enough* n*.*

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

$$\boldsymbol{\vartheta}\mathbf{H}_{n,0}^*\boldsymbol{\vartheta}^\top = \frac{1}{n_T}\sum_{i=1}^n \sum_{s=t_1}^{t_2} \boldsymbol{\vartheta}\mathbf{F}_{is}\mathbf{F}_{is}^\top\boldsymbol{\vartheta}^\top = \|\boldsymbol{g}_{\boldsymbol{\vartheta}}\|_n^2,$$

where $\boldsymbol{g}(\mathbf{Z}, \mathbf{X}, \mathbf{W}, \mathbf{U}; \boldsymbol{\vartheta}) = \boldsymbol{\alpha}^\top\mathbf{Z} + \boldsymbol{\xi}^\top\boldsymbol{\Phi}(\mathbf{X}) + \boldsymbol{\theta}^{*\top}\widetilde{\mathbf{B}}(\mathbf{W}, \mathbf{U})$. By Lemmas B.3 and B.6, we have

$$c(1 - R_n)\|\boldsymbol{\vartheta}\|_2^2 \le (1 - R_n)\|\boldsymbol{g}_{\boldsymbol{\vartheta}}\|^2 \le \|\boldsymbol{g}_{\boldsymbol{\vartheta}}\|_n^2,$$
$$\|\boldsymbol{g}_{\boldsymbol{\vartheta}}\|_n^2 \le (1 + R_n)\|\boldsymbol{g}_{\boldsymbol{\vartheta}}\|^2 \le C(1 + R_n)\|\boldsymbol{\vartheta}\|_2^2.$$

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

$$c\|\boldsymbol{\vartheta}\|_2^2 \le \boldsymbol{\vartheta}\mathbf{H}_{n,0}^*\boldsymbol{\vartheta} \le C\|\boldsymbol{\vartheta}\|_2^2, \tag{B.13}$$

almost surely, for large enough n.

The (M, M')th element of $\mathbf{P}$ is $P_{M,M'} = \int \nabla^2_{u_1} B^*_M(\boldsymbol{u}) \nabla^2_{u_1} B^*_{M'}(\boldsymbol{u}) + \nabla^2_{u_2} B^*_M(\boldsymbol{u}) \nabla^2_{u_2} B^*_{M'}(\boldsymbol{u}) + 2\nabla^2_{u_1u_2} B^*_M(\boldsymbol{u}) \nabla^2_{u_1u_2} B^*_{M'}(\boldsymbol{u}) d\boldsymbol{u}$. By Theorem 2.19 in Lai and Schumaker (2007), we have

$$P_{M,M'} \asymp \begin{cases} |\triangle|^{-4}, & \lceil M/d^* \rceil = \lceil M'/d^* \rceil, \\ 0, & \lceil M/d^* \rceil \neq \lceil M'/d^* \rceil. \end{cases} \tag{B.14}$$

Then, by the Assumption (A6), we have

$$\boldsymbol{\vartheta} \mathbb{D}_\lambda \boldsymbol{\vartheta}^\top = O\{\lambda |\triangle|^{-4} n^{-1} \|\boldsymbol{\vartheta}\|_2^2\} = o(\|\boldsymbol{\vartheta}\|_2^2). \tag{B.15}$$

The desired result follows (B.13) and (B.15). ■

B.4. *Consistency of Penalized Quasi-likelihood Estimators*

By Lemmas B.4 and B.5, there exist

$$\widetilde{\gamma}_k(x_k) = \boldsymbol{\Phi}_k{}^\top(x_k) \widetilde{\boldsymbol{\xi}}_k, \quad \widetilde{\beta}_\ell(\boldsymbol{u}) = \widetilde{\mathbf{B}}^\top(\boldsymbol{u}) \widetilde{\boldsymbol{\theta}}^*_\ell, \tag{B.16}$$

which are the best approximation to γ_k's and β with the approximation rate at $\|\gamma_k - \widetilde{\gamma}_k\|_\infty \leq C_k \|\gamma_k^{(\varrho+1)}\|_\infty h^{\varrho+1}$ and $\|\beta_\ell - \widetilde{\beta}_\ell\|_\infty \leq C_{\beta_\ell} |\beta_\ell|_{d+1,\infty} |\triangle|^{d+1}$. Denote $\widetilde{\boldsymbol{\xi}} = (\widetilde{\boldsymbol{\xi}}_1^\top, \cdots, \widetilde{\boldsymbol{\xi}}_{p_2}^\top)^\top$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_0^{*\top}, \cdots, \widetilde{\boldsymbol{\theta}}_{p_3}^{*\top})^\top$.

Denote that $\eta_{is}(\boldsymbol{\vartheta}) = \mathbf{Z}_{is}^\top \boldsymbol{\alpha} + \boldsymbol{\Phi}(\mathbf{X}_i)^\top \boldsymbol{\xi} + \widetilde{\mathbf{B}}(\mathbf{W}_{is}, \mathbf{U}_i)^\top \boldsymbol{\theta}^* = \mathbf{F}_{is} \boldsymbol{\vartheta}$, which is a function with respect to subject i at time s, where $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\xi}^\top, \boldsymbol{\theta}^{*\top})^\top$. We have

$$\begin{aligned} \nabla L_n^{\mathrm{P}}(\boldsymbol{\vartheta}) &= -\sum_{i=1}^{n} \sum_{s=t_1}^{t_2} q_1\left(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}\right) \mathbf{F}_{is} + \mathbb{D}_\lambda \boldsymbol{\vartheta}, \\ \nabla^2 L_n^{\mathrm{P}}(\boldsymbol{\vartheta}) &= -\sum_{i=1}^{n} \sum_{s=t_1}^{t_2} q_2(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}) \mathbf{F}_{is} \mathbf{F}_{is}^\top + \mathbb{D}_\lambda. \end{aligned}$$

For the notation simplicity, we denote $\eta_{is}^0 = \sum_{j=1}^{p_1} Z_{isj} \alpha_j + \sum_{k=1}^{p_2} \gamma_k(X_{ik}) + \sum_{\ell=0}^{p_3} \beta_\ell(\mathbf{U}_i) W_{is\ell}$, $\widehat{\eta}_{is} = \eta_{is}(\widehat{\boldsymbol{\vartheta}})$ and $\widetilde{\eta}_{is} = \eta_{is}(\widetilde{\boldsymbol{\vartheta}})$.

LEMMA B.9 *Under Assumptions (A1)–(A7), we have*

$$\begin{aligned} \left| \frac{1}{n_T} \nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}) \right| &= O_{a.s} \left\{ \left(\frac{\log n}{n} \right)^{1/2} + h^{\varrho+3/2} + |\triangle|^{d+2} + h^{\varrho+1} |\triangle| + h^{1/2} |\triangle|^{d+1} + \frac{\lambda}{n |\triangle|^3} \right\}, \\ \left\| \frac{1}{n_T} \nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}) \right\| &= O_{a.s} \left\{ (h^{-1/2} + |\triangle|^{-1}) \left(\frac{\log n}{n} \right)^{1/2} + h^{\varrho+1} + |\triangle|^{d+1} + \frac{\lambda}{n |\triangle|^4} \right\}. \end{aligned}$$

Proof. Let $\widetilde{\eta}(\mathbf{Z}, \mathbf{X}, \mathbf{W}, \mathbf{U}; \widetilde{\boldsymbol{\vartheta}}) = \mathbf{Z}^\top \widetilde{\boldsymbol{\alpha}} + \boldsymbol{\Phi}(\mathbf{X})^\top \widetilde{\boldsymbol{\xi}} + \widetilde{\mathbf{B}}(\mathbf{W}, \mathbf{U})^\top \widetilde{\boldsymbol{\theta}}^*$, then $\|\widetilde{\eta} - \eta\|_\infty = O(h^{\varrho+1} +$

$|\triangle|^{d+1}$). By Assumption (A2), we have

$$\begin{aligned}
n_T^{-1}\nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}) &= -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_1\left(\widetilde{\eta}_{is}, Y_{is}\right)\mathbf{F}_{is} + \frac{1}{n_T}\mathbb{D}_\lambda\widetilde{\boldsymbol{\vartheta}} \\
&= -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \frac{Y_{is}-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\widetilde{\eta}_{is}))}\dot{g}^{-1}(\widetilde{\eta}_{is})\mathbf{F}_{is} + \frac{1}{n_T}\mathbb{D}_\lambda\widetilde{\boldsymbol{\vartheta}} \\
&= -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \frac{Y_{is}-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\left\{1+o(1)\right\}\mathbf{F}_{is} + \frac{1}{n_T}\mathbb{D}_\lambda\widetilde{\boldsymbol{\vartheta}} \\
&= \mathbf{V}_v\left\{1+o(1)\right\} + \mathbf{V}_b\left\{1+o(1)\right\} + \mathbf{V}_p,
\end{aligned}$$

where

$$\mathbf{V}_v = -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \frac{Y_{is}-g^{-1}(\eta_{is}^0)}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\mathbf{F}_{is},$$

$$\mathbf{V}_b = -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\mathbf{F}_{is},\ \mathbf{V}_p = \frac{1}{n_T}\mathbb{D}_\lambda\widetilde{\boldsymbol{\vartheta}}. \tag{B.17}$$

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

$$\mathrm{E}\left[\frac{Y_{is}-g^{-1}(\eta_{is}^0)}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\right] = 0, \quad \mathrm{E}\left[\frac{Y_{is}-g^{-1}(\eta_{is}^0)}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\right]^2 = O(1).$$

According to Assumption (A3) and applying the Bernstein inequality, for any $k = 1,\ldots,p_2$ and $J \in \mathcal{J}$, we obtain

$$\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \frac{Y_{is}-g^{-1}(\eta_{is}^0)}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik}) = O_{\text{a.s.}}\left\{\left(\frac{\log n}{n}\right)^{1/2}\right\}. \tag{B.18}$$

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

$$\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \frac{Y_{is}-g^{-1}(\eta_{is}^0)}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)W_{is\ell}B_M^*(\mathbf{U}_i) = O_{\text{a.s.}}\left\{\left(\frac{\log n}{n}\right)^{1/2}\right\}. \tag{B.19}$$

For the vector $\mathbf{V}_b$, we focus on the term

$$n_T^{-1}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik}).$$

We write

$$\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})=\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\xi_{is}$$
$$+\mathrm{E}\left[\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\right],$$

where

$$\xi_{is}=\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{n\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})-\mathrm{E}\left[\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{n\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\right].$$

Note that,

$$\mathrm{E}\left[\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\right]=O\left\{\left(h^{\varrho+1}+|\triangle|^{d+1}\right)h^{1/2}\right\}. \tag{B.20}$$

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

$$\mathrm{E}|\xi_{is}|^r\leq\left\{Cn^{-1}|g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})|\|\Phi_{kJ}\|_\infty\right\}^{r-2}\mathrm{E}|\xi_{is}|^2\leq\left\{Cn^{-1}(h^{\varrho+1}+|\triangle|^{d+1})h^{-1/2}\right\}^{r-2}\mathrm{E}|\xi_{is}|^2,$$

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

$$\begin{aligned}\mathrm{E}|\xi_{is}|^2=&\mathrm{E}\left|\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{n\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\right|^2-\left[\mathrm{E}\left\{\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{n\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\right\}\right]^2\\=&O\left\{n^{-2}(h^{2\varrho+2}+|\triangle|^{2d+2})\right\}.\end{aligned}$$

Applying the Bernstein inequality, we have

$$P\left\{\left|\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\xi_{is}\right|\geq\delta(h^{\varrho+1}+|\triangle|^{d+1})n^{-1/2}\log^{1/2}n\right\}\leq 2\exp\left\{\frac{-\delta^2\log n}{4+Ch^{-1/2}n^{-1/2}\log^{1/2}n}\right\}.$$

Consequently,

$$\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\xi_{is}=O_{\text{a.s.}}\left\{(h^{\varrho+1}+|\triangle|^{d+1})n^{-1/2}\log^{1/2}n\right\}. \tag{B.21}$$

Combining (B.20) and (B.21), we obtain

$$\begin{aligned}&\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)\Phi_{kJ}(X_{ik})\\&=O_{\text{a.s.}}\left\{(h^{\varrho+1}+|\triangle|^{d+1})n^{-1/2}\log^{1/2}n+\left(h^{\varrho+1}+|\triangle|^{d+1}\right)h^{1/2}\right\}.\end{aligned} \tag{B.22}$$

Similarly,

$$\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\frac{g^{-1}(\eta_{is}^0)-g^{-1}(\widetilde{\eta}_{is})}{\sigma^2 V(g^{-1}(\eta_{is}^0))}\dot{g}^{-1}(\eta_{is}^0)W_{is\ell}B_M^*(\mathbf{U}_i)$$
$$= O_{\text{a.s.}}\left\{(h^{\varrho+1}+|\triangle|^{d+1})n^{-1/2}\log^{1/2}n+\left(h^{\varrho+1}+|\triangle|^{d+1}\right)|\triangle|\right\}. \tag{B.23}$$

For the vector $\mathbf{V}_p$, by (B.14) and $|\widetilde{\boldsymbol{\theta}}^*|\asymp|\triangle|$, then we have

$$|\mathbf{V}_p| = O(\lambda n^{-1}|\triangle|^{-3}), \qquad \|\mathbf{V}_p\| = O(\lambda n^{-1}|\triangle|^{-4}). \tag{B.24}$$

Combining (B.18), (B.19), (B.22), (B.23) and (B.24), we obtain

$$\left|\frac{1}{n_T}\nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}})\right| = O_{a.s}\left\{\left(\frac{\log n}{n}\right)^{1/2}+h^{\varrho+3/2}+|\triangle|^{d+2}+h^{\varrho+1}|\triangle|+h^{1/2}|\triangle|^{d+1}+\frac{\lambda}{n|\triangle|^3}\right\},$$
$$\left\|\frac{1}{n_T}\nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}})\right\| = O_{a.s}\left\{(h^{-1/2}+|\triangle|^{-1})\left(\frac{\log n}{n}\right)^{1/2}+h^{\varrho+1}+|\triangle|^{d+1}+\frac{\lambda}{n|\triangle|^4}\right\}.$$

Therefore, Lemma B.9 has been established. ■

LEMMA B.10 *If $\bar{\boldsymbol{\vartheta}}$ is the vector that satisfies $\left\|\bar{\boldsymbol{\vartheta}}-\widetilde{\boldsymbol{\vartheta}}\right\| = O_{\text{a.s.}}(h^{1/2}+|\triangle|)$, then, under Assumptions (A2), (A3), and (A5) – (A7), there exists constants c and C, such that*

$$c\mathbf{I}\le\left\{n_T^{-1}\nabla^2 L_n^{\mathrm{P}}(\bar{\boldsymbol{\vartheta}})\right\}^{-1}\le C\mathbf{I},$$

almost surely, for large enough n.

Proof. Let $\bar{\eta}_{is}=\eta_{is}(\bar{\boldsymbol{\vartheta}})$, we have

$$\begin{aligned}
n^{-1}\nabla^2 L_n^{\mathrm{P}}(\bar{\boldsymbol{\vartheta}}) &= -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\bar{\eta}_{is},Y_{is})\mathbf{F}_{is}\mathbf{F}_{is}^{\top}+n_T^{-1}\mathbb{D}_\lambda\\
&= -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\{Y_{is}-g^{-1}(\eta_{is}^0)\}\rho_1'(\bar{\eta}_{is})\\
&\quad+\{-g^{-1}(\bar{\eta}_{is})+g^{-1}(\eta_{is}^0)\}\rho_1'(\bar{\eta}_{is})-\rho_2(\bar{\eta}_{is})\mathbf{F}_{is}\mathbf{F}_{is}^{\top}+n_T^{-1}\mathbb{D}_\lambda\\
&= -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\left[\rho_2\{g^{-1}(\bar{\eta}_i)\}\mathbf{F}_{is}\mathbf{F}_{is}^{\top}\right]\{1+o_{a.s}(1)\}+n_T^{-1}\mathbb{D}_\lambda.
\end{aligned}$$

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

$$c\mathbf{I}\le\left[\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\rho_2(\bar{\eta}_i)\mathbf{F}_{is}\mathbf{F}_{is}^{\top}+n_T^{-1}\mathbb{D}_\lambda\right]\le C\mathbf{I},$$

almost surely, for large enough n. ■

Proof of Theorem 3.1. We now prove that

$$\left\|\widehat{\boldsymbol{\vartheta}}-\widetilde{\boldsymbol{\vartheta}}\right\| = O_{\mathrm{a.s.}}\left\{(h^{-1/2}+|\triangle|^{-1})\left(\frac{\log n}{n}\right)^{1/2}+h^{\varrho+1}+|\triangle|^{d+1}+\frac{\lambda}{n|\triangle|^{4}}\right\}. \tag{B.25}$$

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

$$\widehat{\boldsymbol{\vartheta}}-\widetilde{\boldsymbol{\vartheta}} = -\left\{\nabla^2 L_n^{\mathrm{P}}(\bar{\boldsymbol{\vartheta}})\right\}^{-1}\nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}),$$

where $\bar{\boldsymbol{\vartheta}}$ is some value between $\widehat{\boldsymbol{\vartheta}}$ and $\widetilde{\boldsymbol{\vartheta}}$. Then (B.25) is obtained from Lemmas B.9 and B.10. Theorem 3.1 is obtained from Lemma B.2 and Assumption (A6). ∎

For any $\Phi\in\mathcal{U}_k$, $\psi\in\mathbb{S}_d^r$, , one has $\|\Phi\|_\infty\le Ch^{-1/2}\|\Phi\|$, $\|\psi\|_\infty\le C|\triangle|^{-1}\|\psi\|$. Then

$$\sum_{k=1}^{p_2}\|\widehat{\gamma}_k-\widetilde{\gamma}_k\|_\infty+\sum_{\ell=0}^{p_3}\|\widehat{\beta}_\ell-\widetilde{\beta}_\ell\|_\infty = O_{\mathrm{a.s.}}\left\{(h^{-1}+|\triangle|^{-2})\left(\frac{\log n}{n}\right)^{1/2}+h^{\varrho+1/2}+|\triangle|^{d}+\frac{\lambda}{n}|\triangle|^{-5}\right\}.$$

Notice that $\|\widehat{\gamma}_k-\gamma_k^0\|_\infty\le\|\widehat{\gamma}_k-\widetilde{\gamma}_k\|_\infty+\|\widetilde{\gamma}_k-\gamma_k^0\|_\infty$, for $k=1,\ldots,p_2$, and $\|\widehat{\beta}_\ell-\beta_\ell^0\|_\infty\le\|\widehat{\beta}_\ell-\widetilde{\beta}_\ell\|_\infty+\|\widetilde{\beta}_\ell-\beta_\ell^0\|_\infty$, for $\ell=0,\ldots,p_3$. Consequently,

$$\sum_{k=1}^{p_2}\|\widehat{\gamma}_k-\gamma_k^0\|_\infty+\sum_{\ell=0}^{p_3}\|\widehat{\beta}_\ell-\beta_\ell^0\|_\infty = O_{\mathrm{a.s.}}\left\{(h^{-1}+|\triangle|^{-2})\left(\frac{\log n}{n}\right)^{1/2}+h^{\varrho+1/2}+|\triangle|^{d}+\frac{\lambda}{n|\triangle|^{5}}\right\}.$$

B.5. *Normality for linear coefficients*

We denote the score function and hessian matrix by

$$\mathbf{S}_n(\boldsymbol{\vartheta}) = -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_1\left(\eta_{is}(\boldsymbol{\vartheta}),Y_{is}\right)\mathbf{F}_{is}+\mathbb{D}_\lambda\boldsymbol{\vartheta},$$

$$\mathbf{H}_n(\boldsymbol{\vartheta}) = \frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\mathbf{F}_{is}\mathbf{F}_{is}^\top+\mathbb{D}_\lambda,$$

respectively, where $\mathbb{D}_\lambda$ is defined in (B.11).

Let $\widetilde{\boldsymbol{\vartheta}}^0 = (\boldsymbol{\alpha}^{0\top},\widetilde{\boldsymbol{\xi}}^\top,\widetilde{\boldsymbol{\theta}}^{*\top})^\top$ and $\widetilde{\boldsymbol{\varrho}} = (\widetilde{\boldsymbol{\xi}}^\top,\widetilde{\boldsymbol{\theta}}^{*\top})^\top$, and $\boldsymbol{\Xi}_{is}\equiv\boldsymbol{\Xi}(\mathbf{X}_i,\mathbf{U}_i) = \left(\boldsymbol{\Phi}(\mathbf{X}_i)^\top,\widetilde{\mathbf{B}}(\mathbf{W}_{is},\mathbf{U}_i)^\top\right)^\top$, then we can rewrite

$$\mathbf{S}_n(\boldsymbol{\vartheta}) = \begin{pmatrix}\mathbf{S}_{n,\boldsymbol{\alpha}}(\boldsymbol{\vartheta})\\ \mathbf{S}_{n,\boldsymbol{\varrho}}(\boldsymbol{\vartheta})\end{pmatrix} = -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_1\left(\eta_{is}(\boldsymbol{\vartheta}),Y_{is}\right)\begin{pmatrix}\mathbf{Z}_{is}\\ \boldsymbol{\Phi}(\boldsymbol{\Xi}_{is})\end{pmatrix}+\begin{pmatrix}\mathbf{0}\\ \mathbb{D}_\lambda^*\boldsymbol{\varrho}\end{pmatrix}, \tag{B.26}$$

$$\begin{aligned}\mathbf{H}_n(\boldsymbol{\vartheta}) &= \begin{pmatrix}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\vartheta}) & \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\boldsymbol{\vartheta})\\ \mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\alpha}}(\boldsymbol{\vartheta}) & \mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}(\boldsymbol{\vartheta})\end{pmatrix} \qquad (\mathrm{B.27})\\ &= \begin{pmatrix}\frac{1}{n_T}\sum_{i,s}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\mathbf{Z}_{is}\mathbf{Z}_{is}^\top & \frac{1}{n_T}\sum_{i,s}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\mathbf{Z}_{is}\boldsymbol{\Xi}_{is}^\top\\ \frac{1}{n_T}\sum_{i,s}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\boldsymbol{\Xi}_{is}\mathbf{Z}_{is}^\top & \frac{1}{n_T}\sum_{i,s}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\boldsymbol{\Xi}_{is}\boldsymbol{\Xi}_{is}^\top\end{pmatrix}+\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbb{D}_\lambda^*\end{pmatrix},\end{aligned}$$

where

$$\mathbb{D}_\lambda^* = \lambda \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \end{pmatrix}, \quad \mathbb{D}_\lambda = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbb{D}_\lambda^* \end{pmatrix}.$$

The inverse of $\mathbf{H}_n(\boldsymbol{\vartheta})$ can be represented as

$$\mathbf{H}_n^{-1}(\boldsymbol{\vartheta}) \equiv \mathbb{V} = \begin{pmatrix} \mathbf{V}_{11} & -\mathbf{V}_{11}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}^{-1} \\ -\mathbf{V}_{22}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}^\top \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1} & \mathbf{V}_{22} \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{V}_{11}^{-1} =& \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}} - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}^{-1}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}^\top = \left\{ \frac{1}{n_T} \sum_{i,s} q_2(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}) \mathbf{Z}_{is}\mathbf{Z}_{is}^\top \right\} \\
& - \left\{ \frac{1}{n_T} \sum_{i,s} q_2(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}) \mathbf{Z}_{is}\boldsymbol{\Xi}_{is}^\top \right\} \left\{ \frac{1}{n_T} \sum_{i,s} q_2(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}) \boldsymbol{\Xi}_{is}\boldsymbol{\Xi}_{is}^\top + \mathbb{D}_\lambda^* \right\}^{-1} \\
& \times \left\{ \frac{1}{n_T} \sum_{i,s} q_2(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}) \boldsymbol{\Xi}_{is}\mathbf{Z}_{is}^\top \right\},
\end{aligned}$$

and $\mathbf{V}_{22}^{-1} = \mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}} - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}^\top \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}$.

Then, by the mean value theorem, we obtain

$$\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0 = -\left\{ \nabla^2 L_n^{\mathrm{P}}(\bar{\boldsymbol{\vartheta}}) \right\}^{-1} \nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}^0),$$

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0 = \left(\mathbf{1}^\top \ \mathbf{0}^\top \right) \left(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0 \right) \\
&= -\left(\mathbf{1}^\top \ \ \mathbf{0}^\top \right) \begin{pmatrix} \mathbf{V}_{11} & -\mathbf{V}_{11}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}^{-1} \\ -\mathbf{V}_{22}\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\alpha}}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1} & \mathbf{V}_{22} \end{pmatrix} \times \\
&\qquad \left\{ \begin{pmatrix} \frac{1}{n_T}\sum_{i,s} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\mathbf{Z}_{is} \\ \frac{1}{n_T}\sum_{i,s} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\boldsymbol{\Xi}_{is} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbb{D}_\lambda^* \widetilde{\boldsymbol{\varrho}} \end{pmatrix} \right\} \\
&= -\mathbf{V}_{11} \left(\mathbf{I} \ \ -\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}^{-1} \right) \left\{ \begin{pmatrix} \frac{1}{n_T}\sum_{i,s} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\mathbf{Z}_{is} \\ \frac{1}{n_T}\sum_{i,s} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\boldsymbol{\Xi}_{is} + \mathbb{D}_\lambda^* \widetilde{\boldsymbol{\varrho}} \end{pmatrix} \right\} \\
&= -\mathbf{V}_{11} \left[\left\{ \frac{1}{n_T} \sum_{i,s} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is}) \mathbf{Z}_{is} \right\} - \left\{ \frac{1}{n_T} \sum_{i,s} q_2(\eta_{is}(\bar{\boldsymbol{\vartheta}}), Y_{is}) \mathbf{Z}_{is}\boldsymbol{\Xi}_{is}^\top \right\} \right. \\
&\qquad \left. \times \left\{ \frac{1}{n_T} \sum_{i,s} q_2(\eta_{is}(\bar{\boldsymbol{\vartheta}}), Y_{is}) \boldsymbol{\Xi}_{is}\boldsymbol{\Xi}_{is}^\top + \mathbb{D}_\lambda^* \right\}^{-1} \left\{ \frac{1}{n_T} \sum_{i,s} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is}) \boldsymbol{\Xi}_{is} + \mathbb{D}_\lambda^* \widetilde{\boldsymbol{\varrho}} \right\} \right]
\end{aligned}$$

Recall that

$$\nabla L_n^{\mathrm{P}}(\boldsymbol{\vartheta}) = -\sum_{i,s} q_1\left(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}\right)\mathbf{F}_{is} + \mathbb{D}_\lambda \boldsymbol{\vartheta},$$

$$\nabla^2 L_n^{\mathrm{P}}(\boldsymbol{\vartheta}) = -\sum_{i,s} q_2(\eta_{is}(\boldsymbol{\vartheta}), Y_{is})\mathbf{F}_{is}\mathbf{F}_{is}^\top + \mathbb{D}_\lambda.$$

$$\frac{\partial}{\partial \vartheta_j}\nabla^2 L_n^{\mathrm{P}}(\boldsymbol{\vartheta}) = -\sum_{i,s} q_3(\eta_{is}(\boldsymbol{\vartheta}), Y_{is})\left(\mathbf{F}_{is}\mathbf{F}_{is}^\top\right) F_{isj}.$$

Let d_n be the length of the vector $\boldsymbol{\vartheta}$. By abuse of notation, we have

$$(\boldsymbol{\vartheta} - \widetilde{\boldsymbol{\vartheta}}^0)^\top\, \nabla^3 L_n^{\mathrm{P}}\left(\boldsymbol{\vartheta}\right)\big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}(\boldsymbol{\vartheta} - \widetilde{\boldsymbol{\vartheta}}^0) = \left[(\boldsymbol{\vartheta} - \widetilde{\boldsymbol{\vartheta}}^0)^\top \left\{\frac{\partial^3 L_n^{\mathrm{P}}\left(\boldsymbol{\vartheta}\right)}{\partial\vartheta_j \partial\vartheta_k \partial\vartheta_l}\right\}_{k=1,l=1}^{d_n\ \ d_n}\Bigg|_{\vartheta_j=\vartheta_j^*} (\boldsymbol{\vartheta} - \widetilde{\boldsymbol{\vartheta}}^0)\right]_{j=1}^{d_n}$$

where $\boldsymbol{\vartheta}^* = \{\vartheta_j^*\}_{j=1}^{d_n}$, $\vartheta_j^* = t_j\vartheta_j + (1-t_j)\widetilde{\vartheta}_j^0$ for some t_j, $j=1,\ldots,d_n$.

By Taylor expansion,

$$\nabla L_n^{\mathrm{P}}(\widehat{\boldsymbol{\vartheta}}) - \nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}^0) = \nabla^2 L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}^0)(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0) + \frac{1}{2}(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0)^\top\, \nabla^3 L_n^{\mathrm{P}}\left(\boldsymbol{\vartheta}\right)\big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0).$$

Since $\nabla L_n^{\mathrm{P}}(\widehat{\boldsymbol{\vartheta}}) = \mathbf{0}$,

$$-\nabla L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}^0) = \nabla^2 L_n^{\mathrm{P}}(\widetilde{\boldsymbol{\vartheta}}^0)(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0) + \frac{1}{2}(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0)^\top\, \nabla^3 L_n^{\mathrm{P}}\left(\boldsymbol{\vartheta}\right)\big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}\,(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0).$$

According to the Cauchy-Schwarz inequality, one has

$$\left\|\frac{1}{n_T}(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0)^\top \nabla^3 L_n^{\mathrm{P}}\left(\boldsymbol{\vartheta}\right)(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0)\right\|^2 \le \|\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0\|^4 \frac{1}{n_T^2}\sum_{j,k,l=1}^{d_n}\left\{\frac{\partial^3 L_n^{\mathrm{P}}\left(\boldsymbol{\vartheta}\right)}{\partial\vartheta_j\partial\vartheta_k\partial\vartheta_l}\right\}^2.$$

Equation (B.25) implies that

$$\begin{aligned}
&\left\|\frac{1}{n_T}(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0)^\top \nabla^3 L_n^{\mathrm{P}}\left(\boldsymbol{\vartheta}\right)(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0)\right\|^2 \le \|\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0\|^4 \times O_P(d_n^3) \\
&= O_{\text{a.s.}}\left\{(h^{-2} + |\triangle|^{-4})\left(\frac{\log n}{n}\right)^2 + h^{4(\varrho+1)} + |\triangle|^{4(d+1)} + \frac{\lambda^4}{n^4|\triangle|^{16}}\right\} \times O_P(d_n^3) \\
&= o_P(n^{-1}).
\end{aligned}$$

Similarly, we have

$$-\left\{\mathbf{S}_n(\widetilde{\boldsymbol{\vartheta}}^0) + O_P\left(n^{-1}d_n\right)\right\} = \left\{\mathbf{H}_n(\widetilde{\boldsymbol{\vartheta}}^0) + o_P(1)\right\}(\widehat{\boldsymbol{\vartheta}} - \widetilde{\boldsymbol{\vartheta}}^0) + o_P(n^{-1/2}),$$

where $\mathbf{S}_n(\boldsymbol{\vartheta})$ and $\mathbf{H}_n(\boldsymbol{\vartheta})$ are defined in (B.26) and (B.27). Thus,

$$\begin{aligned}\begin{pmatrix}\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\\ \widehat{\boldsymbol{\varrho}}-\widetilde{\boldsymbol{\varrho}}\end{pmatrix} = -&\left[\begin{pmatrix}\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) & \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\\ \mathbf{H}^{\top}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0) & \mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\end{pmatrix} + o_P(1)\right]^{-1}\\ &\times\left[\mathbf{S}_n(\widetilde{\boldsymbol{\vartheta}}^0) + O_P\left(n^{-1}d_n\right)\right] + o_P(n^{-1/2}),\end{aligned}$$

which leads to

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0 = &\left\{\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}^{\top}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\right\}^{-1}\left(\mathbf{I}\quad \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\right)\mathbf{S}_n(\widetilde{\boldsymbol{\vartheta}}^0)\\ &+ O_P\left(n^{-1}d_n\right) + o_P(n^{-1/2}).\end{aligned}$$

Note that

$$\left(\mathbf{I}\quad \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\rho}}\left(\boldsymbol{\vartheta}\right)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}\left(\boldsymbol{\vartheta}\right)\right)\mathbf{S}_n\left(\boldsymbol{\vartheta}\right) = \mathbf{S}_{n,\boldsymbol{\alpha}}(\boldsymbol{\vartheta}) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\left(\boldsymbol{\vartheta}\right)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}\left(\boldsymbol{\vartheta}\right)\mathbf{S}_{n,\boldsymbol{\varrho}}(\boldsymbol{\vartheta}).$$

Hence, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)$ is the same as that of

$$\begin{aligned}&\sqrt{n}\left\{\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}^{\top}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\right\}^{-1}\\ &\times\left\{\mathbf{S}_{n,\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\left(\widetilde{\boldsymbol{\vartheta}}^0\right)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}\left(\widetilde{\boldsymbol{\vartheta}}^0\right)\mathbf{S}_{n,\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\right\}.\end{aligned}$$

The desired result follows from Lemmas B.11 and B.12.

LEMMA B.11 *Under Assumptions (A1)–(A8),*

$$\begin{aligned}&\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0)\\ &\qquad = -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}\rho_2(\eta^0_{is})(\mathbf{Z}_{is}-\widehat{\mathbf{Z}}_{is})(\mathbf{Z}_{is}-\widehat{\mathbf{Z}}_{is})^{\top}\{1+o_P(1)\},\end{aligned}$$

$$\mathbf{S}_{n,\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\left(\boldsymbol{\vartheta}^0\right)\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}\left(\boldsymbol{\vartheta}^0\right)\mathbf{S}_{n,\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0) = \frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}q_1(\eta^0_{is},Y_{is})(\mathbf{Z}_{is}-\widehat{\mathbf{Z}}_{is})\{1+o_P(1)\}$$

where $\widetilde{\boldsymbol{\vartheta}}^0 = (\boldsymbol{\alpha}^{0\top},\widetilde{\boldsymbol{\varrho}}^{\top})^{\top}$ *and* $\widehat{\mathbf{Z}}_{is}$ *is given in (B.28).*

Proof. According to the notations in (B.26) and (B.27), one has

$$\begin{aligned}&\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\vartheta}) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\boldsymbol{\vartheta})\mathbf{H}^{-1}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}(\boldsymbol{\vartheta})\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\alpha}}(\boldsymbol{\vartheta})\\ &\qquad = \left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\mathbf{Z}_{is}\mathbf{Z}^{\top}_{is}\right\} - \left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\mathbf{Z}_{is}\boldsymbol{\Xi}^{\top}_{is}\right\}\\ &\qquad\quad\times\left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\boldsymbol{\Xi}_{is}\boldsymbol{\Xi}^{\top}_{is} + \mathbb{D}^*_{\lambda}\right\}^{-1}\left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2}q_2(\eta_{is}(\boldsymbol{\vartheta}),Y_{is})\boldsymbol{\Xi}_{is}\mathbf{Z}^{\top}_{is}\right\}.\end{aligned}$$

Note that

$$\begin{aligned}
\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) &= \frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\mathbf{Z}_{is}\mathbf{Z}_{is}^{\top} \\
&= \frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}^0, Y_{is})\mathbf{Z}_{is}\mathbf{Z}_{is}^{\top}(1 + O(h^{\varrho+1} + |\triangle|^{d+1})) \to -\mathrm{E}\{\rho_2(\eta_{is}^0)\mathbf{Z}_{is}\mathbf{Z}_{is}^{\top}\},
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0) &= \frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\mathbf{Z}_{is}\boldsymbol{\Xi}_{is}^{\top} \\
&= \frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}^0, Y_{is})\mathbf{Z}_{is}\boldsymbol{\Xi}_{is}^{\top}\{1 + O(h^{\varrho+1} + |\triangle|^{d+1})\},
\end{aligned}$$

Therefore, letting

$$\widehat{\mathbf{Z}}_{is} = \boldsymbol{\Xi}_{is}\left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}^0, Y_{is})\boldsymbol{\Xi}_{is}\boldsymbol{\Xi}_{is}^{\top} + \mathbb{D}_{\lambda}^{*}\right\}^{-1}\left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}^0, Y_{is})\boldsymbol{\Xi}_{is}\mathbf{Z}_{is}^{\top}\right\}, \tag{B.28}$$

we have

$$\begin{aligned}
&\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0)\,\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}^{-1}(\widetilde{\boldsymbol{\vartheta}}^0)\mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}^{\top}(\widetilde{\boldsymbol{\vartheta}}^0) \\
&= \left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}^0, Y_{is})(\mathbf{Z}_{is} - \widehat{\mathbf{Z}}_{is})(\mathbf{Z}_{is} - \widehat{\mathbf{Z}}_{is})^{\top}\right\}\{1 + o_P(1)\} \\
&\to \lim_{n\to\infty} -\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} \mathrm{E}\left\{\rho_2(\eta_{is}^0)(\mathbf{Z}_{is} - \widehat{\mathbf{Z}}_{is})(\mathbf{Z}_{is} - \widehat{\mathbf{Z}}_{is})^{\top}\right\} \\
&= -\mathrm{E}\{\rho_2(\eta_{is}^0)\widetilde{\mathbf{Z}}^{\top}\widetilde{\mathbf{Z}}\},
\end{aligned}$$

as in (B.1). Thus,

$$\begin{aligned}
&\mathbf{S}_{n,\boldsymbol{\alpha}}(\widetilde{\boldsymbol{\vartheta}}^0) - \mathbf{H}_{n,\boldsymbol{\alpha}\boldsymbol{\varrho}}\left(\widetilde{\boldsymbol{\vartheta}}^0\right)\mathbf{H}_{n,\boldsymbol{\varrho}\boldsymbol{\varrho}}^{-1}\left(\widetilde{\boldsymbol{\vartheta}}^0\right)\mathbf{S}_{n,\boldsymbol{\varrho}}(\widetilde{\boldsymbol{\vartheta}}^0) \\
&= -\left[\left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\mathbf{Z}_{is}\right\} - \left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\mathbf{Z}_{is}\boldsymbol{\Xi}_{is}^{\top}\right\}\right. \\
&\qquad \left.\times\left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\boldsymbol{\Xi}_{is}\boldsymbol{\Xi}_{is}^{\top} + \mathbb{D}_{\lambda}^{*}\right\}^{-1}\left\{\frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_1(\eta_{is}(\widetilde{\boldsymbol{\vartheta}}^0), Y_{is})\boldsymbol{\Xi}_{is} + \mathbb{D}_{\lambda}^{*}\widetilde{\boldsymbol{\varrho}}\right\}\right] \\
&= \frac{1}{n_T}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_1(\eta_{is}^0, Y_{is})(\mathbf{Z}_{is} - \widehat{\mathbf{Z}}_{is})\{1 + o_P(1)\}.
\end{aligned}$$

The desired result follows. ■

To study the asymptotic properties of $\widehat{\boldsymbol{\alpha}}$, we consider the case that β_ℓ^0 and γ_k^0 can be estimated at reasonable accuracy, for example, we can approximate γ_k^0 and β_ℓ^0 by the spline smoother $\widetilde{\gamma}_{0,k}$ and $\widetilde{\beta}_\ell^0$ in (B.16). We begin our proof by replacing γ_k^0 and β_ℓ^0 with $\widetilde{\gamma}_{0,k}$ and $\widetilde{\beta}_\ell^0$, respectively, and defining an *intermediate* estimator for $\boldsymbol{\alpha}^0$.

Let $\widetilde{\boldsymbol{\alpha}}$ be the minimizing solution of

$$\widetilde{L}_n(\boldsymbol{\alpha};\widetilde{\boldsymbol{\varrho}}) = \sum_{i=1}^{n}\sum_{s=t_1}^{t_2} L\left[g^{-1}\left\{\boldsymbol{\alpha}^\top \mathbf{Z}_{is} + \sum_{k=1}^{p_2}\widetilde{\gamma}_k^0(X_{ik}) + \sum_{\ell=0}^{p_3}\widetilde{\beta}_\ell^0(\mathbf{U}_i)W_{is\ell}\right\}, Y_{is}\right].$$

LEMMA B.12 *Under Assumptions (A2)–(A8), as $n\to\infty$,*

$$\sqrt{n}(\widetilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) \longrightarrow N(\mathbf{0}, \mathbf{A}^{-1}\boldsymbol{\Sigma}_1\mathbf{A}^{-1}),$$

where $\mathbf{A} = -\mathrm{E}\left\{\rho_2(\eta_{is}^0)\mathbf{Z}_{is}\mathbf{Z}_{is}^\top\right\}$, *and* $\boldsymbol{\Sigma}_1 = \mathrm{E}\left\{q_1^2(\eta_{is}^0, Y_{is})\mathbf{Z}_{is}\mathbf{Z}_{is}^\top\right\}$.

Proof. Let $\nabla\widetilde{L}_n$ and $\nabla^2\widetilde{L}_n$ be the gradient vector and Hessian matrix of $\widetilde{L}_n(\boldsymbol{\alpha};\widetilde{\boldsymbol{\varrho}}_0)$. By Taylor's expansion,

$$\nabla\widetilde{L}_n(\widetilde{\boldsymbol{\alpha}}) - \nabla\widetilde{L}_n(\boldsymbol{\alpha}^0) = \nabla^2\widetilde{L}_n(\boldsymbol{\alpha})(\widetilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0) + \frac{1}{2}(\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}_0)^\top \left.\frac{\partial\nabla\widetilde{L}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}(\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}_0),$$

where $\boldsymbol{\alpha}^* = \{\alpha_j^*\}_j$, $\alpha_j^* = t_j\widetilde{\alpha_j} + (1-t_j)\alpha_j^0$ for some $t_j\in[0,1]$. Note that $\nabla\widetilde{L}_n(\widetilde{\boldsymbol{\alpha}}) = \mathbf{0}$ since $\widetilde{\boldsymbol{\alpha}}$ is the minimizer of $\widetilde{L}_n(\boldsymbol{\alpha})$, thus we have

$$-n_T^{-1}\nabla\widetilde{L}_n(\boldsymbol{\alpha}^0) = n_T^{-1}\nabla^2\widetilde{L}_n\left(\boldsymbol{\alpha}^*\right)\left(\widetilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right) + \frac{1}{2n}(\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}_0)^\top \left.\frac{\partial\nabla\widetilde{L}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}(\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}_0).$$

According to the Cauchy-Schwarz inequality, one has

$$\left\|n_T^{-1}(\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}_0)^\top\frac{\partial\nabla\widetilde{L}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top}(\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}_0)\right\|^2 = O_P(d_n^2/n^2)O_p(d_n^3) = o_P(n^{-1}).$$

By Wang et al. (2014),

$$n_T^{-1}\nabla\widetilde{L}_n(\boldsymbol{\alpha}) = \widetilde{\mathbf{S}}_n(\boldsymbol{\alpha}) + O_P\left(n^{-1}d_n\right) = \mathbf{0},\ n_T^{-1}\nabla^2\widetilde{L}_n(\boldsymbol{\alpha}) = \widetilde{\mathbf{H}}_n(\boldsymbol{\alpha}) + o_P(1),$$

where $\widetilde{\mathbf{S}}_n(\boldsymbol{\alpha})$ and $\widetilde{\mathbf{H}}_n(\boldsymbol{\alpha})$ are defined in

$$\widetilde{\mathbf{H}}_n(\boldsymbol{\alpha}) = \left\{n_T^{-1}\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_2(\eta_{is}(\boldsymbol{\vartheta}), Y_{is})\mathbf{Z}_{is}\mathbf{Z}_{is}^\top\right\}^{-1},\quad n_T^{-1}\widetilde{\mathbf{S}}_n(\boldsymbol{\alpha}) = \left\{\sum_{i=1}^{n}\sum_{s=t_1}^{t_2} q_1\left(\eta_{is}(\boldsymbol{\vartheta}), Y_{is}\right)\mathbf{Z}_{is}\right\}.$$

Thus,

$$-\left\{\widetilde{\mathbf{S}}_n(\boldsymbol{\alpha}^0) + O_P\left(n^{-1}d_n\right)\right\} = \left\{\widetilde{\mathbf{H}}_n\left(\boldsymbol{\alpha}^0\right) + o_P(1)\right\}(\widetilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0) + o_P(n^{-1/2}).$$

Thus, the asymptotic distribution of $\sqrt{n}(\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0)$ is the same as the asymptotic distribution $\sqrt{n}\widetilde{\mathbf{H}}_n^{-1}(\boldsymbol{\alpha}^0)\widetilde{\mathbf{S}}_n(\boldsymbol{\alpha}^0)$. Law of large numbers implies that $\widetilde{\mathbf{H}}_n\left(\boldsymbol{\alpha}^0\right) \to \boldsymbol{A}$ in probability, and

$$\mathrm{Var}\left\{q_1(\eta_{is}^0, Y_{is})\mathbf{Z}_{is}\right\} = \mathrm{E}\left\{q_1^2(\eta_{is}^0, Y_{is})\mathbf{Z}_{is}\mathbf{Z}_{is}^\top\right\} = \boldsymbol{\Sigma}_1.$$

We can derive the asymptotic normality of $\sqrt{n}\widetilde{\mathbf{H}}_n^{-1}(\boldsymbol{\alpha}^0)\widetilde{\mathbf{S}}_n(\boldsymbol{\alpha}^0)$ by using the Cramér-Wold device and checking the Linderberg Condition. ■

References

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438):477–489.

Lai, M. J. and Schumaker, L. L. (2007). *Spline Functions on Triangulations*. Cambridge: Cambridge University Press, 1st edition.

Lai, M. J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica*, 23(3):1399–1417.

Liu, R., Yang, L., and Härdle, W. K. (2013). Oracally efficient two-step estimation of generalized additive model. *Journal of the American Statistical Association*, 108(502):619–631.

Mu, J., Wang, G., and Wang, L. (2020). Spatial autoregressive partially linear varying coefficient models. *Journal of Nonparametric Statistics*, 32(2):428–451.

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

Wang, L. and Cao, G. (2018). Efficient estimation for generalized partially linear single-index models. *Bernoulli*, 24:1101–1127.

Wang, L., Liu, X., Liang, H., and Carroll, R. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics*, 39:1827–1851.

Wang, L., Wang, G., Lai, M. J., and Gao, L. (2020). Efficient estimation of partially linear models for data on complicated domains by bivariate penalized splines over triangulations. *Statistica Sinica*, 30:347–369.

Wang, L., Xue, L., Qu, A., Liang, H., et al. (2014). Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *The Annals of Statistics*, 42(2):592–624.

Yu, S., Wang, G., Wang, L., Liu, C., and Yang, L. (2020). Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association*, 115(530):761–774.