SPATIOTEMPORAL AUTOREGRESSIVE PARTIALLY LINEAR VARYING COEFFICIENT MODELS

Shan Yu\textsuperscript{1}, Yueying Wang\textsuperscript{2}, Li Wang\textsuperscript{3} and Lei Gao\textsuperscript{3}

\textsuperscript{1}University of Virginia, Charlottesville, VA 22904, USA.

\textsuperscript{2}Iowa State University, Ames, IA 50011, USA.

\textsuperscript{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 $a$, denote $\|a\|$ its Euclidean norm. For a matrix $A = (a_{ij})$, denote $\|A\|_\infty = \max_i \sum_j |a_{ij}|$, and let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be the smallest eigenvalue and largest eigenvalue of matrix $A$, 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 $\Omega \times T$, let $\|g\|_{\infty, \Omega \times T}$ and $\|g\|_{L^2, \Omega \times T}$, for notational simplicity. For a vector valued function $g = (g_1, \ldots, g_{p_2})^\top$, denote $\|g\|_{L^2} = \left\{ \sum_{k=1}^{p_2} \|g_k\|_{L^2}^2 \right\}^{1/2}$ and $\|g\|_{\infty} = \max_{1 \leq k \leq p_2} \|g_k\|_\infty$. Further denote $\|g\|_{\nu, \infty} = \max_{1 \leq k \leq p_2} |g_k|_{\nu, \infty}$, where $|g_k|_{\nu, \infty} = \max_{|a_1 + a_2 + a_3| = \nu} \|\nabla_{a_1}^{a_3} g_k\|_{\infty}$. Denote $T(\varphi, d, r)(\mathcal{E}) = \{(g_1, \ldots, g_{p_2}) : g_k \in T(\varphi, d, r)(\mathcal{E}), 1 \leq k \leq p_2\}$. For any discrete set $A$, let $|A|$ be the cardinality of $A$.

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

$$\langle g^{(1)}, g^{(2)} \rangle_n = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{p_2} X_{ik} g_k^{(1)}(S_{i1}, S_{i2}, T_i) \right\} \left\{ \sum_{k=1}^{p_2} X_{ik} g_k^{(2)}(S_{i1}, S_{i2}, T_i) \right\}. \quad (A.1)$$
and theoretical inner product as
\[
\langle g^{(1)}, g^{(2)} \rangle = E \left\{ \sum_{k=1}^{p_2} X_k g^{(1)}_k(S_1, S_2, T) \right\} \left\{ \sum_{k=1}^{p_2} X_k g^{(2)}_k(S_1, S_2, T) \right\}, \tag{A.2}
\]
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
\[
\langle g^{(1)}, g^{(2)} \rangle_{f_1} = p_2 \sum_{k=1}^{p_2} \int_{\Omega \times T} \left\{ (\nabla^2_{s_1} g^{(1)}_k)(\nabla^2_{s_1} g^{(2)}_k) + (\nabla^2_{s_2} g^{(1)}_k)(\nabla^2_{s_2} g^{(2)}_k) \right\} ds_1 ds_2 dt,
\]
\[
\langle g^{(1)}, g^{(2)} \rangle_{f_2} = p_2 \sum_{k=1}^{p_2} \int_{\Omega \times T} (\nabla^2_{t} g^{(1)}_k)(\nabla^2_{t} g^{(2)}_k) ds_1 ds_2 dt.
\]
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 \(A\) and \(B\) are bounded in row and columns sums, then the matrix \(AB\) is bounded in row and columns sums.

P2. Suppose the elements \(a_{n,ij}\) of an \(n \times n\) matrix \(A_n\) are \(O(e^{-1})\) uniformly for all \(i,j\). If an \(n \times n\) matrix \(B_n\) is uniformly bounded in row and columns sums, then the elements of \(A_nB_n\) and \(B_nA_n\) are \(O(e^{-1})\) uniformly for all \(i,j\).

P3. If the elements in \(Z\) are uniformly bounded, and \(\lim_{n \to \infty} n^{-1}Z^TZ\) exists and is nonsingular, then \(Z(Z^TZ)^{-1}Z^T\) and \(I_n - Z(Z^TZ)^{-1}Z^T\) are uniformly bounded in both row and column sums.

P4. Let \(A_n\) be an \(n \times n\) matrix. We have \(E(\epsilon^T A_n \epsilon) = \sigma_0^2 \text{tr}(A_n)\), and \(\text{Var}(\epsilon^T A_n \epsilon) = (m_4 - 3\sigma_0^4) \sum_{i=1}^{n} a_{ii} + \sigma_0^4 \{ \text{tr}(A_n A_n^T) + \text{tr}(A_n^2) \}\), where \(m_4 = E\epsilon_i^4\).

### A.2 Assumptions

In this section, we state the technical assumptions used in the main paper. Let \(G = W\Sigma^{-1}_0\).

(A1) Let \(h_b = \pi_b - \pi_{b-1}\) be the distance between two adjacent knots, satisfying that \(C_1 \leq \max h_b / \min h_b \leq C_2\). Denote \(h = \max_{1 \leq b \leq N+1} h_b\).

(A2) Let \(\rho_\tau\) be the radius of the largest disk contained in \(\tau\). The triangulation \(\Delta\) is \(\delta\)-quasi-uniform, that is, there exists a positive constant \(\delta\) such that the triangulation \(\Delta\) satisfies \(|\Delta|/\rho_\tau \leq \delta\), for all \(\tau \in \Delta\).
(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) \(\{(Z_i, X_i, S_i, T_i, \epsilon_i)\}_{i=1}^{n}\) 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 \(\text{E}((Z_i^\top, X_i^\top) (Z_i^\top, X_i^\top)|S_i = s, T_i = t)\). 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 \(s \in \Omega, t \in \mathcal{T}\).

(A5) Let \(g^*_r(s_1, s_2, t) = \{g_{r,1}^*(s_1, s_2, t), \ldots, g_{r,p_2}^*(s_1, s_2, t)\}^\top\) be the vector of functions that minimizes \(\text{E}(Z_r - X_r^\top g(S_r, T_r))^2\) over \(g(s_1, s_2, t) = \{g_1(s_1, s_2, t), \ldots, g_{p_2}(s_1, s_2, t)\}^\top\). For \(g^*(s_1, s_2, t) = \{g_{r,1}^*(s_1, s_2, t), \ldots, g_{p_2}^*(s_1, s_2, t)\}^\top, n^{-1}\text{E}(Z - X g^*(S, T))\}^\top\{Z - X g^*(S, T)\}\) is positive definite.

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

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

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

(A9) The matrix \(\Xi_0\) is nonsingular. The matrices \(W\) and \(\Xi_0^{-1}\) are uniformly bounded in both row and column sums. Let \(\varpi\) be a compact parameter space and \(\alpha_0\) be an interior point of \(\varpi\). For any \(\alpha \in \varpi, \Xi(\alpha)^{-1}\) exists and is uniformly bounded in both row or column sums.

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

(A11) \(\lim_{n \to \infty} n^{-1}(Z, \Xi, G\mu_0)^\top(Z, \Xi, G\mu_0)\) exists and is nonsingular.

(A12) \(\lim_{n \to \infty} n^{-1}\text{E}\{\{G\mu_0\}^\top(I_n - \Pi_{2,A})^\top(I_n - \Pi_{2,A})G\mu_0\} \neq 0,\) where \(\Pi_{2,A}\) 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 \cite{Huang2003} and \cite{Xue2006}. 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, $X$ and $Z$ are assumed functionally unrelated as in Assumption (A5). Assumption (A6) describes a smoothness condition commonly used in the non-parametric 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 $G_{\mu_0}$ and explanatory variables are not asymptotically multicollinear. Assumption (A12) ensures the uniqueness of the maximizer of the profiled likelihood function of $\alpha$. 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 $\Delta$. 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 $|\Delta|$ is a $\delta$-quasi-uniform triangulation of a polygonal domain $\Omega$, and $\psi(\cdot) \in W^{d+1,\infty}(\Omega)$.

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

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

Lemma [A.1] shows that, under some regularity conditions, $S_0^d(\Delta)$ and $S_\exists^d(\Delta)$, $d \geq 3r + 2$, have full approximation power.

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

$$c_0 \left\{ \sum_{b=1}^{N+\varrho} a_b^2 (\pi_b - \pi_{b-\varrho}) \right\}^{1/2} \leq \|s\|_{L_2} \leq \left\{ \sum_{b=1}^{N+\varrho} a_b^2 (\pi_b - \pi_{b-\varrho}) \right\}^{1/2}.$$  \hspace{1cm} (A.3)

**Lemma A.3** (Lemma 1, Lai and Wang (2013)). Let $\{B_m\}_{m \in \mathcal{M}}$ be the basis for $S_\exists^d(\Delta)$ 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

$$C_1 |\Delta|^2 \sum_{m \in \mathcal{M}} |b_m|^2 \leq \left\| \sum_{m \in \mathcal{M}} b_m B_m \right\|_{L^2(\Omega)}^2 \leq C_2 |\Delta|^2 \sum_{m \in \mathcal{M}} |b_m|^2.$$ \hspace{1cm} (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 $\xi_{u,b}^q = (-1)^{q-1}(q-1)! \phi_{\varrho}^q((\varrho - q)!)/(\varrho - q)!$, where $\phi_{\varrho}(t) = \prod_{s=1}^{\varrho-1}(t-\pi_{b+s})$. Let $\{c_q, q \leq q \leq q-1\}$ be the coefficients of the $q$th Bernoulli polynomial $P_q$, i.e. $P_q(x) = x^q + \sum_{q=0}^{q-1} c_q x^q$. For $1 \leq b \leq N + \varrho$ and a sufficiently smooth function $g(s_1,s_2,t) \in \Omega \times T$, we define the linear operator as

$$u_b(g)(s_1,s_2) = \sum_{q=0}^{q-1} c_{q,b,\varrho} g(s_1,s_2,t)\left|_{s=\pi_b}^{\pi_{b-\varrho}} \right.$$  

where $\omega$ 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 $g(s_1,s_2,t) \in \mathbb{W}^{d+1,\infty}(\Omega) \otimes C^0(T)$,

$$\sup_{t \in T} \left| \nabla_t^{\varrho+s} \left\{ \nabla_{s_1}^{\varrho} \nabla_{s_2}^{\varrho} g(s_1,s_2,t) - \sum_{b=1}^{N+\varrho} u_b(\nabla_{s_1}^{\varrho} \nabla_{s_2}^{\varrho} g)(s_1,s_2) U_b(t) \right\} \right| \leq C_1 \sup_{t \in T} \left| \nabla_{s_1}^{\varrho} \nabla_{s_2}^{\varrho} g(s_1,s_2,t) \right| h^{\varrho-\varrho} \leq C_1 \left\| \nabla_{s_1}^{\varrho} \nabla_{s_2}^{\varrho} g \right\| \infty h^{\varrho-\varrho},$$ \hspace{1cm} (A.5)

where $h = \max_{1 \leq b \leq N+\varrho} h_b$. $u_b(\nabla_{s_1}^{\varrho} \nabla_{s_2}^{\varrho} g)(s_1,s_2)$ is a function of $(s_1,s_2)$ and $u_b(\nabla_{s_1}^{\varrho} \nabla_{s_2}^{\varrho} g)(s_1,s_2) = \nabla_{s_1}^{\varrho} \nabla_{s_2}^{\varrho} u_b(g)(s_1,s_2)$. According to Lemma [A.1] there exists $u_b^{\star}(g)(s_1,s_2) = \sum_{m \in \mathcal{M}} c_{b,m} B_m(s_1,s_2)$.
such that, under Assumption (A2) in the main paper and \(g(s_1, s_2, t) \in W_{d+1,\infty}(\Omega) \otimes C^2(\mathcal{T})\),

\[
\sup_{(s_1, s_2) \in \Omega} \left| \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} \left\{ u_b(g)(s_1, s_2) - \sum_{m \in \mathcal{M}} \kappa_{b,m} B_m(s_1, s_2) \right\} \right|
\]

\[
= \sup_{(s_1, s_2) \in \Omega} \left| u_b(\nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} g)(s_1, s_2) - \sum_{m \in \mathcal{M}} \kappa_{b,m} \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} B_m(s_1, s_2) \right|
\]

\[
\leq C_2 \max_{a_1+a_2=d+1} \| \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} g \|_\infty |\Delta|^{d+1-a_1-a_2},
\]

which implies that

\[
\nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} \nabla_t^{a_3} g(s_1, s_2, t) - \sum_{b=1}^{N+\rho} u_b(\nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} g)(s_1, s_2) \nabla_t^{a_3} U_b(t)
\]

\[
= \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} \nabla_t^{a_3} g(s_1, s_2, t) - \sum_{b=1}^{N+\rho} \left\{ \sum_{m \in \mathcal{M}} \kappa_{b,m} \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} B_m(s_1, s_2) + O(\|\Delta\|^{d+1-a_1-a_2}) \right\} \nabla_t^{a_3} U_b(t)
\]

\[
= \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} \nabla_t^{a_3} g(s_1, s_2, t) - \sum_{b=1}^{N+\rho} \sum_{m \in \mathcal{M}} \kappa_{b,m} \nabla_{s_1}^{a_1} \nabla_{s_2}^{a_2} B_m(s_1, s_2) \nabla_t^{a_3} U_b(t) + O(\|\Delta\|^{d+1-a_1-a_2}).
\]

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

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

Thus, Lemma 1 has been established. \qed

Proof of Lemma 3 Note that the \(|\mathcal{J}|\)-dimensional vector \(\gamma = \{\gamma_1, \ldots, \gamma_{|\mathcal{J}|}\}\) can be written as vector \(\{\gamma_1, \ldots, \gamma_{|\mathcal{J}|}, N+\rho, \gamma_2, \ldots, \gamma_{|\mathcal{M}|}, N+\rho\}\), where \(\gamma_{m,b} = \gamma(m-1)_{|\mathcal{M}|+b}\). By (A.4), it is easy to see that

\[
\int_{\Omega} |g(s_1, s_2, t)|^2 ds_1 ds_2 \geq C_1 |\Delta|^2 \sum_{m \in \mathcal{M}} \left| \sum_{b=1}^{N+\rho} U_b(t) \gamma_{m,b}^* \right|^2. \]

Therefore, by Assumptions (A1) – (A2), and the stability properties in (A.3) and (A.4), we have

\[
\|g\|_{L^2}^2 = \int_{\Omega \times \mathcal{T}} |g(s_1, s_2, t)|^2 ds_1 ds_2 dt \geq C_1 |\Delta|^2 \sum_{m \in \mathcal{M}} \int_{\mathcal{T}} \left| \sum_{b=1}^{N+\rho} U_b(t) \gamma_{m,b}^* \right|^2 dt
\]

\[
\geq C_{1,\rho} |\Delta|^2 \sum_{m \in \mathcal{M}} \sum_{b=1}^{N+\rho} \gamma_{m,b}^* (\pi_b - \pi_{b-\rho}) \geq C_{1,\rho} \rho h |\Delta|^2 \sum_{m \in \mathcal{M}} \sum_{b=1}^{N+\rho} \gamma_{m,b}^* = C_{1,\rho} \rho h |\Delta|^2 \sum_{j \in \mathcal{J}} \gamma_j^2.
\]

Similarly, by (A.4),

\[
\int_{\Omega} |g(s_1, s_2, t)|^2 ds_1 ds_2 \leq C_2 |\Delta|^2 \sum_{m \in \mathcal{M}} \left| \sum_{b=1}^{N+\rho} U_b(t) \gamma_{m,b}^* \right|^2.
\]

Then, it follows directly from (A.3) that

\[
\int_{\Omega \times \mathcal{T}} |g(s_1, s_2, t)|^2 ds_1 ds_2 dt \leq C_2 |\Delta|^2 \sum_{m \in \mathcal{M}} \int_{\mathcal{T}} \left| \sum_{b=1}^{N+\rho} U_b(t) \gamma_{m,b}^* \right|^2 dt
\]

\[
\leq C_2 |\Delta|^2 \sum_{m \in \mathcal{M}} \sum_{b=1}^{N+\rho} \gamma_{m,b}^* (\pi_t - \pi_{t-\rho}) \leq C_{2,\rho} |\Delta|^2 h \sum_{m \in \mathcal{M}} \sum_{b=1}^{N+\rho} \gamma_{m,b}^* = C_{2,\rho} |\Delta|^2 h \sum_{j \in \mathcal{J}} \gamma_j^2.
\]
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 \( T^{(e, d, r)}(E) \), and \( g_k^{(1)} = \sum_{j \in J} \gamma_k^{(1)}(s_1, s_2, t) \), \( g_k^{(2)} = \sum_{j \in J} \gamma_k^{(2)}(s_1, s_2, t) \). Suppose Assumptions (A1) – (A8) hold, then

\[
R_n = \sup_{g^{(1)}, g^{(2)} \in T^{(e, d, r)}(E)_{p_2}} \left\{ \frac{\langle g^{(1)}, g^{(2)} \rangle_n - \langle g^{(1)}, g^{(2)} \rangle}{\|g^{(1)}\| \|g^{(2)}\|} \right\} = O_{a.s.} \left\{ h^{-1/2} \|\Delta\|^{-1} n^{-1/2} (\log n)^{1/2} \right\}.
\]

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

**Proof.** It is easy to see

\[
\langle g^{(1)}, g^{(2)} \rangle_n = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{p_2} \sum_{j \in J} \gamma_k^{(1)}(s_1, s_2, T_i) \sum_{j' \in J} \gamma_k^{(2)}(s_1, s_2, T_i) \langle \psi_j, \psi_{j'} \rangle_{n,k,k'} \right\} \leq 1 \sum_{i=1}^{n} \sum_{j \in J} X_{ik} X_{ik'} \psi_j(S_{i1}, S_{i2}, T_i) \psi_j'(S_{i1}, S_{i2}, T_i).
\]

Similarly, we have

\[
\langle g^{(1)}, g^{(2)} \rangle = \sum_{k, k' = 1}^{p_2} \sum_{j, j' \in J} \gamma_k^{(1)}(s_1, s_2, t) \gamma_{k'}^{(2)}(s_1, s_2, t) \langle \psi_j, \psi_{j'} \rangle_{k, k', t},
\]

where \( \langle \psi_j, \psi_{j'} \rangle_{k, k', t} = E \{ X_{ik} X_{ik'} \psi_j(S_{i1}, S_{i2}, T_i) \psi_j'(S_{i1}, S_{i2}, T_i) \}. \)

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 \( |\ell(j) - \ell(j')| > 0 \), \( \psi_j(s_1, s_2, t) \psi_j'(s_1, s_2, t) = 0 \). Therefore, we have

\[
\langle g^{(1)}, g^{(2)} \rangle_n - \langle g^{(1)}, g^{(2)} \rangle = \sum_{k, k' = 1}^{p_2} \sum_{j, j' \in J} \gamma_k^{(1)}(s_1, s_2, t) \gamma_{k'}^{(2)}(s_1, s_2, t) \langle \psi_j, \psi_{j'} \rangle_{n,k,k'} - \sum_{k, k' = 1}^{p_2} \sum_{j, j' \in J} \gamma_k^{(1)}(s_1, s_2, t) \gamma_{k'}^{(2)}(s_1, s_2, t) \langle \psi_j, \psi_{j'} \rangle_{k, k', t}
\]

\[
= \sum_{k, k' = 1}^{p_2} \sum_{|\ell(j) - \ell(j')| \leq \# \{ \}|m(j) - m(j')| \leq (d+2)(d+1)/2 \} \langle \psi_j, \psi_{j'} \rangle_{n,k,k'} - \langle \psi_j, \psi_{j'} \rangle_{k, k', t}
\]

\[
\leq C \left\{ \sum_{k, j} \gamma_k^{(1)}(s_1, s_2, t) \sum_{k', j'} \gamma_{k'}^{(2)}(s_1, s_2, t) \right\}^{1/2} \max_{j, j' \in J} \langle \psi_j, \psi_{j'} \rangle_{n,k,k'} - \langle \psi_j, \psi_{j'} \rangle_{k, k', t}.
\]
It follows Lemma 2 that, for \( v = 1, 2 \),
\[
C_1 h |\Delta|^2 \sum_{k,j} \gamma_{kj}^{(v)} \leq \| g^{(v)} \|^2 \leq C_2 h |\Delta|^2 \sum_{k,j} \gamma_{kj}^{(v)},
\]
which implies that
\[
C_1 h |\Delta|^2 \left( \sum_{k,j} \gamma_{kj}^{(1)2} \sum_{k',j'} \gamma_{k'j'}^{(2)2} \right)^{1/2} \leq \| g^{(1)} \| \| g^{(2)} \| \leq C_2 h |\Delta|^2 \left( \sum_{k,j} \gamma_{kj}^{(1)2} \sum_{k',j'} \gamma_{k'j'}^{(2)2} \right)^{1/2}.
\]
Combining (A.7) and (A.8), we have
\[
R_n \leq C^{-1} h^{-1} |\Delta|^{-2} \max_{k,k'=1,...,p_2} | (\langle \psi_j, \psi_{j'} \rangle_{n,kk'} - \langle \psi_j, \psi_{j'} \rangle_{kk'} |.
\]
Note that \( X_{ik} \) are uniformly bounded. Hence, it is straightforward to obtain
\[
\max_{k,k'=1,...,p_2} | (\langle \psi_j, \psi_{j'} \rangle_{n,kk'} - \langle \psi_j, \psi_{j'} \rangle_{kk'} | = O_{a.s.} \left\{ n^{-1/2} h^{1/2} |\Delta|(\log n)^{1/2} \right\}.
\]

Lemma A.4 follows.

**Appendix B. Proof of the Main Results in Section 3**

For the matrices \( X_{\psi^*} \) and \( Z \) given in (2.6), we denote
\[
V \equiv \left( \begin{array}{cc} V_{11} & V_{12} \\ V_{21} & V_{22} \end{array} \right) = \left( \begin{array}{cc} Z^T Z & Z^T X_{\psi^*} \\ X_{\psi^*}^T Z & X_{\psi^*}^T X_{\psi^*} \end{array} \right) + P_\Lambda,
\]
where \( P_\Lambda \) is defined in (2.7). Then, the inverse of \( V \) can be represented as
\[
V^{-1} \equiv U = \left( \begin{array}{cc} U_{11} & -U_{11} V_{12} V_{22}^{-1} \\ -U_{22} V_{21} V_{11}^{-1} & U_{22} \end{array} \right),
\]
where \( U_{11} = V_{11} - V_{12} V_{22}^{-1} V_{21} \), and \( U_{22} = V_{22} - V_{21} V_{11}^{-1} V_{12} \).

Denote
\[
P_{\lambda_1} = \lambda_1 I_{p_2} \otimes (Q_2^T P_1 Q_2), \quad P_{\lambda_2} = \lambda_2 I_{p_2} \otimes (Q_2^T P_2 Q_2),
\]
and
\[
\Gamma_\Lambda = X_{\psi^*}^T X_{\psi^*} + P_{\lambda_1} + P_{\lambda_2},
\]
then, it is clear that \( V_{22} = \Gamma_\Lambda \). Next, we denote
\[
\Pi_Z = Z(Z^T Z)^{-1} Z^T, \quad \Pi_{D,\Lambda} = D(D^T D + P_\Lambda)^{-1} D^T,
\]
\[
\Pi_{X_{\psi^*},\Lambda} = X_{\psi^*} \Gamma_\Lambda^{-1} X_{\psi^*}^T, \quad H_{X_{\psi^*},\Lambda} = I_n - \Pi_{X_{\psi^*},\Lambda},
\]
and
\[
(\Pi)_{X_{\psi^*},\Lambda} = X_{\psi^*} \Gamma_\Lambda^{-1} X_{\psi^*}, \quad H_{X_{\psi^*},\Lambda} = I_n - (\Pi)_{X_{\psi^*},\Lambda}.
\]
Then, we have

\[
U_{11}^{-1} = Z^\top(I_n - \Pi_{X_\psi, \Lambda})Z, \quad U_{22}^{-1} = X_\psi^\top(I_n - \Pi_Z)X_\psi + \mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2}, \tag{B.4}
\]

where \(\mathbb{P}_{\lambda_1}, \mathbb{P}_{\lambda_2},\) and \(\Gamma_{\Lambda}\) in \((\text{B.1})\) and \((\text{B.2})\).

Recall that

\[
L_n (\sigma^2, \alpha) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) + \log(||\xi(\alpha)||) - \frac{1}{2\sigma^2} ||Y(\alpha) - \mathbb{D}\xi(\alpha)||^2,
\]

\[
L_n (\alpha) = -\frac{n}{2} \log(2\pi) + 1 \} - \frac{n}{2} \log(\sigma(\alpha)) + \log(||\xi(\alpha)||),
\]

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

\[
\sigma^{*2}(\alpha) = n^{-1}E||Y(\alpha) - \mathbb{D}\hat{\xi}(\alpha)||^2. \tag{B.5}
\]

It is straightforward to verify that

\[
Q_n(\alpha) \equiv \max_{\sigma^2} E \{L_n(\sigma^2, \alpha)\} = -\frac{n}{2} \log(2\pi) + 1 \} - \frac{n}{2} \log(\sigma^{*2}(\alpha)) + \log(||\xi(\alpha)||). \tag{B.6}
\]

## B.1 Proof of Theorem 1

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

**Proof.** It is straightforward to obtain that

\[
n^{-1} \{L_n(\alpha) - Q_n(\alpha)\} = -2^{-1} \{\log(\hat{\sigma}(\alpha)) - \log(\sigma^{*2}(\alpha))\},
\]

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

\[
(I_n - \Pi_{D, \Lambda})\xi(\alpha)Y = (I_n - \Pi_{D, \Lambda})I_n + (\alpha_0 - \alpha)G\mu_0 + (I_n - \Pi_{D, \Lambda})\xi(\alpha)\xi_0^{-1}\epsilon
\]

\[
= (\alpha_0 - \alpha)(I_n - \Pi_{D, \Lambda})G\mu_0 + (I_n - \Pi_{D, \Lambda})\mu_0 + (I_n - \Pi_{D, \Lambda})\xi(\alpha)\xi_0^{-1}\epsilon.
\]

Thus, \(\hat{\sigma}(\alpha)\) in \((\text{2.9})\) can be written as

\[
\hat{\sigma}(\alpha) = \frac{1}{n} ||(I_n - \Pi_{D, \Lambda})\xi(\alpha)Y||^2 + \frac{2(\alpha_0 - \alpha)}{n} \mu_0^\top(I_n - \Pi_{D, \Lambda})^\top(I_n - \Pi_{D, \Lambda})G\mu_0
\]

\[
+ \frac{2}{n} \mu_0^\top(I_n - \Pi_{D, \Lambda})^\top(I_n - \Pi_{D, \Lambda})\xi(\alpha)\xi_0^{-1}\epsilon + \frac{2}{n} \mu_0^\top(I_n - \Pi_{D, \Lambda})^\top(I_n - \Pi_{D, \Lambda})\xi(\alpha)\xi_0^{-1}\epsilon.
\]
According to Lemma [C.4], we have

\[
\hat{\sigma}^2(\alpha) = \frac{(\alpha_0 - \alpha)^2}{n} \| (I_n - \Pi_{D,A})(G\mu_0) \|^2 + \frac{1}{n} \| (I_n - \Pi_{D,A})\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \|^2 \\
+ \frac{2(\alpha_0 - \alpha)}{n} (G\mu_0)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \\
+ \frac{2}{n} \mu_0^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon + o_P(1). 
\]  
(B.7)

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

\[
\sigma^2(\alpha) = \frac{\sigma_0^2}{n} \mathrm{tr} \left\{ (\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} \\
+ \frac{1}{n} \mathbb{E} \| (I_n - \Pi_{D,A}) (I_n + (\alpha_0 - \alpha)G) \mu_0 \|^2 \\
= \frac{\sigma_0^2}{n} \mathrm{tr} \left\{ (\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} - \frac{2\sigma_0^2}{n} \mathrm{tr} \left\{ (\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)^\top \Pi_{D,A}^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} \\
+ \frac{\sigma_0^2}{n} \mathrm{tr} \left\{ (\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)^\top \Pi_{D,A}^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} + \frac{(\alpha_0 - \alpha)^2}{n} \mathbb{E} \| (I_n - \Pi_{D,A}) G \mu_0 \|^2 \\
+ \frac{1}{n} \mathbb{E} \| (I_n - \Pi_{D,A}) \mu_0 \|^2 + \frac{2(\alpha_0 - \alpha)}{n} \mathbb{E} \left\{ \mu_0^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A}) G \mu_0 \right\}.
\]

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

\[
\mathrm{tr} \left\{ (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \Pi_{D,A}^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} = O_P(|\Delta|^{-2} h^{-1}), 
\]  
(B.8)

\[
\mathrm{tr} \left\{ (\Pi_{D,A} \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \Pi_{D,A} \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} = O_P(|\Delta|^{-2} h^{-1}). 
\]  
(B.9)

Then, it implies

\[
\sigma^2(\alpha) = \frac{\sigma_0^2}{n} \mathrm{tr} \left\{ (\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} + \frac{(\alpha_0 - \alpha)^2}{n} \mathbb{E} \| (I_n - \Pi_{D,A}) G \mu_0 \|^2 \\
+ \frac{1}{n} \mathbb{E} \| (I_n - \Pi_{D,A}) \mu_0 \|^2 + \frac{2(\alpha_0 - \alpha)}{n} \mathbb{E} \left\{ \mu_0^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A}) G \mu_0 \right\} + o_P(1). 
\]  
(B.10)

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

\[
\hat{\sigma}^2(\alpha) - \sigma^2(\alpha) = \frac{(\alpha_0 - \alpha)^2}{n} \left[ \| (I_n - \Pi_{D,A})(G\mu_0) \|^2 - \mathbb{E} \| (I_n - \Pi_{D,A})(G\mu_0) \|^2 \right] \\
+ \frac{1}{n} \mathbb{E} \| (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon)^\top \{ (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \Pi_{D,A} \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \} + \frac{1}{n} \mathbb{E} \| \Pi_{D,A} \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \|^2 \\
- \frac{\sigma_0^2}{n} \mathrm{tr} \left\{ (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \right\} + \frac{2(\alpha_0 - \alpha)}{n} (G\mu_0)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \\
+ \frac{2}{n} \mu_0^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon + o_P(1).
\]

By Property P1 in Appendix A and Assumption (A9), \( (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1} \) is bounded uniformly, for any \(\alpha \in \mathbb{R}\), both in row and column sums. Then, \( \text{Var} \left[ \epsilon^\top \{ (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \} \right] = O(n) \). Thus,

\[
\sup_{\alpha \in \mathbb{R}} \left| \frac{1}{n} \mathbb{E} \left[ \epsilon^\top \{ (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \} - \frac{1}{n} \mathbb{E} \left[ \epsilon^\top \{ (\tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1})^\top \tilde{\Sigma}(\alpha)\tilde{\Sigma}_0^{-1}\epsilon \} \right] \right] \right| = o_P(1), 
\]  
(B.11)
and $E\{\epsilon^\top (\Xi(\alpha)\Xi_0^{-1})^\top \Xi(\alpha)\Xi_0^{-1} \epsilon\} = \sigma_0^2 \text{tr} [ (\Xi(\alpha)\Xi_0^{-1})^\top \Xi(\alpha)\Xi_0^{-1} ]$. Therefore,

$$\sup_{\alpha \in \mathbb{W}} \left\{ \frac{1}{n} \epsilon^\top (\Xi(\alpha)\Xi_0^{-1})^\top \Xi(\alpha)\Xi_0^{-1} \epsilon - \frac{\sigma_0^2}{n} \text{tr} [ (\Xi(\alpha)\Xi_0^{-1})^\top \Xi(\alpha)\Xi_0^{-1} ] \right\} = o_P(1). \quad (B.12)$$

Similarly, we have

$$\sup_{\alpha \in \mathbb{W}} \left\{ \frac{(\alpha_0 - \alpha)^2}{n} \| (I_n - \Pi_{D,A})(G\mu_0) \| ^2 - \frac{(\alpha_0 - \alpha)^2}{n} E \left\{ \| (I_n - \Pi_{D,A})(G\mu_0) \| ^2 \right\} \right\} = o_P(1).$$

Note that

$$(\alpha_0 - \alpha)(G\mu_0)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\Xi(\alpha)\Xi_0^{-1} \epsilon$$

$$= (\alpha_0 - \alpha)\epsilon^\top (I_n + (\alpha_0 - \alpha)G)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A}) G\mu_0$$

$$= (\alpha_0 - \alpha)\epsilon^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A}) G\mu_0$$

$$+ (\alpha_0 - \alpha)^2 (G\epsilon)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A}) G\mu_0.$$

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

$$\sup_{\alpha \in \mathbb{W}} \left\{ \frac{2(\alpha_0 - \alpha)}{n} (G\mu_0)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\Xi(\alpha)\Xi_0^{-1} \epsilon \right\} = o_P(1). \quad (B.13)$$

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

$$\sup_{\alpha \in \mathbb{W}} \left\{ \frac{2}{n} \mu_0^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})\Xi(\alpha)\Xi_0^{-1} \epsilon \right\}$$

$$= \sup_{\alpha \in \mathbb{W}} \left\{ \frac{2}{n} \mu_0^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})(I_n + (\alpha_0 - \alpha)G)\epsilon \right\} = o_P(1).$$

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

$$n^{-1} \epsilon^\top (\Xi(\alpha)\Xi_0^{-1})^\top \Pi_{D,A} \Xi(\alpha)\Xi_0^{-1} \epsilon = O_P(n^{-1}|\Delta|^{-2}h^{-1}),$$

$$n^{-1} \| \Pi_{D,A} \Xi(\alpha)\Xi_0^{-1} \epsilon \| ^2 = O_P(n^{-1}|\Delta|^{-2}h^{-1}).$$

It follows that

$$\sup_{\alpha \in \mathbb{W}} \left\{ n^{-1} \epsilon^\top (\Xi_0^{-1})^\top \Xi(\alpha)^\top \Pi_{D,A} \Xi(\alpha)\Xi_0^{-1} \epsilon \right\} \asymp \sup_{\alpha \in \mathbb{W}} \left\{ n^{-1} \| \Pi_{D,A} \Xi(\alpha)\Xi_0^{-1} \epsilon \| ^2 \right\} = o_P(1). \quad (B.14)$$

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

$$\sup_{\alpha \in \mathbb{W}} \left| \log \hat{\sigma}^2(\alpha) - \log \sigma^*^2(\alpha) \right| = \sup_{\alpha \in \mathbb{W}} \left| \frac{1}{\hat{\sigma}^2(\alpha)} |\hat{\sigma}^2(\alpha) - \sigma^*^2(\alpha)| \right| = o_P(1),$$

where $\hat{\sigma}^2(\alpha)$ is between $\hat{\sigma}^2(\alpha)$ and $\hat{\sigma}^2(\alpha)$. Therefore, $\sup_{\alpha \in \mathbb{W}} \left| n^{-1} \{ L_n(\alpha) - Q_n(\alpha) \} \right| = o_P(1)$.

Lemma [B.1] is established. \qed
Proof of Theorem 1. The proof follows from the following steps:

(i) \( \sup_{\alpha \in \mathbb{R}} |n^{-1} L_n(\alpha) - n^{-1} Q_n(\alpha)| = o_p(1) \); (ii) for any \( \alpha \in \mathbb{R} \), \( Q_n(\alpha) - Q_n(\alpha_0) \leq 0 \); (iii) for any \( \varepsilon > 0 \), \( \limsup_{n \to \infty} \max_{\alpha \in \mathcal{N}_\varepsilon(\alpha_0)} n^{-1} \{ Q_n(\alpha) - Q_n(\alpha_0) \} < 0 \), where \( \mathcal{N}_\varepsilon(\alpha_0) \) is the open neighborhood of \( \alpha_0 \) with radius \( \varepsilon \) 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 3.1 in Appendix B.

Proof of Step (ii). Consider a standard SAR model: \( Y = \alpha W Y + \epsilon \), where \( \epsilon \sim N(0, \sigma^2 I_n) \). Then the corresponding log-likelihood function is \( L_n(\alpha) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) + \log|\Sigma(\alpha)| - \frac{1}{\sigma^2} \| \Sigma(\alpha) Y \|^2 \). Notice that

\[
Q_n^*(\alpha) = \max_{\sigma^2} E_{(\alpha_0, \sigma_0^2)} \{ L_n(\sigma^2, \alpha) \} = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2(\alpha)) + \log|\Sigma(\alpha)|,
\]

where \( \sigma^2(\alpha) = \frac{\sigma^2}{n} \text{tr} \left\{ (\Sigma(\alpha) \Sigma_0^{-1})^T (\Sigma(\alpha) \Sigma_0^{-1}) \right\} \) and \( E_{(\alpha_0, \sigma_0^2)} \) represents the expectation under the model \( Y = \alpha_0 W Y + \epsilon \), where \( \epsilon \sim N(0, \sigma_0^2 I_n) \). By Jensen’s inequality, we have

\[
Q_n^*(\alpha) \leq E_{(\alpha_0, \sigma_0^2)} \max_{\sigma^2} \{ L_n(\sigma^2, \alpha) \} \leq E_{(\alpha_0, \sigma_0^2)} \max_{\sigma^2, \alpha} \{ L_n(\sigma^2, \alpha) \} = Q_n^*(\alpha_0).
\]

It is straightforward to obtain

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

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

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

\[
\lim_{n \to \infty} \frac{1}{n} \{ Q_n(\alpha_n) - Q_n(\alpha_0) \} = 0, \quad \lim_{n \to \infty} \alpha_n = \bar{\alpha} \neq \alpha_0,
\]

(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 \( \alpha \). Thus, by (B.15), we have

\[
\lim_{n \to \infty} \frac{1}{n} |Q_n(\bar{\alpha}) - Q_n(\alpha_0)| = \lim_{n \to \infty} \frac{1}{n} |Q_n(\alpha_n) + Q_n(\alpha_n) - Q_n(\alpha_0)| \\
\leq \lim_{n \to \infty} \frac{1}{n} |Q_n(\alpha_n) - Q_n(\bar{\alpha})| + \lim_{n \to \infty} \frac{1}{n} |Q_n(\alpha_n) - Q_n(\alpha_0)| = 0.
\]
Proof of Theorem 2.

Note that, for any $\alpha$ and $\sigma$ defined in Section 2.4. Because $\hat{\alpha}$ is the maximizer of (B.18). Notice that $(\hat{\alpha}, \tilde{\sigma}, \tilde{\eta})$ be the maximizer of (B.18). Therefore, $\lim_{n \to \infty} n^{-1} \{Q_n(\alpha_n) - Q_n(\alpha_0)\} = 0$ implies

$$\lim_{n \to \infty} \frac{1}{n} \{\log(\sigma^2(\hat{\alpha})) - \log(\sigma^2(\tilde{\alpha}))\} = 0. \tag{B.16}$$

Thus, (B.16) leads to $\lim_{n \to \infty} \{\sigma^2(\hat{\alpha}) - \sigma^2(\tilde{\alpha})\} = 0$, which implies $\lim_{n \to \infty} n^{-1} \{I_n - \Pi_{D,\Lambda}\} \{I_n - (\alpha_0 - \bar{\alpha})G\} \mu_0 = 0$. Therefore, $\lim_{n \to \infty} n^{-1} \{G\mu_0\}^T (I_n - \Pi_{D,\Lambda})^T (I_n - \Pi_{D,\Lambda}) G \mu_0 = 0$. Then, (B.17) conflicts with Assumption (A12). Thus, $\alpha_0$ is unique. \qed

B.2 Proof of Theorem 2

Proof of Theorem 2 Define

$$\ell_n(\alpha, \sigma^2, \eta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) + \log(\|\chi\|) - \frac{1}{2\sigma^2} \|Y(\alpha) - \chi \eta - \chi \psi \cdot \theta(\alpha, \eta)\|^2, \tag{B.18}$$

where $\theta(\alpha, \eta) = \Gamma^{-1}_\chi \{Y(\alpha) - \chi \eta\}$. Denote $\kappa = (\alpha, \sigma^2, \eta)^T$, $\kappa_0 = (\alpha_0, \sigma^2_0, \eta_0)^T$ and let $\hat{\kappa} = (\hat{\alpha}, \hat{\sigma}, \hat{\eta})^T$ be the maximizer of (B.18). Notice that $(\hat{\alpha}, \hat{\sigma}, \hat{\eta})^T$ is equal to the TPST estimator defined in Section 2.4. Because $\hat{\kappa}$ is the maximizer of (B.18), $\nabla_\kappa \ell_n(\kappa)|_{\kappa = \hat{\kappa}} = 0$ and

$$n^{1/2} (\hat{\kappa} - \kappa_0) = \left( -n^{-1} \nabla_\kappa \ell_n(\kappa)|_{\kappa = \kappa_0} \right)^{-1} \left( n^{-1/2} \nabla_\kappa \ell_n(\kappa)|_{\kappa = \kappa_0} \right). \tag{B.19}$$

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

$$\nabla_\alpha \ell_n(\alpha, \sigma^2, \eta) = -\text{tr} \{\chi \psi^{-1} W\} + \frac{1}{\sigma^2} \{Y(\alpha) - \chi \eta - \chi \psi \cdot \theta(\alpha, \eta)\} H_{\chi \psi^{-1} \Lambda} W Y,$$

$$\nabla_\sigma \ell_n(\alpha, \sigma^2, \eta) = -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4} \{Y(\alpha) - \chi \eta - \chi \psi \cdot \theta(\alpha, \eta)\}^T \{Y(\alpha) - \chi \eta - \chi \psi \cdot \theta(\alpha, \eta)\},$$

$$\nabla_\eta \ell_n(\alpha, \sigma^2, \eta) = -\frac{1}{\sigma^2} \{Y(\alpha) - \chi \eta - \chi \psi \cdot \theta(\alpha, \eta)\}^T H_{\chi \psi^{-1} \Lambda} Z, \tag{B.20}$$
where $H_{x_{\theta},\Lambda}$ is given in \[B.3\]. In addition, we have the following second order derivatives of the profile log-likelihood function:

$$
\nabla_{\alpha}^2 \ell_n(\alpha, \sigma^2, \eta) = -\text{tr}(G^2) - \frac{1}{\sigma^2} (WY)^\top H_{x_{\theta},\Lambda}^\top H_{x_{\theta},\Lambda} WY,
$$

$$
\nabla_{\sigma}^2 \ell_n(\alpha, \sigma^2, \eta) = \frac{n}{2\sigma^4} \{ Y(\alpha) - Z\eta - X_{\theta}(\alpha, \eta) \} \mathbb{T} \{ Y(\alpha) - Z\eta - X_{\theta}(\alpha, \eta) \},
$$

$$
\nabla_{\eta} \nabla_{\eta} \ell_n(\alpha, \sigma^2, \eta) = -\frac{1}{\sigma^2} \mathbb{Z}^\top H_{x_{\theta},\Lambda}^\top H_{x_{\theta},\Lambda} \mathbb{Z},
$$

$$
\nabla_{\alpha} \nabla_{\alpha} \ell_n(\alpha, \sigma^2, \eta) = -\frac{1}{\sigma^2} \mathbb{Z}^\top H_{x_{\theta},\Lambda}^\top H_{x_{\theta},\Lambda} WY,
$$

$$
\nabla_{\eta} \nabla_{\eta} \ell_n(\alpha, \sigma^2, \eta) = -\frac{1}{\sigma^2} \mathbb{Z}^\top H_{x_{\theta},\Lambda}^\top \{ Y(\alpha) - Z\eta - X_{\theta}(\alpha, \eta) \} WY,
$$

$$
\nabla_{\sigma} \nabla_{\sigma} \ell_n(\alpha, \sigma^2, \eta) = -\frac{1}{\sigma^2} \mathbb{Z}^\top H_{x_{\theta},\Lambda}^\top \{ Y(\alpha) - Z\eta - X_{\theta}(\alpha, \eta) \} WY.
$$

(B.21)

We first study the asymptotical normality of $n^{-1/2} \nabla_{\kappa} \ell_n(\kappa)|_{\kappa=\kappa_0}$. According to Lemma 1 for $\beta_k(s_1, s_2, t)$, $k = 1, \ldots, p_2$, there exists $\theta^*_k$ such that $\sup_{(s_1, s_2, t) \in \Omega \times T} |\beta_k(s_1, s_2, t) - \psi^*(s_1, s_2, t)| = O(h^0 + |\Delta|^{d+1})$. Denote that $\theta^* = (\theta_1^T, \ldots, \theta_{p_2}^T)^T$. Notice that

$$
Y(\alpha_0) - Z\eta_0 - X_{\theta}(\alpha_0, \eta_0) = H_{x_{\theta},\Lambda}(\epsilon + \varsigma) + \eta,
$$

where $\varsigma = (\varsigma_1, \ldots, \varsigma_n)^T = \{ \sum_{k=1}^{p_2} X_{ik}\beta_k(s_1, s_2, T_i) - \sum_{k=1}^{p_2} X_{ik}\psi^*(s_1, s_2, T_i)^T \theta^*_k \}_{i=1}^n$ and $\eta = (t_1, \ldots, t_n)^T = X_{\theta}(P_{\lambda_1} + P_{\lambda_2})\theta^*$. Note that by Lemma C.2 in Appendix B, we have $\varsigma_i = O(h^0 + |\Delta|^{d+1})$ and $t_i = O_P(\lambda_1|\Delta|^{-1/2}h^{1/2} + \lambda_2|\Delta|^{h/2})$, $i = 1, \ldots, n$. Therefore, $\nabla_{\alpha} \ell_n(\alpha, \sigma^2, \eta)|_{\kappa=\kappa_0}$, $\nabla_{\sigma} \ell_n(\alpha, \sigma^2, \eta)|_{\kappa=\kappa_0}$, and $\nabla_{\eta} \ell_n(\alpha, \sigma^2, \eta)|_{\kappa=\kappa_0}$ are dominated by

$$
-\text{tr}(G) + \sigma_0^{-2} \epsilon^\top H_{x_{\theta},\Lambda}^\top H_{x_{\theta},\Lambda} G\mu_0 + \sigma_0^{-2} \epsilon^\top H_{x_{\theta},\Lambda}^\top H_{x_{\theta},\Lambda} G\epsilon,
$$

$$
-0.5\sigma_0^{-2} n + 0.5\sigma_0^{-4} \epsilon^\top H_{x_{\theta},\Lambda}^\top H_{x_{\theta},\Lambda} \epsilon, \text{ and } -\sigma_0^{-2} \epsilon^\top H_{x_{\theta},\Lambda}^\top H_{x_{\theta},\Lambda} \mathbb{Z},
$$

respectively. By Lemma C.7 in Appendix B and \[B.20\],

$$
\frac{1}{\sqrt{n}} \nabla_{\alpha} \ell_n(\alpha, \sigma^2, \eta)|_{\kappa=\kappa_0} = -\frac{1}{\sqrt{n}} \text{tr}(G) + \frac{1}{\sqrt{n}\sigma_0^2} \epsilon^\top H_{x_{\theta},\Lambda}^\top G\mu_0 + \frac{1}{\sqrt{n}\sigma_0^2} \epsilon^\top G\epsilon + o_P(1),
$$

$$
\frac{1}{\sqrt{n}} \nabla_{\sigma} \ell_n(\alpha, \sigma^2, \eta)|_{\kappa=\kappa_0} = -\frac{\sqrt{n}}{2\sigma_0^2} + \frac{1}{2\sqrt{n}\sigma_0^2} \epsilon^\top \epsilon + o_P(1),
$$

$$
\frac{1}{\sqrt{n}} \nabla_{\eta} \ell_n(\alpha, \sigma^2, \eta)|_{\kappa=\kappa_0} = -\frac{1}{\sqrt{n}\sigma_0^2} \epsilon^\top H_{x_{\theta},\Lambda} \mathbb{Z} + o_P(1).
$$

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 \( n^{-1}(\mu_0 - Z\eta_0)^T H_{\psi_r^*}^{\top} A H_{\psi_r^*} A (\mu_0 - Z\eta_0) = O(|\Delta|^{2d+2} + \eta^2) \). Similar to the above discussion, by Lemma [C.7] in Appendix B and \([B.21]\), we have

\[
\begin{align*}
    n^{-1} \nabla^2 \ell_n(\alpha, \sigma^2, \eta)|_{\kappa = \kappa_0} &= -\text{tr}(G^2) - \frac{1}{\sigma_0^2}(G\mu_0)^T H_{\psi_r^*}^{\top} A H_{\psi_r^*} A G\mu_0 - \frac{1}{\sigma_0^2} \epsilon G^\top G\epsilon + o_P(1), \\
    n^{-1} \nabla^2 \ell_n(\alpha, \sigma^2, \eta)|_{\kappa = \kappa_0} &= \frac{n}{2\sigma_0^2} - \frac{1}{\sigma_0^2} \epsilon G^\top G\epsilon + o_P(1), \\
    n^{-1} \nabla_\eta \nabla_\eta^\top \ell_n(\alpha, \sigma^2, \eta)|_{\kappa = \kappa_0} &= -\frac{1}{\sigma_0^2} Z^\top H_{\psi_r^*}^{\top} A H_{\psi_r^*} A Z + o_P(1), \\
    n^{-1} \nabla_\alpha \nabla_\sigma \ell_n(\alpha, \sigma^2, \eta)|_{\kappa = \kappa_0} &= -\frac{1}{\sigma_0^2} \epsilon G^\top G\epsilon + o_P(1), \\
    n^{-1} \nabla_\sigma \ell_n(\alpha, \sigma^2, \eta)|_{\kappa = \kappa_0} &= -\frac{1}{\sigma_0^2} \epsilon G^\top G\epsilon + o_P(1).
\end{align*}
\]

Then, it is straightforward to show that

\[
\begin{align*}
    E\{n^{-1} \nabla_\kappa \ell_n(\kappa)|_{\kappa = \kappa_0} \nabla_\kappa^\top \ell_n(\kappa)|_{\kappa = \kappa_0} \} &= \Sigma_n + \Omega_n + o(1), \\
    E\{n^{-1} \nabla_\kappa \nabla_\kappa^\top \ell_n(\kappa)|_{\kappa = \kappa_0} \} &= -\Sigma_n + o(1).
\end{align*}
\]

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 \( \bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_{p_2})^\top \) satisfies that \( \|\bar{\beta} - \beta_0\|_{L_2} = O_P \left( |\Delta|^{d+1} + \eta^2 + \lambda_1 n^{-1} |\Delta|^{-3} \eta^{-1/2} + \lambda_2 n^{-1} |\Delta|^{-1} \eta^{-5/2} + n^{-1/2} |\Delta|^{-1} \eta^{-1/2} \right) \).

**Proof.** We first decompose \( \mu_0 \) as follows

\[
\mu_0 = \mu_c + \mu_v = (Z_1^\top \eta_0, \ldots, Z_n^\top \eta_0)^\top + \{X_1^\top \beta_0(S_{11}, S_{12}, T_1), \ldots, X_n^\top \beta_0(S_{n1}, S_{n2}, T_n)\}^\top.
\]

Then, we can write

\[
\bar{\theta} = U_{22} X_{\psi_r^*}^\top (I_n - \Pi_2)(\mu_c + \mu_v + \epsilon) = U_{22} X_{\psi_r^*}^\top (I_n - \Pi_2)\mu_v + U_{22} X_{\psi_r^*}^\top (I_n - \Pi_2)\epsilon = \bar{\theta}_\mu + \bar{\theta}_\epsilon.
\]

According to Lemma 1 in the main paper, for \( \beta_{k0}(s_1, s_2, t) \), \( k = 1, \ldots, p_2 \), there exists \( \theta^*_k \) such that \( \sup_{(s_1, s_2, t) \in \Omega \times T} |\beta_{k0}(s_1, s_2, t) - \psi^*(s_1, s_2, t)^\top \theta^*_k| = O(\eta^2 + |\Delta|^{d+1}) \). Denote that \( \theta^* = (\theta_1^* \top, \ldots, \theta_{p_2}^* \top)^\top \),
\[ \beta_k^* = \psi^*(s_1, s_2, t) \mathbf{1}_k^T, \text{ and } \beta^*(s_1, s_2, t) = \{\beta_1^*(s_1, s_2, t), \ldots, \beta_{p_2}^*(s_1, s_2, t)\}^T. \] 

Denote \( \theta^* = \{X_{\psi}^T (I_n - \Pi_Z)X_{\psi}^*)^{-1}X_{\psi}^T (I_n - \Pi_Z)\mu_{\psi}^* \), where

\[
\mu_{\psi}^* = \{X_1^T \beta^*(S_{11}, S_{12}, T_1), \ldots, X_n^T \beta^*(S_{n1}, S_{n2}, T_n)\}^T.
\]

Then, we have the following decomposition: \( \dot{\theta} - \theta^* = \dot{\theta}_\mu - \theta^* + \dot{\theta}_\nu \). Notice that \( \dot{\theta}_\mu - \theta^* = U_{22}X_{\psi}^T (I_n - \Pi_Z)(\mu_{\nu} - X_{\psi}^T \theta^*) \). Hence, \( \|\dot{\theta}_\mu - \theta^*\| \leq \|U_{22}X_{\psi}^T (I_n - \Pi_Z)(\mu_{\nu} - X_{\psi}^T \theta^*)\| + \|U_{22}P_{\lambda_2} \theta^*\| \) holds. Lemma \( C.1 \) implies that

\[
\|U_{22}X_{\psi}^T (I_n - \Pi_Z)(\mu_{\nu} - X_{\psi}^T \theta^*)\| \leq Cn^{-1}|\triangle|^{-2}h^{-1}\|X_{\psi}^T (I_n - \Pi_Z)(\mu_{\nu} - X_{\psi}^T \theta^*)\|
\leq Cn^{-1}|\triangle|^{-2}h^{-1}\left(\sum_{k=1}^{p_2} \sum_{j \in J} \left(\sum_{i=1}^n |X_{ik}\psi_j(S_{i1}, S_{i2}, T_i)|\right)^2\right)^{1/2}
\leq O_p(|\triangle|^{-1}h^{-1/2}|\triangle|^{-d+1} + \|X_{\psi}\|)
\]

(B.22)

According to Lemmas \( C.1 \) and \( C.2 \) we have

\[
\|U_{22}P_{\lambda} \theta^*\| \leq Cn^{-1}|\triangle|^{-2}h^{-1}\|P_{\lambda} \theta^*\| = Cn^{-1}|\triangle|^{-2}h^{-1}\left\{\sum_{j} (e_j^T P_{\lambda} \theta^*)^2\right\}^{1/2}
= O_p(\lambda_1 n^{-1}|\triangle|^{-4}h^{-1} + \lambda_2 n^{-1}|\triangle|^{-2}h^{-3})
\]

(B.23)

where \( e_j \) is a vector with \( j \)-th element being one and the rest of elements being zero. Now we derive the order of \( \|\dot{\theta}_\nu\| \). Observe that

\[
\|\dot{\theta}_\nu\| = U_{22}X_{\psi}^T (I_n - \Pi_Z)\epsilon \leq n^{-1}|\triangle|^{-2}h^{-1}\|X_{\psi}^T \epsilon\|
\leq n^{-1}|\triangle|^{-2}h^{-1}\left[\sum_{k=1}^{p_2} \sum_{j \in J} \left(\sum_{i=1}^n X_{ik}\psi_j(S_{i1}, S_{i2}, T_i)\epsilon_i\right)^2\right]^{1/2}
= O_p(n^{-1/2}|\triangle|^{-2}h^{-1}).
\]

(B.24)

Combining (B.22) - (B.24) and Lemma 2, we established Theorem 3. \( \square \)

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

\[
\{Y(\alpha_0) - Z\eta - X_{\psi} \theta\}^T \{Y(\alpha_0) - Z\eta - X_{\psi} \theta\} + \theta^T P_{\lambda_1} \theta + \theta^T P_{\lambda_2} \theta,
\]

and \( \theta = (\theta_1^T, \ldots, \theta_{p_2}^T)^T \). Therefore, the oracle estimators of \( \beta_k(s_1, s_2, t) \) is

\[
\tilde{\beta}_k(s_1, s_2, t) = \psi(s_1, s_2, t)^T Q_2 \dot{\theta}_k.
\]
Notice that $Y(\bar{\alpha}_0) = \Xi_0^{-1}(\mu_0 + \epsilon) = \mu_0 + \epsilon$. Then, we have

$$\bar{\eta} = U_{11}Z^T(I_n - X_{\psi}^T\Gamma^{-1}_X X_{\psi}^T)(\mu_0 + \epsilon), \quad \bar{\theta} = U_{22}X_{\psi}^T(I_n - \Pi_Z)(\mu_0 + \epsilon).$$

Notice that $\bar{\theta} = (\bar{\theta}_1^T, \ldots, \bar{\theta}_{p_2}^T)^T = U_{22}X_{\psi}^T(I_n - \Pi_Z)\Xi_0 Y$ and $\hat{\theta} = U_{22}X_{\psi}^T(I_n - \Pi_Z)\Xi(\hat{\alpha}) Y$. For any $k = 1, \ldots, p_2$, let $\tilde{\beta}_k(s_1, s_2, t) = \psi^*(s_1, s_2, t)^T \bar{\theta}_k$, and $\hat{\beta}_k(s_1, s_2, t) = \psi^*(s_1, s_2, t)^T \hat{\theta}_k$. Then, by Theorem 2, Lemma C.1 and Lemma C.6 in Appendix B, we obtain that

$$\|\bar{\theta} - \hat{\theta}\|^2 \asymp n^{-1}\|U_{22}X_{\psi}^T(I_n - \Pi_Z)G\Xi_0 Y\|^2 \asymp n^{-3}|J|^2\|X_{\psi}^T(I_n - \Pi_Z)G(\mu_0 + \epsilon)\|^2 \asymp n^{-3}|J|^2\|X_{\psi}^T(I_n - \Pi_Z)G\mu_0\|^2 \asymp n^{-1}|J| + n^{-2}|J|^2.$$ 

Also, Theorem B.1 in Appendix B implies that

$$\|\hat{\beta} - \beta_0\| \leq |\Delta|\frac{1}{2}\|\bar{\theta} - \hat{\theta}\| + \|\beta_0 - \bar{\beta}\| \leq O_P\left(|\Delta|d + h^6 + \lambda_1 n^{-1}|\Delta|^{-3}h^{-1/2} + \lambda_2 n^{-1}|\Delta|^{-5/2} + n^{-1/2}|\Delta|^{-1}h^{-1/2}\right).$$

Theorem 3 is established. □
Appendix C. Additional Technical Lemmas

**Lemma C.1.** Under Assumptions (A1) – (A8), for $\Gamma_A$ in ([B.2]) and $U_{22}$ 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 $n \to \infty$,

$$c_1 n |\Delta|^2 h \leq \lambda_{\min}(\Gamma_A) \leq \lambda_{\max}(\Gamma_A) \leq C_1 \left( n |\Delta|^2 h + \lambda_1 |\Delta|^{-2} h + \lambda_2 |\Delta|^2 h^{-3} \right),$$

$$c_2 n |\Delta|^2 h \leq \lambda_{\min}(U_{22}^{-1}) \leq \lambda_{\max}(U_{22}^{-1}) \leq C_2 \left( n |\Delta|^2 h + \lambda_1 |\Delta|^{-2} h + \lambda_2 |\Delta|^2 h^{-3} \right).$$

(C.1)

(C.2)

If $\lambda_1 n^{-1} |\Delta|^{-4} \to 0$ and $\lambda_1 n^{-1} |\Delta|^{-4} \to 0$ as $n \to \infty$, then $\lambda_{\max}(\Gamma_A) \leq C_1 n |\Delta|^2 h$ and $\lambda_{\max}(U_{22}^{-1}) \leq C_2 n |\Delta|^2 h$ hold.

**Proof.** It is easy to see that, for any vector $\theta = (\theta_1^T, \ldots, \theta_{p_2}^T)^T$,

$$n^{-1} \theta^T \Gamma_A \theta = \|g_\gamma\|_n^2 + \frac{\lambda_1}{n} \sum_{k=1}^{p_2} \left| \sum_{j, j' \in J} \gamma_k \langle \psi_j, \psi_{j'} \rangle f_1 \right|^2 + \frac{\lambda_2}{n} \sum_{j, j' \in J} \left| \sum_{k=1}^{p_2} \gamma_k \langle \psi_j, \psi_{j'} \rangle f_2 \right|^2,$$

where $\gamma = (\gamma_1, \ldots, \gamma_{p_2})^T = Q_A \theta$ with $\gamma_k = (\gamma_{kj}, j \in J)^T$ and $g_\gamma = (g_{\gamma_1}, \ldots, g_{\gamma_{p_2}})$ with $g_{\gamma_k}$ being a spline function with coefficient $\gamma_k$. Using the Markov’s inequality, we have

$$\sum_{k=1}^{p_2} \left| \sum_{j \in J} \gamma_k \psi_j \right|^2 _{f_1} \leq \frac{C}{|\Delta|^{-4}} \sum_{k=1}^{p_2} \left| \sum_{j \in J} \gamma_k \psi_j \right|^2 _{f_2} \leq C |\Delta|^{-2} h \|\gamma\|_2^2,$$

$$\sum_{k=1}^{p_2} \left| \sum_{j \in J} \gamma_k \psi_j \right|^2 _{f_2} \leq \frac{C}{|\Delta|^{-4}} \sum_{k=1}^{p_2} \left| \sum_{j \in J} \gamma_k \psi_j \right|^2 _{f_2} \leq C |\Delta|^{-2} h^{-3} \|\gamma\|_2^2.$$

Therefore, by Lemma A.4, the largest eigenvalue of the matrix $\Gamma_A$ satisfies that

$$\lambda_{\max}(\Gamma_A) \leq C \left\{ (1 + R_n) |\Delta|^2 h + \lambda_1 n^{-1} |\Delta|^{-2} h + \lambda_2 n^{-1} |\Delta|^2 h^{-3} \right\},$$

Consequently, we have $\lambda_{\max}(\Gamma_A) \leq C_1 \left( |\Delta|^2 h + \lambda_1 n^{-1} |\Delta|^{-2} h + \lambda_2 n^{-1} |\Delta|^2 h^{-3} \right)$ with probability approaching one, for some positive constant $C_1$. Using Lemma A.4 again, it is easy to obtain $\|g_\gamma\|_n^2 = (1 - R_n) \|\gamma\|_2^2 \geq c(1 - R_n) |\Delta|^2 h \|\gamma\|_2^2$. Therefore, the smallest eigenvalue of $\Gamma_A$ is greater than $c(1 - R_n) |\Delta|^2 h = c_1 |\Delta|^2 h$. Let

$$\Upsilon^T = \begin{pmatrix} I & 0 \\ -|\Delta|^{-1} h^{-1/2} X_{\phi}^T Z (Z^T Z)^{-1} & I \end{pmatrix},$$
where \( I \) is the identity matrix and \( 0 \) is a matrix with zeros. Denote \( D^* = (Z, |\Delta|^{-1} h^{-1/2} X_{\psi^*})^T \), then we have
\[
\Upsilon^T D^* D^* \Upsilon = \begin{pmatrix} Z^T Z & 0 \\ 0 & |\Delta|^{-2} h^{-1} X_{\psi^*}^T (I_n - \Pi_Z) X_{\psi^*} \end{pmatrix}.
\]
Lemma C.1 and Assumption (A4) implies that there exist \( c_2 \) and \( C_2 \) such that \( \lambda_{\min}(D^* D^*) \geq c_2 n \) and \( \lambda_{\max}(D^* D^*) \leq C_2 n \). According to the fact that the eigenvalues of \( \Upsilon \) are equal to one and properties of spectral radius of a squared matrix, we have
\[
\lambda_{\max}(\Upsilon^T D^* D^* \Upsilon) = \rho(\Upsilon^T D^* D^* \Upsilon) \leq \rho(\Upsilon^T) \rho(D^* D^*) \rho(\Upsilon) \leq C_2 n,
\]
\[
\lambda_{\min}(\Upsilon^T D^* D^* \Upsilon) = \{\lambda_{\max}(\Upsilon^{-1}(D^* D^*)^{-1}(\Upsilon^T)^{-1})\}^{-1} \geq (\lambda_{\min}(D^* D^*))^{-1} \geq c_2 n,
\]
where \( \rho(\cdot) \) represents spectral radius of a squared matrix. Therefore,
\[
c_2 n |\Delta|^2 h \geq \lambda_{\min}(X_{\psi^*}^T (I_n - \Pi_Z) X_{\psi^*}) \leq \lambda_{\max}(X_{\psi^*}^T (I_n - \Pi_Z) X_{\psi^*}) \leq C_2 n |\Delta|^2 h.
\]
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 \( \theta^*_k \) such that
\[
\sup_{(s_1, s_2, t) \in \Omega \times T} |\beta_{k0}(s_1, s_2, t) - \psi^*(s_1, s_2, t)^T \theta^*_k| = O(h^d + |\Delta|^{d+1}). \tag{C.3}
\]

Let \( \theta^* = (\theta^*_1, \ldots, \theta^*_p)^T \) and \( \beta^*_k(s_1, s_2, t) = \psi^*(s_1, s_2, t)^T \theta^*_k \). In addition, denote \( \beta^*(s_1, s_2, t) = \{\beta^*_1(s_1, s_2, t), \ldots, \beta^*_p(s_1, s_2, t)\}^T \).

**Lemma C.2.** Let \( \theta^* = (\theta^*_1, \ldots, \theta^*_p)^T \), where \( \theta^*_k \)'s are defined in (C.3). Under Assumptions (A1) - (A8), for any vector \( a \) with \( \|a\| = 1 \), we have
\[
\|a^T (P_{\lambda_1} + P_{\lambda_2}) \theta^*\| = O\{\lambda_1(|\Delta|^{-1} h^{1/2} + |\Delta|^d h^{1/2} + |\Delta|^{-1} h^{(d+1)/2}) + \lambda_2(|\Delta| h^{-3/2} + |\Delta| h^{-1/2} + |\Delta| h^{-(d+1)/2} + |\Delta| h^{-(d+1)/2})\}
\]
where \( P_{\lambda_1} \) and \( P_{\lambda_2} \) are given in (B.1).

**Proof.** Let \( a = (a^T, \ldots, a^T) \), then we have
\[
a^T (P_{\lambda_1} + P_{\lambda_2}) \theta^* = \sum_{k=1}^{p_2} \lambda_1 a_k^T Q_1 Q_1^T P_1 Q_2 \theta^*_k + \sum_{k=1}^{p_2} \lambda_2 a_k^T Q_2 Q_2^T P_2 Q_2 \theta^*_k \tag{C.4}
\]
\[
= \sum_{k=1}^{p_2} \lambda_1 \langle \psi_{a_k}, \beta^*_k \rangle_{f_1} + \sum_{k=1}^{p_2} \lambda_2 \langle \psi_{a_k}, \beta^*_k \rangle_{f_2} \leq \sum_{k=1}^{p_2} \lambda_1 \|\psi_{a_k}\|_{f_1} \|\beta^*_k\|_{f_1} + \sum_{k=1}^{p_2} \lambda_1 \|\psi_{a_k}\|_{f_2} \|\beta^*_k\|_{f_2}.
\]
where $\psi_{a_k}(s_1, s_2, t) = \psi^*(s_1, s_2, t)^T a_k$ is a tensor-product spline function. By Markov inequality,
\[ \|\psi_{a_k}\|_f \leq C_1 |\Delta|^{-4} \|\psi_{a_k}\|_2 \leq |\Delta|^{-2} h \|a_k\|^2, \quad \|\psi_{a_k}\|_f \leq C_2 h^{-4} \|\psi_{a_k}\|_2 \leq |\Delta|^2 h^{-3} \|a_k\|^2. \] (C.5)

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

Lemma C.3. Under Assumptions (A1) – (A8), $\Pi_{\mathcal{B}, \Lambda}$ is bounded in both column and row sums, and the elements in $\Pi_{\mathcal{B}, \Lambda}$ are $O_P(n^{-1}|\Delta|^{-2}h^{-1})$ uniformly for all $i, j$.

Proof. Notice that
\[
\Pi_{\mathcal{B}, \Lambda} = \begin{pmatrix} Z \ 1 \end{pmatrix} \begin{pmatrix} U_{11} & U_{12} V_{22}^{-1} \\ -U_{21} V_{21}^{-1} & U_{22} \end{pmatrix} \begin{pmatrix} Z^T \\ X_{\psi^*} \end{pmatrix}
\]
\[
= ZU_{11}Z^T + X_{\psi^*} U_{22}X_{\psi^*}^T - ZU_{11}Z^T X_{\psi^*}(X_{\psi^*}^T X_{\psi^*})^{-1}X_{\psi^*}^T,
\] (C.6)

where $U_{11}$ and $U_{22}$ are defined in (B.4). Note that $n^{-1}U_{11}^{-1}$ is a consistent estimator of $n^{-1}E[Z^T X^T g^*(S_1, S_2, T)] \{Z^T X^T g^*(S_1, S_2, T)\}^T$, which is defined in the Assumption (A5). Assumption (A5) implies that there exists a constant $C$ such that, for any $i = 1, \ldots, n$, $\sum_{j=1}^n |Z_i^T U_{11} Z_j| \leq C$ and $\sum_{j=1}^n |Z_i^T U_{11} Z_j| \leq C$. Thus, $ZU_{11}Z^T$ is bounded both in row and column sums. Next, we prove $X_{\psi^*}(X_{\psi^*}^T X_{\psi^*})^{-1}X_{\psi^*}^T$ is bounded both in row and column sums. Observe that
\[
\sum_{i=1}^n \{X_i \otimes \psi^*(S_1, S_2, T_i)\}^T (X_{\psi^*}^T X_{\psi^*})^{-1} \{X_i \otimes \psi^*(S_1, S_2, T_i)\} 
\]
\[
\leq \sum_{(S_1, S_2, T_i) \in e_{(i)}} n^{-1} |\Delta|^{-2} h^{-1} C \|X_i \otimes \psi^*(S_1, S_2, T_i)\| \|X_i \otimes \psi^*(S_1, S_2, T_i)\| 
\]
\[
+ \sum_{(S_1, S_2, T_i) \notin e_{(i)}} c \sum_{k, k'=1}^{p_2} \sum_{j, j'=1}^{(|J|)} X_{ik} X_{ik'} \psi_j(S_1, S_2, T_i) \psi_{j'}(S_1, S_2, T_i) \leq C,
\]
where $e_{(i)}$ represents the triangular prism contains $(S_1, S_2, T_i)$. Hence, $X_{\psi^*}(X_{\psi^*}^T X_{\psi^*})^{-1}X_{\psi^*}^T$ is bounded both in row and column sums. Similarly, we can prove $X_{\psi^*} U_{22}X_{\psi^*}^T$ is bounded both in row and column sums. By (C.6), we have $\Pi_{\mathcal{B}, \Lambda}$ and $X_{\psi^*} U_{22}X_{\psi^*}^T, \Pi_Z$ are bounded in both column and row sums.

Also, the elements of $ZU_{11}Z^T$ and $\Pi_Z$ are $O_P(n^{-1})$ uniformly for all $i, j$. The elements of $X_{\psi^*} U_{22}X_{\psi^*}^T$ and $X_{\psi^*}(X_{\psi^*}^T X_{\psi^*})^{-1}X_{\psi^*}^T$ are $O_P(n^{-1}|\Delta|^{-2}h^{-1})$ uniformly for all $i, j$. Applying the property in Lee (2004), we have the elements in $\Pi_{\mathcal{B}, \Lambda}$ are $O_P(n^{-1}|\Delta|^{-2}h^{-1})$ uniformly for all $i, j$. Lemma C.3 is established.
Lemma C.4. Under Assumptions (A1) – (A10), we have

\[ n^{-1} \mu_0^\top (I_n - \Pi_{\mathcal{D},\Lambda})^\top (I_n - \Pi_{\mathcal{D},\Lambda}) \mu_0 \]
\[ = O_P \{ \| \Delta \|^{2d+2} + h^{2p} + n^{-1} (\lambda_1^2 |\Delta|^{-6}h^{-1} + \lambda_2^2 |\Delta|^{-2}h^{-5}) \}, \quad (C.7) \]
\[ n^{-1} \mu_0^\top G^\top (I_n - \Pi_{\mathcal{D},\Lambda})^\top (I_n - \Pi_{\mathcal{D},\Lambda}) \mu_0 \]
\[ = O_P \{ \| \Delta \|^{d+1} + h^{p} + n^{-1/2} (\lambda_1 |\Delta|^{-3}h^{-1/2} + \lambda_2 |\Delta|^{-1}h^{-5/2}) \}, \quad (C.8) \]
\[ n^{-1} \epsilon^\top (I_n - \Pi_{\mathcal{D},\Lambda})^\top (I_n - \Pi_{\mathcal{D},\Lambda}) \mu_0 \]
\[ = O_P \{ n^{-1/2} \| \Delta \|^{d+1} + n^{-1/2} h^{p} + n^{-1} (\lambda_1 |\Delta|^{-3}h^{-1/2} + \lambda_2 |\Delta|^{-1}h^{-5/2}) \}, \quad (C.9) \]
\[ n^{-1} \epsilon^\top G (I_n - \Pi_{\mathcal{D},\Lambda})^\top (I_n - \Pi_{\mathcal{D},\Lambda}) \mu_0 \]
\[ = O_P \{ n^{-1/2} \| \Delta \|^{d+1} + n^{-1/2} h^{p} + n^{-1} (\lambda_1 |\Delta|^{-3}h^{-1/2} + \lambda_2 |\Delta|^{-1}h^{-5/2}) \}, \quad (C.10) \]

where \( \Pi_{\mathcal{D},\Lambda} \) is defined in \( (B.3) \).

Proof. Notice that \( \mu_0^\top (I_n - \Pi_{\mathcal{D},\Lambda})^\top (I_n - \Pi_{\mathcal{D},\Lambda}) \mu_0 = \| (I_n - \Pi_{\mathcal{D},\Lambda}) \mu_0 \|^2 \) and

\[ (I_n - \Pi_{\mathcal{D},\Lambda}) \mu_0 = (I_n - \Pi_{\mathcal{D},\Lambda}) \{ Z_i^\top \eta_0 + X_i^\top \beta_0(S_{i1}, S_{i2}, T_i) \}_{i=1}^n \]
\[ = (I_n - \Pi_{\mathcal{D},\Lambda}) \{ Z_i^\top \eta_0 + X_i^\top \beta^*(S_{i1}, S_{i2}, T_i) + X_i^\top \beta_0(S_{i1}, S_{i2}, T_i) - X_i^\top \beta^*(S_{i1}, S_{i2}, T_i) \}_{i=1}^n \]
\[ = (I_n - \Pi_{\mathcal{D},\Lambda}) \{ X_i^\top \beta_0(S_{i1}, S_{i2}, T_i) - X_i^\top \beta^*(S_{i1}, S_{i2}, T_i) \}_{i=1}^n + \mathbb{D} \mathbb{V}^{-1} \mathbb{P}_\Lambda (\eta_0^\top, \theta^*_{\psi})^\top, \]

where \( \theta^* = (\theta_1^\top, \ldots, \theta_{\psi_{T_2}}^\top)^\top \) with \( \theta_{\psi_i}^* \)'s in \( (C.3) \), then we have

\[ \mathbb{D} \mathbb{V}^{-1} \mathbb{P}_\Lambda (\eta_0^\top, \theta^*)^\top = (Z \ X_{\psi}) \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2} \end{pmatrix} \begin{pmatrix} \eta_0 \\ \theta^* \end{pmatrix} \]
\[ = \left( Z \{ Z^\top (I_n - \Pi_{X_{\psi}, \Lambda}) Z \}^{-1} Z^\top X_{\psi} (X_{\psi^*} X_{\psi^*})^{-1} + X_{\psi} (X_{\psi^*} (I_n - \Pi_{\mathcal{Z}}) X_{\psi^*} + \mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2})^{-1} \right) \times (\mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2}) \theta^* = \{ (b_{11} + b_{12})^\top (\mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2}) \theta^* \}_{i=1}^n, \]

where \( \Pi_{X_{\psi}, \Lambda} \) and \( \Pi_{\mathcal{Z}} \) are defined in \( (B.3) \), and

\[ b_{11} = Z_i^\top \{ Z^\top (I_n - \Pi_{X_{\psi}, \Lambda}) Z \}^{-1} Z^\top X_{\psi} (X_{\psi^*} X_{\psi^*})^{-1}, \]
\[ b_{12} = (X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i))^\top \{ X_{\psi^*}^\top (I_n - \Pi_{\mathcal{Z}}) X_{\psi^*} + \mathbb{P}_{\lambda_1} + \mathbb{P}_{\lambda_2} \}^{-1}. \]

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

\[ \| b_{11} \| \leq C n^{-1} |\Delta|^{-2} h^{-1} \| Z_i^\top \{ Z^\top (I_n - \Pi_{X_{\psi}, \Lambda}) Z \}^{-1} Z^\top X_{\psi} \| = O_P (n^{-1} |\Delta|^{-1} h^{-1/2}), \]
\[ \| b_{12} \| \leq C n^{-1} |\Delta|^{-2} h^{-1} \| X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i) \| = O_P (n^{-1} |\Delta|^{-2} h^{-1}). \]
By Lemma C.2 and Assumption (A8), we have
\[
\|n^{-1/2}DV^{-1}F_{\lambda}(\eta_0^T, \theta^T)^T\| \\
\leq n^{-1/2} \left\{ \sum_{i=1}^{n} \|b_1^T(P_{\lambda_1} + P_{\lambda_2})\|_2^2 \right\}^{1/2} + n^{-1/2} \left\{ \sum_{i=1}^{n} \|b_2^T(P_{\lambda_1} + P_{\lambda_2})\theta^*\|_2^2 \right\}^{1/2} \\
= O_P(\lambda n^{-1}|\Delta|^{-3}h^{-1/2} + \lambda_2 n^{-1}|\Delta|^{-1}h^{-5/2}).
\]

According to the SVD decomposition of matrices \( D \) and \( F_{\lambda} \), we can prove that \( \lambda_{\min}(\Pi_{DA}) \geq 0 \). Thus, we have
\[
\| (I_n - \Pi_{DA}) \{ X_1^T \beta_0(S_{t1}, S_{t2}, T_i) - X_i^T \beta^*(S_{t1}, S_{t2}, T_i) \} \| \leq \lambda_{\max}(I_n - \Pi_{DA}) \| \{ X_1^T \beta_0(S_{t1}, S_{t2}, T_i) - X_i^T \beta^*(S_{t1}, S_{t2}, T_i) \} \| \\
= \{1 - \lambda_{\min}(\Pi_{DA})\} \| \{ X_1^T \beta_0(S_{t1}, S_{t2}, T_i) - X_i^T \beta^*(S_{t1}, S_{t2}, T_i) \} \|^n \| \\
= O(n^{1/2}(|\Delta|^{d+1} + \epsilon^6)).
\]

Combining (C.11) and (C.12), we obtain (C.7). By Cauchy Schwarz inequality, we have
\[
\|n^{-1} \mu_0^TG^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})\mu_0\| \leq n^{-1}\|I_n - \Pi_{DA}\|G\mu_0\|\|I_n - \Pi_{DA}\|^T\|I_n - \Pi_{DA}\|\mu_0\| \\
= O_P\{|\Delta|^{d+1} + \epsilon^6 + n^{-1/2}(\lambda_1|\Delta|^{-4}h^{-2} + \lambda_2|\Delta|^{-2}h^{-3})\},
\]
which yields (C.8). By the properties of \( \epsilon_i \), we have \( E\{\epsilon^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})\mu_0\} = 0 \), and
\[
E\{\epsilon^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})\mu_0\}^2 = E\|I_n - \Pi_{DA}\|^T(I_n - \Pi_{DA})\mu_0\|^2 \\
= O(n|\Delta|^{2d+2} + nh^2\epsilon + \lambda_1^2|\Delta|^{-8}h^{-2} + \lambda_2^2|\Delta|^{-4}h^{-6}).
\]
Thus, we obtain (C.9).

Similarly, we have \( E\{\epsilon^T G^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})\mu_0\} = 0 \). According to Lemma C.3, \( I_n - \Pi_{DA} \) is bounded both in row and column sums. Then, \( G(I_n - \Pi_{DA}) \) is bounded both in row and column sums, and \( \text{Var}\{\epsilon^T G^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})\mu_0\} = E\{\|G^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})\mu_0\|^2\} = O(n|\Delta|^{2d+2} + nh^2\epsilon + \lambda_1^2|\Delta|^{-8}h^{-2} + \lambda_2^2|\Delta|^{-4}h^{-6}) \), which yields (C.10).

\( \square \)

**Lemma C.5.** Under Assumptions (A1) – (A10), we have
\[
n^{-1}\epsilon^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})G\mu_0 = O_P(1), \tag{C.13}
n^{-1}\epsilon^T G^T(I_n - \Pi_{DA})^T(I_n - \Pi_{DA})G\mu_0 = O_P(1), \tag{C.14}
\]
where \( \Pi_{DA} \) is defined in (B.3).
Proof. Note that \((I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})G\) is bounded both in row and columns sums. Then
\[
E\{\epsilon^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})G\epsilon\} = 0, \quad \text{Var}\{\epsilon^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})G\epsilon\} = E\| (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})G\epsilon\|^2 = O(n),
\]
which implies \((C.13)\). Similarly, \(E\{(G\epsilon)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})G\epsilon\} = 0\) and \(\text{Var}\{(G\epsilon)^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})G\epsilon\} = E\|G^\top (I_n - \Pi_{D,A})^\top (I_n - \Pi_{D,A})G\epsilon\|^2\) = \(O(n)\). Therefore, \((C.14)\) holds. \(\square\)

Lemma C.6. Under Assumptions (A1) – (A10), if \(C_n\) is uniformly bounded both in row sums and column sums in absolute values, then we have \(\|n^{-1}X^\top \phi C_n\mu_0\|^2 = O_P(\|\triangle\|^2 h)\).

Proof. Note that
\[
X^\top \phi C_n\mu_0 = \{X_1 \otimes \phi^*(S_{11}, S_{12}, T_1), \ldots, X_n \otimes \phi^*(S_{n1}, S_{n2}, T_n)\} \left(\sum_{i=1}^n c_{n,ii} \mu_{0i}, \ldots, \sum_{i=1}^n c_{n,ni} \mu_{0i}\right)^\top
\]
where \(X_i \otimes \phi^*(S_{11}, S_{12}, T_i) = \{X_i \otimes \phi(S_{11}, S_{12}, T_i)^\top\}(I_{p_2} \otimes Q_2)\). Then we have
\[
n^{-2}(X^\top \phi C_n\mu_0)^\top (X^\top \phi C_n\mu_0)
= \left[n^{-1} \sum_{i=1}^n \{X_i \otimes \phi^*(S_{11}, S_{12}, T_i)^\top\} \sum_{i'=1}^n c_{n,ii'} \mu_{0i'}\right] \left[n^{-1} \sum_{s=1}^n X_s \otimes \phi^*(S_{11}, S_{12}, T_i) \sum_{s'=1}^n c_{n,ss'} \mu_{0s'}\right]
= \left[n^{-1} \sum_{i=1}^n \{X_i \otimes \phi(S_{11}, S_{12}, T_i)^\top\} \sum_{i'=1}^n c_{n,ii'} \mu_{0i'}\right] (I_{p_2} \otimes Q_2) (I_{p_2} \otimes Q_2^\top)
\]
\[
\leq C_1 n^{-2} \sum_{i=1}^n \sum_{s=1}^n |\{X_i \otimes \phi(S_{11}, S_{12}, T_i)^\top\} \{X_s \otimes \phi(S_{11}, S_{12}, T_i)^\top\}| \sum_{i'=1}^n \sum_{s'=1}^n |c_{n,ii'}| |c_{n,ss'}| \approx \frac{1}{n} \sum_{i=1}^n \psi(S_{i1}, S_{i2}, T_i)^\top \frac{1}{n} \sum_{i'=1}^n \psi(S_{i'1}, T_{i'}) \approx O_P(\|\triangle\|^2 h).
\]
Lemma \((C.6)\) has been established. \(\square\)

Lemma C.7. Under Assumptions (A1) – (A12), we have
\[
n^{-1/2} \epsilon^\top \Pi_{X,\phi^*} \Lambda H_{X,\phi^*} \Lambda G \mu_0 = o_P(1), \quad (C.15)
n^{-1/2} \epsilon^\top \Pi_{X,\phi^*} \Lambda H_{X,\phi^*} \Lambda Z = o_P(1), \quad (C.16)
n^{-1/2} \epsilon^\top \Pi_{X,\phi^*} \Lambda \epsilon = o_P(1), \quad n^{-1/2} \epsilon^\top \Pi_{X,\phi^*}^\top \Pi_{X,\phi^*} \Lambda \epsilon = o_P(1), \quad (C.17)
n^{-1/2} \epsilon^\top \Pi_{X,\phi^*} \Lambda G \epsilon = o_P(1), \quad n^{-1/2} \epsilon^\top \Pi_{X,\phi^*}^\top \Pi_{X,\phi^*} \Lambda G \epsilon = o_P(1). \quad (C.18)
\]
Proof. First, we prove that (C.15) holds. Note that $E(\epsilon^\top \Pi_{\Psi_{\psi}^* \Lambda} H_{\Psi_{\psi}^* \Lambda} G \mu_0) = 0$, and

$$\text{Var}^{1/2}(\epsilon^\top \Pi_{\Psi_{\psi}^* \Lambda} H_{\Psi_{\psi}^* \Lambda} G \mu_0) = \|\Pi_{\Psi_{\psi}^* \Lambda} H_{\Psi_{\psi}^* \Lambda} G \mu_0\|$$

$$= \|((X_{\psi}^* \Gamma_A^{-1} X_{\psi}^* - X_{\psi}^* \Gamma_A^{-1} X_{\psi}^* X_{\psi}^* \Gamma_A^{-1} X_{\psi}^*) G \mu_0)\|$$

$$= \|X_{\psi}^* \Gamma_A^{-1} \Pi_{\Lambda} \Gamma_A^{-1} X_{\psi}^* G \mu_0\| = \left[\sum_{i=1}^n \{X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i) \Gamma_A^{-1} \Pi_{\Lambda} \Gamma_A^{-1} X_{\psi}^* G \mu_0\}^2\right]^{1/2}$$

$$\leq C(\lambda_1 n^{-2}|\Delta|^{-6} h^{-1} + \lambda_2 n^{-2}|\Delta|^{-2} h^{-5}) \left\{\sum_{i=1}^n \|X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i)\|^2 \|X_{\psi}^* G \mu_0\|^2\right\}^{1/2}$$

$$= O_P(\lambda_1 n^{-1/2}|\Delta|^{-5} h^{-1} + \lambda_2 n^{-1/2}|\Delta|^{-1} h^{-9/2}).$$

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

Note that $E(\epsilon^\top \Pi_{\Psi_{\psi}^* \Lambda} \epsilon) = \text{tr}(\Pi_{\Psi_{\psi}^* \Lambda})$, $E(\epsilon^\top \Pi_{\Psi_{\psi}^* \Lambda}^\top \Pi_{\Psi_{\psi}^* \Lambda} \epsilon) = \text{tr}(\Pi_{\Psi_{\psi}^* \Lambda}^\top \Pi_{\Psi_{\psi}^* \Lambda})$. According to Lemma C.1, we have

$$\text{tr}(\Pi_{\Psi_{\psi}^* \Lambda}) = \sum_{i=1}^n \{X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i)\}^\top \Gamma_A^{-1} \{X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i)\} \leq C|\Delta|^{-2} h^{-1},$$

$$\text{tr}(\Pi_{\Psi_{\psi}^* \Lambda}^\top \Pi_{\Psi_{\psi}^* \Lambda}) = \sum_{i=1}^n \{X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i)\}^\top \Gamma_A^{-1} X_{\psi}^* \Gamma_A^{-1} \{X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i)\}$$

$$\leq C|\Delta|^{-2} h^{-1}.$$ 

Therefore, by Property P4 in Appendix A,

$$E|n^{-1/2} \epsilon^\top \Pi_{\Psi_{\psi}^* \Lambda} \epsilon| = O(n^{-1/2}|\Delta|^{-2} h^{-1}), \quad E|n^{-1/2} \epsilon^\top \Pi_{\Psi_{\psi}^* \Lambda}^\top \Pi_{\Psi_{\psi}^* \Lambda} \epsilon| = O(n^{-1/2}|\Delta|^{-2} h^{-1}).$$

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

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

$$\text{Var}(\epsilon^\top \Pi_{\Psi_{\psi}^* \Lambda} G) = (m_4 - 3\sigma_0^2) \sum_{i=1}^n (\Pi_{\Psi_{\psi}^* \Lambda} G)_{ii}^2 + \sigma_0^4 \text{tr}(\Pi_{\Psi_{\psi}^* \Lambda} G \Pi_{\Psi_{\psi}^* \Lambda} G)$$

$$+ \sigma_0^4 \text{tr}(G^\top \Pi_{\Psi_{\psi}^* \Lambda}^\top \Pi_{\Psi_{\psi}^* \Lambda} G),$$

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

$$(\Pi_{\Psi_{\psi}^* \Lambda})_{ij} = \{X_i \otimes \psi^*(S_{i1}, S_{i2}, T_i)\}^\top \Gamma_A^{-1} \{X_j \otimes \psi^*(S_{j1}, S_{j2}, T_j)\} = O_P(n^{-1}|\Delta|^{-2} h^{-1}).$$
Hence, \( \sum_{i=1}^{n} (\Pi_{X_{\psi}^\ast, \Lambda} G)^2_{ii} \leq C_1 n^{-2} |\Delta|^{-4} h^{-2} \sum_{i=1}^{n} (\sum_{j=1}^{n} (G)_{ij})^2 \leq C_2 n^{-1} |\Delta|^{-4} h^{-2} \). In addition, we have \((\Pi_{X_{\psi}^\ast, \Lambda} G)_{ij} = O_P(n^{-1} |\Delta|^{-2} h^{-1})\), and

\[
\text{tr}(\Pi_{X_{\psi}^\ast, \Lambda} G \Pi_{X_{\psi}^\ast, \Lambda} G) = \sum_{i=1}^{n} \sum_{k=1}^{n} (\Pi_{X_{\psi}^\ast, \Lambda} G)_{ik} (\Pi_{X_{\psi}^\ast, \Lambda} G)_{ki} = O_P(|\Delta|^{-4} h^{-2}),
\]

\[
\text{tr}(G^\top \Pi_{X_{\psi}^\ast, \Lambda}^\top \Pi_{X_{\psi}^\ast, \Lambda} G) = \sum_{i=1}^{n} \sum_{k=1}^{n} (\Pi_{X_{\psi}^\ast, \Lambda} G)_{ik} (\Pi_{X_{\psi}^\ast, \Lambda} G)_{ki} = O_P(|\Delta|^{-4} h^{-2}).
\]

Thus, \( \text{Var}(\epsilon^\top \Pi_{X_{\psi}^\ast, \Lambda} G \epsilon) = O(|\Delta|^{-4} h^{-2}) \). Also, \( E(\epsilon^\top \Pi_{X_{\psi}^\ast, \Lambda} G \epsilon) = \text{tr}(\Pi_{X_{\psi}^\ast, \Lambda} G) = O(|\Delta|^{-2} h^{-1}) \), and Assumption (A8) implies (C.18). Similarly, the second part in (C.18) can be established. \( \square \)
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.

Figure D.1: Boxplots of MISEs of the estimators of $\beta_k$ 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 $\triangle_q$, $q = 1, 2, 3$.

![Figure D.2: Triangulations on the horseshoe domain considered in Simulation Study 2.](image)

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 $\beta^D$ 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.

<table>
<thead>
<tr>
<th>$(n_s,n_T)$</th>
<th>$\sigma_0$</th>
<th>Method</th>
<th>$\sigma_0$</th>
<th>$\sigma^2$</th>
<th>$\eta_{00}$</th>
<th>$\eta_{01}$</th>
<th>$\eta_{02}$</th>
<th>$\beta_{01}$</th>
<th>MSE ($\times 10^{-3}$)</th>
<th>MISE ($\times 10^{-3}$)</th>
<th>Computing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>(100,30)</td>
<td>0.5</td>
<td>TPST($\xi_1$)</td>
<td>0.34</td>
<td>0.06</td>
<td>34.77</td>
<td>0.10</td>
<td>0.08</td>
<td>7.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>0.34</td>
<td>0.06</td>
<td>34.64</td>
<td>0.10</td>
<td>0.08</td>
<td>7.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>0.34</td>
<td>0.06</td>
<td>34.94</td>
<td>0.10</td>
<td>0.08</td>
<td>7.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>0.35</td>
<td>0.07</td>
<td>35.22</td>
<td>0.10</td>
<td>0.08</td>
<td>7.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>0.35</td>
<td>0.07</td>
<td>35.64</td>
<td>0.10</td>
<td>0.08</td>
<td>7.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>0.35</td>
<td>0.07</td>
<td>35.83</td>
<td>0.10</td>
<td>0.08</td>
<td>8.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>TPST($\xi_1$)</td>
<td>1.04</td>
<td>1.08</td>
<td>106.18</td>
<td>0.39</td>
<td>0.33</td>
<td>19.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>1.04</td>
<td>1.07</td>
<td>106.37</td>
<td>0.39</td>
<td>0.33</td>
<td>19.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>1.04</td>
<td>1.07</td>
<td>106.26</td>
<td>0.39</td>
<td>0.33</td>
<td>19.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>1.10</td>
<td>1.14</td>
<td>112.45</td>
<td>0.39</td>
<td>0.33</td>
<td>21.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>1.09</td>
<td>1.17</td>
<td>111.14</td>
<td>0.39</td>
<td>0.33</td>
<td>21.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>1.11</td>
<td>1.20</td>
<td>113.20</td>
<td>0.39</td>
<td>0.33</td>
<td>21.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(100,50)</td>
<td>0.5</td>
<td>TPST($\xi_1$)</td>
<td>0.25</td>
<td>0.03</td>
<td>25.02</td>
<td>0.04</td>
<td>0.04</td>
<td>5.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>0.28</td>
<td>0.03</td>
<td>28.36</td>
<td>0.04</td>
<td>0.04</td>
<td>5.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>0.25</td>
<td>0.03</td>
<td>25.15</td>
<td>0.04</td>
<td>0.04</td>
<td>5.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>0.24</td>
<td>0.04</td>
<td>23.87</td>
<td>0.04</td>
<td>0.04</td>
<td>5.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>0.23</td>
<td>0.04</td>
<td>23.04</td>
<td>0.04</td>
<td>0.04</td>
<td>5.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>0.22</td>
<td>0.04</td>
<td>22.55</td>
<td>0.04</td>
<td>0.04</td>
<td>6.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>TPST($\xi_1$)</td>
<td>0.76</td>
<td>0.55</td>
<td>76.68</td>
<td>0.16</td>
<td>0.16</td>
<td>14.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>0.69</td>
<td>0.55</td>
<td>69.66</td>
<td>0.16</td>
<td>0.16</td>
<td>14.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>0.76</td>
<td>0.55</td>
<td>77.68</td>
<td>0.16</td>
<td>0.16</td>
<td>14.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>0.76</td>
<td>0.57</td>
<td>77.27</td>
<td>0.16</td>
<td>0.16</td>
<td>15.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>0.73</td>
<td>0.58</td>
<td>74.43</td>
<td>0.16</td>
<td>0.16</td>
<td>15.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>0.73</td>
<td>0.59</td>
<td>74.67</td>
<td>0.16</td>
<td>0.16</td>
<td>16.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(200,30)</td>
<td>0.5</td>
<td>TPST($\xi_1$)</td>
<td>0.18</td>
<td>0.03</td>
<td>18.51</td>
<td>0.03</td>
<td>0.05</td>
<td>4.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>0.20</td>
<td>0.03</td>
<td>20.44</td>
<td>0.03</td>
<td>0.05</td>
<td>4.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>0.19</td>
<td>0.03</td>
<td>19.50</td>
<td>0.03</td>
<td>0.05</td>
<td>4.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>0.20</td>
<td>0.03</td>
<td>19.90</td>
<td>0.03</td>
<td>0.05</td>
<td>4.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>0.19</td>
<td>0.03</td>
<td>19.38</td>
<td>0.03</td>
<td>0.04</td>
<td>4.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>0.19</td>
<td>0.03</td>
<td>19.03</td>
<td>0.03</td>
<td>0.04</td>
<td>4.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>TPST($\xi_1$)</td>
<td>0.59</td>
<td>0.51</td>
<td>59.54</td>
<td>0.13</td>
<td>0.18</td>
<td>11.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>0.55</td>
<td>0.51</td>
<td>55.33</td>
<td>0.13</td>
<td>0.18</td>
<td>11.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>0.56</td>
<td>0.51</td>
<td>56.84</td>
<td>0.13</td>
<td>0.18</td>
<td>11.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>0.55</td>
<td>0.53</td>
<td>55.83</td>
<td>0.13</td>
<td>0.18</td>
<td>12.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>0.54</td>
<td>0.54</td>
<td>54.73</td>
<td>0.13</td>
<td>0.18</td>
<td>13.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>0.53</td>
<td>0.55</td>
<td>53.80</td>
<td>0.13</td>
<td>0.18</td>
<td>13.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(200,50)</td>
<td>0.5</td>
<td>TPST($\xi_1$)</td>
<td>0.08</td>
<td>0.02</td>
<td>8.44</td>
<td>0.03</td>
<td>0.02</td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>0.10</td>
<td>0.02</td>
<td>10.17</td>
<td>0.03</td>
<td>0.02</td>
<td>3.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>0.10</td>
<td>0.02</td>
<td>9.62</td>
<td>0.03</td>
<td>0.02</td>
<td>3.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>0.09</td>
<td>0.02</td>
<td>9.33</td>
<td>0.03</td>
<td>0.02</td>
<td>3.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>0.09</td>
<td>0.02</td>
<td>9.16</td>
<td>0.03</td>
<td>0.02</td>
<td>3.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>0.09</td>
<td>0.02</td>
<td>9.27</td>
<td>0.03</td>
<td>0.02</td>
<td>3.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>TPST($\xi_1$)</td>
<td>0.29</td>
<td>0.29</td>
<td>29.5</td>
<td>0.13</td>
<td>0.09</td>
<td>8.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_2$)</td>
<td>0.30</td>
<td>0.28</td>
<td>29.99</td>
<td>0.13</td>
<td>0.09</td>
<td>8.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_3$)</td>
<td>0.29</td>
<td>0.28</td>
<td>28.74</td>
<td>0.13</td>
<td>0.09</td>
<td>8.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_4$)</td>
<td>0.31</td>
<td>0.29</td>
<td>30.68</td>
<td>0.13</td>
<td>0.09</td>
<td>8.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_5$)</td>
<td>0.30</td>
<td>0.30</td>
<td>29.88</td>
<td>0.13</td>
<td>0.09</td>
<td>9.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPST($\xi_6$)</td>
<td>0.29</td>
<td>0.30</td>
<td>29.33</td>
<td>0.13</td>
<td>0.09</td>
<td>9.43</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table D.2: Sources of datasets

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>COVID-19 Related Time-series</td>
<td></td>
</tr>
<tr>
<td>Infections Data</td>
<td>NYT (2020); Atlantic (2020); CSSE (2020); USAFacts (2020)</td>
</tr>
<tr>
<td>Fatality Data</td>
<td>NYT (2020); Atlantic (2020); CSSE (2020); USAFacts (2020)</td>
</tr>
<tr>
<td>Recovery Data</td>
<td>Atlantic (2020)</td>
</tr>
<tr>
<td>Mobility Data</td>
<td></td>
</tr>
<tr>
<td>Bureau of Transportation Statistics</td>
<td>BTS (2020)</td>
</tr>
<tr>
<td>Descartes Labs</td>
<td>Warren and Skillman (2020)</td>
</tr>
<tr>
<td>American Community Survey (ACS) Data</td>
<td></td>
</tr>
<tr>
<td>2012 Economic Census</td>
<td>USCB (2012)</td>
</tr>
<tr>
<td>Homeland Infrastructure Foundation-level Data</td>
<td>DHS (2020)</td>
</tr>
</tbody>
</table>

Bibliography


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

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


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.

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