SPATIOTEMPORAL AUTOREGRESSIVE PARTIALLY LINEAR VARYING COEFFICIENT MODELS

Shan Yu[1], Yueying Wang[2], Li Wang[3] and Lei Gao[3]

[1] *University of Virginia, Charlottesville, VA 22904, USA.*

[2] *Iowa State University, Ames, IA 50011, USA.*

[3] *George Mason University, Fairfax, VA 22030, USA.*

Supplementary Material

Appendix A. Notations, Assumptions and Preliminary Results

A.1 Notations

Let C, C_1, C_2, etc. be generic constants, which may be different even in the same line. For a real valued vector $\boldsymbol{a}$, denote $\|\boldsymbol{a}\|$ its Euclidean norm. For a matrix $\mathbf{A} = (a_{ij})$, denote $\|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$, and let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ be the smallest eigenvalue and largest eigenvalue of matrix $\mathbf{A}$, respectively. For a function g on a domain $\mathcal{D}$, $\|g\|_{\infty,\mathcal{D}}$ and $\|g\|_{L_2,\mathcal{D}}$ are the supremum norm and L_2 norm defined in the Section 3.1 in the main paper. For a 3D function $g(s_1, s_2, t)$ on $\Omega \times \mathcal{T}$, let $\|g\|_\infty = \|g\|_{\infty,\Omega\times\mathcal{T}}$ and $\|g\|_{L_2} = \|g\|_{L_2,\Omega\times\mathcal{T}}$, for notational simplicity. For a vector valued function $\boldsymbol{g} = (g_1, \ldots, g_{p_2})^\top$, denote $\|\boldsymbol{g}\|_{L_2} = \{\sum_{k=1}^{p_2} \|g_k\|_{L_2}^2\}^{1/2}$ and $\|\boldsymbol{g}\|_\infty = \max_{1\le k\le p_2} \|g_k\|_\infty$. Further denote $\|\boldsymbol{g}\|_{\nu,\infty} = \max_{1\le k\le p_2} |g_k|_{\nu,\infty}$, where $|g_k|_{\nu,\infty} = \max_{|a_1+a_2+a_3|=\nu} \|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}\nabla_t^{a_3} g_k\|_\infty$. Denote $\mathbb{T}^{(\varrho,d,r)}(\mathcal{E})^{(p_2)} = \{(g_1, \ldots, g_{p_2}) : g_k \in \mathbb{T}^{(\varrho,d,r)}(\mathcal{E}), 1 \le k \le p_2\}$. For any discrete set $\mathcal{A}$, let $|\mathcal{A}|$ be the cardinality of $\mathcal{A}$.

For any two vectors of functions $\boldsymbol{g}^{(1)}(s_1, s_2, t) = \{g_1^{(1)}(s_1, s_2, t), \ldots, g_{p_2}^{(1)}(s_1, s_2, t)\}^\top$, $\boldsymbol{g}^{(2)}(s_1, s_2, t) = \{g_1^{(2)}(s_1, s_2, t), \ldots, g_{p_2}^{(2)}(s_1, s_2, t)\}^\top$, define their empirical inner product as

$$\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle_n = \frac{1}{n}\sum_{i=1}^{n} \left\{\sum_{k=1}^{p_2} X_{ik} g_k^{(1)}(S_{i1}, S_{i2}, T_i)\right\} \left\{\sum_{k=1}^{p_2} X_{ik} g_k^{(2)}(S_{i1}, S_{i2}, T_i)\right\}, \tag{A.1}$$

and theoretical inner product as

$$\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle = \mathrm{E}\left\{\sum_{k=1}^{p_2} X_k g_k^{(1)}(S_1, S_2, T)\right\}\left\{\sum_{k=1}^{p_2} X_k g_k^{(2)}(S_1, S_2, T)\right\}, \tag{A.2}$$

where (S_1, S_2, T) has the joint density function $f_{(S_1,S_2,T)}$. Denote the corresponding empirical and theoretical norms as $\|\cdot\|_n$ and $\|\cdot\|$, respectively. Define

$$\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle_{f_1} = \sum_{k=1}^{p_2} \int_{\Omega\times\mathcal{T}} \left\{(\nabla_{s_1}^2 g_k^{(1)})(\nabla_{s_1}^2 g_k^{(2)}) + (\nabla_{s_2}^2 g_k^{(1)})(\nabla_{s_2}^2 g_k^{(2)})\right\} ds_1 ds_2 dt,$$

$$\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle_{f_2} = \sum_{k=1}^{p_2} \int_{\Omega\times\mathcal{T}} (\nabla_t^2 g_k^{(1)})(\nabla_t^2 g_k^{(2)}) ds_1 ds_2 dt.$$

We use $\|\cdot\|_{f_1}$ and $\|\cdot\|_{f_2}$ to represent their induced norms.

Next, we present some statements which summarize a few basic properties used in the proof of theoretical results. See Appendix A in Lee (2004).

- P1. Suppose matrices $\mathbf{A}$ and $\mathbf{B}$ are bounded in row and columns sums, then the matrix $\mathbf{AB}$ is bounded in row and columns sums.
- P2. Suppose the elements $a_{n,ij}$ of an $n \times n$ matrix $\mathbf{A}_n$ are $O(e_n^{-1})$ uniformly for all i, j. If an $n \times n$ matrix $\mathbf{B}_n$ is uniformly bounded in row and columns sums, then the elements of $\mathbf{A}_n\mathbf{B}_n$ and $\mathbf{B}_n\mathbf{A}_n$ are $O(e_n^{-1})$ uniformly for all i, j.
- P3. If the elements in $\mathbb{Z}$ are uniformly bounded, and $\lim_{n\to\infty} n^{-1}\mathbb{Z}^\top\mathbb{Z}$ exists and is nonsingular, then $\mathbb{Z}(\mathbb{Z}^\top\mathbb{Z})^{-1}\mathbb{Z}^\top$ and $\mathbf{I}_n - \mathbb{Z}(\mathbb{Z}^\top\mathbb{Z})^{-1}\mathbb{Z}^\top$ are uniformly bounded in both row and column sums.
- P4. Let $\mathbf{A}_n$ be an $n \times n$ matrix. We have $\mathrm{E}(\boldsymbol{\epsilon}^\top\mathbf{A}_n\boldsymbol{\epsilon}) = \sigma_0^2\mathrm{tr}(\mathbf{A}_n)$, and $\mathrm{Var}(\boldsymbol{\epsilon}^\top\mathbf{A}_n\boldsymbol{\epsilon}) = (m_4 - 3\sigma_0^2)\sum_{i=1}^n a_{ii} + \sigma_0^4\{\mathrm{tr}(\mathbf{A}_n\mathbf{A}_n^\top) + \mathrm{tr}(\mathbf{A}_n^2)\}$, where $m_4 = \mathrm{E}\epsilon_i^4$.

A.2 Assumptions

In this section, we state the technical assumptions used in the main paper. Let $\mathbf{G} = \mathbf{W}\boldsymbol{\Xi}_0^{-1}$.

- (A1) Let $h_b = \pi_b - \pi_{b-1}$ be the distance between two adjoint knots, satisfying that $C_1 \le \max h_b / \min h_b \le C_2$. Denote $h = \max_{1\le b\le N+1} h_b$.
- (A2) Let ρ_τ be the radius of the largest disk contained in τ. The triangulation $\triangle$ is δ-quasi-uniform, that is, there exists a positive constant δ such that the triangulation $\triangle$ satisfies $|\triangle|/\rho_\tau \le \delta$, for all $\tau \in \triangle$.

(A3) The joint density function of (S_1, S_2, T), $f_{(S_1,S_2,T)}(\cdot,\cdot,\cdot)$, is bounded away from zero and infinity over $\Omega \times \mathcal{T}$.

(A4) $\{(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{S}_i, T_i, \epsilon_i)\}_{i=1}^n$ are independently and identically distributed. The random variables $Z_{i\ell}$ and X_{ik} are uniformly bounded for $i = 1, \ldots, n$, $\ell = 0, \ldots, p_1$, and $k = 1, \ldots, p_2$. Denote that $\lambda_1(s_1, s_2, t) \leq \cdots \leq \lambda_{1+p_1+p_2}(s_1, s_2, t)$ be the the eigenvalues of $\mathrm{E}\{(\mathbf{Z}_i^\top, \mathbf{X}_i^\top)^\top(\mathbf{Z}_i^\top, \mathbf{X}_i^\top)|\mathbf{S}_i = \boldsymbol{s}, T_i = t\}$. There are positive constants C_1 and C_2 such that $C_1 \leq \lambda_1(s_1, s_2, t) \leq \cdots \leq \lambda_{1+p_1+p_2}(s_1, s_2, t) \leq C_2$ for all $\boldsymbol{s} \in \Omega$, $t \in \mathcal{T}$.

(A5) Let $\mathbf{g}_\ell^*(s_1, s_2, t) = \{g_{\ell,1}^*(s_1, s_2, t), \ldots, g_{\ell,p_2}^*(s_1, s_2, t)\}^\top$ be the vector of functions that minimizes $\mathrm{E}\{Z_{i\ell} - \mathbf{X}_i^\top \mathbf{g}(\mathbf{S}_i, T_i)\}^2$ over $\mathbf{g}(s_1, s_2, t) = \{g_1(s_1, s_2, t), \ldots, g_{p_2}(s_1, s_2, t)\}^\top$. For $\mathbf{g}^*(s_1, s_2, t) = \{\mathbf{g}_0^*(s_1, s_2, t), \ldots, \mathbf{g}_{p_1}^*(s_1, s_2, t)\}^\top$, $n^{-1}\mathrm{E}\{\mathbb{Z} - \mathbb{X}\mathbf{g}^*(S_1, S_2, T)\}^\top\{\mathbb{Z} - \mathbb{X}\mathbf{g}^*(S_1, S_2, T)\}$ is positive definite.

(A6) For any $k = 1, \ldots, p_2$, the coefficient function $\beta_k \in \mathbb{W}^{d+1,\infty}(\Omega) \otimes \mathbb{C}^{\varrho-2}(\mathcal{T})$.

(A7) There is a constant C such that $\mathrm{E}|\epsilon_i|^{4+\delta} \leq C < \infty$ for some $\delta > 0$.

(A8) The triangulation size $|\triangle|$ and the subinterval size h for the univariate splines satisfies that $n^{1/2}|\triangle|^2 h \to \infty$, $n|\triangle|^{2d+2} \to 0$, and $nh^{2\varrho} \to 0$. Roughness parameters λ_1 and λ_2 satisfy $\lambda_1 n^{1/2}|\triangle|^{-1}h^{1/2} \to 0$, $\lambda_2 n^{1/2}|\triangle|h^{-3/2} \to 0$, $\lambda_1 n^{-1}|\triangle|^{-4} \to 0$ and $\lambda_2 n^{-1}h^{-4} \to 0$.

(A9) The matrix $\mathbf{\Xi}_0$ is nonsingular. The matrices $\mathbf{W}$ and $\mathbf{\Xi}_0^{-1}$ are uniformly bounded in both row and column sums. Let ϖ be a compact parameter space and α_0 be an interior point of ϖ. For any $\alpha \in \varpi$, $\mathbf{\Xi}(\alpha)^{-1}$ exists and is uniformly bounded in both row or column sums.

(A10) The elements w_{ij} of $\mathbf{W}$ are at most of order e_n^{-1}, uniformly in all i, j, and $\lim_{n\to\infty} e_n/n = 0$.

(A11) $\lim_{n\to\infty} n^{-1}(\mathbb{Z}, \mathbb{X}, \mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbb{Z}, \mathbb{X}, \mathbf{G}\boldsymbol{\mu}_0)$ exists and is nonsingular.

(A12) $\lim_{n\to\infty} n^{-1}\mathrm{E}\left\{(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\right\} \neq 0$, where $\mathbf{\Pi}_{\mathbb{D},\Lambda}$ is in (B.3).

The above assumptions are mild conditions that can be satisfied in many practical situations. Assumptions (A1) – (A2) are commonly used in the literature of spline approximation. Assumption (A1) requires that the knot sequence for the univariate spline have bounded mesh ratio, which was used in Huang (2003) and Xue and Yang (2006). Assumption (A2) suggests the use of the quasi-uniform triangulations for bivariate splines; see Lai and Wang (2013). Assumption (A3) guarantees that the observations are randomly scattered in the domain. Assumption (A4) is to ensure the non-multicollinearity of covariates. In addition, $\mathbb{X}$ and $\mathbb{Z}$ are assumed functionally unrelated as in Assumption (A5). Assumption (A6) describes a smoothness condition commonly used in the nonparametric estimation literature; see, for instance, Liu and Yang (2010), Lai and Wang (2013) and especially Mu et al. (2018). Note that the tensor-product spline spaces is a type of sieve spaces. Assumption (A1), (A2) and (A6) satisfy the typical assumptions on the sieve spaces; see Conditions 3.1 – 3.5 in Chen (2007). The purpose of Assumption (A8) specifies the requirement of the number of knots for univariate splines and the number of triangles for bivariate splines via triangulation. Assumptions (A7), (A9) and (A10) are routinely used in the SAR model literature; see, for example, Assumptions 2–7 in Lee (2004). These assumptions provide the necessary requirements of the weight matrix and disturbances for SAR models. Assumption (A11) requires that the generated regressors $\mathbf{G}\boldsymbol{\mu}_0$ and explanatory variables are not asymptotically multicollinear. Assumption (A12) ensures the uniqueness of the maximizer of the profiled likelihood function of α. Similar assumptions have been used in Lee (2004).

A.3 Properties of tensor-product splines

We first cite some important results for univariate polynomial splines and bivariate splines over a triangulation. Lemma A.1 states the approximation error of the bivariate splines over a triangulation $\triangle$. Lemmas A.2 and A.3 present the stability property of univariate polynomial splines and bivariate splines over a triangulation, respectively.

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

(i) For bi-integer (a_1, a_2) *with* $0 \le a_1 + a_2 \le d$*, there exists a spline* $\psi^*(\cdot) \in \mathbb{S}_d^0(\triangle)$ *such that* $\|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}(\psi - \psi^*)\|_\infty \le C|\triangle|^{d+1-a_1-a_2}|\psi|_{d+1,\infty}$*, where* C *is a constant depending on* d*, and the shape parameter* δ*.*

(ii) For bi-integer (a_1, a_2) *with* $0 \le a_1 + a_2 \le d$*, there exists a spline function* $\psi^{**}(\cdot) \in \mathbb{S}_d^r(\triangle)$ *(*$d \ge 3r + 2$*) such that* $\|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}(\psi - \psi^{**})\|_\infty \le C|\triangle|^{d+1-a_1-a_2}|\psi|_{d+1,\infty}$*, where* C *is a constant depending on* d*,* r*, and the shape parameter* δ*.*

Lemma A.1 shows that, under some regularity conditions, $\mathbb{S}_d^0(\triangle)$ and $\mathbb{S}_d^r(\triangle)$, $d \geq 3r+2$, have full approximation power.

Lemma A.2 (Page 155, De Boor (2001)). *There exists a constant $c_0 > 0$, depending only on the order ϱ, such that for any $s(t) = \sum_{b=1}^{N+\varrho} a_\ell U_b(t) \in \mathcal{U}^\varrho(\boldsymbol{\pi})$,*

$$c_0 \Big\{ \sum_{b=1}^{N+\varrho} a_b^2 (\pi_b - \pi_{b-\varrho}) \Big\}^{1/2} \leq \|s\|_{L_2} \leq \Big\{ \sum_{b=1}^{N+\varrho} a_\ell^2 (\pi_b - \pi_{b-\varrho}) \Big\}^{1/2}. \tag{A.3}$$

Lemma A.3 (Lemma 1, Lai and Wang (2013)). *Let $\{B_m\}_{m\in\mathcal{M}}$ be the basis for $\mathbb{S}_d^r(\triangle)$ constructed in Lai and Schumaker (2007), where $\mathcal{M}$ stands for an index set. Under Assumption (A2), there exist positive constants C_1, C_2 depending on d and r such that*

$$C_1 |\triangle|^2 \sum_{m\in\mathcal{M}} |b_m|^2 \leq \Big\| \sum_{m\in\mathcal{M}} b_m B_m \Big\|_{L^2(\Omega)}^2 \leq C_2 |\triangle|^2 \sum_{m\in\mathcal{M}} |b_m|^2. \tag{A.4}$$

Next, we provide the proofs of Lemmas 1 and 2 in the main paper.

Proof of Lemma 1. For any $1 \leq b \leq N+\varrho$, let $\xi_{u,b}^q = (-1)^{q-1}(q-1)!\phi_u^{(\varrho-q)}(\pi_b)/(\varrho-1)!$, where $\phi_u(t) = \prod_{s=1}^{\varrho-1}(t - \pi_{b+s})$. Let $\{c_{q,\varrho}, 0 \leq q \leq \varrho-1\}$ be the coefficients of the ϱth Bernoulli polynomial P_ϱ, i.e. $P_\varrho(x) = x^\varrho + \sum_{q=0}^{\varrho-1} c_{q,\varrho} x^q$. For $1 \leq b \leq N+\varrho$ and a sufficiently smooth function $g(s_1,s_2,t) \in \Omega \times \mathcal{T}$, we define the linear operator as

$$u_b(g)(s_1,s_2) = \sum_{q=0}^{\varrho-1} \xi_{u,\omega\varrho}^{q+1} \left\{ \frac{\nabla_t^q g(s_1,s_2,t)|_{t=\pi_b}}{q!} - \frac{\nabla_t^\varrho g(s_1,s_2,t)|_{t=\pi_b}}{\varrho!} c_{q,\rho} h_b^{\varrho-q} \right\},$$

where ω is the integer satisfying $(\omega-1)\varrho < b \leq \omega\varrho$, and $h_b = \pi_b - \pi_{b-1}$. According to Theorem 1 in Barrow and Smith (1979), we have, for any given $(s_1,s_2) \in \Omega$, under Assumption (A1) in the main paper and $g(s_1,s_2,t) \in \mathbb{W}^{d+1,\infty}(\Omega) \otimes \mathbb{C}^{\varrho-2}(\mathcal{T})$,

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} \left| \nabla_t^{a_3} \left\{ \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} g(s_1,s_2,t) - \sum_{b=1}^{N+\varrho} u_b(\nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} g)(s_1,s_2) U_b(t) \right\} \right| \\
&\quad \leq C_1 \sup_{t\in\mathcal{T}} \left| \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} \nabla_t^{a_3} g(s_1,s_2,t) \right| h^{\varrho-a_3} \leq C_1 \|\nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} \nabla_t^{a_3} g\|_\infty h^{\varrho-a_3},
\end{aligned} \tag{A.5}$$

where $h = \max_{1\leq b\leq N+\varrho} h_b$. $u_b(\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2} g)(s_1,s_2)$ is a function of (s_1,s_2) and $u_b(\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2} g)(s_1,s_2) = \nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2} u_b(g)(s_1,s_2)$. According to Lemma A.1, there exists $u_b^*(g)(s_1,s_2) = \sum_{m\in\mathcal{M}} \kappa_{b,m} B_m(s_1,s_2)$

such that, under Assumption (A2) in the main paper and $g(s_1,s_2,t)\in\mathbb{W}^{d+1,\infty}(\Omega)\otimes\mathbb{C}^{\varrho-2}(\mathcal{T})$,

$$\begin{aligned}&\sup_{(s_1,s_2)\in\Omega}\left|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}\left\{u_b(g)(s_1,s_2)-\sum_{m\in\mathcal{M}}\kappa_{b,m}B_m(s_1,s_2)\right\}\right|\\&=\sup_{(s_1,s_2)\in\Omega}\left|u_b(\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}g)(s_1,s_2)-\sum_{m\in\mathcal{M}}\kappa_{b,m}\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}B_m(s_1,s_2)\right|\\&\le C_2\max_{\substack{a_1+a_2=d+1,\\0\le a_3\le\varrho}}\|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}\nabla_t^{a_3}g\|_\infty|\triangle|^{d+1-a_1-a_2},\end{aligned}$$

which implies that

$$\begin{aligned}&\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}\nabla_t^{a_3}g(s_1,s_2,t)-\sum_{b=1}^{N+\varrho}u_b(\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}g)(s_1,s_2)\nabla_t^{a_3}U_b(t)\\&=\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}\nabla_t^{a_3}g(s_1,s_2,t)-\sum_{b=1}^{N+\varrho}\left\{\sum_{m\in\mathcal{M}}\kappa_{b,m}\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}B_m(s_1,s_2)+O(|\triangle|^{d+1-a_1-a_2})\right\}\nabla_t^{a_3}U_b(t)\\&=\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}\nabla_t^{a_3}g(s_1,s_2,t)-\sum_{b=1}^{N+\varrho}\sum_{m\in\mathcal{M}}\kappa_{b,m}\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}B_m(s_1,s_2)\nabla_t^{a_3}U_b(t)+O(|\triangle|^{d+1-a_1-a_2}).\end{aligned}\tag{A.6}$$

Combining (A.5) and (A.6), we have

$$\sup_{s_1,s_2\in\Omega,t\in\mathcal{T}}\left|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}\nabla_t^{a_3}\left\{g(s_1,s_2,t)-\sum_{b=1}^{N+\varrho}\sum_{m\in\mathcal{M}}\kappa_{b,m}B_m(s_1,s_2)U_b(t)\right\}\right|=O(h^{\varrho-a_3}+|\triangle|^{d+1-a_1-a_2}).$$

Thus, Lemma 1 has been established. □

Proof of Lemma 2. Note that the $|\mathcal{J}|$-dimensional vector $\boldsymbol{\gamma}=\{\gamma_1,\ldots,\gamma_{|\mathcal{J}|}\}$ can be written as vector $\{\gamma_{1,1}^*,\ldots,\gamma_{1,N+\varrho}^*,\gamma_{2,1}^*,\ldots,\gamma_{|\mathcal{M}|,N+\varrho}^*\}$, where $\gamma_{m,b}^*=\gamma_{(m-1)|\mathcal{M}|+b}$. By (A.4), it is easy to see that $\int_\Omega|g(s_1,s_2,t)|^2ds_1ds_2\ge C_1|\triangle|^2\sum_{m\in\mathcal{M}}\left|\sum_{b=1}^{N+\varrho}U_b(t)\gamma_{m,b}^*\right|^2$. Therefore, by Assumptions (A1) – (A2), and the stability properties in (A.3) and (A.4), we have

$$\begin{aligned}\|g\|_{L^2}^2&=\int_{\Omega\times\mathcal{T}}|g(s_1,s_2,t)|^2ds_1ds_2dt\ge C_1|\triangle|^2\sum_{m\in\mathcal{M}}\int_{\mathcal{T}}\left|\sum_{b=1}^{N+\varrho}U_b(t)\gamma_{m,b}^*\right|^2dt\\&\ge C_{1,\varrho}|\triangle|^2\sum_{m\in\mathcal{M}}\sum_{b=1}^{N+\varrho}\gamma_{m,b}^{*2}(\pi_b-\pi_{b-\varrho})\ge C_{1,\varrho}\varrho h|\triangle|^2\sum_{m\in\mathcal{M}}\sum_{b=1}^{N+\varrho}\gamma_{m,b}^{*2}=C_{1,\varrho}\varrho h|\triangle|^2\sum_{j\in\mathcal{J}}\gamma_j^2.\end{aligned}$$

Similarly, by (A.4), $\int_\Omega|g(s_1,s_2,t)|^2ds_1ds_2\le C_2|\triangle|^2\sum_{m\in\mathcal{M}}\left|\sum_{b=1}^{N+\varrho}U_b(t)\gamma_{m,b}^*\right|^2$. Then, it follows directly from (A.3) that

$$\begin{aligned}&\int_{\Omega\times\mathcal{T}}|g(s_1,s_2,t)|^2ds_1ds_2dt\le C_2|\triangle|^2\sum_{m\in\mathcal{M}}\int_{\mathcal{T}}\left|\sum_{b=1}^{N+\varrho}U_b(t)\gamma_{m,b}^*\right|^2dt\\&\le C_2|\triangle|^2\sum_{m\in\mathcal{M}}\sum_{b=1}^{N+\varrho}\gamma_{m,b}^{*2}(\pi_\ell-\pi_{\ell-\varrho})\le C_2\varrho|\triangle|^2h\sum_{m\in\mathcal{M}}\sum_{b=1}^{N+\varrho}\gamma_{m,b}^{*2}=C_2\varrho|\triangle|^2h\sum_{j\in\mathcal{J}}\gamma_j^2.\end{aligned}$$

Therefore, Lemma 2 has been verified. □

The next lemma shows that the theoretical inner product defined in (A.2) can be approximated by the empirical inner product defined in (A.1) uniformly over the estimation spaces.

Lemma A.4. *For* $k = 1, \ldots, p_2$, $g_k^{(1)}(s_1, s_2, t)$, $g_k^{(2)}(s_1, s_2, t)$ *are the spline functions in* $\mathbb{T}^{(\varrho,d,r)}(\mathcal{E})$, *and* $g_k^{(1)} = \sum_{j\in\mathcal{J}} \gamma_{kj}^{(1)} \psi_j(s_1, s_2, t)$, $g_k^{(2)} = \sum_{j\in\mathcal{J}} \gamma_{kj}^{(2)} \psi_j(s_1, s_2, t)$. *Suppose Assumptions (A1) – (A8) hold, then*

$$R_n = \sup_{\boldsymbol{g}^{(1)},\boldsymbol{g}^{(2)}\in\mathbb{T}^{(\varrho,d,r)}(\mathcal{E})^{(p_2)}} \left| \frac{\left\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \right\rangle_n - \left\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \right\rangle}{\|\boldsymbol{g}^{(1)}\| \, \|\boldsymbol{g}^{(2)}\|} \right| = O_{a.s.} \left\{ h^{-1/2} |\triangle|^{-1} n^{-1/2} (\log n)^{1/2} \right\}.$$

Consequently, if $h^{-1}|\triangle|^{-2}n^{-1}(\log n) \to 0$ *as* $n \to \infty$, *then* $R_n = o_{a.s.}(1)$.

Proof. It is easy to see

$$\begin{aligned}
\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle_n &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{p_2} \sum_{j\in\mathcal{J}} \gamma_{kj}^{(1)} X_{ik} \psi_j\left(S_{1i}, S_{2i}, T_i\right) \right\} \left\{ \sum_{k'=1}^{p_2} \sum_{j'\in\mathcal{J}} \gamma_{k'j'}^{(2)} X_{ik'} \psi_{j'}\left(S_{1i}, S_{2i}, T_i\right) \right\} \\
&= \sum_{k,k'=1}^{p_2} \sum_{j,j'\in\mathcal{J}} \gamma_{kj}^{(1)} \gamma_{k'j'}^{(2)} \left\langle \psi_j, \psi_{j'} \right\rangle_{n,kk'},
\end{aligned}$$

where $\langle \psi_j, \psi_{j'} \rangle_{n,kk'} = \frac{1}{n} \sum_{i=1}^{n} X_{ik} X_{ik'} \psi_j(S_{i1}, S_{i2}, T_i) \psi_{j'}(S_{i1}, S_{i2}, T_i)$. Similarly, we have

$$\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle = \sum_{k,k'=1}^{p_2} \sum_{j,j'\in\mathcal{J}} \gamma_{kj}^{(1)} \gamma_{k'j'}^{(2)} \langle \psi_j, \psi_{j'} \rangle_{kk'},$$

$$\|\boldsymbol{g}^{(1)}\|^2 = \sum_{k,k'=1}^{p_2} \sum_{j,j'\in\mathcal{J}} \gamma_{kj}^{(1)} \gamma_{k'j'}^{(1)} \langle \psi_j, \psi_{j'} \rangle_{kk'}, \ \ \|\boldsymbol{g}^{(2)}\|^2 = \sum_{k,k'=1}^{p_2} \sum_{j,j'\in\mathcal{J}} \gamma_{kj}^{(2)} \gamma_{k'j'}^{(2)} \langle \psi_j, \psi_{j'} \rangle_{kk'},$$

where $\langle \psi_j, \psi_{j'} \rangle_{kk'} = \mathrm{E}\{X_{ik} X_{ik'} \psi_j(S_{i1}, S_{i2}, T_i) \psi_{j'}(S_{i1}, S_{i2}, T_i)\}$.

Given the spline basis function $\psi_j(s_1, s_2, t)$, the indices $m(j)$ and $\ell(j)$ satisfy $\psi_j(s_1, s_2, t) = B_{m(j)}(s_1, s_2) U_{\ell(j)}(t)$. If $|m(j) - m(j')| > (d+2)(d+1)/2$ or $|\ell(j) - \ell(j')| > \varrho$, $\psi_j(s_1, s_2, t)\psi_{j'}(s_1, s_2, t) = 0$. Therefore, we have

$$\begin{aligned}
\langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle_n - \langle \boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)} \rangle &= \sum_{k,k'=1}^{p_2} \sum_{j,j'\in\mathcal{J}} \gamma_{kj}^{(1)} \gamma_{k'j'}^{(2)} \left\langle \psi_j, \psi_{j'} \right\rangle_{n,kk'} - \sum_{k,k'=1}^{p_2} \sum_{j,j'\in\mathcal{J}} \gamma_{kj}^{(1)} \gamma_{k'j'}^{(2)} \left\langle \psi_j, \psi_{j'} \right\rangle_{kk'} \\
&= \sum_{k.k'=1}^{p_2} \sum_{\substack{|\ell(j)-\ell(j')|\leq\varrho \\ |m(j)-m(j')|\leq(d+2)(d+1)/2}} \gamma_{kj}^{(1)} \gamma_{k'j'}^{(2)} \left\{ \left\langle \psi_j, \psi_{j'} \right\rangle_{n,kk'} - \left\langle \psi_j, \psi_{j'} \right\rangle_{kk'} \right\} \\
&\leq C \left\{ \sum_{k,j} {\gamma_{kj}^{(1)}}^2 \sum_{k',j'} {\gamma_{k'j'}^{(2)}}^2 \right\}^{1/2} \max_{\substack{k,k'=1,\ldots,p_2 \\ j,j'\in\mathcal{J}}} \left| \langle \psi_j, \psi_{j'} \rangle_{n,kk'} - \langle \psi_j, \psi_{j'} \rangle_{kk'} \right|. \qquad \text{(A.7)}
\end{aligned}$$

It follows Lemma 2 that, for $v = 1,2$, $C_1 h|\triangle|^2 \sum_{k,j} {\gamma_{kj}^{(v)}}^2 \le \|\boldsymbol{g}^{(v)}\|^2 \le C_2 h|\triangle|^2 \sum_{k,j} {\gamma_{kj}^{(v)}}^2$, which implies that

$$C_1 h|\triangle|^2 \left\{ \sum_{k,j} {\gamma_{kj}^{(1)}}^2 \sum_{k',j'} {\gamma_{k'j'}^{(2)}}^2 \right\}^{1/2} \le \|\boldsymbol{g}^{(1)}\|\|\boldsymbol{g}^{(2)}\| \le C_2 h|\triangle|^2 \left\{ \sum_{k,j} {\gamma_{kj}^{(1)}}^2 \sum_{k',j'} {\gamma_{k'j'}^{(2)}}^2 \right\}^{1/2}. \tag{A.8}$$

Combining (A.7) and (A.8), we have

$$R_n \le C^{-1} h^{-1} |\triangle|^{-2} \max_{\substack{k,k'=1,\ldots,p_2 \\ j,j' \in \mathcal{J}}} |\langle \psi_j, \psi_{j'} \rangle_{n,kk'} - \langle \psi_j, \psi_{j'} \rangle_{kk'}|.$$

Note that X_{ik} are uniformly bounded. Hence, it is straightforward to obtain

$$\max_{\substack{k,k'=1,\ldots,p_2 \\ j,j' \in \mathcal{J}}} |\langle \psi_j, \psi_{j'} \rangle_{n,kk'} - \langle \psi_j, \psi_{j'} \rangle_{kk'}| = O_{a.s.} \left\{ n^{-1/2} h^{1/2} |\triangle| (\log n)^{1/2} \right\}.$$

Lemma A.4 follows. □

Appendix B. Proof of the Main Results in Section 3

For the matrices $\mathbb{X}_{\boldsymbol{\psi}^*}$ and $\mathbb{Z}$ given in (2.6), we denote

$$\mathbb{V} \equiv \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} = \begin{pmatrix} \mathbb{Z}^\top \mathbb{Z} & \mathbb{Z}^\top \mathbb{X}_{\boldsymbol{\psi}^*} \\ \mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{Z} & \mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{X}_{\boldsymbol{\psi}^*} \end{pmatrix} + \mathbb{P}_\Lambda,$$

where $\mathbb{P}_\Lambda$ is defined in (2.7). Then, the inverse of $\mathbb{V}$ can be represented as

$$\mathbb{V}^{-1} \equiv \mathbb{U} = \begin{pmatrix} \mathbf{U}_{11} & -\mathbf{U}_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1} \\ -\mathbf{U}_{22}\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & \mathbf{U}_{22} \end{pmatrix},$$

where $\mathbf{U}_{11}^{-1} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$, and $\mathbf{U}_{22}^{-1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$.

Denote

$$\mathbb{P}_{\lambda_1} = \lambda_1 \mathbf{I}_{p_2} \otimes (\mathcal{Q}_2^\top \mathbf{P}_1 \mathcal{Q}_2), \ \mathbb{P}_{\lambda_2} = \lambda_2 \mathbf{I}_{p_2} \otimes (\mathcal{Q}_2^\top \mathbf{P}_2 \mathcal{Q}_2), \tag{B.1}$$

and

$$\boldsymbol{\Gamma}_\Lambda = \mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{X}_{\boldsymbol{\psi}^*} + \mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2}, \tag{B.2}$$

then, it is clear that $\mathbf{V}_{22} = \boldsymbol{\Gamma}_\Lambda$. Next, we denote

$$\begin{aligned} &\boldsymbol{\Pi}_{\mathbb{Z}} = \mathbb{Z}(\mathbb{Z}^\top \mathbb{Z})^{-1}\mathbb{Z}^\top, \ \ \boldsymbol{\Pi}_{\mathbb{D},\Lambda} = \mathbb{D}(\mathbb{D}^\top \mathbb{D} + \mathbb{P}_\Lambda)^{-1}\mathbb{D}^\top, \\ &\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda} = \mathbb{X}_{\boldsymbol{\psi}^*} \boldsymbol{\Gamma}_\Lambda^{-1} \mathbb{X}_{\boldsymbol{\psi}^*}^\top, \ \ \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda} = \mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}. \end{aligned} \tag{B.3}$$

Then, we have

$$\mathbf{U}_{11}^{-1} = \mathbb{Z}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{X}_{\psi^*},\Lambda})\mathbb{Z},\ \mathbf{U}_{22}^{-1} = \mathbb{X}_{\psi^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})\mathbb{X}_{\psi^*} + \mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2}, \tag{B.4}$$

where $\mathbb{P}_{\lambda_2}$, $\mathbb{P}_{\lambda_2}$, and $\mathbf{\Gamma}_\Lambda$ in (B.1) and (B.2).

Recall that

$$L_n\left(\sigma^2,\alpha\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \log(|\mathbf{\Xi}(\alpha)|) - \frac{1}{2\sigma^2}\|\mathbf{Y}(\alpha) - \mathbb{D}\widehat{\boldsymbol{\xi}}(\alpha)\|^2,$$
$$L_n\left(\alpha\right) = -\frac{n}{2}\{\log(2\pi)+1\} - \frac{n}{2}\log\{\widehat{\sigma}^2(\alpha)\} + \log(|\mathbf{\Xi}(\alpha)|),$$

where $\widehat{\boldsymbol{\xi}}(\alpha) = \mathbb{V}^{-1}\mathbb{D}^\top\mathbf{\Xi}(\alpha)\mathbf{Y}$ and $\widehat{\sigma}^2(\alpha) = n^{-1}\|\mathbf{Y}(\alpha) - \mathbb{D}\widehat{\boldsymbol{\xi}}(\alpha)\|^2$. Denote

$$\sigma^{*2}(\alpha) = n^{-1}\mathrm{E}\|\mathbf{Y}(\alpha) - \mathbb{D}\widehat{\boldsymbol{\xi}}(\alpha)\|^2. \tag{B.5}$$

It is straightforward to verify that

$$Q_n(\alpha) \equiv \max_{\sigma^2}\mathrm{E}\left\{L_n(\sigma^2,\alpha)\right\} = -\frac{n}{2}\{\log(2\pi)+1\} - \frac{n}{2}\log\{\sigma^{*2}(\alpha)\} + \log(|\mathbf{\Xi}(\alpha)|). \tag{B.6}$$

B.1 Proof of Theorem 1

Lemma B.1. *Under Assumptions (A1) – (A10),* $\sup_{\alpha\in\varpi}|n^{-1}L_n(\alpha) - n^{-1}Q_n(\alpha)| = o_P(1)$*, where* $L_n(\alpha)$ *and* $Q_n(\alpha)$ *are defined in (2.10) and (B.6).*

Proof. It is straight forward to obtain that

$$n^{-1}\left\{L_n(\alpha) - Q_n(\alpha)\right\} = -2^{-1}\left\{\log\widehat{\sigma}^2(\alpha) - \log\sigma^{*2}(\alpha)\right\},$$

where $\widehat{\sigma}^2(\alpha)$ and $\sigma^{*2}(\alpha)$ are given in (2.9) and (B.5), respectively. For $\widehat{\sigma}^2(\alpha)$, note that

$$\begin{aligned}(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{Y} &= (\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})[\mathbf{I}_n + (\alpha_0-\alpha)\mathbf{G}]\boldsymbol{\mu}_0 + (\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\\ &= (\alpha_0-\alpha)(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0 + (\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0 + (\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}.\end{aligned}$$

Thus, $\widehat{\sigma}^2(\alpha)$ in (2.9) can be written as

$$\begin{aligned}\widehat{\sigma}^2(\alpha) = \frac{1}{n}\|(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{Y}\|^2 &= \frac{(\alpha_0-\alpha)^2}{n}\|(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})(\mathbf{G}\boldsymbol{\mu}_0)\|^2 + \frac{1}{n}\|(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\|\\ &+ \frac{1}{n}\|(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\|^2 + \frac{2(\alpha_0-\alpha)}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\\ &+ \frac{2(\alpha_0-\alpha)}{n}(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\\ &+ \frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}.\end{aligned}$$

According to Lemma C.4, we have

$$\begin{aligned}\widehat{\sigma}^2(\alpha) &= \frac{(\alpha_0-\alpha)^2}{n}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})(\mathbf{G}\boldsymbol{\mu}_0)\|^2 + \frac{1}{n}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\|^2\\ &+ \frac{2(\alpha_0-\alpha)}{n}(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\\ &+ \frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon} + o_P(1). \end{aligned} \tag{B.7}$$

Next, for $\sigma^{*2}(\alpha)$ in (B.5), we have

$$\begin{aligned}\sigma^{*2}(\alpha) &= \frac{\sigma_0^2}{n}\mathrm{tr}\left\{(\mathbf{\Xi}_0^{-1})^\top\mathbf{\Xi}(\alpha)^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right\}\\ &+ \frac{1}{n}\mathrm{E}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\{\mathbf{I}_n+(\alpha_0-\alpha)\mathbf{G}\}\boldsymbol{\mu}_0\|^2\\ &= \frac{\sigma_0^2}{n}\mathrm{tr}\left\{(\mathbf{\Xi}_0^{-1})^\top\mathbf{\Xi}(\alpha)^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right\} - \frac{2\sigma_0^2}{n}\mathrm{tr}\left\{(\mathbf{\Xi}_0^{-1})^\top\mathbf{\Xi}(\alpha)^\top\mathbf{\Pi}_{\mathbb{D},\Lambda}^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right\}\\ &+ \frac{\sigma_0^2}{n}\mathrm{tr}\left\{(\mathbf{\Xi}_0^{-1})^\top\mathbf{\Xi}(\alpha)^\top\mathbf{\Pi}_{\mathbb{D},\Lambda}^\top\mathbf{\Pi}_{\mathbb{D},\Lambda}\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right\} + \frac{(\alpha_0-\alpha)^2}{n}\mathrm{E}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\|^2\\ &+ \frac{1}{n}\mathrm{E}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\|^2 + \frac{2(\alpha_0-\alpha)}{n}\mathrm{E}\left\{\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\right\}.\end{aligned}$$

Notice that by Lemma C.3 and Property P2 in Section B.1.2, we have

$$\mathrm{tr}\left[\{\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Pi}_{\mathbb{D},\Lambda}^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right] = O_P(|\triangle|^{-2}h^{-1}), \tag{B.8}$$

$$\mathrm{tr}\left[\{\mathbf{\Pi}_{\mathbb{D},\Lambda}\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Pi}_{\mathbb{D},\Lambda}\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right] = O_P(|\triangle|^{-2}h^{-1}). \tag{B.9}$$

Then, it implies

$$\begin{aligned}\sigma^{*2}(\alpha) &= \frac{\sigma_0^2}{n}\mathrm{tr}\left\{(\mathbf{\Xi}_0^{-1})^\top\mathbf{\Xi}(\alpha)^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right\} + \frac{(\alpha_0-\alpha)^2}{n}\mathrm{E}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\|^2\\ &+ \frac{1}{n}\mathrm{E}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\|^2 + \frac{2(\alpha_0-\alpha)}{n}\mathrm{E}\left\{\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\right\} + o_P(1).\end{aligned} \tag{B.10}$$

Combining (B.7) and (B.10), we have

$$\begin{aligned}\widehat{\sigma}^2(\alpha) - \sigma^{*2}(\alpha) &= \frac{(\alpha_0-\alpha)^2}{n}\left[\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})(\mathbf{G}\boldsymbol{\mu}_0)\|^2 - \mathrm{E}\|(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})(\mathbf{G}\boldsymbol{\mu}_0)\|^2\right]\\ &+ \frac{1}{n}\|\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\|^2 - \frac{2}{n}\boldsymbol{\epsilon}^\top\{\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Pi}_{\mathbb{D},\Lambda}\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon} + \frac{1}{n}\|\mathbf{\Pi}_{\mathbb{D},\Lambda}\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\|^2\\ &- \frac{\sigma_0^2}{n}\mathrm{tr}\left[\{\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\right] + \frac{2(\alpha_0-\alpha)}{n}(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\\ &+ \frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon} + o_P(1).\end{aligned}$$

By Property P1 in Appendix A and Assumption (A9), $\{\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}$ is bounded uniformly, for any $\alpha \in \varpi$, both in row and column sums. Then, $\mathrm{Var}\left[\boldsymbol{\epsilon}^\top\{\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\right] = O(n)$. Thus,

$$\sup_{\alpha\in\varpi}\left|\frac{1}{n}\boldsymbol{\epsilon}^\top\{\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon} - \frac{1}{n}\mathrm{E}\left[\boldsymbol{\epsilon}^\top\{\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\}^\top\mathbf{\Xi}(\alpha)\mathbf{\Xi}_0^{-1}\boldsymbol{\epsilon}\right]\right| = o_P(1), \tag{B.11}$$

and $\mathrm{E}\{\boldsymbol{\epsilon}^\top\{\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\}^\top\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon}\} = \sigma_0^2\mathrm{tr}\left[\{\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\}^\top\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\right]$. Therefore,

$$\sup_{\alpha\in\varpi}\left|\frac{1}{n}\boldsymbol{\epsilon}^\top\{\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\}^\top\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon} - \frac{\sigma_0^2}{n}\mathrm{tr}\left[\{\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\}^\top\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\right]\right| = o_P(1). \tag{B.12}$$

Similarly, we have

$$\sup_{\alpha\in\varpi}\left|\frac{(\alpha_0-\alpha)^2}{n}\|(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})(\mathbf{G}\boldsymbol{\mu}_0)\|^2 - \frac{(\alpha_0-\alpha)^2}{n}\mathrm{E}\left\{\|(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})(\mathbf{G}\boldsymbol{\mu}_0)\|^2\right\}\right| = o_P(1).$$

Note that

$$\begin{aligned}
&(\alpha_0-\alpha)(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon}\\
&\quad= (\alpha_0-\alpha)\boldsymbol{\epsilon}^\top\{\mathbf{I}_n + (\alpha_0-\alpha)\mathbf{G}\}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\\
&\quad= (\alpha_0-\alpha)\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\\
&\qquad+ (\alpha_0-\alpha)^2(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0.
\end{aligned}$$

Thus, by (C.13) and (C.14), we have

$$\sup_{\alpha\in\varpi}\left\{\frac{2(\alpha_0-\alpha)}{n}(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon}\right\} = o_P(1). \tag{B.13}$$

Similarly, by (C.9) and (C.13), we obtain

$$\begin{aligned}
&\sup_{\alpha\in\varpi}\left\{\frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon}\right\}\\
&\quad= \sup_{\alpha\in\varpi}\left\{\frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})(\mathbf{I}_n + (\alpha_0-\alpha)\mathbf{G})\boldsymbol{\epsilon}\right\} = o_P(1).
\end{aligned}$$

Also, (B.8) and (B.9) imply

$$\begin{aligned}
n^{-1}\boldsymbol{\epsilon}^\top\{\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\}^\top\boldsymbol{\Pi}_{\mathbb{D},\Lambda}\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon} &= O_P(n^{-1}|\triangle|^{-2}h^{-1}),\\
n^{-1}\|\boldsymbol{\Pi}_{\mathbb{D},\Lambda}\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon}\|^2 &= O_P(n^{-1}|\triangle|^{-2}h^{-1}).
\end{aligned}$$

It follows that

$$\sup_{\alpha\in\varpi}\left\{n^{-1}\boldsymbol{\epsilon}^\top(\boldsymbol{\Xi}_0^{-1})^\top\boldsymbol{\Xi}(\alpha)^\top\boldsymbol{\Pi}_{\mathbb{D},\Lambda}\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon}\right\} \asymp \sup_{\alpha\in\varpi}\left\{n^{-1}\|\boldsymbol{\Pi}_{\mathbb{D},\Lambda}\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\boldsymbol{\epsilon}\|^2\right\} = o_P(1). \tag{B.14}$$

By (B.11) – (B.14), $\widehat{\sigma}^2(\alpha) - \sigma^{*2}(\alpha) = o_P(1)$ uniformly in $\alpha \in \varpi$, then we can use the mean value theorem,

$$\sup_{\alpha\in\varpi}\left|\log\widehat{\sigma}^2(\alpha) - \log\sigma^{*2}(\alpha)\right| = \sup_{\alpha\in\varpi}\frac{1}{\check{\sigma}^2(\alpha)}\left|\widehat{\sigma}^2(\alpha) - \sigma^{*2}(\alpha)\right| = o_P(1),$$

where $\check{\sigma}^2(\alpha)$ is between $\widehat{\sigma}^2(\alpha)$ and $\widehat{\sigma}_n^{*2}(\alpha)$. Therefore, $\sup_{\alpha\in\varpi}\left|n^{-1}\left\{L_n(\alpha) - Q_n(\alpha)\right\}\right| = o_P(1)$. Lemma B.1 is established. □

Proof of Theorem 1. The proof follows from the following steps:
(i) $\sup_{\alpha\in\varpi}|n^{-1}L_n(\alpha)-n^{-1}Q_n(\alpha)|=o_P(1)$; (ii) for any $\alpha\in\varpi$, $Q_n(\alpha)-Q_n(\alpha_0)\le 0$; (iii) for any $\varepsilon>0$, $\limsup_{n\to\infty}[\max_{\alpha\in\overline{\mathcal{N}}_\varepsilon(\alpha_0)} n^{-1}\{Q_n(\alpha)-Q_n(\alpha_0)\}]<0$, where $\mathcal{N}_\varepsilon(\alpha_0)$ is the open neighborhood of α_0 with radius ε and $\overline{\mathcal{N}}_\varepsilon(\alpha_0)$ is the closure of $\mathcal{N}_\varepsilon(\alpha_0)$.

Proof of Step (i). See the proof in Lemma B.1 in Appendix B.

Proof of Step (ii). Consider a standard SAR model: $\mathbf{Y}=\alpha\mathbf{W}\mathbf{Y}+\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I}_n)$. Then the corresponding log-likelihood function is $L^*(\sigma^2,\lambda)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)+\log(|\boldsymbol{\Xi}(\alpha)|)-\frac{1}{\sigma^2}\|\boldsymbol{\Xi}(\alpha)\mathbf{Y}\|^2$. Notice that

$$Q_n^*(\alpha)=\max_{\sigma^2}E_{(\sigma_0^2,\alpha_0)}\left\{L^*(\sigma^2,\alpha)\right\}=-\frac{n}{2}\{\log(2\pi)+1\}-\frac{n}{2}\log\{\sigma^2(\alpha)\}+\log(|\boldsymbol{\Xi}(\alpha)|),$$

where $\sigma^2(\alpha)=\frac{\sigma_0^2}{n}\mathrm{tr}\left[\{\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\}^\top\boldsymbol{\Xi}(\alpha)\boldsymbol{\Xi}_0^{-1}\right]$ and $\mathrm{E}_{(\sigma_0^2,\alpha_0)}$ represents the expectation under the model $\mathbf{Y}=\alpha_0\mathbf{W}\mathbf{Y}+\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}\sim N(\mathbf{0},\sigma_0^2\mathbf{I}_n)$. By Jensen's inequality, we have

$$Q_n^*(\alpha)\le\mathrm{E}_{(\sigma_0^2,\alpha_0)}\max_{\sigma^2}\left\{L^*(\sigma^2,\alpha)\right\}\le\mathrm{E}_{(\sigma_0^2,\alpha_0)}\max_{\sigma^2,\alpha}\left\{L^*(\sigma^2,\alpha)\right\}=Q_n^*(\alpha_0).$$

It is straightforward to obtain

$$\begin{aligned}\frac{1}{n}Q_n(\alpha)-\frac{1}{n}Q_n(\alpha_0)=&\frac{1}{n}Q_n^*(\alpha)-\frac{1}{n}Q_n^*(\alpha_0)-\frac{1}{2}\left[\log\{\sigma^{*2}(\alpha)\}-\log\{\sigma^2(\alpha)\}\right]\\&+\frac{1}{2}\left[\log\{\sigma^{*2}(\alpha_0)\}-\log\{\sigma^2(\alpha_0)\}\right].\end{aligned}$$

According to (B.10) in Appendix B, we have $\sigma^{*2}(\alpha)-\sigma^2(\alpha)>0$ for $\alpha\ne\alpha_0$ and the difference between $\sigma^{*2}(\alpha_0)$ and $\sigma^2(\alpha_0)$ is negligible. Therefore, we could verify that $n^{-1}Q_n(\alpha)\le n^{-1}Q_n(\alpha_0)$.

Proof of Step (iii). We prove the uniqueness of α_0 by contradiction. If the uniqueness of α_0 doesn't hold, then there exists $\varepsilon>0$ and a sequence $\{\alpha_n\}\in\overline{\mathcal{N}}_\varepsilon(\alpha_0)$, such that

$$\lim_{n\to\infty}\frac{1}{n}\{Q_n(\alpha_n)-Q_n(\alpha_0)\}=0,\quad\lim_{n\to\infty}\alpha_n=\widetilde{\alpha}\ne\alpha_0,\tag{B.15}$$

Similar to the argument in the proof of Theorem 3.1 in Lee (2004), $\{n^{-1}Q_n(\alpha)\}$ is uniform equicontinuous of α. Thus, by (B.15), we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}|Q_n(\widetilde{\alpha})-Q_n(\alpha_0)|&=\lim_{n\to\infty}\frac{1}{n}|Q_n(\widetilde{\alpha})-Q_n(\alpha_n)+Q_n(\alpha_n)-Q_n(\alpha_0)|\\&\le\lim_{n\to\infty}\frac{1}{n}|Q_n(\alpha_n)-Q_n(\widetilde{\alpha})|+\lim_{n\to\infty}\frac{1}{n}|Q_n(\alpha_n)-Q_n(\alpha_0)|=0.\end{aligned}$$

Note that, for any $\alpha \in \varpi$,

$$\begin{aligned} Q_n(\alpha_0) - Q_n(\alpha) &= \frac{n}{2}\{\log(\sigma^{*2}(\alpha)) - \log(\sigma^{*2}(\alpha_0))\} + \log(|\boldsymbol{\Xi}_0|) - \log(|\boldsymbol{\Xi}(\alpha)|) \\ &= \frac{n}{2}\{\log(\sigma^{*2}(\alpha)) - \log(\sigma^{2}(\alpha))\} + \frac{n}{2}\{\log(\sigma^{2}(\alpha_0)) - \log(\sigma^{*2}(\alpha_0))\} \\ &\quad + \frac{n}{2}\{\log(\sigma^{2}(\alpha)) - \log(\sigma^{2}(\alpha_0))\} + \log(|\boldsymbol{\Xi}_0|) - \log(|\boldsymbol{\Xi}(\alpha)|). \end{aligned}$$

According to the fact $Q_n^*(\alpha_0) \geq Q_n^*(\alpha)$, we have

$$\frac{n}{2}\{\log(\sigma^{2}(\alpha)) - \log(\sigma^{2}(\alpha_0))\} + \log(|\boldsymbol{\Xi}_0|) - \log(|\boldsymbol{\Xi}(\alpha)|) \geq 0, \quad \text{for all } \alpha \in \varpi,$$

and $\sigma^{*2}(\alpha) - \sigma^2(\alpha) \geq 0$, for all $\alpha \in \varpi$. Then, $\lim_{n\to\infty} n^{-1}\{Q_n(\alpha_n) - Q_n(\alpha_0)\} = 0$ implies

$$\lim_{n\to\infty} \log \sigma^{*2}(\widetilde{\alpha}) - \log \sigma^2(\widetilde{\alpha}) = 0, \tag{B.16}$$

$$\lim_{n\to\infty} \frac{1}{n}\left[\{\log(\sigma^2(\widetilde{\alpha})) - \log(\sigma^2(\alpha_0))\} + \log(|\boldsymbol{\Xi}_0|) - \log(|\boldsymbol{\Xi}(\widetilde{\alpha})|)\right] = 0. \tag{B.17}$$

Thus, (B.16) leads to $\lim_{n\to\infty}\{\sigma^{*2}(\widetilde{\alpha}) - \sigma^2(\widetilde{\alpha})\} = 0$, which implies $\lim_{n\to\infty} n^{-1}\|(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\{\mathbf{I}_n - (\alpha_0 - \widetilde{\alpha})\mathbf{G}\}\boldsymbol{\mu}_0\|^2 = 0$. Therefore, $\lim_{n\to\infty} n^{-1}(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0 = 0$. Then, (B.17) conflicts with Assumption (A12). Thus, α_0 is unique. □

B.2 Proof of Theorem 2

Proof of Theorem 2. Define

$$\ell_n(\alpha, \sigma^2, \boldsymbol{\eta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \log\{|\boldsymbol{\Xi}(\alpha)|\} - \frac{1}{2\sigma^2}\|\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha, \boldsymbol{\eta})\|^2, \tag{B.18}$$

where $\boldsymbol{\theta}(\alpha, \boldsymbol{\eta}) = \boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta}\}$. Denote $\boldsymbol{\kappa} = (\alpha, \sigma^2, \boldsymbol{\eta}^\top)^\top$, $\boldsymbol{\kappa}_0 = (\alpha_0, \sigma_0^2, \boldsymbol{\eta}_0^\top)^\top$ and let $\widehat{\boldsymbol{\kappa}} = (\widehat{\alpha}, \widehat{\sigma}^2, \widehat{\boldsymbol{\eta}}^\top)^\top$ be the maximizer of (B.18). Notice that $(\widehat{\alpha}, \widehat{\sigma}^2, \widehat{\boldsymbol{\eta}}^\top)$ is equal to the TPST estimator defined in Section 2.4. Because $\widehat{\boldsymbol{\kappa}}$ is the maximizer of (B.18), $\nabla_{\boldsymbol{\kappa}}\ell_n(\boldsymbol{\kappa})|_{\boldsymbol{\kappa}=\widehat{\boldsymbol{\kappa}}} = \mathbf{0}$ and

$$n^{1/2}(\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0) = \left\{-n^{-1}\nabla_{\boldsymbol{\kappa}}\nabla_{\boldsymbol{\kappa}'}\ell_n(\boldsymbol{\kappa})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}\right\}^{-1}\left\{n^{-1/2}\nabla_{\boldsymbol{\kappa}}\ell_n(\boldsymbol{\kappa})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}\right\}. \tag{B.19}$$

We calculate the first order derivatives of (B.18):

$$\begin{aligned} \nabla_\alpha \ell_n(\alpha, \sigma^2, \boldsymbol{\eta}) &= -\mathrm{tr}\left\{\boldsymbol{\Xi}(\alpha)^{-1}\mathbf{W}\right\} + \frac{1}{\sigma^2}\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha, \boldsymbol{\eta})\}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{W}\mathbf{Y}, \\ \nabla_{\sigma^2} \ell_n(\alpha, \sigma^2, \boldsymbol{\eta}) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha, \boldsymbol{\eta})\}^\top\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha, \boldsymbol{\eta})\}, \\ \nabla_{\boldsymbol{\eta}} \ell_n(\alpha, \sigma^2, \boldsymbol{\eta}) &= -\frac{1}{\sigma^2}\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha, \boldsymbol{\eta})\}^\top\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbb{Z}, \end{aligned} \tag{B.20}$$

where $\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}$ is given in (B.3). In addition, we have the following second order derivatives of the profile log-likelihood function:

$$
\begin{aligned}
&\nabla^2_{\alpha}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta}) = -\mathrm{tr}(\mathbf{G}^2) - \frac{1}{\sigma^2}(\mathbf{WY})^\top \mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{WY},\\
&\nabla^2_{\sigma^2}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta}) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha,\boldsymbol{\eta})\}^\top\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha,\boldsymbol{\eta})\},\\
&\nabla_{\boldsymbol{\eta}}\nabla_{\boldsymbol{\eta}'}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta}) = -\frac{1}{\sigma^2}\mathbb{Z}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbb{Z},\\
&\nabla_{\alpha}\nabla_{\sigma^2}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta}) = -\frac{1}{\sigma^4}\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha,\boldsymbol{\eta})\}^\top\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{WY},\\
&\nabla_{\alpha}\nabla_{\boldsymbol{\eta}}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta}) = -\frac{1}{\sigma^2}\mathbb{Z}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{WY},\\
&\nabla_{\sigma^2}\nabla_{\boldsymbol{\eta}}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta}) = -\frac{1}{\sigma^4}\mathbb{Z}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\{\mathbf{Y}(\alpha) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha,\boldsymbol{\eta})\}.
\end{aligned}
\tag{B.21}
$$

We first study the asymptotical normality of $n^{-1/2}\nabla_{\boldsymbol{\kappa}}\ell_n(\boldsymbol{\kappa})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}$. According to Lemma 1, for $\beta_k(s_1,s_2,t)$, $k=1,\ldots,p_2$, there exists $\boldsymbol{\theta}^*_k$ such that $\sup_{(s_1,s_2,t)\in\Omega\times\mathcal{T}}|\beta_k(s_1,s_2,t) - \boldsymbol{\psi}^*(s_1,s_2,t)^\top\boldsymbol{\theta}^*_k| = O(h^{\varrho} + |\triangle|^{d+1})$. Denote that $\boldsymbol{\theta}^* = (\boldsymbol{\theta}^{*\top}_1,\ldots,\boldsymbol{\theta}^{*\top}_{p_2})^\top$. Notice that

$$
\mathbf{Y}(\alpha_0) - \mathbb{Z}\boldsymbol{\eta}_0 - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}(\alpha_0,\boldsymbol{\eta}_0) = \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}(\boldsymbol{\epsilon}+\boldsymbol{\varsigma}) + \boldsymbol{\iota},
$$

where $\boldsymbol{\varsigma} = (\varsigma_1,\ldots,\varsigma_n)^\top = \left\{\sum_{k=1}^{p_2}X_{ik}\beta_k(S_{i1},S_{i2},T_i) - \sum_{k=1}^{p_2}X_{ik}\boldsymbol{\psi}^*(S_{i1},S_{i2},T_i)^\top\boldsymbol{\theta}^*_k\right\}_{i=1}^n$ and $\boldsymbol{\iota} = (\iota_1,\ldots,\iota_n)^\top = \mathbb{X}_{\boldsymbol{\psi}^*}(\mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2})\boldsymbol{\theta}^*$. Note that by Lemma C.2 in Appendix B, we have $\varsigma_i = O_P(h^{\varrho} + |\triangle|^{d+1})$ and $\iota_i = O_P(\lambda_1|\triangle|^{-1}h^{1/2} + \lambda_2|\triangle|h^{-3/2})$, $i=1,\ldots,n$. Therefore, $\nabla_{\alpha}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}$, $\nabla_{\sigma^2}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}$, and $\nabla_{\boldsymbol{\eta}}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}$ are dominated by

$$
\begin{gathered}
-\mathrm{tr}(\mathbf{G}) + \sigma_0^{-2}\boldsymbol{\epsilon}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0 + \sigma_0^{-2}\boldsymbol{\epsilon}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\epsilon},\\
-0.5\sigma_0^{-2}n + 0.5\sigma_0^{-4}\boldsymbol{\epsilon}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon}, \text{ and } -\sigma_0^{-2}\boldsymbol{\epsilon}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbb{Z},
\end{gathered}
$$

respectively. By Lemma C.7 in Appendix B and (B.20),

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\nabla_{\alpha}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} = -\frac{1}{\sqrt{n}}\mathrm{tr}(\mathbf{G}) + \frac{1}{\sqrt{n}\sigma_0^2}\boldsymbol{\epsilon}^\top\mathbf{H}^\top_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0 + \frac{1}{\sqrt{n}\sigma_0^2}\boldsymbol{\epsilon}^\top\mathbf{G}\boldsymbol{\epsilon} + o_P(1),\\
&\frac{1}{\sqrt{n}}\nabla_{\sigma^2}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} = -\frac{\sqrt{n}}{2\sigma_0^2} + \frac{1}{2\sqrt{n}\sigma_0^4}\boldsymbol{\epsilon}^\top\boldsymbol{\epsilon} + o_P(1),\\
&\frac{1}{\sqrt{n}}\nabla_{\boldsymbol{\eta}}\ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} = -\frac{1}{\sqrt{n}\sigma_0^2}\boldsymbol{\epsilon}^\top\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbb{Z} + o_P(1).
\end{aligned}
$$

According to Theorem 1 in Kelejian and Prucha (2001), the central limit theorem for linear-quadratic functions can be applied.

Under Assumptions (A1) – (A8), we have $n^{-1}(\boldsymbol{\mu}_0-\mathbb{Z}\boldsymbol{\eta}_0)^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}(\boldsymbol{\mu}_0-\mathbb{Z}\boldsymbol{\eta}_0) = O(|\triangle|^{2d+2}+h^{2\rho})$. Similar to the above discussion, by Lemma C.7 in Appendix B and (B.21), we have

$$
\begin{aligned}
n^{-1}\nabla_\alpha^2 \ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} &= -\mathrm{tr}(\mathbf{G}^2) - \frac{1}{\sigma_0^2}(\mathbf{G}\boldsymbol{\mu}_0)^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0 - \frac{1}{\sigma_0^2}\boldsymbol{\epsilon}^\top \mathbf{G}^\top \mathbf{G}\boldsymbol{\epsilon} + o_P(1),\\
n^{-1}\nabla_{\sigma^2}^2 \ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} &= \frac{n}{2\sigma_0^4} - \frac{1}{\sigma_0^6}\boldsymbol{\epsilon}^\top\boldsymbol{\epsilon} + o_P(1),\\
n^{-1}\nabla_{\boldsymbol{\eta}}\nabla_{\boldsymbol{\eta}'} \ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} &= -\frac{1}{\sigma_0^2}\mathbb{Z}^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbb{Z} + o_P(1),\\
n^{-1}\nabla_\alpha\nabla_{\sigma^2} \ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} &= -\frac{1}{\sigma_0^4}\boldsymbol{\epsilon}^\top \mathbf{G}\boldsymbol{\epsilon} + o_P(1),\\
n^{-1}\nabla_\alpha\nabla_{\boldsymbol{\eta}} \ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} &= -\frac{1}{\sigma_0^2}\mathbb{Z}^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0 + o_P(1),\\
n^{-1}\nabla_{\sigma^2}\nabla_{\boldsymbol{\eta}} \ell_n(\alpha,\sigma^2,\boldsymbol{\eta})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0} &= -\frac{1}{\sigma_0^4}\mathbb{Z}^\top \mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top \boldsymbol{\epsilon} + o_P(1).
\end{aligned}
$$

Then, it is straightforward to show that

$$
\begin{aligned}
\mathrm{E}\left\{n^{-1}\nabla_{\boldsymbol{\kappa}}\ell_n(\boldsymbol{\kappa})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}\nabla_{\boldsymbol{\kappa}'}\ell_n(\boldsymbol{\kappa})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}\right\} &= \boldsymbol{\Sigma}_n + \boldsymbol{\Omega}_n + o(1),\\
\mathrm{E}\left\{n^{-1}\nabla_{\boldsymbol{\kappa}}\nabla_{\boldsymbol{\kappa}'}\ell_n(\boldsymbol{\kappa})|_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}\right\} &= -\boldsymbol{\Sigma}_n + o(1).
\end{aligned}
$$

By (B.19), Theorem 2 is established. □

B.3 Proof of Theorem 3

Theorem B.1 below shows the estimation consistency of the oracle estimator.

Theorem B.1. *Under Assumptions (A1) – (A8), the oracle estimator* $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1,\ldots,\bar{\beta}_{p_2})^\top$ *satisfies that* $\|\bar{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_{L_2} = O_P\left(|\triangle|^{d+1}+h^{\varrho}+\lambda_1 n^{-1}|\triangle|^{-3}h^{-1/2}+\lambda_2 n^{-1}|\triangle|^{-1}h^{-5/2}+n^{-1/2}|\triangle|^{-1}h^{-1/2}\right)$.

Proof. We first decompose $\boldsymbol{\mu}_0$ as follows

$$
\boldsymbol{\mu}_0 = \boldsymbol{\mu}_{\mathrm{C}} + \boldsymbol{\mu}_{\mathrm{V}} = (\mathbf{Z}_1^\top\boldsymbol{\eta}_0,\ldots,\mathbf{Z}_n^\top\boldsymbol{\eta}_0)^\top + \left\{\mathbf{X}_1^\top\boldsymbol{\beta}_0(S_{11},S_{12},T_1),\ldots,\mathbf{X}_n^\top\boldsymbol{\beta}_0(S_{n1},S_{n2},T_n)\right\}^\top.
$$

Then, we can write

$$
\bar{\boldsymbol{\theta}} = \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{Z}})(\boldsymbol{\mu}_{\mathrm{C}}+\boldsymbol{\mu}_{\mathrm{V}}+\boldsymbol{\epsilon}) = \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{Z}})\boldsymbol{\mu}_{\mathrm{V}} + \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{Z}})\boldsymbol{\epsilon} = \bar{\boldsymbol{\theta}}_\mu + \bar{\boldsymbol{\theta}}_\epsilon.
$$

According to Lemma 1 in the main paper, for $\beta_{k0}(s_1,s_2,t)$, $k=1,\ldots,p_2$, there exists $\boldsymbol{\theta}_k^*$ such that $\sup_{(s_1,s_2,t)\in\Omega\times\mathcal{T}}|\beta_{k0}(s_1,s_2,t)-\boldsymbol{\psi}^*(s_1,s_2,t)^\top\boldsymbol{\theta}_k^*| = O(h^{\varrho}+|\triangle|^{d+1})$. Denote that $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top},\ldots,\boldsymbol{\theta}_{p_2}^{*\top})^\top$,

$\beta_k^* = \boldsymbol{\psi}^*(s_1,s_2,t)^\top \boldsymbol{\theta}_k^*$, and $\boldsymbol{\beta}^*(s_1,s_2,t) = \{\beta_1^*(s_1,s_2,t),\ldots,\beta_{p_2}^*(s_1,s_2,t)\}^\top$. Denote $\boldsymbol{\theta}^* = \{\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})\mathbb{X}_{\boldsymbol{\psi}^*}\}^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})\boldsymbol{\mu}_{\mathrm{V}}^*$, where

$$\boldsymbol{\mu}_{\mathrm{V}}^* = \{\mathbf{X}_1^\top \boldsymbol{\beta}^*(S_{11},S_{12},T_1),\ldots,\mathbf{X}_n^\top \boldsymbol{\beta}^*(S_{n1},S_{n2},T_n)\}^\top.$$

Then, we have the following decomposition: $\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \bar{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}^* + \bar{\boldsymbol{\theta}}_\epsilon$. Notice that $\bar{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}^* = \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})(\boldsymbol{\mu}_{\mathrm{V}} - \mathbb{X}_{\boldsymbol{\psi}^*}^\top\boldsymbol{\theta}^*) - \mathbf{U}_{22}\mathbb{P}_{\lambda_1}\boldsymbol{\theta}^* - \mathbf{U}_{22}\mathbb{P}_{\lambda_2}\boldsymbol{\theta}^*$. Hence, $\|\bar{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}^*\| \le \|\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})(\boldsymbol{\mu}_{\mathrm{V}} - \mathbb{X}_{\boldsymbol{\psi}^*}^\top\boldsymbol{\theta}^*)\| + \|\mathbf{U}_{22}\mathbb{P}_{\lambda_1}\boldsymbol{\theta}^*\| + \|\mathbf{U}_{22}\mathbb{P}_{\lambda_2}\boldsymbol{\theta}^*\|$ holds. Lemma C.1 implies that

$$\begin{aligned}
\|\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})(\boldsymbol{\mu}_{\mathrm{V}} - \mathbb{X}_{\boldsymbol{\psi}^*}^\top\boldsymbol{\theta}^*)\| &\le n^{-1}|\triangle|^{-2}h^{-1}\|\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})(\boldsymbol{\mu}_{\mathrm{V}} - \mathbb{X}_{\boldsymbol{\psi}^*}^\top\boldsymbol{\theta}^*)\| \\
&\le Cn^{-1}|\triangle|^{-2}h^{-1}(|\triangle|^{d+1} + h^{\varrho})\left[\sum_{k=1}^{p_2}\sum_{j\in\mathcal{J}}\left\{\sum_{i=1}^n |X_{ik}\psi_j(S_{i1},S_{i2},T_i)|\right\}^2\right]^{1/2} \\
&= O_P\{|\triangle|^{-1}h^{-1/2}(|\triangle|^{d+1} + h^{\varrho})\}. \qquad \text{(B.22)}
\end{aligned}$$

According to Lemmas C.1 and C.2, we have

$$\begin{aligned}
\|\mathbf{U}_{22}\mathbb{P}_\Lambda\boldsymbol{\theta}^*\| \le & Cn^{-1}|\triangle|^{-2}h^{-1}\|\mathbb{P}_\Lambda\boldsymbol{\theta}^*\| = Cn^{-1}|\triangle|^{-2}h^{-1}\left\{\sum_j(\boldsymbol{e}_j^\top\mathbb{P}_\Lambda\boldsymbol{\theta}^*)^2\right\}^{1/2} \\
= & O_P(\lambda_1 n^{-1}|\triangle|^{-4}h^{-1} + \lambda_2 n^{-1}|\triangle|^{-2}h^{-3}), \qquad \text{(B.23)}
\end{aligned}$$

where $\boldsymbol{e}_j$ is a vector with jth element being one and the rest of elements being zero. Now we derive the order of $\|\bar{\boldsymbol{\theta}}_\epsilon\|$. Observe that

$$\begin{aligned}
\|\bar{\boldsymbol{\theta}}_\epsilon\| &= \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})\boldsymbol{\epsilon} \le n^{-1}|\triangle|^{-2}h^{-1}\|\mathbb{X}_{\boldsymbol{\psi}^*}^\top\boldsymbol{\epsilon}\| \\
&\le n^{-1}|\triangle|^{-2}h^{-1}\left[\sum_{k=1}^{p_2}\sum_{j\in\mathcal{J}}\left\{\sum_{i=1}^n X_{ik}\psi_j(S_{i1},S_{i2},T_i)\epsilon_i\right\}^2\right]^{1/2} = O_P(n^{-1/2}|\triangle|^{-2}h^{-1}). \qquad \text{(B.24)}
\end{aligned}$$

Combining (B.22) – (B.24) and Lemma 2, we established Theorem B.1. □

Proof of Theorem 3. To study the consistency of the estimators of coefficient functions, we first consider oracle estimators $\bar{\beta}_k, k = 1,\ldots,p_2$. For a given $\Lambda = (\lambda_1,\lambda_2)$, let $(\bar{\boldsymbol{\eta}},\bar{\boldsymbol{\theta}})$ be the minimizer of the following penalized minimization problem

$$\{\mathbf{Y}(\alpha_0) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}\}^\top\{\mathbf{Y}(\alpha_0) - \mathbb{Z}\boldsymbol{\eta} - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\theta}\} + \boldsymbol{\theta}^\top\mathbb{P}_{\lambda_1}\boldsymbol{\theta} + \boldsymbol{\theta}^\top\mathbb{P}_{\lambda_2}\boldsymbol{\theta},$$

and $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\theta}}_1^\top,\ldots,\bar{\boldsymbol{\theta}}_{p_2}^\top)^\top$. Therefore, the oracle estimators of $\beta_k(s_1,s_2,t)$ is

$$\bar{\beta}_k(s_1,s_2,t) = \boldsymbol{\psi}(s_1,s_2,t)^\top\mathcal{Q}_2\bar{\boldsymbol{\theta}}_k.$$

Notice that $\mathbf{Y}(\alpha_0) = \mathbf{\Xi}_0\mathbf{\Xi}_0^{-1}(\boldsymbol{\mu}_0 + \boldsymbol{\epsilon}) = \boldsymbol{\mu}_0 + \boldsymbol{\epsilon}$. Then, we have

$$\bar{\boldsymbol{\eta}} = \mathbf{U}_{11}\mathbb{Z}^\top(\mathbf{I}_n - \mathbb{X}_{\boldsymbol{\psi}^*}\mathbf{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top)(\boldsymbol{\mu}_0 + \boldsymbol{\epsilon}),\ \bar{\boldsymbol{\theta}} = \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})(\boldsymbol{\mu}_0 + \boldsymbol{\epsilon}).$$

Notice that $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\theta}}_1^\top, \ldots, \bar{\boldsymbol{\theta}}_{p_2}^\top)^\top = \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})\mathbf{\Xi}_0\mathbf{Y}$ and $\widehat{\boldsymbol{\theta}} = \mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})\mathbf{\Xi}(\widehat{\alpha})\mathbf{Y}$. For any $k = 1, \ldots, p_2$, let $\bar{\beta}_k(s_1, s_2, t) = \boldsymbol{\psi}^*(s_1, s_2, t)^\top\bar{\boldsymbol{\theta}}_k$, and $\widehat{\beta}_k(s_1, s_2, t) = \boldsymbol{\psi}^*(s_1, s_2, t)^\top\widehat{\boldsymbol{\theta}}_k$. Then, by Theorem 2, Lemma C.1 and Lemma C.6 in Appendix B, we obtain that

$$\begin{aligned}\|\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}\|^2 &\asymp n^{-1}\|\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})\mathbf{G}\mathbf{\Xi}_0\mathbf{Y}\|^2 \asymp n^{-3}|\mathcal{J}|^2\|\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})\mathbf{G}(\boldsymbol{\mu}_0 + \boldsymbol{\epsilon})\|^2 \\ &\asymp n^{-3}|\mathcal{J}|^2\|\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})\mathbf{G}\boldsymbol{\mu}_0\|^2 + n^{-3}|\mathcal{J}|^2\|\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{Z}})\mathbf{G}\boldsymbol{\epsilon}\|^2 \asymp n^{-1}|\mathcal{J}| + n^{-2}|\mathcal{J}|^2.\end{aligned}$$

Also, Theorem B.1 in Appendix B implies that

$$\begin{aligned}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| &\le |\triangle|h^{1/2}\|\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}\| + \|\boldsymbol{\beta}_0 - \bar{\boldsymbol{\beta}}\| \\ &= O_P\left(|\triangle|^{d+1} + h^{\varrho} + \lambda_1 n^{-1}|\triangle|^{-3}h^{-1/2} + \lambda_2 n^{-1}|\triangle|^{-1}h^{-5/2} + n^{-1/2}|\triangle|^{-1}h^{-1/2}\right).\end{aligned}$$

Theorem 3 is established. □

Appendix C. Additional Technical Lemmas

Lemma C.1. *Under Assumptions (A1) – (A8), for* $\mathbf{\Gamma}_\Lambda$ *in (B.2) and* $\mathbf{U}_{22}$ *in (B.4) in the main paper, there exist constants* $0 < c_1 < C_1 < \infty$ *and* $0 < c_2 < C_2 < \infty$*, such that, with probability approaching one, as* $n \to \infty$*,*

$$c_1 n|\triangle|^2 h \le \lambda_{\min}(\mathbf{\Gamma}_\Lambda) \le \lambda_{\max}(\mathbf{\Gamma}_\Lambda) \le C_1\left(n|\triangle|^2 h + \lambda_1|\triangle|^{-2}h + \lambda_2|\triangle|^2 h^{-3}\right), \tag{C.1}$$

$$c_2 n|\triangle|^2 h \le \lambda_{\min}(\mathbf{U}_{22}^{-1}) \le \lambda_{\max}(\mathbf{U}_{22}^{-1}) \le C_2\left(n|\triangle|^2 h + \lambda_1|\triangle|^{-2}h + \lambda_2|\triangle|^2 h^{-3}\right). \tag{C.2}$$

If $\lambda_1 n^{-1}|\triangle|^{-4} \to 0$ *and* $\lambda_1 n^{-1}|\triangle|^{-4} \to 0$ *as* $n \to \infty$*, then* $\lambda_{\max}(\mathbf{\Gamma}_\Lambda) \le C_1 n|\triangle|^2 h$ *and* $\lambda_{\max}(\mathbf{U}_{22}^{-1}) \le C_2 n|\triangle|^2 h$ *hold.*

Proof. It is easy to see that, for any vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \cdots, \boldsymbol{\theta}_{p_2}^\top)^\top$,

$$\begin{aligned} n^{-1}\boldsymbol{\theta}^\top \mathbf{\Gamma}_\Lambda \boldsymbol{\theta} &= \|\boldsymbol{g}_{\boldsymbol{\gamma}}\|_n^2 + \frac{\lambda_1}{n}\sum_{k=1}^{p_2} \boldsymbol{\gamma}_k^\top [\langle \psi_j, \psi_{j'}\rangle_{f_1}]_{j,j'\in\mathcal{J}} \boldsymbol{\gamma}_k + \frac{\lambda_2}{n}\sum_{k=1}^{p_2} \boldsymbol{\gamma}_k^\top [\langle \psi_j, \psi_{j'}\rangle_{f_2}]_{j,j'\in\mathcal{J}} \boldsymbol{\gamma}_k \\ &= \|\boldsymbol{g}_{\boldsymbol{\gamma}}\|_n^2 + \frac{\lambda_1}{n}\sum_{k=1}^{p_2}\left\|\sum_{j\in\mathcal{J}} \gamma_{kj}\psi_j\right\|_{f_1}^2 + \frac{\lambda_2}{n}\sum_{k=1}^{p_2}\left\|\sum_{j\in\mathcal{J}} \gamma_{kj}\psi_j\right\|_{f_2}^2, \end{aligned}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p_2})^\top = \mathcal{Q}_2\boldsymbol{\theta}$ with $\boldsymbol{\gamma}_k = (\gamma_{kj}, j \in \mathcal{J})^\top$ and $\boldsymbol{g}_{\boldsymbol{\gamma}} = (g_{\boldsymbol{\gamma}_1}, \ldots, g_{\boldsymbol{\gamma}_{p_2}})$ with $g_{\boldsymbol{\gamma}_k}$ being a spline function with coefficient $\boldsymbol{\gamma}_k$. Using the Markov's inequality, we have

$$\sum_{k=1}^{p_2}\left\|\sum_{j\in\mathcal{J}} \gamma_{kj}\psi_j\right\|_{f_1}^2 \le \frac{C}{|\triangle|^4}\sum_{k=1}^{p_2}\left\|\sum_{j\in\mathcal{J}} \gamma_{kj}\psi_j\right\|_{L_2}^2 \le C|\triangle|^{-2}h\|\boldsymbol{\gamma}\|^2,$$

$$\sum_{k=1}^{p_2}\left\|\sum_{j\in\mathcal{J}} \gamma_{kj}\psi_j\right\|_{f_2}^2 \le \frac{C}{h^4}\sum_{k=1}^{p_2}\left\|\sum_{j\in\mathcal{J}} \gamma_{kj}\psi_j\right\|_{L_2}^2 \le C|\triangle|^{2}h^{-3}\|\boldsymbol{\gamma}\|^2.$$

Therefore, by Lemma A.4, the largest eigenvalue of the matrix $\mathbf{\Gamma}_\Lambda$ satisfies that

$$\lambda_{\max}(\mathbf{\Gamma}_\Lambda) \le C\left\{(1+R_n)|\triangle|^2 h + \lambda_1 n^{-1}|\triangle|^{-2}h + \lambda_2 n^{-1}|\triangle|^2 h^{-3}\right\}.$$

Consequently, we have $\lambda_{\max}(\mathbf{\Gamma}_\Lambda) \le C_1\left(|\triangle|^2 h + \lambda_1 n^{-1}|\triangle|^{-2}h + \lambda_2 n^{-1}|\triangle|^2 h^{-3}\right)$ with probability approaching one, for some positive constant C_1. Using Lemma A.4 again, it is easy to obtain $\|\boldsymbol{g}_{\boldsymbol{\gamma}}\|_n^2 = (1-R_n)\|\boldsymbol{g}_{\boldsymbol{\gamma}}\|^2 \ge c(1-R_n)|\triangle|^2 h\|\boldsymbol{\gamma}\|^2$. Therefore, the smallest eigenvalue of $\mathbf{\Gamma}_\Lambda$ is greater than $c(1-R_n)|\triangle|^2 h = c_1|\triangle|^2 h$. Let

$$\mathbf{\Upsilon}^\top = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -|\triangle|^{-1}h^{-1/2}\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{Z}(\mathbb{Z}^\top\mathbb{Z})^{-1} & \mathbf{I} \end{pmatrix},$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ is a matrix with zeros. Denote $\mathbb{D}^* = (\mathbb{Z}, |\triangle|^{-1}h^{-1/2}\mathbb{X}_{\boldsymbol{\psi}^*})^\top$, then we have

$$\boldsymbol{\Upsilon}^\top \mathbb{D}^{*\top}\mathbb{D}^*\boldsymbol{\Upsilon} = \begin{pmatrix} \mathbb{Z}^\top\mathbb{Z} & \mathbf{0} \\ \mathbf{0} & |\triangle|^{-2}h^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})\mathbb{X}_{\boldsymbol{\psi}^*} \end{pmatrix}.$$

Lemma C.1 and Assumption (A4) implies that there exist c_2 and C_2 such that $\lambda_{\min}(\mathbb{D}^{*\top}\mathbb{D}^*) \geq c_2 n$ and $\lambda_{\max}(\mathbb{D}^{*\top}\mathbb{D}^*) \leq C_2 n$. According to the fact that the eigenvalues of $\boldsymbol{\Upsilon}$ are equal to one and properties of spectral radius of a squared matrix, we have

$$\lambda_{\max}(\boldsymbol{\Upsilon}^\top \mathbb{D}^{*\top}\mathbb{D}^*\boldsymbol{\Upsilon}) = \rho(\boldsymbol{\Upsilon}^\top \mathbb{D}^{*\top}\mathbb{D}^*\boldsymbol{\Upsilon}) \leq \rho(\boldsymbol{\Upsilon}^\top)\rho(\mathbb{D}^{*\top}\mathbb{D}^*)\rho(\boldsymbol{\Upsilon}) \leq C_2 n,$$
$$\lambda_{\min}(\boldsymbol{\Upsilon}^\top \mathbb{D}^{*\top}\mathbb{D}^*\boldsymbol{\Upsilon}) = \left\{\lambda_{\max}(\boldsymbol{\Upsilon}^{-1}(\mathbb{D}^{*\top}\mathbb{D}^*)^{-1}(\boldsymbol{\Upsilon}^\top)^{-1})\right\}^{-1} \geq \left\{\lambda_{\min}^{-1}(\mathbb{D}^{*\top}\mathbb{D}^*)\right\}^{-1} \geq c_2 n,$$

where $\rho(\cdot)$ represents spectral radius of a squared matrix. Therefore,

$$c_2 n|\triangle|^2 h \leq \lambda_{\min}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})\mathbb{X}_{\boldsymbol{\psi}^*}) \leq \lambda_{\max}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{Z}})\mathbb{X}_{\boldsymbol{\psi}^*}) \leq C_2 n|\triangle|^2 h.$$

Similar to the discussion about the eigenvalues of (C.1), we obtain (C.2). □

According to Lemma 1 in the main paper, for $\beta_{k0}(s_1, s_2, t)$, $k = 1, \ldots, p_2$, there exists $\boldsymbol{\theta}_k^*$ such that

$$\sup_{(s_1,s_2,t)\in\Omega\times\mathcal{T}} \left|\beta_{k0}(s_1, s_2, t) - \boldsymbol{\psi}^*(s_1, s_2, t)^\top\boldsymbol{\theta}_k^*\right| = O(h^\varrho + |\triangle|^{d+1}). \tag{C.3}$$

Let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top}, \ldots, \boldsymbol{\theta}_{p_2}^{*\top})^\top$ and $\beta_k^*(s_1, s_2, t) = \boldsymbol{\psi}^*(s_1, s_2, t)^\top\boldsymbol{\theta}_k^*$. In addition, denote $\boldsymbol{\beta}^*(s_1, s_2, t) = \{\beta_1^*(s_1, s_2, t), \ldots, \beta_{p_2}^*(s_1, s_2, t)\}^\top$.

Lemma C.2. *Let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top}, \ldots, \boldsymbol{\theta}_{p_2}^{*\top})^\top$, where $\boldsymbol{\theta}_k^*$'s are defined in (C.3). Under Assumptions (A1) – (A8), for any vector $\boldsymbol{a}$ with $\|\boldsymbol{a}\| = 1$, we have*

$$\begin{aligned}&\|\boldsymbol{a}^\top(\mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2})\boldsymbol{\theta}^*\| \\ &\quad = O\{\lambda_1(|\triangle|^{-1}h^{1/2} + |\triangle|^d h^{1/2} + |\triangle|^{-1}h^{\varrho+1/2}) + \lambda_2(|\triangle|h^{-3/2} + |\triangle|^{d+2}h^{-3/2} + |\triangle|h^{\varrho-7/2})\},\end{aligned}$$

where $\mathbb{P}_{\lambda_1}$ and $\mathbb{P}_{\lambda_2}$ are given in (B.1).

Proof. Let $\boldsymbol{a} = (\boldsymbol{a}_1^\top, \ldots, \boldsymbol{a}_{p_2}^\top)^\top$, then we have

$$\begin{aligned}\boldsymbol{a}^\top(\mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2})\boldsymbol{\theta}^* &= \sum_{k=1}^{p_2}\lambda_1\boldsymbol{a}_k^\top\mathcal{Q}_2^\top\mathbf{P}_1\mathcal{Q}_2\boldsymbol{\theta}_k^* + \sum_{k=1}^{p_2}\lambda_2\boldsymbol{a}_k^\top\mathcal{Q}_2^\top\mathbf{P}_2\mathcal{Q}_2\boldsymbol{\theta}_k^* \qquad \text{(C.4)}\\ &= \sum_{k=1}^{p_2}\lambda_1\langle\psi_{\boldsymbol{a}_k}, \beta_k^*\rangle_{f_1} + \sum_{k=1}^{p_2}\lambda_2\langle\psi_{\boldsymbol{a}_k}, \beta_k^*\rangle_{f_2} \leq \sum_{k=1}^{p_2}\lambda_1\|\psi_{\boldsymbol{a}_k}\|_{f_1}\|\beta_k^*\|_{f_1} + \sum_{k=1}^{p_2}\lambda_1\|\psi_{\boldsymbol{a}_k}\|_{f_2}\|\beta_k^*\|_{f_2},\end{aligned}$$

where $\psi_{\boldsymbol{a}_k}(s_1, s_2, t) = \boldsymbol{\psi}^*(s_1, s_2, t)^\top \boldsymbol{a}_k$ is a tensor-product spline function. By Markov inequality,

$$\|\psi_{\boldsymbol{a}_k}\|_{f_1}^2 \le C_1 |\triangle|^{-4} \|\psi_{\boldsymbol{a}_k}\|_{L_2}^2 \asymp |\triangle|^{-2} h \|\boldsymbol{a}_k\|^2, \ \|\psi_{\boldsymbol{a}_k}\|_{f_2}^2 \le C_2 h^{-4} \|\psi_{\boldsymbol{a}_k}\|_{L_2}^2 \asymp |\triangle|^2 h^{-3} \|\boldsymbol{a}_k\|^2. \tag{C.5}$$

Combining Lemma 1, (C.4), (C.5) yields Lemma C.2. □

Lemma C.3. *Under Assumptions (A1) – (A8), $\boldsymbol{\Pi}_{\mathbb{D},\Lambda}$ is bounded in both column and row sums, and the elements in $\boldsymbol{\Pi}_{\mathbb{D},\Lambda}$ are $O_P(n^{-1}|\triangle|^{-2}h^{-1})$ uniformly for all i, j.*

Proof. Notice that

$$\begin{aligned}\boldsymbol{\Pi}_{\mathbb{D},\Lambda} &= \begin{pmatrix} \mathbb{Z} & \mathbb{X}_{\boldsymbol{\psi}^*} \end{pmatrix} \begin{pmatrix} \mathbf{U}_{11} & -\mathbf{U}_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1} \\ -\mathbf{U}_{22}\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & \mathbf{U}_{22} \end{pmatrix} \begin{pmatrix} \mathbb{Z}^\top \\ \mathbb{X}_{\boldsymbol{\psi}^*}^\top \end{pmatrix} \\ &= \mathbb{Z}\mathbf{U}_{11}\mathbb{Z}^\top - \mathbb{X}_{\boldsymbol{\psi}^*}\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top \boldsymbol{\Pi}_{\mathbb{Z}} + \mathbb{X}_{\boldsymbol{\psi}^*}\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top - \mathbb{Z}\mathbf{U}_{11}\mathbb{Z}^\top \mathbb{X}_{\boldsymbol{\psi}^*}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{X}_{\boldsymbol{\psi}^*})^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top, \end{aligned} \tag{C.6}$$

where $\mathbf{U}_{11}$ and $\mathbf{U}_{22}$ are defined in (B.4). Note that $n^{-1}\mathbf{U}_{11}^{-1}$ is a consistent estimator of $n^{-1}\mathrm{E}\{\mathbb{Z} - \mathbb{X}^\top \mathbf{g}^*(S_1, S_2, T)\}\{\mathbb{Z} - \mathbb{X}^\top \mathbf{g}^*(S_1, S_2, T)\}^\top$, which is defined in the Assumption (A5). Assumption (A5) implies that there exists a constant C such that, for any $i = 1, \ldots, n$, $\sum_{j=1}^n |\mathbf{Z}_i^\top \mathbf{U}_{11}\mathbf{Z}_j| \le C$ and $\sum_{j=1}^n |\mathbf{Z}_j^\top \mathbf{U}_{11}\mathbf{Z}_i| \le C$. Thus, $\mathbb{Z}\mathbf{U}_{11}\mathbb{Z}^\top$ is bounded both in row and column sums. Next, we prove $\mathbb{X}_{\boldsymbol{\psi}^*}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{X}_{\boldsymbol{\psi}^*})^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top$ is bounded both in row and column sums. Observe that

$$\begin{aligned} &\sum_{t=1}^n \{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\}^\top (\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{X}_{\boldsymbol{\psi}^*})^{-1} \{\mathbf{X}_t \otimes \boldsymbol{\psi}^*(S_{t1}, S_{t2}, T_t)\} \\ &\quad \le \sum_{(S_{t1}, S_{t2}, T_t) \in e_{(i)}} n^{-1}|\triangle|^{-2}h^{-1}C\|\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\|\|\mathbf{X}_t \otimes \boldsymbol{\psi}^*(S_{t1}, S_{t2}, T_t)\| \\ &\quad\quad + \sum_{(S_{t1}, S_{t2}, T_t) \notin e_{(i)}} c \sum_{k,k'=1}^{p_2} \sum_{j,j'=1}^{|\mathcal{J}|} X_{ik}X_{tk'}\psi_j(S_{i1}, S_{i2}, T_i)\psi_{j'}(S_{t1}, S_{t2}, T_t) \le C, \end{aligned}$$

where $e_{(i)}$ represents the triangular prism contains (S_{i1}, S_{i2}, T_i). Hence, $\mathbb{X}_{\boldsymbol{\psi}^*}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{X}_{\boldsymbol{\psi}^*})^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top$ is bounded both in row and column sums. Similarly, we can prove $\mathbb{X}_{\boldsymbol{\psi}^*}\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top$ is bounded both in row and column sums. By (C.6), we have $\boldsymbol{\Pi}_{\mathbb{D},\Lambda}$ and $\mathbb{X}_{\boldsymbol{\psi}^*}\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top \boldsymbol{\Pi}_{\mathbb{Z}}$ are bounded in both column and row sums.

Also, the elements of $\mathbb{Z}\mathbf{U}_{11}\mathbb{Z}^\top$ and $\boldsymbol{\Pi}_{\mathbb{Z}}$ are $O_P(n^{-1})$ uniformly for all i, j. The elements of $\mathbb{X}_{\boldsymbol{\psi}^*}\mathbf{U}_{22}\mathbb{X}_{\boldsymbol{\psi}^*}^\top$ and $\mathbb{X}_{\boldsymbol{\psi}^*}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbb{X}_{\boldsymbol{\psi}^*})^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top$ are $O_P(n^{-1}|\triangle|^{-2}h^{-1})$ uniformly for all i, j. Applying the property in Lee (2004), we have the elements in $\boldsymbol{\Pi}_{\mathbb{D},\Lambda}$ are $O_P(n^{-1}|\triangle|^{-2}h^{-1})$ uniformly for all i, j. Lemma C.3 is established. □

Lemma C.4. *Under Assumptions (A1) – (A10), we have*

$$n^{-1}\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0 = O_P\{|\triangle|^{2d+2}+h^{2\varrho}+n^{-1}(\lambda_1^2|\triangle|^{-6}h^{-1}+\lambda_2^2|\triangle|^{-2}h^{-5})\}, \tag{C.7}$$

$$n^{-1}\boldsymbol{\mu}_0^\top\mathbf{G}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0 = O_P\{|\triangle|^{d+1}+h^{\varrho}+n^{-1/2}(\lambda_1|\triangle|^{-3}h^{-1/2}+\lambda_2|\triangle|^{-1}h^{-5/2})\}, \tag{C.8}$$

$$n^{-1}\boldsymbol{\epsilon}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0 = O_P\{n^{-1/2}|\triangle|^{d+1}+n^{-1/2}h^{\varrho}+n^{-1}(\lambda_1|\triangle|^{-3}h^{-1/2}+\lambda_2|\triangle|^{-1}h^{-5/2})\}, \tag{C.9}$$

$$n^{-1}\boldsymbol{\epsilon}^\top\mathbf{G}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0 = O_P\{n^{-1/2}|\triangle|^{d+1}+n^{-1/2}h^{\varrho}+n^{-1}(\lambda_1|\triangle|^{-3}h^{-1/2}+\lambda_2|\triangle|^{-1}h^{-5/2})\}, \tag{C.10}$$

where $\boldsymbol{\Pi}_{\mathbb{D},\Lambda}$ *is defined in (B.3).*

Proof. Notice that $\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0=\|(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\|^2$ and

$$\begin{aligned}(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0 &= (\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\left\{\mathbf{Z}_i^\top\boldsymbol{\eta}_0+\mathbf{X}_i^\top\boldsymbol{\beta}_0(S_{i1},S_{i2},T_i)\right\}_{i=1}^n\\ &= (\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\left\{\mathbf{Z}_i^\top\boldsymbol{\eta}_0+\mathbf{X}_i^\top\boldsymbol{\beta}^*(S_{i1},S_{i2},T_i)+\mathbf{X}_i^\top\boldsymbol{\beta}_0(S_{i1},S_{i2},T_i)-\mathbf{X}_i^\top\boldsymbol{\beta}^*(S_{i1},S_{i2},T_i)\right\}_{i=1}^n\\ &= (\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\left\{\mathbf{X}_i^\top\boldsymbol{\beta}_0(S_{i1},S_{i2},T_i)-\mathbf{X}_i^\top\boldsymbol{\beta}^*(S_{i1},S_{i2},T_i)\right\}_{i=1}^n+\mathbb{D}\mathbb{V}^{-1}\mathbb{P}_\Lambda(\boldsymbol{\eta}_0^\top,{\boldsymbol{\theta}^*}^\top)^\top,\end{aligned}$$

where $\boldsymbol{\theta}^*=(\boldsymbol{\theta}_1^{*\top},\ldots,\boldsymbol{\theta}_{p_2}^{*\top})^\top$ with $\boldsymbol{\theta}_k^*$'s in (C.3), then we have

$$\begin{aligned}\mathbb{D}\mathbb{V}^{-1}\mathbb{P}_\Lambda(\boldsymbol{\eta}_0^\top,{\boldsymbol{\theta}^*}^\top)^\top &= (\mathbb{Z}\ \ \mathbb{X}_{\phi^*})\begin{pmatrix}\mathbf{U}_{11} & \mathbf{U}_{12}\\ \mathbf{U}_{21} & \mathbf{U}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbb{P}_{\lambda_1}+\mathbb{P}_{\lambda_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{\eta}_0\\ \boldsymbol{\theta}^*\end{pmatrix}\\ &= \Big[\mathbb{Z}\left\{\mathbb{Z}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda})\mathbb{Z}\right\}^{-1}\mathbb{Z}^\top\mathbb{X}_{\boldsymbol{\psi}^*}(\mathbb{X}_{\boldsymbol{\psi}^{*\top}}\mathbb{X}_{\boldsymbol{\psi}^*})^{-1}+\mathbb{X}_{\boldsymbol{\psi}^*}\{\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{Z}})\mathbb{X}_{\boldsymbol{\psi}^*}+\mathbb{P}_{\lambda_1}+\mathbb{P}_{\lambda_2}\}^{-1}\Big]\\ &\quad\times(\mathbb{P}_{\lambda_1}+\mathbb{P}_{\lambda_2})\boldsymbol{\theta}^*=\{(\boldsymbol{b}_{i1}+\boldsymbol{b}_{i2})^\top(\mathbb{P}_{\lambda_1}+\mathbb{P}_{\lambda_2})\boldsymbol{\theta}^*\}_{i=1}^n,\end{aligned}$$

where $\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}$ and $\boldsymbol{\Pi}_{\mathbb{Z}}$ are defined in (B.3), and

$$\begin{aligned}\boldsymbol{b}_{i1} &= \mathbf{Z}_i^\top\left\{\mathbb{Z}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda})\mathbb{Z}\right\}^{-1}\mathbb{Z}^\top\mathbb{X}_{\boldsymbol{\psi}^*}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top\mathbb{X}_{\boldsymbol{\psi}^*})^{-1},\\ \boldsymbol{b}_{i2} &= \{\mathbf{X}_i\otimes\boldsymbol{\psi}^*(\mathbf{S}_{i1},\mathbf{S}_{i2},T_i)\}^\top\{\mathbb{X}_{\boldsymbol{\psi}^*}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{Z}})\mathbb{X}_{\boldsymbol{\psi}^*}+\mathbb{P}_{\lambda_1}+\mathbb{P}_{\lambda_2}\}^{-1}.\end{aligned}$$

Lemma C.1 yields that, for any $i=1,\ldots,n$,

$$\begin{aligned}\|\boldsymbol{b}_{i1}\| &\le Cn^{-1}|\triangle|^{-2}h^{-1}\|\mathbf{Z}_i^\top\left\{\mathbb{Z}^\top(\mathbf{I}_n-\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda})\mathbb{Z}\right\}^{-1}\mathbb{Z}^\top\mathbb{X}_{\boldsymbol{\psi}^*}\|=O_P(n^{-1}|\triangle|^{-1}h^{-1/2}),\\ \|\boldsymbol{b}_{i2}\| &\le Cn^{-1}|\triangle|^{-2}h^{-1}\|\mathbf{X}_i\otimes\boldsymbol{\psi}^*(\mathbf{S}_{i1},\mathbf{S}_{i2},T_i)\|=O_P(n^{-1}|\triangle|^{-2}h^{-1}).\end{aligned}$$

By Lemma C.2 and Assumption (A8), we have

$$\begin{aligned}
&\|n^{-1/2}\mathbb{D}\mathbb{V}^{-1}\mathbb{P}_\Lambda(\boldsymbol{\eta}_0^\top, \boldsymbol{\theta}^{*\top})^\top\| \\
&\le n^{-1/2}\left\{\sum_{i=1}^n \|\boldsymbol{b}_{i1}^\top(\mathbb{P}_{\lambda_1}+\mathbb{P}_{\lambda_2})\boldsymbol{\theta}^*\|^2\right\}^{1/2} + n^{-1/2}\left\{\sum_{i=1}^n \|\boldsymbol{b}_{i2}^\top(\mathbb{P}_{\lambda_1}+\mathbb{P}_{\lambda_2})\boldsymbol{\theta}^*\|^2\right\}^{1/2} \qquad (C.11)\\
&= O_P\{\lambda_1 n^{-1}|\triangle|^{-3}h^{-1/2} + \lambda_2 n^{-1}|\triangle|^{-1}h^{-5/2}\}.
\end{aligned}$$

According to the SVD decomposition of matrices $\mathbb{D}$ and $\mathbb{P}_\Lambda$, we can prove that $\lambda_{\min}(\boldsymbol{\Pi}_{\mathbb{D},\Lambda}) \ge 0$. Thus, we have

$$\begin{aligned}
&\|(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\left\{\mathbf{X}_i^\top\boldsymbol{\beta}_0(S_{i1},S_{i2},T_i) - \mathbf{X}_i^\top\boldsymbol{\beta}^*(S_{i1},S_{i2},T_i)\right\}_{i=1}^n\| \\
&\le \lambda_{\max}(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\|\left\{\mathbf{X}_i^\top\boldsymbol{\beta}_0(S_{i1},S_{i2},T_i) - \mathbf{X}_i^\top\boldsymbol{\beta}^*(S_{i1},S_{i2},T_i)\right\}_{i=1}^n\| \\
&= \{1-\lambda_{\min}(\boldsymbol{\Pi}_{\mathbb{D},\Lambda})\}\|\left\{\mathbf{X}_i^\top\boldsymbol{\beta}_0(S_{i1},S_{i2},T_i) - \mathbf{X}_i^\top\boldsymbol{\beta}^*(S_{i1},S_{i2},T_i)\right\}_{i=1}^n\| \\
&= O\{n^{1/2}(|\triangle|^{d+1}+h^{\varrho})\}. \qquad (C.12)
\end{aligned}$$

Combining (C.11) and (C.12), we obtain (C.7). By Cauchy Schwarz inequality, we have

$$\begin{aligned}
\|n^{-1}\boldsymbol{\mu}_0^\top\mathbf{G}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\| &\le n^{-1}\|(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\|\|(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\| \\
&= O_P\{|\triangle|^{d+1} + h^{\varrho} + n^{-1/2}(\lambda_1|\triangle|^{-4}h^{-1} + \lambda_2|\triangle|^{-2}h^{-3})\},
\end{aligned}$$

which yields (C.8). By the properties of ϵ_i, we have $\mathrm{E}\{\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\} = 0$, and

$$\begin{aligned}
\mathrm{E}\{\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\}^2 &= \mathrm{E}\|(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\|^2 \\
&= O(n|\triangle|^{2d+2} + nh^{2\varrho} + \lambda_1^2|\triangle|^{-8}h^{-2} + \lambda_2^2|\triangle|^{-4}h^{-6}).
\end{aligned}$$

Thus, we obtain (C.9).

Similarly, we have $\mathrm{E}\{\boldsymbol{\epsilon}^\top\mathbf{G}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\} = 0$. According to Lemma C.3, $(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})$ is bounded both in row and column sums. Then, $\mathbf{G}(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})$ is bounded both in row and column sums, and $\mathrm{Var}\{\boldsymbol{\epsilon}^\top\mathbf{G}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\} = \mathrm{E}\left\{\|\mathbf{G}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\boldsymbol{\mu}_0\|^2\right\} = O(n|\triangle|^{2d+2} + nh^{2\varrho} + \lambda_1^2|\triangle|^{-8}h^{-2} + \lambda_2^2|\triangle|^{-4}h^{-6})$, which yields (C.10). □

Lemma C.5. *Under Assumptions (A1) – (A10), we have*

$$n^{-1}\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0 = o_P(1), \qquad (C.13)$$

$$n^{-1}\boldsymbol{\epsilon}^\top\mathbf{G}^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \boldsymbol{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0 = o_P(1), \qquad (C.14)$$

where $\boldsymbol{\Pi}_{\mathbb{D},\Lambda}$ *is defined in (B.3).*

Proof. Note that $(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}$ is bounded both in row and columns sums. Then $\mathrm{E}\{\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\} = 0$, $\mathrm{Var}\{\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\} = \mathrm{E}\|(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\|^2 = O(n)$, which implies (C.13). Similarly, $\mathrm{E}\{(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\} = 0$ and $\mathrm{Var}\{(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\} = \mathrm{E}\{\|\mathbf{G}^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})^\top(\mathbf{I}_n - \mathbf{\Pi}_{\mathbb{D},\Lambda})\mathbf{G}\boldsymbol{\mu}_0\|^2\} = O(n)$. Therefore, (C.14) holds. □

Lemma C.6. *Under Assumptions (A1) – (A10), if $\mathbf{C}_n$ is uniformly bounded both in row sums and column sums in absolute values, then we have $\|n^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbf{C}_n\boldsymbol{\mu}_0\|^2 = O_P(|\triangle|^2 h)$.*

Proof. Note that

$$\begin{aligned}
\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbf{C}_n\boldsymbol{\mu}_0 &= \{\mathbf{X}_1 \otimes \boldsymbol{\psi}^*(S_{11}, S_{12}, T_1), \ldots, \mathbf{X}_n \otimes \boldsymbol{\psi}^*(S_{n1}, S_{n2}, T_n)\}\left(\sum_{i=1}^n c_{n,1i}\mu_{0i}, \ldots, \sum_{i=1}^n c_{n,ni}\mu_{0i}\right)^\top \\
&= \sum_{i=1}^n \mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i) \sum_{i'=1}^n c_{n,ii'}\mu_{0,i'},
\end{aligned}$$

where $\mathbf{X}_i^\top \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)^\top = \{\mathbf{X}_i^\top \otimes \boldsymbol{\psi}(S_{i1}, S_{i2}, T_i)^\top\}(\mathbf{I}_{p_2} \otimes \mathcal{Q}_2)$. Then we have

$$\begin{aligned}
&n^{-2}(\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbf{C}_n\boldsymbol{\mu}_0)^\top(\mathbb{X}_{\boldsymbol{\psi}^*}^\top \mathbf{C}_n\boldsymbol{\mu}_0) \\
&= \left[n^{-1}\sum_{i=1}^n \{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\}^\top \sum_{i'=1}^n c_{n,ii'}\mu_{0,i'}\right]\left[n^{-1}\sum_{s=1}^n \mathbf{X}_s \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\sum_{s'=1}^n c_{n,ss'}\mu_{0,s'}\right] \\
&= \left[n^{-1}\sum_{i=1}^n \{\mathbf{X}_i \otimes \boldsymbol{\psi}(S_{i1}, S_{i2}, T_i)\}^\top \sum_{i'=1}^n c_{n,ii'}\mu_{0,i'}\right](\mathbf{I}_{p_2} \otimes \mathcal{Q}_2)(\mathbf{I}_{p_2} \otimes \mathcal{Q}_2^\top) \\
&\quad \times \left[n^{-1}\sum_{s=1}^n \mathbf{X}_s \otimes \boldsymbol{\psi}(S_{i1}, S_{i2}, T_i)\sum_{s'=1}^n c_{n,ss'}\mu_{0,s'}\right] \\
&\le C_1 n^{-2}\sum_{i=1}^n\sum_{s=1}^n |\{\mathbf{X}_i \otimes \boldsymbol{\psi}(S_{i1}, S_{i2}, T_i)\}^\top\{\mathbf{X}_s \otimes \boldsymbol{\psi}(S_{i1}, S_{i2}, T_i)\}|\sum_{i'=1}^n\sum_{s'=1}^n |c_{n,ii'}||c_{n,ss'}| \\
&\asymp \frac{1}{n}\sum_{i=1}^n \boldsymbol{\psi}(S_{i1}, S_{i2}, T_i)^\top \frac{1}{n}\sum_{i'=1}^n \boldsymbol{\psi}(\mathbf{S}_{i'}, T_{i'}) \asymp O_P(|\triangle|^2 h).
\end{aligned}$$

Lemma C.6 has been established. □

Lemma C.7. *Under Assumptions (A1) – (A12), we have*

$$n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0 = o_P(1), \tag{C.15}$$

$$n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbb{Z} = o_P(1), \tag{C.16}$$

$$n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon} = o_P(1),\ \ n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon} = o_P(1), \tag{C.17}$$

$$n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\epsilon} = o_P(1),\ \ n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top \mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\epsilon} = o_P(1). \tag{C.18}$$

Proof. First, we prove that (C.15) holds. Note that $\mathrm{E}(\boldsymbol{\epsilon}^\top \boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0) = 0$, and

$$\begin{aligned}
\mathrm{Var}^{1/2}&(\boldsymbol{\epsilon}^\top \boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0) = \|\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{H}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\mu}_0\| \\
&= \|(\mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top - \mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top\mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top)\mathbf{G}\boldsymbol{\mu}_0\| \\
&= \|\mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{P}_\Lambda\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top\mathbf{G}\boldsymbol{\mu}_0\| = \left[\sum_{i=1}^n \left\{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{P}_\Lambda\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top\mathbf{G}\boldsymbol{\mu}_0\right\}^2\right]^{1/2} \\
&\le C(\lambda_1 n^{-2}|\triangle|^{-6}h^{-1} + \lambda_2 n^{-2}|\triangle|^{-2}h^{-5})\left\{\sum_{i=1}^n \|\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\|^2\|\mathbb{X}_{\boldsymbol{\psi}^*}^\top\mathbf{G}\boldsymbol{\mu}_0\|^2\right\}^{1/2} \\
&= O_P(\lambda_1 n^{-1/2}|\triangle|^{-5}h^{-1/2} + \lambda_2 n^{-1/2}|\triangle|^{-1}h^{-9/2}).
\end{aligned}$$

Thus, (C.15) holds. Similarly, we can obtain (C.16).

Note that $\mathrm{E}|\boldsymbol{\epsilon}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon}| = \mathrm{tr}(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda})$, $\mathrm{E}|\boldsymbol{\epsilon}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon}| = \mathrm{tr}(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda})$. According to Lemma C.1, we have

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}) &= \sum_{i=1}^n \{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\}^\top\boldsymbol{\Gamma}_\Lambda^{-1}\{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\} \le C|\triangle|^{-2}h^{-1}, \\
\mathrm{tr}(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}) &= \sum_{i=1}^n \{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\}^\top\boldsymbol{\Gamma}_\Lambda^{-1}\mathbb{X}_{\boldsymbol{\psi}^*}^\top\mathbb{X}_{\boldsymbol{\psi}^*}\boldsymbol{\Gamma}_\Lambda^{-1}\{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\} \\
&\le C|\triangle|^{-2}h^{-1}.
\end{aligned}$$

Therefore, by Property P4 in Appendix A,

$$\mathrm{E}|n^{-1/2}\boldsymbol{\epsilon}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon}| = O(n^{-1/2}|\triangle|^{-2}h^{-1}),\ \mathrm{E}|n^{-1/2}\boldsymbol{\epsilon}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon}| = O(n^{-1/2}|\triangle|^{-2}h^{-1}).$$

By Assumption (A8), $n^{-1/2}\boldsymbol{\epsilon}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon} = o_P(1)$ and $n^{-1/2}\boldsymbol{\epsilon}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\boldsymbol{\epsilon} = o_P(1)$, which yields (C.17).

It remains to show that (C.18) holds. Note that, by Property P4 in Appendix A,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{\epsilon}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\epsilon}) = (m_4 - 3\sigma_0^4)&\sum_{i=1}^n(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ii}^2 + \sigma_0^4\mathrm{tr}(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}) \\
&+ \sigma_0^4\mathrm{tr}(\mathbf{G}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}),
\end{aligned}$$

and $(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ii}$ is the (i,i)th entry of matrix $\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}$. Let $(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda})_{ij}$ be the (i,j)th entry of matrix $\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}$. Lemma C.1 implies that

$$(\boldsymbol{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda})_{ij} = \{\mathbf{X}_i \otimes \boldsymbol{\psi}^*(S_{i1}, S_{i2}, T_i)\}^\top\boldsymbol{\Gamma}_\Lambda^{-1}\{\mathbf{X}_j \otimes \boldsymbol{\psi}^*(S_{j1}, S_{j2}, T_j)\} = O_P(n^{-1}|\triangle|^{-2}h^{-1}).$$

Hence, $\sum_{i=1}^{n}(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ii}^2 \le C_1 n^{-2}|\triangle|^{-4}h^{-2}\sum_{i=1}^{n}(\sum_{j=1}^{n}(\mathbf{G})_{ij})^2 \le C_2 n^{-1}|\triangle|^{-4}h^{-2}$. In addition, we have $(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ij} = O_P(n^{-1}|\triangle|^{-2}h^{-1})$, and

$$\operatorname{tr}(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}) = \sum_{i=1}^{n}\sum_{k=1}^{n}(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ik}(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ki} = O_P(|\triangle|^{-4}h^{-2}),$$
$$\operatorname{tr}(\mathbf{G}^\top\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}^\top\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}) = \sum_{i=1}^{n}\sum_{k=1}^{n}(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ik}(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G})_{ik} = O_P(|\triangle|^{-4}h^{-2}).$$

Thus, $\operatorname{Var}(\boldsymbol{\epsilon}^\top\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\epsilon}) = O(|\triangle|^{-4}h^{-2})$. Also, $\operatorname{E}\left(\boldsymbol{\epsilon}^\top\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}\boldsymbol{\epsilon}\right) = \operatorname{tr}(\mathbf{\Pi}_{\mathbb{X}_{\boldsymbol{\psi}^*},\Lambda}\mathbf{G}) = O(|\triangle|^{-2}h^{-1})$, and Assumption (A8) implies (C.18). Similarly, the second part in (C.18) can be established. □

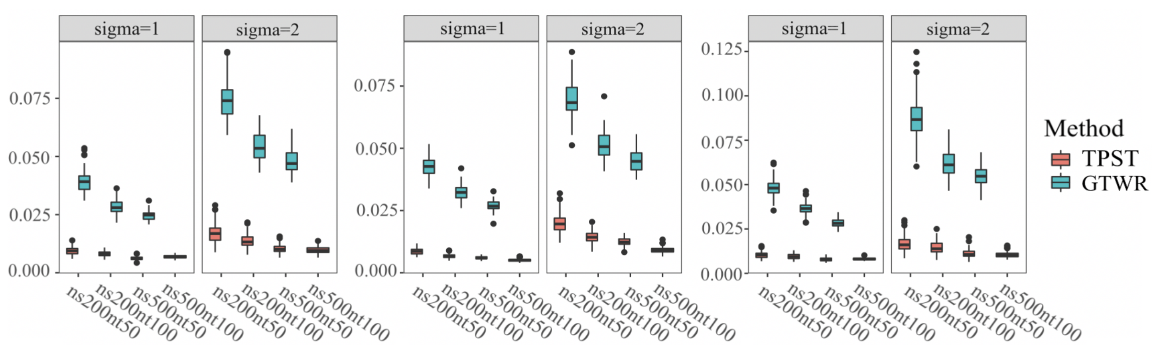
Appendix D. More on Implementation and Numerical Results

D.1 Choosing the optimal block size

For spatiotemporal process, there are two kinds of block sizes: the spatial block size and the temporal block size, see Valavi et al. (2019). In the spatial sense, the range, which is a crucial parameter in the variogram, can be used to define the optimal separation distance between the testing and training sets. We use the empirical variogram to estimate the range. At each observed time point, we calculate the empirical variogram and get the corresponding range estimator. Let b_S be the 0.7 quantile of the estimated ranges. The spatial block size is set to be $4b_S^2$. In the temporal sense, we calculate the empirical autocorrelation function (ACF) at each fixed location and find the smallest time point where the empirical ACF is approximate to zero. The temporal block size b_T is set to the 0.7 quantile of these estimated time points. In our proposed method, we first use the random CV to fit the model. Following the above procedures, we use the residuals to decide the block sizes. Then, we refit the model using block CV.

D.2 Additional simulation results

Figure D.1 (a)–(c) show the box plots of the MISEs of the estimators of the varying coefficients.



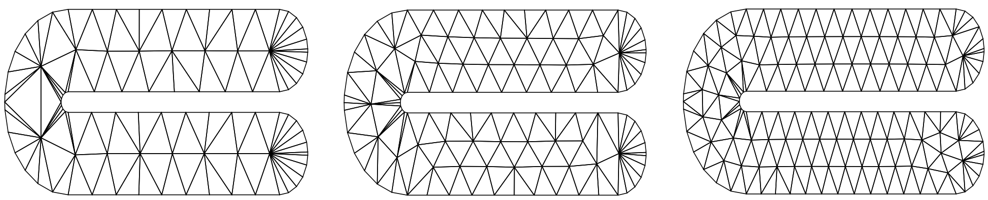
(a) β_0 (b) β_1 (c) β_2

Figure D.1: Boxplots of MISEs of the estimators of β_k in Simulation Study 1.

Figure D.3 shows the sequences of spatial plots of the true coefficient functions evaluated at time points $t = 0.0,\ 0.17,\ 0.50,\ 0.83$ and 1.0. It also depicts the sequences of the estimated coefficient functions via the TPST and GTWR methods based on a typical run in Simulation Study 1 with

$n_S = 200$, $n_T = 50$ and $\sigma_0 = 1$. These TPST plots are obtained using $\varrho = 3$, $d = 2$, $r = 1$ and the triangular prism shown in Figure 2 in the main paper.

Figure D.4 illustrates sequences of spatial plots of the TPST estimated coefficient functions for a typical run in Simulation Study 2, where the plots are based on six different triangular prisms with $n_S = 100$, $n_T = 30$ and $\sigma_0 = 0.5$. It is hard to tell the difference among the plots produced by different triangular prisms, which implies the effect of different triangular prisms is negligible. Figure D.2 show $\triangle_q$, $q = 1, 2, 3$.



$\triangle_1$ (109 triangles) $\triangle_2$ (154 triangles) $\triangle_3$ (219 triangles)

Figure D.2: Triangulations on the horseshoe domainconsidered in Simulation Study 2.

Table D.1 reports the mean squared error (MSE) for the TPST estimators of the constant parameters and the mean integrated squared error (MISE) of the TPST estimator of the varying coefficient function in Simulation Study 2.

D.3 Data and results in the COVID-19 study

Figure D.5 presents the triangulation used in the application example.

Tables D.2 and D.3 show the sources and the detailed explanation of the county-level predictors used in the modeling of COVID-19 infection and death counts.

The fitted varying coefficient functions of β^D in the death model are shown in Figure D.6, in which Figure D.6 (a)–(f) present the estimated coefficient function maps of at six different days from April to June.

Figures D.7 (a)–(d) in the Supplementary Materials show example cases when the traditional SIR model does not work. Without integrating the nearby information, the county-level prediction of SIR is sensitive to observed data of each county. For example, in Figures D.7 (b) and (d), there are jumps in the cumulative infected cases, which leads to severe over-predictions in the following seven days.

Table D.1: Mean squared error (MSE) and mean squared integrated error (MISE) of the estimators of the constant (functional) parameters and the average computing time in Simulation Study 2.

(n_S, n_T)	σ_0	Method	MSE ($\times 10^{-3}$)					MISE ($\times 10^{-3}$)
			α_0	σ^2	η_{00}	η_{01}	η_{02}	β_{01}
(100,30)	0.5	TPST($\mathcal{E}_1$)	0.34	0.06	34.77	0.10	0.08	7.21
		TPST($\mathcal{E}_2$)	0.34	0.06	34.64	0.10	0.08	7.45
		TPST($\mathcal{E}_3$)	0.34	0.06	34.94	0.10	0.08	7.38
		TPST($\mathcal{E}_4$)	0.35	0.07	35.22	0.10	0.08	7.68
		TPST($\mathcal{E}_5$)	0.35	0.07	35.64	0.10	0.08	7.97
		TPST($\mathcal{E}_6$)	0.35	0.07	35.83	0.10	0.08	8.08
	1.0	TPST($\mathcal{E}_1$)	1.04	1.08	106.18	0.39	0.33	19.09
		TPST($\mathcal{E}_2$)	1.04	1.07	106.37	0.39	0.33	19.45
		TPST($\mathcal{E}_3$)	1.04	1.07	106.26	0.39	0.33	19.35
		TPST($\mathcal{E}_4$)	1.10	1.14	112.45	0.39	0.33	20.56
		TPST($\mathcal{E}_5$)	1.09	1.17	111.14	0.39	0.33	21.37
		TPST($\mathcal{E}_6$)	1.11	1.20	113.20	0.39	0.33	21.85
(100,50)	0.5	TPST($\mathcal{E}_1$)	0.25	0.03	25.02	0.04	0.04	5.39
		TPST($\mathcal{E}_2$)	0.28	0.03	28.36	0.04	0.04	5.64
		TPST($\mathcal{E}_3$)	0.25	0.03	25.15	0.04	0.04	5.55
		TPST($\mathcal{E}_4$)	0.24	0.04	23.87	0.04	0.04	5.77
		TPST($\mathcal{E}_5$)	0.23	0.04	23.04	0.04	0.04	5.99
		TPST($\mathcal{E}_6$)	0.22	0.04	22.55	0.04	0.04	6.06
	1.0	TPST($\mathcal{E}_1$)	0.76	0.55	76.68	0.16	0.16	14.10
		TPST($\mathcal{E}_2$)	0.69	0.55	69.66	0.16	0.16	14.34
		TPST($\mathcal{E}_3$)	0.76	0.55	77.68	0.16	0.16	14.27
		TPST($\mathcal{E}_4$)	0.76	0.57	77.27	0.16	0.16	15.10
		TPST($\mathcal{E}_5$)	0.73	0.58	74.43	0.16	0.16	15.70
		TPST($\mathcal{E}_6$)	0.73	0.59	74.67	0.16	0.16	16.02
(200,30)	0.5	TPST($\mathcal{E}_1$)	0.18	0.03	18.51	0.03	0.05	4.44
		TPST($\mathcal{E}_2$)	0.20	0.03	20.44	0.03	0.05	4.63
		TPST($\mathcal{E}_3$)	0.19	0.03	19.50	0.03	0.05	4.57
		TPST($\mathcal{E}_4$)	0.20	0.03	19.90	0.03	0.05	4.74
		TPST($\mathcal{E}_5$)	0.19	0.03	19.38	0.03	0.04	4.90
		TPST($\mathcal{E}_6$)	0.19	0.03	19.03	0.03	0.04	4.97
	1.0	TPST($\mathcal{E}_1$)	0.59	0.51	59.54	0.13	0.18	11.71
		TPST($\mathcal{E}_2$)	0.55	0.51	55.33	0.13	0.18	11.90
		TPST($\mathcal{E}_3$)	0.56	0.51	56.84	0.13	0.18	11.85
		TPST($\mathcal{E}_4$)	0.55	0.53	55.83	0.13	0.18	12.54
		TPST($\mathcal{E}_5$)	0.54	0.54	54.73	0.13	0.18	13.03
		TPST($\mathcal{E}_6$)	0.53	0.55	53.80	0.13	0.18	13.30
(200,50)	0.5	TPST($\mathcal{E}_1$)	0.08	0.02	8.44	0.03	0.02	3.25
		TPST($\mathcal{E}_2$)	0.10	0.02	10.17	0.03	0.02	3.41
		TPST($\mathcal{E}_3$)	0.10	0.02	9.62	0.03	0.02	3.36
		TPST($\mathcal{E}_4$)	0.09	0.02	9.33	0.03	0.02	3.49
		TPST($\mathcal{E}_5$)	0.09	0.02	9.16	0.03	0.02	3.62
		TPST($\mathcal{E}_6$)	0.09	0.02	9.27	0.03	0.02	3.67
	1.0	TPST($\mathcal{E}_1$)	0.29	0.29	29.5	0.13	0.09	8.20
		TPST($\mathcal{E}_2$)	0.30	0.28	29.99	0.13	0.09	8.37
		TPST($\mathcal{E}_3$)	0.29	0.28	28.74	0.13	0.09	8.32
		TPST($\mathcal{E}_4$)	0.31	0.29	30.68	0.13	0.09	8.85
		TPST($\mathcal{E}_5$)	0.30	0.30	29.88	0.13	0.09	9.21
		TPST($\mathcal{E}_6$)	0.29	0.30	29.33	0.13	0.09	9.43

Table D.2: Sources of datasets

Data Type	Source
COVID-19 Related Time-series	
Infections Data	NYT (2020); Atlantic (2020); CSSE (2020); USAFacts (2020)
Fatality Data	NYT (2020); Atlantic (2020); CSSE (2020); USAFacts (2020)
Recovery Data	Atlantic (2020)
Mobility Data	
Bureau of Transportation Statistics	BTS (2020)
Descartes Labs	Warren and Skillman (2020)
American Community Survey (ACS) Data	
2005-2009 ACS 5-year Estimates	USCB (2018)
2012 Economic Census	USCB (2012)
Homeland Infrastructure Foundation-level Data	DHS (2020)

Bibliography

Atlantic (2020), "The COVID Tracking Project Data," Available at `https://covidtracking.com/api`.

Barrow, D. and Smith, P. (1979), "Efficient L2 approximation by splines," *Numerische Mathematik*, 33, 101–114.

BTS (2020), "Trips by Distance," Available at `https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv`.

Chen, X. (2007), "Large sample sieve estimation of semi-nonparametric models," *Handbook of econometrics*, 6, 5549–5632.

CSSE, J. H. U. (2020), "2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository," Available at `https://github.com/CSSEGISandData/COVID-19`.

De Boor, C. (2001), *A practical guide to splines*, Springer.

DHS (2020), "Homeland infrastructure foundation-level data," Available at `https://hifld-geoplatform.opendata.arcgis.com/datasets/6ac5e325468c4cb9b905f1728d6fbf0f_0/data`.

Table D.3: County-level predictors used in the modeling of COVID-19 infection and death counts.

Covariates	Description
Demographic Characteristics	
AA	Percent of African American population
HL	Percent of Hispanic or Latino population
PD*	Population density per square mile of land area
Old	Aged people (age $\geq$ 65 years) rate per capita
Sex	Ratio of male over female
Socioeconomic Status	
Affluence	A measure of more economically privileged areas, including:
	Percent of households with income over $75,000
	Percent of adults obtaining bachelor's degree or higher
	Percent of employed persons in management, professional and related occupations
	Median value of owner-occupied housing units
Disadvantage	A measure for conditions of economic disadvantage, including:
	Percent of households with public assistance income
	Percent of households with female householder and no husband present
	Civilian labor force unemployment rate
Gini	Gini coefficient, a measure of economic inequality and wealth distribution
Rural/urban Factor	
Urban	Urban rate
Healthcare Infrastructure	
NHIC	Percent of persons under 65 years without health insurance
EHPC	Local government expenditures for health per capita
TBed*	Total bed counts per 1000 population
Mobility	Change in number of trips since March 2, 2020

Huang, J. Z. (2003), "Local asymptotics for polynomial spline regression," *The Annals of Statistics*, 31, 1600–1635.

Kelejian, H. H. and Prucha, I. R. (2001), "On the asymptotic distribution of the Moran I test statistic with applications," *Journal of Econometrics*, 104, 219–257.

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

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

Lee, L.-F. (2004), "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

Liu, R. and Yang, L. (2010), "Spline-backfitted kernel smoothing of additive coefficient model," *Econometric Theory*, 26, 29–59.

Mu, J., Wang, G., and Wang, L. (2018), "Estimation and inference in spatially varying coefficient models," *Environmetrics*, 29, e2485.

NYT (2020), "Coronavirus (COVID-19) Data in the United States," Available at `https://github.com/nytimes/covid-19-data`.

USAFacts (2020), "Coronavirus Locations: COVID-19 Map by County and State," Available at `https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/`. (Accessed: 30th May 2020).

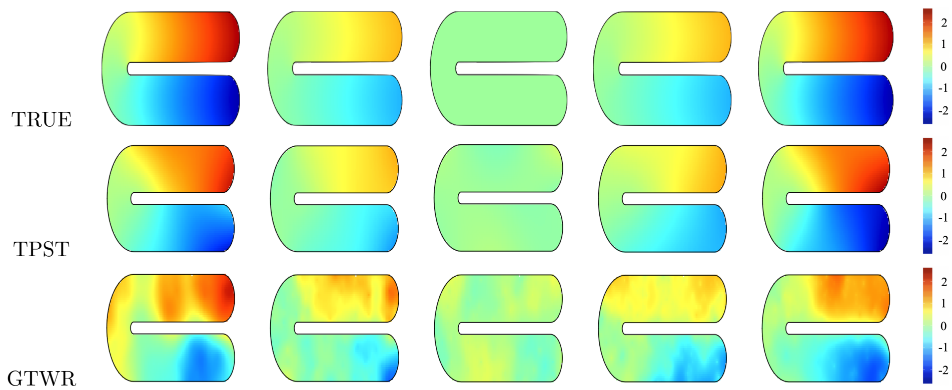
USCB (2012), "2012 Economic Census," Available at `https://data.census.gov/cedsci/table?q=government%20expenditures%20for%20health&hidePreview=false&tid=ECNGRANT2012.EC1262SXSB3&t=Government%3AHealth&vintage=2012`.

— (2018), "2005-2009 American Community Survey 5-year estimates," Available at `https://data.census.gov/cedsci/table?q=gini%20coefficient&hidePreview=false&tid=ACSDT1Y2018.B19083&vintage=2018`.

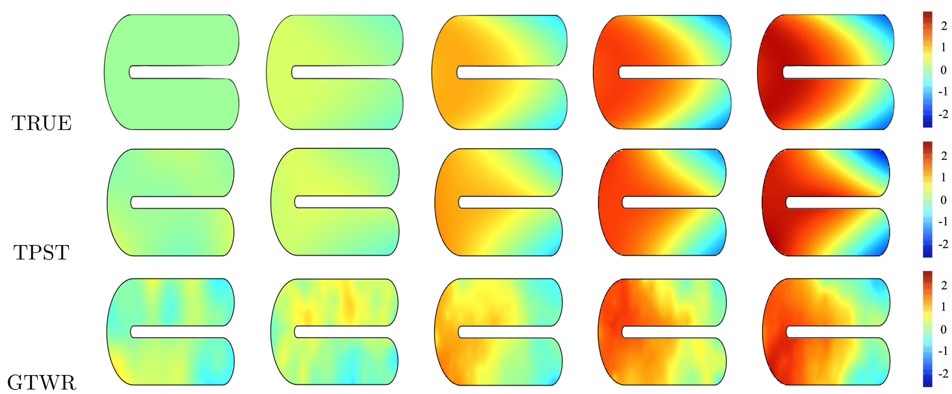
Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G. (2019), "blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models," *Methods in Ecology and Evolution*, 10, 225–232.

Warren, M. S. and Skillman, S. W. (2020), "Mobility changes in response to COVID-19," *arXiv preprint arXiv:2003.14228*.

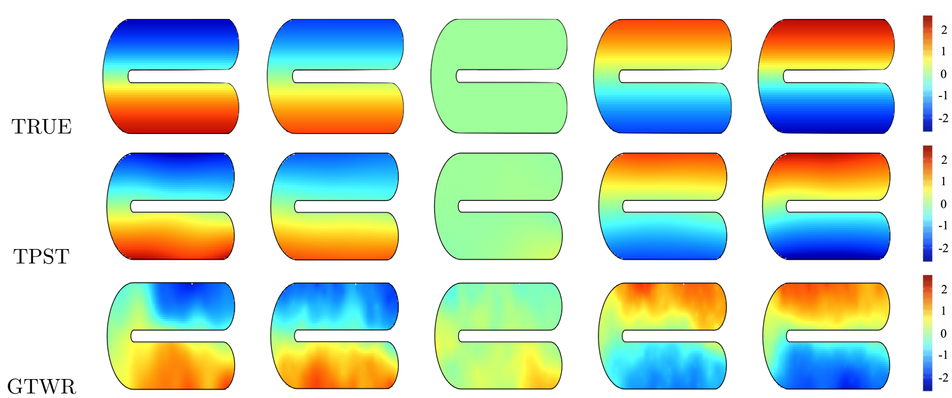
Xue, L. and Yang, L. (2006), "Additive coefficient modeling via polynomial spline," *Statistica Sinica*, 1423–1446.



(a) β_{00}



(b) β_{01}



(c) β_{02}

Figure D.3: Sequences of spatial plots of the coefficient functions for Simulation Study 1.

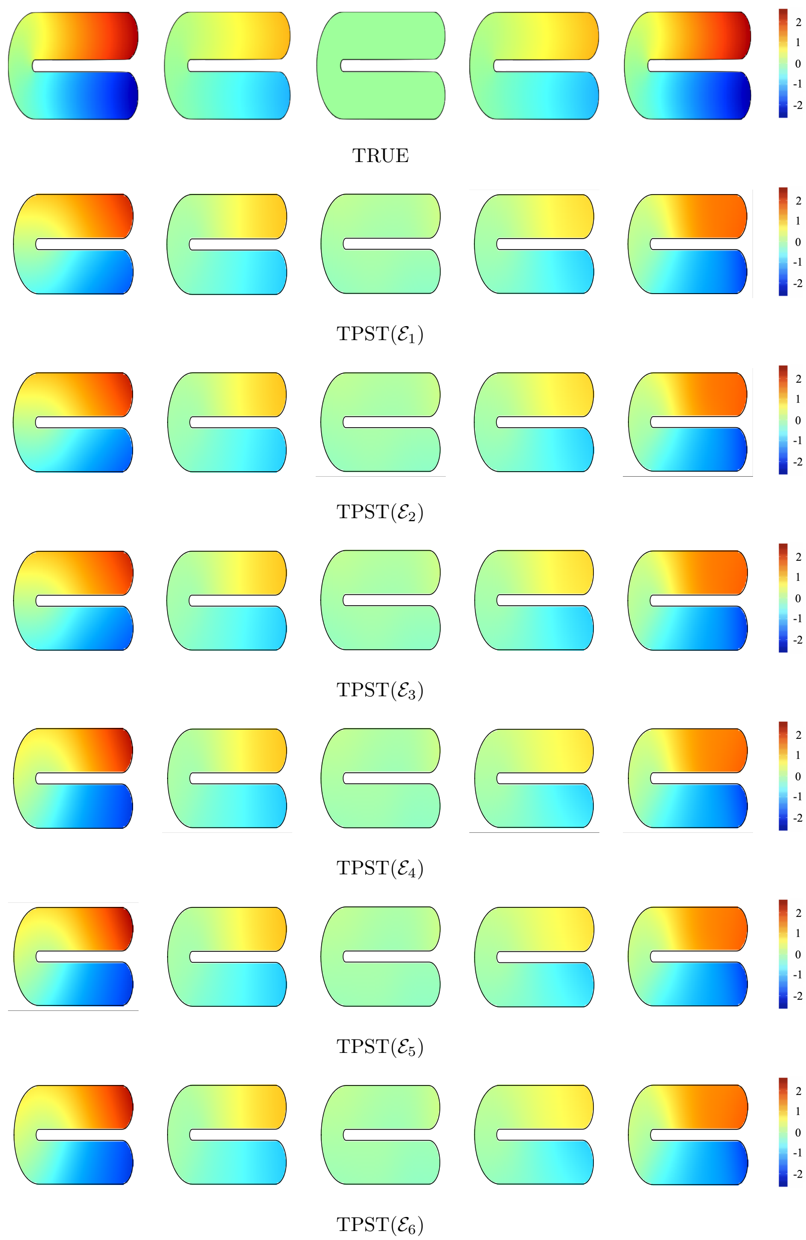


Figure D.4: Sequences of spatial plots of the coefficient functions for Simulation Study 2 using TPST with different triangular prisms.

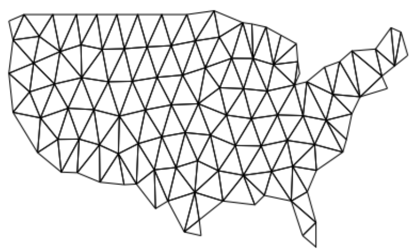
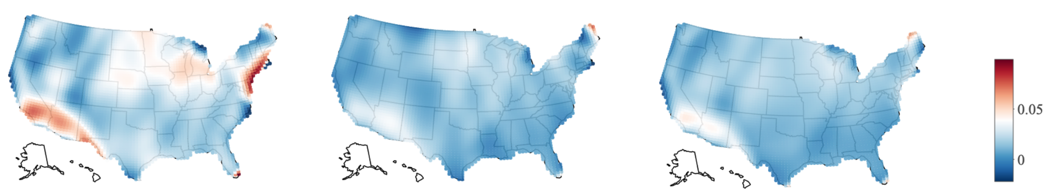
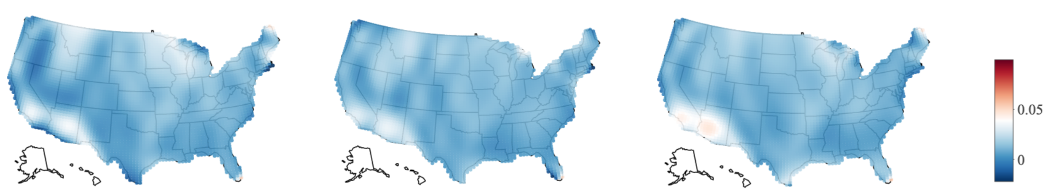


Figure D.5: Triangulation used in the TPST.

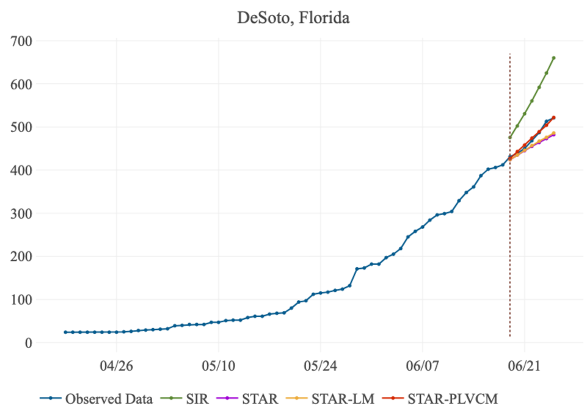


(a) April 25 (b) May 5 (c) May 15

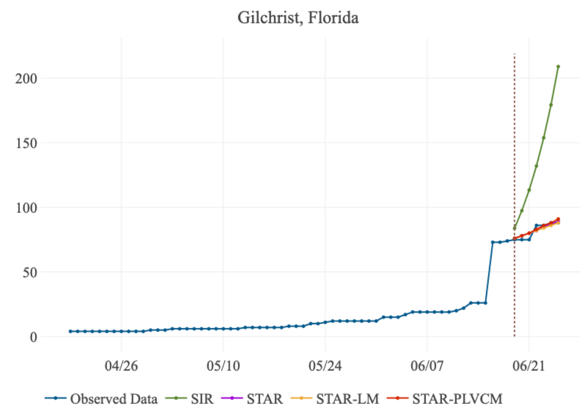


(d) May 25 (e) June 5 (f) June 15

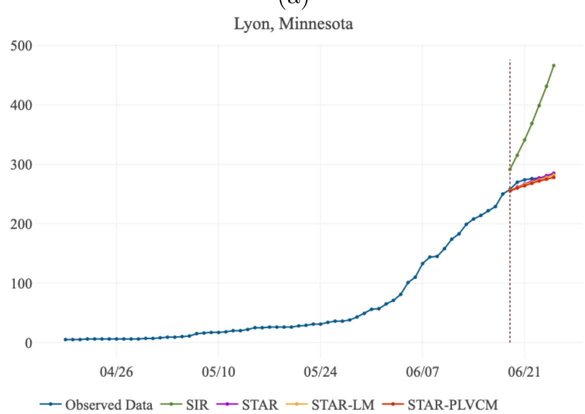
Figure D.6: Spatial plots of the estimated coefficient functions in death model.



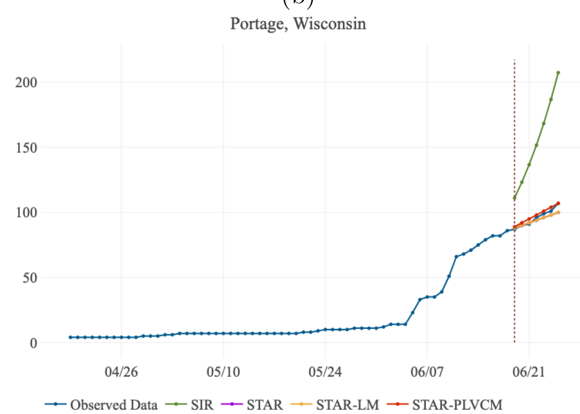
(a)



(b)



(c)



(d)

Figure D.7: Plots of prediction performance using STAR-PLVCM, STAR-LM, STAR, and SIR for the cumulative infected cases in DeSoto County, Florida (a), Gilchrist County, Florida (b), Lyon County, Minnesota (c), and Portage County, Wisconsin (d). The prediction starts from June 19 and is based on the training data from April 19 to June 18.