Supplementary materials of the article 'Spatial Autoregressive Partially Linear Varying Coefficient Models'

In this Web Supplement, we include in Sections S.1 – S.4 some preliminary lemmas and detailed proofs of the theoretical results in the main article. Section S.6 presents more results from simulation studies in the main article.

S.1. *Preliminary Lemmas*

LEMMA S.1 (Theorem 2.7, Lai and Schumaker (2007)) *Let $\{B_j\}_{j\in\mathcal{J}}$ be the Bernstein polynomial basis for spline space $\mathcal{S}$ with smoothness r and degree d, where $\mathcal{J}$ is an index set. Then there exist positive constants C_1, C_2 depending on r, d and the shape parameter π such that*

$$C_1|\triangle|^2\sum_{j\in\mathcal{J}}c_j^2 \le \left\|\sum_{j\in\mathcal{J}}c_jB_j\right\|_{L_2}^2 \le C_2|\triangle|^2\sum_{j\in\mathcal{J}}c_j^2.$$

LEMMA S.2 (Theorem 1.3, Lai and Schumaker (2007)) *Assume $g(\cdot)\in W^{\ell+1,\infty}(\Omega)$. For bi-integer (α_1,α_2) with $0\le\alpha_1+\alpha_2\le\nu$, there exist an absolute constant C depending on r and π and unique spline fit $g^*(\cdot)\in\mathcal{S}$ such that $\|D_{x_1}^{\alpha_1}D_{x_2}^{\alpha_2}(g-g^*)\|_\infty\le C|\triangle|^{\ell+1-\alpha_1-\alpha_2}|g|_{\ell+1,\infty}$.*

LEMMA S.3 *If two sequences of $n\times n$ matrices $\{\mathbf{F}_n\}$ and $\{\mathbf{C}_n\}$ are uniformly bounded both in row sums and column sums, then $\{\mathbf{F}_n\mathbf{C}_n\}$ are uniformly bounded both in row sums and column sums.*

Proof. Let $\mathbf{F}_n=(f_{n,ij})$, $\mathbf{C}_n=(c_{n,ij})$, and $\mathbf{P}_n=\mathbf{F}_n\mathbf{C}_n=(p_{n,ij})$. Note that for any $i=1,\ldots,n$,

$$\sum_{j=1}^n|p_{n,ij}|=\sum_{j=1}^n|\sum_{k=1}^n f_{n,ik}c_{n,kj}|\le\sum_{k=1}^n|f_{n,ik}|\sum_{j=1}^n|c_{n,kj}|=O(1).$$

So $\mathbf{P}_n$ are bounded in row sums. Similarly we can prove the result still holds for column sums. ■

In the following, denote

$$\begin{aligned}
\mathbf{V}&=\begin{pmatrix}\mathbf{V}_{11}&\mathbf{V}_{12}\\ \mathbf{V}_{21}&\mathbf{V}_{22}\end{pmatrix}=\begin{pmatrix}\mathbf{Z}^\top\mathbf{Z}&\mathbf{Z}^\top\boldsymbol{\mathcal{X}}_B\\ \boldsymbol{\mathcal{X}}_B^\top\mathbf{Z}&\boldsymbol{\mathcal{X}}_B^\top\boldsymbol{\mathcal{X}}_B\end{pmatrix}\\
&=\begin{pmatrix}\sum_{i=1}^n\mathbf{Z}_i\mathbf{Z}_i^\top&\sum_{i=1}^n\mathbf{Z}_i\{\mathbf{X}_i\otimes\mathbf{B}^*(\mathbf{U}_i)\}^\top\\ \sum_{i=1}^n\{\mathbf{X}_i\otimes\mathbf{B}^*(\mathbf{U}_i)\}\mathbf{Z}_i^\top&\sum_{i=1}^n\{\mathbf{X}_i\otimes\mathbf{B}^*(\mathbf{U}_i)\}\{\mathbf{X}_i\otimes\mathbf{B}^*(\mathbf{U}_i)\}^\top\end{pmatrix},\\
\mathbf{V}^{-1}\equiv\mathbf{U}&=\begin{pmatrix}\mathbf{U}_{11}&-\mathbf{U}_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\\ -\mathbf{U}_{22}\mathbf{V}_{21}\mathbf{V}_{11}^{-1}&\mathbf{U}_{22}\end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{U}_{11}^{-1}&=\mathbf{V}_{11}-\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}=\mathbf{Z}^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{Z},\\
\mathbf{U}_{22}^{-1}&=\mathbf{V}_{22}-\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}=\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n-\mathbf{P}_{\mathbf{Z}})\boldsymbol{\mathcal{X}}_B.
\end{aligned}\tag{S.1}$$

Then, the minimizer of (7) can be obtained in the following forms:

$$\widehat{\gamma}(\alpha) = \mathbf{U}_{11}\mathbf{Z}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{T}(\alpha)\mathbf{Y},\ \widehat{\boldsymbol{\theta}}(\alpha) = \mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\mathbf{T}(\alpha)\mathbf{Y}.$$

LEMMA S.4 *Under Assumptions (A1) – (A6), if $K_n \log n/n \to 0$, the eigenvalues of both $n\mathbf{U}_{11}$ and $nK_n^{-1}\mathbf{U}_{22}$ are bounded below and above except on an event with probability going to zero.*

Proof. Note that all the eigenvalues of $\mathbf{P}_{\boldsymbol{\mathcal{X}}_B}$ and $\mathbf{P}_{\mathbf{Z}}$ are either 1 or 0, thus, by (S.1), we have

$$\begin{aligned}(n\mathbf{U}_{11})^{-1} &= n^{-1}\mathbf{Z}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{Z} \asymp n^{-1}\mathbf{Z}^\top\mathbf{Z},\\ (nK_n^{-1}\mathbf{U}_{22})^{-1} &= n^{-1}K_n\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\mathcal{X}}_B \asymp n^{-1}K_n\boldsymbol{\mathcal{X}}_B^\top\boldsymbol{\mathcal{X}}_B.\end{aligned}$$

According to Assumption (A2), the eigenvalues of $n^{-1}\mathbf{Z}^\top\mathbf{Z}$ is bounded below and above except on an event with probability going to zero. By Lemma C.5 in Mu, Wang, and Wang (2018), the the eigenvalues of $n^{-1}K_n\boldsymbol{\mathcal{X}}_B^\top\boldsymbol{\mathcal{X}}_B$ is bounded below and above except on an event with probability going to zero. Therefore, the eigenvalues of $n\mathbf{U}_{11}$ and $nK_n^{-1}\mathbf{U}_{22}$ are bounded below and above except on an event with probability going to zero. ■

In the following, denote

$$\boldsymbol{\Phi}^* = (\mathbf{Z}, \boldsymbol{\mathcal{X}}_B)\mathbf{K} = (\mathbf{Z}, \sqrt{K_n}\boldsymbol{\mathcal{X}}_B), \tag{S.2}$$

where $\mathbf{K} = \mathrm{diag}(\mathbf{I}_{p_1}, \sqrt{K_n}\mathbf{I}_m)$ is a block diagonal matrix with $m = p_2|\mathcal{J}|$. Let $\mathbf{V}_n^* = \boldsymbol{\Phi}^{*\top}\boldsymbol{\Phi}^*$.

LEMMA S.5 *Under (A1) – (A6), except on an event with probability approaching to zero, there exist positive constants C_3 and C_4 such that all the eigenvalues of $n^{-1}\mathbf{V}^*$ fall between C_3 and C_4, and $n^{-1}\mathbf{V}^*$ is nonsingular.*

Proof. Since $\mathbf{V}_n^*$ is symmetric, and $n^{-1}\mathbf{Z}^\top\mathbf{Z}$ is non-singular, then $n^{-1}\mathbf{V}^*$ is congruent with matrix

$$\overline{\mathbf{V}}_n^* = \begin{pmatrix} n^{-1}\mathbf{Z}^\top\mathbf{Z} & \mathbf{0} \\ \mathbf{0} & n^{-1}K_n\mathbf{U}_{22}^{-1}\end{pmatrix},$$

since $\mathbf{C}\mathbf{V}_n^*\mathbf{C}^\top = \overline{\mathbf{V}}_n^*$, where

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} & -\sqrt{K_n}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\boldsymbol{\mathcal{X}}_B \\ \mathbf{0} & \mathbf{I}\end{pmatrix}.$$

Then we can apply Lemma S.4 and obtain that all the eigenvalues of $\overline{\mathbf{V}}_n^*$ are bounded by two positive constants with probability approaching 1. By the properties of congruent matrices, all the eigenvalues of $\mathbf{V}_n^*$ are bounded away from zero and infinity except on an event whose probability goes to zero. The desired result follows. ■

LEMMA S.6 *Under the Assumptions of Lemma S.5, except on an event whose probability tends to zero, $\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}}$ are uniformly bounded in both row sums and column sums.*

Proof. Note that under Assumption (A2), $n^{-1}\mathbf{Z}^\top\mathbf{Z}$ exists and nonsingular. Then by Lemma A.5 from Lee (2004), $\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}}$ satisfies the results. The desired result follows. ■

Next, let

$$\widetilde{\boldsymbol{\theta}} = \mathbf{V}_{22}^{-1} \sum_{i=1}^{n} [\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)]\, \mathbf{X}_i^\top \boldsymbol{\beta}_0(\mathbf{U}_i), \quad \widetilde{\beta}_k(\boldsymbol{u}) = \mathbf{B}^*(\boldsymbol{u})^\top \widetilde{\boldsymbol{\theta}}_k,\ k = 1, 2, \ldots, p. \tag{S.3}$$

LEMMA S.7 *For $\widetilde{\boldsymbol{\beta}} = (\widetilde{\beta}_1, \ldots, \widetilde{\beta}_{p_2})^\top$ defined in (S.3), under the same Assumptions of Lemma S.5, we have $\|\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}\|_{L_2}^2 = O_P(K_n^{-(\ell+1)})$.*

Proof. For any $k = 1, \ldots, p_2$, there exists β_k^* such that $\|\beta_k^* - \beta_{0k}\|_\infty = O(|\triangle|^{\ell+1}|\beta_{0k}|_{\ell+1,\infty})$. Define $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_p^*)^\top$, where $\beta_k^* \in \mathcal{S}_d^r(\triangle)$. Let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top}, \cdots, \boldsymbol{\theta}_p^{*\top})^\top$ with $\boldsymbol{\theta}_k^* = (\theta_{k1}^*, \cdots, \theta_{kJ}^*)^\top$ be such that $\beta_k^*(\boldsymbol{u}) = \mathbf{B}^*(\boldsymbol{u})^\top \boldsymbol{\theta}_k^*$. Note that

$$\boldsymbol{\theta}^* = \mathbf{V}_{n,22}^{-1} \frac{1}{n} \sum_{i=1}^{n} \{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\}\, \mathbf{X}_i^\top \boldsymbol{\beta}^*(\mathbf{U}_i). \tag{S.4}$$

By Lemma C.1 and Lemma C.8 from Mu et al. (2018), $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = O_P(\sqrt{K_n}|\triangle|^{\ell+1})$, thus, $\|\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}\| \asymp |\triangle| \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = O_P(|\triangle|^{\ell+1})$, i.e. $\|\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}\|_{L_2}^2 = O_P(K_n^{-(\ell+1)})$. ■

We denote

$$\boldsymbol{\mu}_0 = \boldsymbol{\mu}_{\mathrm{C}} + \boldsymbol{\mu}_{\mathrm{V}} = (\mathbf{Z}_1^\top \boldsymbol{\gamma}_0, \ldots, \mathbf{Z}_n^\top \boldsymbol{\gamma}_0)^\top + (\mathbf{X}_1^\top \boldsymbol{\beta}_0(\mathbf{U}_1), \ldots, \mathbf{X}_n^\top \boldsymbol{\beta}_0(\mathbf{U}_n))^\top, \tag{S.5}$$

and let

$$\widetilde{\boldsymbol{\gamma}}_\mu = \mathbf{U}_{11}\mathbf{Z}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\mu}_0, \tag{S.6}$$

$$\widetilde{\boldsymbol{\gamma}}_\epsilon = \mathbf{U}_{11}\mathbf{Z}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\epsilon}. \tag{S.7}$$

LEMMA S.8 *Under Assumptions (A1) – (A6), we have $\|\widetilde{\boldsymbol{\gamma}}_\mu - \boldsymbol{\gamma}_0\| = O_P(|\triangle|^{2(\ell+1)})$.*

Proof. Denote $\mathbf{Z} = (\overline{\mathbf{Z}}_1, \ldots, \overline{\mathbf{Z}}_{p_1})$. We can write $\widetilde{\boldsymbol{\gamma}}_\mu - \boldsymbol{\gamma}_0$ as the following:

$$\widetilde{\boldsymbol{\gamma}}_\mu - \boldsymbol{\gamma}_0 = \mathbf{U}_{11}\mathbf{Z}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\mu}_{\mathrm{V}} = n\mathbf{U}_{11}\mathbf{M},$$

where $\mathbf{M} = (M_1, \ldots, M_{p_1})^\top$, and for $l = 1, \ldots, p_1$, $M_l = n^{-1}\overline{\mathbf{Z}}_l^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\mu}_{\mathrm{V}}$. Using Lemma S.7, we have

$$(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\mu}_{\mathrm{V}} = \begin{pmatrix} \mathbf{X}_1^\top\{\boldsymbol{\beta}_0(\mathbf{U}_1) - \widetilde{\boldsymbol{\beta}}(\mathbf{U}_1)\} \\ \vdots \\ \mathbf{X}_n^\top\{\boldsymbol{\beta}_0(\mathbf{U}_n) - \widetilde{\boldsymbol{\beta}}(\mathbf{U}_n)\} \end{pmatrix}.$$

Let a_l be the coordinate mapping function that maps $\mathbf{Z}_i$ to its l-th component, that is, $a_l(\mathbf{Z}_i) = Z_{il}$. Let $\mathbf{g}_l^*(\boldsymbol{u}) = (g_{l,1}^*(\boldsymbol{u}), \ldots, g_{l,p_2}^*(\boldsymbol{u}))^\top$ where $g_{l,k}^* \in L_2(\Omega)$ is the function that minimizes

$$\mathrm{E}\left[\{Z_{il} - \mathbf{X}_i^\top \mathbf{g}_l(\mathbf{U}_i)\}\right]^2 = \|Z_{il} - \mathbf{X}_i^\top \mathbf{g}_l(\mathbf{U}_i)\|_{L_2}^2,$$

and also note that $\mathbf{g}_l^* = \arg\min \|a_l - \mathbf{X}^\top \mathbf{g}\|_{L_2}^2$. That is, $\mathbf{X}^\top \mathbf{g}_l^*$ is the orthogonal projection of a_l onto $\{\mathbf{X}^\top \mathbf{g} : g_k \in L_2\}$. Let $Z_{il}^* = \mathbf{X}_i^\top \mathbf{g}_l^*(\mathbf{U}_i)$ and $\overline{\mathbf{Z}}_l^* = (Z_{1l}^*, \ldots, Z_{nl}^*)^\top$, then

$$\begin{aligned} M_l =& \frac{1}{n}(\overline{\mathbf{Z}}_l^\top - \overline{\mathbf{Z}}_l^{*\top})(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\mu}_{\mathrm{V}} + \frac{1}{n}\overline{\mathbf{Z}}_l^{*\top}(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\mu}_{\mathrm{V}} \\ =& \frac{1}{n}\sum_{i=1}^n (Z_{il} - A_{il}^*)\sum_{k=1}^{p_2} X_{ik}(\beta_{0k} - \widetilde{\beta}_k)(\mathbf{U}_i) \\ &+ \frac{1}{n}\sum_{i=1}^n \left\{\sum_{k=1}^{p_2} X_{ik}(g_{l,k}^* - \widetilde{g}_{l,k})(\mathbf{U}_i)\sum_{k=1}^{p_2} X_{ik}(\beta_{0k} - \widetilde{\beta}_k)(\mathbf{U}_i)\right\} \\ =& \mathcal{I}_1 + \mathcal{I}_2. \end{aligned}$$

Note that $\mathrm{E}\mathcal{I}_1 = 0$, and by Slutsky's theorem and Lemma S.7,

$$\begin{aligned} \mathrm{Var}(\mathcal{I}_1) &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{E}\left\{(Z_{il} - Z_{il}^*)\sum_{k=1}^{p_2} X_{ik}(\beta_{0k} - \widetilde{\beta}_k)(\mathbf{U}_i)\right\}^2 \\ &\asymp O_P(n^{-1}|\triangle|^{2(\ell+1)})\left\{\mathrm{E}(Z_{il} - Z_{il}^*)^2\right\}^{1/2}, \end{aligned}$$

which implies that $\mathcal{I}_1 = O_P(n^{-1/2}|\triangle|^{\ell+1})$. By using Cauchy-Schwartz inequality,

$$\begin{aligned} |\mathcal{I}_2|^2 \le& \left[n^{-1}\sum_{i=1}^n \left\{\sum_{k=1}^{p_2} X_{ik}(g_{l,k}^* - \widetilde{g}_{l,k})(\mathbf{U}_i)\right\}^2\right]\left[n^{-1}\sum_{i=1}^n \left\{\sum_{k=1}^{p_2} X_{ik}(\beta_{0k} - \widetilde{\beta}_k)(\mathbf{U}_i)\right\}^2\right] \\ =& O_P(|\triangle|^{4(\ell+1)}) = o_P(n^{-1}). \end{aligned}$$

Follow the same discussion for any $l = 1, \ldots, p_1$, the desired result follows. ■

LEMMA S.9 *Let $\breve{\boldsymbol{\theta}} = \mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\mu}_0$. Under Assumptions (A1) – (A6), $\|\breve{\beta}_k - \beta_{0k}\|_{L_2} = O_P(|\triangle|^{\ell+1})$, where $\breve{\beta}_k(\boldsymbol{u}) = \mathbf{B}^*(\boldsymbol{u})^\top\breve{\boldsymbol{\theta}}_k$, $k = 1, \ldots, p_2$.*

Proof. By Lemma S.2, for any $k = 1, \ldots, p_2$, there exists $\beta_k^* \in \mathcal{S}$ such that $\|\beta_{0k} - \beta_k^*\|_\infty = O_P(|\triangle|^{\ell+1}|\beta_{0k}|_{\ell+1,\infty})$, and $\boldsymbol{\theta}^*$ in (S.4) can be rewritten as

$$\boldsymbol{\theta}^* = \mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\mu}_{\mathrm{V}}^*, \tag{S.8}$$

where $\boldsymbol{\mu}_{\mathrm{V}}^*$ is defined in (S.5). Then, we have

$$\|\breve{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = \mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})(\boldsymbol{\mu}_{\mathrm{V}} - \boldsymbol{\mu}_{\mathrm{V}}^*). \tag{S.9}$$

By Lemma S.4,

$$\|\mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})(\boldsymbol{\mu}_{\mathrm{V}} - \boldsymbol{\mu}_{\mathrm{V}}^*)\|^2 \asymp \frac{K_n^2}{n^2}\|\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})(\boldsymbol{\mu}_{\mathrm{V}} - \boldsymbol{\mu}_{\mathrm{V}}^*)\|^2.$$

By Lemma S.6, $\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}}$ are uniformly bounded in both rows and columns. Note that the row sums of $\mathbf{Q}_2\mathbf{Q}_2^\top$ are all equal to 1. Let $(\mathbf{c}_1,\dots,\mathbf{c}_n) = \mathbf{I}_n - \mathbf{P}_\mathbf{Z}$, then we have

$$
\begin{aligned}
&\frac{K_n^2}{n^2}\|\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_\mathbf{Z})(\boldsymbol{\mu}_v - \boldsymbol{\mu}_v^*)\|^2 \\
=&\frac{K_n^2}{n^2}\left\{\sum_{s=1}^n\sum_{i=1}^n c_{is}(\beta(\mathbf{U}_s) - \beta^*(\mathbf{U}_s))\mathbf{B}(\mathbf{U}_i)^\top\right\}\mathbf{Q}_2\mathbf{Q}_2^\top\left\{\sum_{s=1}^n\sum_{i=1}^n c_{is}(\beta(\mathbf{U}_s) - \beta^*(\mathbf{U}_s))\mathbf{B}(\mathbf{U}_i)\right\} \\
\leq&\frac{K_n^2}{n^2}c|\triangle|^{2(\ell+1)}\sum_{s=1}^n\sum_{s'=1}^n\sum_{i=1}^n\sum_{i'=1}^n|c_{is}c_{i's'}\mathbf{B}(\mathbf{U}_i)^\top\mathbf{B}(\mathbf{U}_{i'})| \\
\leq&C_2\frac{K_n^2}{n^2}|\triangle|^{2(\ell+1)}\sum_{i=1}^n\sum_{i'=1}^n\mathbf{B}(\mathbf{U}_i)^\top\mathbf{B}(\mathbf{U}_{i'}) = C_2K_n^2|\triangle|^{2(\ell+1)}\left\|\frac{1}{n}\sum_{i=1}^n\mathbf{B}(\mathbf{U}_i)\right\|^2 = O_P(|\triangle|^{2\ell}).
\end{aligned}
$$

Plugging the above into (S.9), we obtain $\|\breve{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \asymp |\triangle|\|\breve{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = O_P(|\triangle|^{\ell+1})$. ■

LEMMA S.10 *Under Assumptions (A1) – (A10),*

$$
\begin{aligned}
n^{-1/2}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_\mathbf{\Phi})\boldsymbol{\mu}_0 &= O_P(n^{1/2}|\triangle|^{2(\ell+1)} + n^{1/2}|\triangle|^{4(\ell+1)}) = o_P(1), && \text{(S.10)}\\
n^{-1/2}\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{P}_\mathbf{\Phi})\boldsymbol{\mu}_0 &= O_P(|\triangle|^{(\ell+1)} + |\triangle|^{2(\ell+1)}) = o_P(1), && \text{(S.11)}\\
n^{-1/2}(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n - \mathbf{P}_\mathbf{\Phi})\boldsymbol{\mu}_0 &= O_P(|\triangle|^{(\ell+1)} + |\triangle|^{2(\ell+1)}) = o_P(1). && \text{(S.12)}
\end{aligned}
$$

Proof. We first prove (S.10). Note that

$$
(\mathbf{I}_n - \mathbf{P}_\mathbf{\Phi})\boldsymbol{\mu}_0 = \mathbf{Z}(\gamma_0 - \widetilde{\gamma}_\mu) + \begin{pmatrix} \mathbf{X}_1^\top\{\boldsymbol{\beta}_0(\mathbf{U}_1) - \widetilde{\boldsymbol{\beta}}(\mathbf{U}_1)\} \\ \vdots \\ \mathbf{X}_n^\top\{\boldsymbol{\beta}_0(\mathbf{U}_n) - \widetilde{\boldsymbol{\beta}}(\mathbf{U}_n)\}\end{pmatrix},
$$

where $\widetilde{\gamma}_\mu$ and $\widetilde{\boldsymbol{\beta}}$ are defined in (S.6) and (S.3). Therefore,

$$
\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_\mathbf{\Phi})\boldsymbol{\mu}_0 = \sum_{i=1}^n\begin{pmatrix}\gamma_0 - \widetilde{\gamma}_\mu \\ (\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\end{pmatrix}^\top\begin{pmatrix}\mathbf{Z}_i \\ \mathbf{X}_i\end{pmatrix}\begin{pmatrix}\mathbf{Z}_i \\ \mathbf{X}_i\end{pmatrix}^\top\begin{pmatrix}\gamma_0 - \widetilde{\gamma}_\mu \\ (\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\end{pmatrix}.
$$

Thus,

$$
E\left\{\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_\mathbf{\Phi})\boldsymbol{\mu}_0\right\} = \sum_{i=1}^n E\left[\begin{pmatrix}\gamma_0 - \widetilde{\gamma}_\mu \\ (\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\end{pmatrix}^\top\boldsymbol{\Sigma}(\mathbf{U}_i)\begin{pmatrix}\gamma_0 - \widetilde{\gamma}_\mu \\ (\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\end{pmatrix}\right],
$$

where, by Assumption (A2), $\boldsymbol{\Sigma}(\boldsymbol{u}) = \mathrm{E}\{(\mathbf{Z}_i^\top, \mathbf{X}_i^\top)^\top(\mathbf{Z}_i^\top, \mathbf{X}_i^\top)|\mathbf{U}_i = \boldsymbol{u}\}$, and the eigenvalues of $\boldsymbol{\Sigma}(\boldsymbol{u})$ are bounded away from 0 and infinity. Therefore, by using Lemmas S.8 and S.9,

$$
\begin{aligned}
E\left\{\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_\mathbf{\Phi})\boldsymbol{\mu}_0\right\} &\asymp \sum_{i=1}^n E\left[\begin{pmatrix}\gamma_0 - \widetilde{\gamma}_\mu \\ (\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\end{pmatrix}^\top\begin{pmatrix}\gamma_0 - \widetilde{\gamma}_\mu \\ (\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\end{pmatrix}\right] \\
&\asymp n\|\gamma_0 - \widetilde{\gamma}_\mu\|^2 + n\|\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}\|^2 = O(n|\triangle|^{2(\ell+1)} + n|\triangle|^{4(\ell+1)}).
\end{aligned}
$$

Then (S.10) follows.

Next we prove (S.11). Note that $\boldsymbol{\epsilon}^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0=\sum_{i=1}^n\epsilon_i\{\mathbf{Z}_i^\top(\gamma_0-\widetilde{\gamma}_\mu)+\mathbf{X}_i^\top(\boldsymbol{\beta}_0-\widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\}$. Thus, $\mathrm{E}[\boldsymbol{\epsilon}^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0]=0$, and

$$\begin{aligned}n^{-1}\mathrm{E}[\boldsymbol{\epsilon}^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0]^2&=n^{-1}\mathrm{E}\left[\sum_{i=1}^n\epsilon_i^2\{\mathbf{Z}_i^\top(\gamma_0-\widetilde{\gamma}_\mu)+\mathbf{X}_i^\top(\boldsymbol{\beta}_0-\widetilde{\boldsymbol{\beta}})(\mathbf{U}_i)\}^2\right]\\&=n^{-1}\sigma_0^2\sum_{i=1}^n\mathrm{E}\left\{\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0\right\}\asymp O(|\triangle|^{2(\ell+1)}+|\triangle|^{4(\ell+1)}).\end{aligned}$$

Then (S.11) follows. Finally, we can show (S.12) in a similar way. Note that

$$\boldsymbol{\epsilon}^\top\mathbf{G}^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0=\sum_{i=1}^n\epsilon_i\mathbf{g}_i^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0,$$

where $\mathbf{g}_i$ is the ith column of $\mathbf{G}$. It is easy to see that $\mathrm{E}[\boldsymbol{\epsilon}^\top\mathbf{G}(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0]=0$. Let g_{ij} denote the (i,j)th entry of $\mathbf{G}$. Under Assumption (A2), we have

$$\begin{aligned}&n^{-1}\mathrm{E}\{\boldsymbol{\epsilon}^\top\mathbf{G}(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0\}^2=n^{-1}\mathrm{E}\sum_{i=1}^n\epsilon_i^2\{\mathbf{g}_i^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0\}^2\\&=n^{-1}\sigma_0^2\sum_{i=1}^n\sum_{s=1}^n\sum_{s'=1}^n g_{si}g_{s'i}\mathrm{E}\left\{\left(\begin{array}{c}\gamma_0-\widetilde{\gamma}_\mu\\(\boldsymbol{\beta}_0-\widetilde{\boldsymbol{\beta}})(\mathbf{U}_s)\end{array}\right)^\top\left(\begin{array}{c}\mathbf{Z}_s\\\mathbf{X}_s\end{array}\right)\left(\begin{array}{c}\mathbf{Z}_{s'}\\\mathbf{X}_{s'}\end{array}\right)^\top\left(\begin{array}{c}\gamma_0-\widetilde{\gamma}_\mu\\(\boldsymbol{\beta}_0-\widetilde{\boldsymbol{\beta}})(\mathbf{U}_{s'})\end{array}\right)\right\}.\\&\le n^{-1}\sigma_0^2\sum_{i=1}^n\sum_{s=1}^n\sum_{s'=1}^n|g_{si}||g_{s'i}|[\mathrm{E}\{\mathbf{Z}_s^\top(\gamma_0-\widetilde{\gamma}_\mu)+\mathbf{X}_i^\top(\boldsymbol{\beta}_0-\widetilde{\boldsymbol{\beta}})(\mathbf{U}_s)\}^2]\\&=n^{-1}\sigma_0^2\sum_{i=1}^n\sum_{s=1}^n\sum_{s'=1}^n|g_{si}||g_{s'i}|O(|\triangle|^{2(\ell+1)}+|\triangle|^{4(\ell+1)})=O(|\triangle|^{2(\ell+1)}+|\triangle|^{4(\ell+1)}),\end{aligned}$$

since $\mathbf{G}$ are uniformly bounded in row sums and column sums by applying Lemma S.3. Therefore (S.12) is satisfied. ■

LEMMA S.11 *Under Assumptions (A1) – (A10),*

$$n^{-1}\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0=o_P(1), \tag{S.13}$$
$$n^{-1}\boldsymbol{\epsilon}^\top(\mathbf{I}_n-\mathbf{P}^*)\mathbf{G}\boldsymbol{\mu}_0=o_P(1), \tag{S.14}$$
$$n^{-1}(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n-\mathbf{P}^*)\mathbf{G}\boldsymbol{\mu}_0=o_P(1), \tag{S.15}$$

where $\mathbf{P}^*=\mathbf{P}_{\mathbf{Z}}$, $\mathbf{P}_{\boldsymbol{\mathcal{X}}_B}$ *or* $\mathbf{P}_{\boldsymbol{\Phi}}$.

Proof. Applying Cauchy-Schwarz inequality, it is easy to obtain that

$$\begin{aligned}&\mathrm{E}[\boldsymbol{\mu}_0^\top\mathbf{G}^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0]\\&=\sum_{s=1}^n\sum_{i=1}^n g_{si}\mathrm{E}\left[\{\mathbf{Z}_i^\top\gamma_0+\mathbf{X}_i^\top\boldsymbol{\beta}_0(\mathbf{U}_i)\}\{\mathbf{Z}_s^\top(\gamma_0-\widetilde{\gamma}_\mu)+\mathbf{X}_s^\top(\boldsymbol{\beta}_0-\widetilde{\boldsymbol{\beta}})(\mathbf{U}_s)\}\right]\\&\le\sum_{s=1}^n\sum_{i=1}^n|g_{si}|[\mathrm{E}\{\mathbf{Z}_i^\top\gamma_0+\mathbf{X}_i^\top\boldsymbol{\beta}_0(\mathbf{U}_i)\}^2]^{1/2}[\mathrm{E}\{\mathbf{Z}_s^\top(\gamma_0-\widetilde{\gamma}_\mu)+\mathbf{X}_s^\top(\boldsymbol{\beta}_0-\widetilde{\boldsymbol{\beta}})(\mathbf{U}_s)\}^2]^{1/2}\end{aligned}$$

Therefore, by Lemma S.8,

$$\begin{aligned}\mathrm{E}[\boldsymbol{\mu}_0^\top \mathbf{G}^\top (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0] \le{}& C\sum_{s=1}^n\sum_{i=1}^n |g_{si}| \times \{O(|\triangle|^{\ell+1} + |\triangle|^{2(\ell+1)} + |\triangle|^{3(\ell+1)/2})\}\\ ={}&O\{n(|\triangle|^{\ell+1} + |\triangle|^{2(\ell+1)})\}.\end{aligned}$$

Similarly, we can obtain that

$$\begin{aligned}&\mathrm{E}[\boldsymbol{\mu}_0^\top \mathbf{G}^\top (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0]^2\\ =&\sum_{i,i'}\sum_{s,s'} g_{si}g_{si'}\mathrm{E}\Big[\{\mathbf{Z}_i^\top\boldsymbol{\gamma}_0 + \mathbf{X}_i^\top\boldsymbol{\beta}_0(\mathbf{U}_i)\}\{\mathbf{Z}_s^\top(\boldsymbol{\gamma}_0 - \widetilde{\boldsymbol{\gamma}}_\mu) + \mathbf{X}_s^\top(\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_s)\}\\ &\times \{\mathbf{Z}_{i'}^\top\boldsymbol{\gamma}_0 + \mathbf{X}_{i'}^\top\boldsymbol{\beta}_0(\mathbf{U}_{i'})\}\{\mathbf{Z}_{s'}^\top(\boldsymbol{\gamma}_0 - \widetilde{\boldsymbol{\gamma}}_\mu) + \mathbf{X}_{s'}^\top(\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_{s'})\}\Big]\\ \le&\sum_{i,i'}\sum_{s,s'}|g_{si}||g_{si'}|[\mathrm{E}\{\mathbf{Z}_i^\top\boldsymbol{\gamma}_0 + \mathbf{X}_i^\top\boldsymbol{\beta}_0(\mathbf{U}_i)\}^2][\mathrm{E}\{\mathbf{Z}_s^\top(\boldsymbol{\gamma}_0 - \widetilde{\boldsymbol{\gamma}}_\mu) + \mathbf{X}_s^\top(\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})(\mathbf{U}_s)\}^2]\\ =&O\{n^2(|\triangle|^{4(\ell+1)} + |\triangle|^{2(\ell+1)})\}.\end{aligned}$$

Then result (S.13) is desired. Next we prove (S.14). Note that $\mathrm{E}\{\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{G}\boldsymbol{\mu}_0\} = 0$. Let $\boldsymbol{\Sigma}_\epsilon = \mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top|\{\mathbf{Z}_i, \mathbf{X}_i, \mathbf{U}_i\}, i = 1, \ldots, n]$. Under Assumption (A2), $\boldsymbol{\Sigma}_\epsilon = \sigma_0^2\mathbf{I}_n$.

$$\begin{aligned}&\mathrm{E}\{\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{G}\boldsymbol{\mu}_0\}^2 = \mathrm{E}\{(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{G}\boldsymbol{\mu}_0\}\\ =&\mathrm{E}\{(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\boldsymbol{\Sigma}_\epsilon(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{G}\boldsymbol{\mu}_0\} = \sigma_0^2\mathrm{E}\{(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{G}\boldsymbol{\mu}_0\}\\ \le&\sigma_0^2\mathrm{E}\{(\mathbf{G}\boldsymbol{\mu}_0)^\top\mathbf{G}\boldsymbol{\mu}_0\} = O(n).\end{aligned}$$

From the similar arguments, the above still holds for $\mathbf{P}^* = \mathbf{P}_{\mathbf{Z}}$ and $\mathbf{P}_{\boldsymbol{\Phi}}$, thus, (S.14) follows.

Finally, we prove (S.15). It is straightforward that $\mathrm{E}\{n^{-1}(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n - \mathbf{P}^*)\mathbf{G}\boldsymbol{\mu}_0\} = 0$. When $\mathbf{P}^* = \mathbf{P}_{\boldsymbol{\Phi}}$, we have

$$\begin{aligned}&\mathrm{E}\{(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n - \mathbf{P}^*)\mathbf{G}\boldsymbol{\mu}_0\}^2 = \mathrm{E}\{(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\mathbf{G}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0\}\\ =&\mathrm{E}\{(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\mathbf{G}^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0\} = \mathrm{E}\sum_{i=1}^n\{\mathbf{g}_i^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0\}^2.\end{aligned}$$

By Assumption (A8) and Lemma S.3, there exists a nonnegative constant M such that for any $i = 1, \ldots, n$, $\sum_{i'=1}^n |g_{ii'}| \le M$. Thus, $\|\mathbf{g}_i\mathbf{g}_i^\top\| \le \|\mathbf{g}_i\|\|\mathbf{g}_i\| = \sum_{i'=1}^n g_{ii'}^2 \le M^2$. Therefore,

$$\{\mathbf{g}_i^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0\}^2 = (\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{g}_i\mathbf{g}_i^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0 \le M(\mathbf{G}\boldsymbol{\mu}_0)^\top\mathbf{G}\boldsymbol{\mu}_0,$$

and $n^{-2}\mathrm{E}\{n^{-1}(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n - \mathbf{P}^*)\mathbf{G}\boldsymbol{\mu}_0\}^2 \le n^{-1}M^2 = O(n^{-1})$.

The above also holds when $\mathbf{P}^* = \mathbf{P}_{\mathbf{Z}}$ and $\mathbf{P}_{\boldsymbol{\mathcal{X}}_B}$. The desired result in (S.15) follows. ∎

LEMMA S.12 *Under Assumptions (A1) – (A10), when $K_n\log(n)/n^{1/2} \to 0$ as $n \to \infty$,*

$$n^{-1/2}\boldsymbol{\epsilon}^\top\mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon} = o_P(1), \tag{S.16}$$
$$n^{-1/2}\boldsymbol{\epsilon}^\top\mathbf{G}^\top\mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon} = o_P(1), \tag{S.17}$$
$$n^{-1/2}\boldsymbol{\epsilon}^\top\mathbf{G}^\top\mathbf{P}_{\boldsymbol{\Phi}}\mathbf{G}\boldsymbol{\epsilon} = o_P(1). \tag{S.18}$$

Proof. First, it is easy to obtain that

$$\begin{aligned}
&\mathrm{E}\left\{n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon}\right\} = n^{-1/2}\mathrm{E}\{\mathrm{tr}(\boldsymbol{\epsilon}^\top \mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon})\} = n^{-1/2}\mathrm{E}\{\mathrm{tr}(\mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top)\} \\
&= n^{-1/2}\mathrm{tr}\{\mathrm{E}(\mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top)\} = n^{-1/2}\mathrm{tr}\{\mathrm{E}\mathbf{P}_{\boldsymbol{\Phi}}\mathrm{E}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\} = n^{-1/2}\sigma_0^2\mathrm{E}\{\mathrm{tr}(\mathbf{P}_{\boldsymbol{\Phi}})\} = o(1),
\end{aligned}$$

therefore, (S.16) holds. Let $\mathbf{P}_{\boldsymbol{\Phi},i}$ denote the ith column of $\mathbf{P}_{\boldsymbol{\Phi}}$, and $\mathbf{g}_i$ denote the ith column of $\mathbf{G}$. Then, we have

$$\begin{aligned}
\mathrm{E}\left\{n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon}\right\} &= n^{-1/2}\sigma_0^2\mathrm{E}\{\mathrm{tr}(\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}})\} = n^{-1/2}\sigma^2\mathrm{E}\left\{\sum_{i=1}^n \mathbf{g}_i^\top \mathbf{P}_{\boldsymbol{\Phi},i}\right\} \\
&= n^{-1/2}\mathrm{E}\left\{\sum_{i=1}^n\sum_{j=1}^n g_{ij}\mathbf{P}_{\boldsymbol{\Phi},ij}\right\} \le n^{-1/2}\sigma_0^2\mathrm{E}\left\{\sum_{i=1}^n\sum_{j=1}^n |g_{ij}||\mathbf{P}_{\boldsymbol{\Phi},ij}|\right\} \\
&\asymp n^{-1/2}\sigma_0^2\frac{K_n}{n}\sum_{i=1}^n\sum_{j=1}^n |g_{ij}| = o(1).
\end{aligned}$$

Next we examine the variance:

$$\begin{aligned}
\mathrm{E}\left\{n^{-1/2}\boldsymbol{\epsilon}^\top \mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon}\right\}^2 =& n^{-1}(\mu_4 - 3\sigma_0^4)\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i^\top \mathbf{P}_{\boldsymbol{\Phi},i})^2 + n^{-1}\sigma_0^4[\mathrm{E}\{\mathrm{tr}(\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}})\}^2 \\
&+ \mathrm{E}\mathrm{tr}(\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}}\mathbf{G}) + \mathrm{E}\mathrm{tr}(\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}}\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}})].
\end{aligned}$$

Note that

$$\begin{aligned}
n^{-1}(\mu_4 - 3\sigma_0^4)\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i^\top \mathbf{P}_{\boldsymbol{\Phi},i})^2 &\le n^{-1}(\mu_4 - 3\sigma_0^4)\mathrm{E}\sum_{i=1}^n \mathbf{P}_{\boldsymbol{\Phi},i}^\top \mathbf{P}_{\boldsymbol{\Phi},i} = o(1), \\
n^{-1}\sigma_0^4\mathrm{E}\{\mathrm{tr}(\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}})\}^2 &\le n^{-1}\sigma_0^4\mathrm{E}\{\sum_{i=1}^n\sum_{j=1}^n |g_{ij}|\mathbf{P}_{\boldsymbol{\Phi},ij}|\}^2 = o(1), \\
n^{-1}\sigma_0^4\mathrm{E}\mathrm{tr}(\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}}\mathbf{G}) &= n^{-1}\sigma_0^4\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i^\top \mathbf{P}_{\boldsymbol{\Phi}}\mathbf{g}_i) = o(1),
\end{aligned}$$

and

$$\begin{aligned}
n^{-1}\sigma_0^4\mathrm{E}\mathrm{tr}(\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}}\mathbf{G}^\top \mathbf{P}_{\boldsymbol{\Phi}}) &\asymp n^{-1}\sigma_0^4\sum_{i=1}^n\sum_{i'=1}^n\sum_{j=1}^n\sum_{j'=1}^n g_{jj'}g_{ii'}\mathbf{P}_{\boldsymbol{\Phi},ij'}\mathbf{P}_{\boldsymbol{\Phi},ji'} \\
&\le n^{-1}\sigma_0^4\frac{K_n^2}{n^2}\sum_{i=1}^n\sum_{i'=1}^n\sum_{j=1}^n\sum_{j'=1}^n |g_{jj'}||g_{ii'}| = o(1).
\end{aligned}$$

So (S.17) is proved. Similarly, we can prove (S.18). ■

LEMMA S.13 *Under Assumptions (A1) – (A11), if* $\mathbf{C}_n$ *is uniformly bounded both in row sums* *and column sums in absolute values, then except on an event whose probability goes to zero,*

$$\left\|n^{-1}\boldsymbol{\mathcal{X}}_B^\top \mathbf{C}_n\boldsymbol{\mu}_0\right\|^2 = O_P(|\triangle|^2), \tag{S.19}$$

$$\left\|n^{-1/2}(\boldsymbol{\Phi}^{*\top}\mathbf{C}_n\boldsymbol{\epsilon})\right\|^2 = O_P(|\triangle|^{-2}), \tag{S.20}$$

where $\boldsymbol{\Phi}^*$ *is in (S.2).*

Proof. First we prove (S.19). Note that

$$\begin{aligned}\boldsymbol{\mathcal{X}}_B^\top \mathbf{C}_n\boldsymbol{\mu}_0 &= \{\mathbf{X}_1\otimes\mathbf{B}^*(\mathbf{U}_1),\ldots,\mathbf{X}_n\otimes\mathbf{B}^*(\mathbf{U}_n)\}\left(\sum_{i=1}^n c_{n,1i}\mu_{0i},\ldots,\sum_{i=1}^n c_{n,ni}\mu_{0i}\right)^\top \\ &= \sum_{i=1}^n \mathbf{X}_i\otimes\mathbf{B}^*(\mathbf{U}_i)\sum_{i'=1}^n c_{n,ii'}\mu_{0,i'},\end{aligned}$$

where $\mathbf{X}_i^\top\otimes\mathbf{B}^*(\mathbf{U}_i)^\top = \{\mathbf{X}_i^\top\otimes\mathbf{B}(\mathbf{U}_i)^\top\}(\mathbf{I}_{p_2}\otimes\mathbf{Q}_2)$. Then we have

$$\begin{aligned}&n^{-2}(\boldsymbol{\mathcal{X}}_B^\top \mathbf{C}_n\boldsymbol{\mu}_0)^\top(\boldsymbol{\mathcal{X}}_B^\top \mathbf{C}_n\boldsymbol{\mu}_0)\\ =&\left[n^{-1}\sum_{i=1}^n\{\mathbf{X}_i\otimes\mathbf{B}^*(\mathbf{U}_i)\}^\top\sum_{i'=1}^n c_{n,ii'}\mu_{0,i'}\right]\left[n^{-1}\sum_{s=1}^n \mathbf{X}_s\otimes\mathbf{B}^*(\mathbf{U}_s)\sum_{s'=1}^n c_{n,ss'}\mu_{0,s'}\right]\\ \le& C_1 n^{-2}\sum_{i=1}^n\sum_{s=1}^n|\{\mathbf{X}_i\otimes\mathbf{B}(\mathbf{U}_i)\}^\top\{\mathbf{X}_s\otimes\mathbf{B}(\mathbf{U}_s)\}|\sum_{i'=1}^n\sum_{s'=1}^n|c_{n,ii'}||c_{n,ss'}|\\ \asymp&\frac{1}{n}\sum_{i=1}\mathbf{B}(\mathbf{U}_i)^\top\frac{1}{n}\sum_{i'=1}^n\mathbf{B}(\mathbf{U}_{i'})\asymp O_P(|\triangle|^2).\end{aligned}$$

Next, we provide the proof of (S.20). By Lemma S.3, $\mathbf{C}_n\mathbf{C}_n^\top$ and $\mathbf{C}_n^\top\mathbf{C}_n$ are uniformly bounded both in row sums and column sums. By Lemma A.10 from Lee (2004),

$$n^{-1/2}(\boldsymbol{\Phi}^{*\top}\mathbf{C}_n\boldsymbol{\epsilon}) = \begin{pmatrix} n^{-1/2}\mathbf{Z}^\top\mathbf{C}_n\boldsymbol{\epsilon}\\ n^{-1/2}\sqrt{K_n}\boldsymbol{\mathcal{X}}_B^\top\mathbf{C}_n\boldsymbol{\epsilon}\end{pmatrix},\quad n^{-1/2}\mathbf{Z}^\top\mathbf{C}_n\boldsymbol{\epsilon} = O_P(1).$$

Next, let $q_{nm} = \sqrt{K_n}\boldsymbol{\mathcal{X}}_B^\top\mathbf{C}_{nm}$, where $\mathbf{C}_{nm}$ is the m-thcolumn of $\mathbf{C}_n$. Then we can obtain

$$\begin{aligned}q_{nm} &= \sqrt{K_n}\sum_{i=1}^n \mathbf{X}_i\otimes\mathbf{B}^*(\mathbf{U}_i)C_{n,im}\\ &= \sqrt{K_n}\left(\sum_i C_{n,im}X_{i1}\mathbf{B}(\mathbf{U}_i)^\top\mathbf{Q}_2,\ldots,\sum_i C_{n,im}X_{ip_2}\mathbf{B}(\mathbf{U}_i)^\top\mathbf{Q}_2\right)^\top,\end{aligned}$$

and $\mathrm{Var}(n^{-1/2}\sqrt{K_n}\boldsymbol{\mathcal{X}}_B^\top\mathbf{C}_n\boldsymbol{\epsilon}) = n^{-1}\sigma^2\sum_{m=1}^n q_{nm}q_{nm}^\top$, where q_{nm} are uniformly bounded. Note that for any k,k',i,i',j,j',

$$\mathrm{E}\{X_{ik}X_{i'k'}B_j(\mathbf{U}_i)B_{j'}(\mathbf{U}_{i'})\}\le\mathrm{E}\{|X_{ik}||X_{i'k'}|\times B_j(\mathbf{U}_i)B_{j'}(\mathbf{U}_{i'})\} = O(|\triangle|^2),$$

which implies q_{nj} are uniformly bounded. By applying Chebyshev's Inequality, the desired result follows. ■

S.2. *Asymptotic Results for Unpenalized Partially Linear Bivariate Spline Estimators without Neighboring Effects*

If $\alpha = 0$, then we have

$$Y_i = \sum_{l=1}^{p_1} Z_{il}^\top \gamma_{0l} + \sum_{k=1}^{p_2} X_{ik}\boldsymbol{\beta}_0(\mathbf{U}_i) + \epsilon_i. \tag{S.21}$$

Thus, $\widehat{\gamma} - \gamma_0 = (\widetilde{\gamma}_\mu - \gamma_0) + \widetilde{\gamma}_\epsilon$, where $\widetilde{\gamma}_\mu$, $\widetilde{\gamma}_\epsilon$ are defined in (S.6) and (S.7).

LEMMA S.14 *Under Assumptions (A1) – (A6), as $n \to \infty$,*

$$[\mathrm{Var}\,(\widetilde{\gamma}_\epsilon | \{(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{U}_i), i = 1, \ldots, n\})]^{-1/2}\, \widetilde{\gamma}_\epsilon \to N(0, \mathbf{I}_{p_1 \times p_1}).$$

Proof. For any $\mathbf{b} \in \mathbb{R}^{p_1}$ with $\|\mathbf{b}\| = 1$, denote that $\mathbf{b}^\top \widetilde{\boldsymbol{\epsilon}} = \sum_{i=1}^n \eta_i \epsilon_i$, where

$$\eta_i^2 = n^{-2}\mathbf{b}^\top (n\mathbf{U}_{11})(\mathbf{Z}_i - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathcal{X}_{B,i})(\mathbf{Z}_i^\top - \boldsymbol{\mathcal{X}}_{B,i}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})(n\mathbf{U}_{11})\mathbf{b},$$

where $\mathcal{X}_{B,i} = \mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)$, and conditioning on $\{(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{U}_i), i = 1, \ldots, n\}$, $\eta_i \epsilon_i^\prime s$ are independent. By Lemma S.5,

$$\max_{1 \le i \le n} \eta_i^2 \le C n^{-2} \max_{1 \le i \le n} \left\{ \|\mathbf{Z}_i\|^2 + \|\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathcal{X}_{B,i}\|^2 \right\}.$$

Note that for any $i \in 1, \ldots, n$, $\mathbf{Z}_i^\top \mathbf{Z}_i = O_P(1)$ under Assumption (A2). Thus, $\|\mathbf{Z}\mathbf{Z}^\top\|_F = \{\sum_{i=1}^n \sum_{i'=1}^n |\mathbf{Z}_i^\top \mathbf{Z}_{i'}|^2\}^{1/2} = O(n)$. Note that $\|\mathbf{Z}\mathbf{Z}^\top\|_2 \le \|\mathbf{Z}\mathbf{Z}^\top\|_F \le \sqrt{p_1}\|\mathbf{Z}\mathbf{Z}^\top\|_2$. Then, we have

$$\begin{aligned}
\|\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathcal{X}_{B,i}\|^2 &= \{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\} \\
&\le Cn\{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\}^\top \mathbf{V}_{22}^{-1}\boldsymbol{\mathcal{X}}_B^\top \boldsymbol{\mathcal{X}}_B \mathbf{V}_{22}^{-1}\{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\} \\
&\asymp n\{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\}^\top \mathbf{V}_{22}^{-1}\{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\} \\
&\asymp K_n \|\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\|^2 \asymp K_n \asymp |\triangle|^{-2}.
\end{aligned}$$

For large n, with probability approaching 1,

$$\max_{1 \le i \le n} \|\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathcal{X}_{B,i}\|^2 = O_P(|\triangle|^{-2}), \quad \max_{1 \le i \le n} \eta_i^2 = O_P(n^{-2}|\triangle|^{-2}).$$

Thus,

$$\begin{aligned}
\sigma_0^2 \sum_{i=1}^n \eta_i^2 &= \mathrm{Var}\left[\mathbf{b}^\top \widetilde{\gamma}_\epsilon | \{(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{U}_i), i = 1, \ldots, n\}\right] = \mathbf{b}^\top \mathbf{U}_{11}\mathbf{Z}^\top (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\mathcal{X}}_B})\mathbf{Z}\mathbf{U}_{11}\mathbf{b} \\
&= n^{-1}\mathbf{b}^\top (n\mathbf{U}_{11}) \left\{ n^{-1} \sum_{i=1}^n (\mathbf{Z}_i - \mathbf{Z}_i^*)(\mathbf{Z}_i - \mathbf{Z}_i^*)^\top \right\} (n\mathbf{U}_{11})\mathbf{b}\sigma_0^2,
\end{aligned}$$

where $\mathbf{Z}_i^*$ is the i-thcolumn of $\mathbf{Z}^\top \mathbf{P}_{\boldsymbol{\mathcal{X}}_B}$. By Lemma S.5, with probability approaching 1, $\sum_{i=1}^n \eta_i^2 \geq c/n$. So $\max_{1\leq i\leq n} \eta_i^2 / \sum_{i=1}^n \eta_i^2 = O_P(n^{-1}|\triangle|^{-2}) = o_P(1)$. Applying Lindeberg-Feller Central Limit Theorem, we have

$$\left(\sigma_0^2 \sum_{i=1}^n \eta_i^2\right)^{-1/2} \sum_{i=1}^n \eta_i \epsilon_i \longrightarrow N(0,1).$$

■

LEMMA S.15 *Under (A1) – (A6), for the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \sigma^{-2}\boldsymbol{\Xi}$ of the estimator $\boldsymbol{\gamma}$, $\boldsymbol{\Xi}$ is defined in (10), $(n\sigma_0^{-2}\boldsymbol{\Xi})^{1/2}(\widehat{\gamma} - \gamma_0) \to N(\mathbf{0}, \mathbf{I}_{p_1\times p_1})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ could be consistently estimated by $(n\widehat{\sigma}^2)^{-1}\sum_{i=1}^n (\mathbf{Z}_i - \mathbf{Z}_i^*)(\mathbf{Z}_i - \mathbf{Z}_i^*)^\top$. That is,*

$$\mathrm{Var}\left(\widetilde{\gamma}_\epsilon | \{(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{U}_i), i = 1,\ldots,n\}\right) = n^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} + o_P(1).$$

Proof. The proof is similar to Theorem 1 in Wang, Wang, Lai, and Gao (2020), thus omitted. ■

LEMMA S.16 *For Model (S.21), under Assumptions (A1) – (A6), for any $k = 1,\ldots,p_2$, the spline estimator $\widehat{\beta}_k(\cdot)$ is consistent and satisfies that $\|\widehat{\beta}_k - \beta_{0k}\|_{L_2} = O_P(n^{-1/2}|\triangle| + |\triangle|^{\ell+1})$.*

Proof. By (S.1), if $\alpha = 0$,

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= \mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})(\boldsymbol{\mu}_c + \boldsymbol{\mu}_{\mathrm{V}} + \boldsymbol{\epsilon}) \\ &= \mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\mu}_{\mathrm{V}} + \mathbf{U}_{22}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\epsilon} \equiv \widetilde{\boldsymbol{\theta}}_\mu + \widetilde{\boldsymbol{\theta}}_\epsilon.\end{aligned}$$

Then, we have the following decomposition:

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \widetilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}^* + \widetilde{\boldsymbol{\theta}}_\epsilon,$$

where $\boldsymbol{\theta}^*$ is defined in (S.8). Then, according to Lemma S.6 and S.13, $n^{-1/2}K_n^{1/2}\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\epsilon} = O_P(1)$. Next, by Lemmas S.4 and S.5, we have

$$\begin{aligned}\|\widetilde{\boldsymbol{\theta}}_\epsilon\|^2 &= \frac{K_n^2}{n^2}\boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\mathcal{X}}_B(nK_n^{-1}\mathbf{U}_{22})(nK_n^{-1}\mathbf{U}_{22})\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\epsilon} \\ &\asymp \frac{K_n^2}{n^2}\|\boldsymbol{\mathcal{X}}_B^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})\boldsymbol{\epsilon}\|^2 = O_P(K_n^2/n).\end{aligned}$$

Thus, by Lemma S.9, we have

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \leq \|\widetilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}^*\| + \|\widetilde{\boldsymbol{\theta}}_\epsilon\| = O_P(|\triangle|^\ell + K_n/\sqrt{n}).$$

Therefore, Lemma S.1 implies that $\|\widehat{\beta}_k - \beta_{0k}\|_{L_2} = O_P(|\triangle|/\sqrt{n} + |\triangle|^{\ell+1})$. ■

S.3. *Proof of Lemma* A.1

Proof. By Lemma A.5 in Lee (2004) and Assumptions (A7) – (A9),

$$\sup_{\alpha\in\mathcal{D}} n^{-1}\mathrm{tr}[\{\mathbf{T}(\alpha)\mathbf{T}^{-1}\}^\top \mathbf{T}(\alpha)\mathbf{T}^{-1}] = O(1). \tag{S.22}$$

Therefore, $\sigma^2(\alpha) = n^{-1}\sigma_0^2 \mathrm{tr}[\{\mathbf{T}(\alpha)\mathbf{T}^{-1}\}^\top \mathbf{T}(\alpha)\mathbf{T}^{-1}]$ is uniformly bounded for $\alpha \in \mathcal{D}$. Recall $l_n^*(\sigma^2, \alpha)$ defined in (A.3) is the log-likelihood function of a standard SAR model:

$$\mathbf{Y} = \alpha \mathbf{W}\mathbf{Y} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Suppose that σ_0^2 and α_0 are the true parameters for this SAR process, then by Jensen's inequality, we have

$$\max_{\sigma^2} E_{(\sigma_0^2,\alpha_0)}[\ell_n^*(\sigma^2,\alpha)] \le E_{(\sigma_0^2,\alpha_0)} l_n^*(\sigma_0^2,\alpha_0), \quad \text{for all } \alpha \in \mathcal{D}$$

i.e., by ignoring the constant term, we have

$$-\frac{n}{2}\log\sigma^2(\alpha) + \log|\mathbf{T}(\alpha)| \le -\frac{n}{2}\log\sigma_0^2 + \log|\mathbf{T}|.$$

That is, for all $\alpha \in \mathcal{D}$

$$\frac{1}{2}\log\sigma^2(\alpha) \ge \frac{1}{2}\log\sigma_0^2 + \frac{1}{n}(\log|\mathbf{T}| - \log|\mathbf{T}(\alpha)|). \tag{S.23}$$

For any $\alpha_1, \alpha_2 \in \mathcal{D}$, by the mean value theorem, there exists an $\overline{\alpha}$ between α_1 and α_2, such that

$$\frac{1}{n}\{\log|\mathbf{T}(\alpha_2)| - \log|\mathbf{T}(\alpha_1)|\} = \frac{1}{n}\mathrm{tr}\{\mathbf{W}\mathbf{T}^{-1}(\overline{\alpha})\}(\alpha_2 - \alpha_1).$$

Under Assumption (A8) and Lemma A.8 in Lee (2004), $\sup_{\alpha\in\mathcal{D}}\left[n^{-1}\mathrm{tr}\{\mathbf{W}\mathbf{T}^{-1}(\alpha)\}\right] = O(\nu_n^{-1})$. So $\frac{1}{n}\log|\mathbf{T}(\alpha)|$ is uniformly equicontinous in $\alpha \in \mathcal{D}$. Thus by (S.23), $\log\sigma^2(\alpha)$ is bounded from below. Combined with (S.22), $\sigma^2(\alpha)$ is uniformly bounded away from 0. Recall (A.2), by Assumptions (A7) – (A10) and Lemma S.10 and Lemma S.11,

$$\widehat{\sigma}^{*2}(\alpha) = \sigma^2(\alpha) + n^{-1}(\alpha_0 - \alpha)^2\left[\mathrm{E}\left\{\mathbf{G}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})(\mathbf{G}\boldsymbol{\mu}_0)\right\}\right] + o(1). \tag{S.24}$$

So $\widehat{\sigma}^{*2}(\alpha) \ge \sigma^2(\alpha) + o(1)$. Therefore, $\widehat{\sigma}^{*2}(\alpha)$ is continuous and uniformly bounded away from zero. So $\{\widehat{\sigma}^*(\alpha)\}^{-2}$ is uniformly bounded. That will result in the uniform equicontinuity of $\log\widehat{\sigma}^{*2}(\alpha)$. Recall that $n^{-1}s_n(\alpha) = -1/2\log\widehat{\sigma}^{*2}(\alpha) + n^{-1}\log|\mathbf{T}(\alpha)| + const$. The uniformly equicontinuity of $n^{-1}s_n(\alpha)$ holds. ■

S.4. *Proof of Lemma* A.2

Proof. It is straight forward to obtain that:

$$\frac{1}{n}\ell_n(\alpha) - \frac{1}{n}s_n(\alpha) = -\frac{1}{2}[\log\widehat{\sigma}^2(\alpha) - \log\widehat{\sigma}^{*2}(\alpha)].$$

Noting that $\mathbf{T}(\alpha)\mathbf{T}^{-1} = (\mathbf{I}_n - \alpha\mathbf{W})\mathbf{T}^{-1} = \mathbf{I}_n + (\alpha_0 - \alpha)\mathbf{G}$, we have

$$\begin{aligned}
(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{Y} &= (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\mathbf{T}\mathbf{Y} \\
&= (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})[\mathbf{I}_n + (\alpha_0 - \alpha)\mathbf{G}](\boldsymbol{\mu}_0 + \boldsymbol{\epsilon}) \\
&= (\alpha_0 - \alpha)(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0 + (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} + (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0,
\end{aligned}$$

Then $\widehat{\sigma}^2(\alpha)$ could be written as

$$\begin{aligned}\widehat{\sigma}^2(\alpha) =& \frac{1}{n}\|(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{Y}\|^2 \\ =& \frac{1}{n}(\alpha_0-\alpha)^2(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})(\mathbf{G}\boldsymbol{\mu}_0) + \frac{2}{n}(\alpha_0-\alpha)(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} \\ &+ \frac{1}{n}\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} + \frac{1}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\mu}_0 \\ &+ \frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} + \frac{2(\alpha_0-\alpha)}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0. \end{aligned} \tag{S.25}$$

By Lemmas S.10 and S.11,

$$\begin{aligned}\widehat{\sigma}^2(\alpha) =& \frac{1}{n}(\alpha_0-\alpha)^2(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})(\mathbf{G}\boldsymbol{\mu}_0) + \frac{2}{n}(\alpha_0-\alpha)(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} \\ &+ \frac{1}{n}\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} \\ &+ \frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} + o_P(1).\end{aligned}$$

According to (S.24), we have

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

By Lemma A.12 in Lee (2004),

$$\sup_{\alpha\in\mathcal{D}}\left|\frac{1}{n}(\alpha_0-\alpha)^2(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})(\mathbf{G}\boldsymbol{\mu}_0) - \frac{(\alpha_0-\alpha)^2}{n}\left[\mathrm{E}\left\{\mathbf{G}\boldsymbol{\mu}_0^\top(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})(\mathbf{G}\boldsymbol{\mu}_0)\right\}\right]\right| = o_P(1),$$

because α appears simply as a quadratic factor in the left handed side. Similarly we have

$$\sup_{\alpha\in\mathcal{D}}\left|\frac{1}{n}\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} - \frac{1}{n}\mathrm{E}\{\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon}\}\right| = o_P(1). \tag{S.26}$$

Note that

$$\mathrm{E}\{\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon}\} = \sigma_0^2\mathrm{tr}\{(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{T}(\alpha)\mathbf{T}^{-1}\},$$

thus,

$$\sup_{\alpha\in\mathcal{D}}\left\{\frac{1}{n}\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} - \sigma^2(\alpha)\right\} = o_P(1). \tag{S.27}$$

Note that

$$\begin{aligned}(\alpha_0-\alpha)(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} &= (\alpha_0-\alpha)\boldsymbol{\epsilon}^\top\{\mathbf{I}_n+(\alpha_0-\alpha)\mathbf{G}\}(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0\\ &= (\alpha_0-\alpha)\boldsymbol{\epsilon}^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\mathbf{G}\boldsymbol{\mu}_0+(\alpha_0-\alpha)^2(\mathbf{G}\boldsymbol{\epsilon})^\top(\mathbf{I}_n-\mathbf{P}_{\mathbf{Z}})\mathbf{G}\boldsymbol{\mu}_0,\end{aligned}$$

where α appears as linear and quadratic in the above. Thus, by Lemmas S.10 and S.11, we have

$$\sup_{\alpha\in\mathcal{D}}\left\{\frac{2}{n}(\alpha_0-\alpha)(\mathbf{G}\boldsymbol{\mu}_0)^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon}\right\}=o_P(1). \tag{S.28}$$

Similarly, we can obtain that

$$\sup_{\alpha\in\mathcal{D}}\left\{\frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon}\right\}=\sup_{\alpha\in\mathcal{D}}\left\{\frac{2}{n}\boldsymbol{\mu}_0^\top(\mathbf{I}_n-\mathbf{P}_{\boldsymbol{\Phi}})(\mathbf{I}_n+(\alpha_0-\alpha)\mathbf{G})\boldsymbol{\epsilon}\right\}=o_P(1). \tag{S.29}$$

Recall $\boldsymbol{\Phi}^*$ defined in (S.2), then we can obtain that$\mathbf{P}_{\boldsymbol{\Phi}^*}=\mathbf{P}_{\boldsymbol{\Phi}}$. Then by Lemma S.5 and Lemma S.13, noting that $\lim_{n\to\infty}n^{-1}K_n=0$, $\mathbf{T}(\alpha),\mathbf{T}$ are matrices with off-diagonal elements non-positive, so by using the properties of Metzler matrix, all the elements of $\mathbf{T}^{-1}$ and $\alpha\mathbf{W}\mathbf{T}^{-1}$ are non-negative.

$$\begin{aligned}\frac{1}{n}\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{P}_{\boldsymbol{\Phi}}\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon} &= \frac{1}{n}\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{P}_{\boldsymbol{\Phi}^*}\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon}\\ &= \frac{1}{n^2}\{\boldsymbol{\Phi}^{*\top}(\mathbf{T}^{-1}-\alpha\mathbf{G})\boldsymbol{\epsilon}\}^\top\left(\frac{\boldsymbol{\Phi}^{*\top}\boldsymbol{\Phi}^*}{n}\right)^{-1}\{\boldsymbol{\Phi}^{*\top}(\mathbf{T}^{-1}-\alpha\mathbf{G})\boldsymbol{\epsilon}\}\\ &\asymp \frac{1}{n}\frac{\{\boldsymbol{\Phi}^{*\top}(\mathbf{T}^{-1}-\alpha\mathbf{G})\boldsymbol{\epsilon}\}}{\sqrt{n}}^\top\frac{\{\boldsymbol{\Phi}^{*\top}(\mathbf{T}^{-1}-\alpha\mathbf{G})\boldsymbol{\epsilon}\}}{\sqrt{n}}\\ &= O_P(K_n/n),\end{aligned}$$

and

$$\sup_{\alpha\in\mathcal{D}}\left\{\frac{1}{n}\boldsymbol{\epsilon}^\top(\mathbf{T}^{-1})^\top\mathbf{T}(\alpha)^\top\mathbf{P}_{\boldsymbol{\Phi}}\mathbf{T}(\alpha)\mathbf{T}^{-1}\boldsymbol{\epsilon}\right\}=o_P(1). \tag{S.30}$$

By (S.26) to (S.30), $\widehat{\sigma}^2(\alpha)-\widehat{\sigma}^{*2}(\alpha)=o_P(1)$ uniformly in $\alpha\in\mathcal{D}$, then we can use the mean value theorem,

$$\sup_{\alpha\in\mathcal{D}}\{|\log\widehat{\sigma}^2(\alpha)-\log\widehat{\sigma}^{*2}(\alpha)|\}=\sup_{\alpha\in\mathcal{D}}\frac{1}{\breve{\sigma}^2(\alpha)}|\widehat{\sigma}^2(\alpha)-\widehat{\sigma}^{*2}(\alpha)|=o_P(1),$$

where $\breve{\sigma}^2(\alpha)$ is between $\widehat{\sigma}^2(\alpha)$ and $\widehat{\sigma}_n^{*2}(\alpha)$. Therefore, $\sup_{\alpha\in\mathcal{D}}n^{-1}|\ell_n(\alpha)-s_n(\alpha)|=o_P(1)$. ■

S.5. *Additional Simulation Studies*

In this section, we conduct additional simulation studies to investigate the proposed weight estimation algorithms presented in Section 4.1 of the main paper.

We consider the horseshoe domain described in Section 5.1 in the paper, and randomly select $n=200,\ 500,\ 1000,\ 2000,\ 5000$ sample locations from the domain uniformly. To show the

Table S.1. Convergence results of Algorithm 1.

n		$S = 5$	$S = 15$	$S = 30$	$S = 50$	$S = 70$
200	δ_w	0.0347	0.0202	0.0135	0.0109	0.0105
	Time (seconds)	17	44	86	136	144
500	δ_w	0.0383	0.0263	0.0208	0.0167	0.0152
	Time (seconds)	20	50	100	162	220
1000	δ_w	0.0270	0.0194	0.0154	0.0133	0.0127
	Time (seconds)	86	224	387	690	923
2000	δ_w	0.0156	0.0122	0.0100	0.0088	0.0081
	Time (seconds)	382	1021	1942	3423	4711
5000*	δ_w	0.0069	0.0063	0.0063	0.0061	0.0061
	Time (seconds)	6320	13805	21827	36378	50121

"*" indicates that the results are based on 10 replications.

performance of the proposed Algorithm 1, we let $w_{ij} = \exp(-10d_{ij})/\sum_{k\neq i}\exp(-10d_{ik})$, and calculate the following measure:

$$\delta_w = \max_{i\neq j}|w_{ij}^{(1)} - w_{ij}^{*}|,$$

where $w_{ij}^{(1)}$ is the weight calculated based on the estimated distance ($d_{ij}^{(1)}$) by Algorithm 1, and w_{ij}^{*} is the weight calculated based on the geodesic distance (d_{ij}^{*}) calculated using the approach proposed by Miller and Wood (2014). We run 100 replications for each of the setting with $n = 200,\ 500,\ 1000,\ 2000$, and 10 replications for $n = 5000$ due to the huge amount of computing power needed by the approach in Miller and Wood (2014).

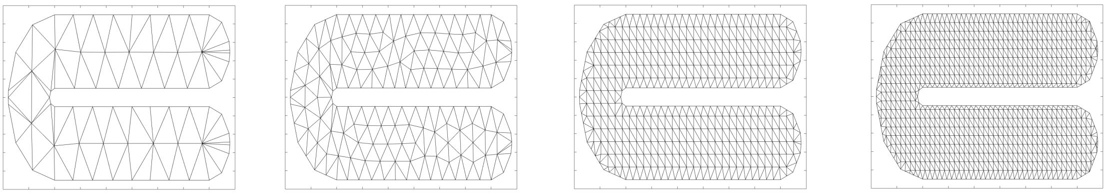
Table S.1 reports the average δ_w values over replications, and the average computing time (seconds per replication) on a regular PC with processor Core i7 @3.3GHz CPU and 16.00GB RAM. From Table S.1, when the number of iteration steps, S, is increasing, we can clearly see the convergence of the results. When $S = 50$, the value of δ_w is close to be a constant. So in the real application, we suggest take $S = 30 \sim 50$ for small sample size and $S = 10 \sim 30$ when sample size $n > 2000$.

Next, we examine the performance of Algorithm 2. For that purpose, we randomly select $n = 1000,\ 2000,\ 5000,\ 10000$ sample locations from the domain uniformly, and calculate the maximum of the difference between the two weights as follows:

$$\delta_w = \max_{i\neq j}|w_{ij}^{(2)} - w_{ij}^{(*)}|,$$

where $w_{ij}^{(2)}$ is the weight calculated based on the estimated distance ($d_{ij}^{(2)}$) by Algorithm 2, and w_{ij}^{*} is the weight calculated based on the geodesic distance (d_{ij}^{*}) calculated using the approach proposed by Miller and Wood (2014). To show the effect of triangulation, we consider four different triangulations; see $\triangle_1$–$\triangle_4$ in Figure S.1.

Table S.2 shows the average δ_w values over replications, and the average computing time (seconds per replication). We run 100 replications for each of the settings with $n = 1000,\ 2000$, and 10 replications for $n = 5000$ due to the huge amount of computing power needed by the approach in Miller and Wood (2014). For $n = 10000$, we only report the computing time of Algorithm 2, because the approach in Miller and Wood (2014) cannot be implemented due to its huge memory requirement. The columns $\triangle_1$–$\triangle_4$ represent the results obtained based on the distance calculated from $\triangle_1$–$\triangle_4$, respectively. From Tables S.1 and S.2, one sees that for $n = 1000$ and 2000, the results based on Algorithm 2 are less accurate than those based on



(a) $\triangle_1$ (87 triangles) (b) $\triangle_2$ (238 triangles) (c) $\triangle_3$ (736 triangles) (d) $\triangle_4$ (1613 triangles)

Figure S.1. Triangulations considered in Algorithm 2.

Table S.2. Performance of Algorithm 2 based on different triangulations.

n		$\triangle_1$	$\triangle_2$	$\triangle_3$	$\triangle_4$
1000	δ_w	0.0297	0.0286	–	–
	Time (seconds)	52	52	–	–
2000	δ_w	0.0226	0.0218	0.0166	–
	Time (seconds)	53	53	111	–
5000*	δ_w	0.0138	0.0136	0.0116	0.0081
	Time (seconds)	64	67	125	657
10000*	δ_w	–	–	–	–
	Time (seconds)	135	137	206	745

"*" indicates that the results are based on 10 replications.

Algorithm 1. However, when $n = 5000$, the results based on Algorithms 1 and 2 are quite comparable. Overall, Algorithm 2 is much faster than Algorithm 1 when $n = 5000$. Remarkably, Algorithm 2 with $n = 10000$ is faster than Algorithm 1 with $n = 2000$, and much faster than Algorithm 1 with $n = 5000$.

From Table S.2, we find that results are more accurate when the triangulation becomes finer, but the computing is also more expensive for fine triangulations. To reduce the computing burden for large n, we recommend using the triangulations with the number of triangles less than $n/3$.

S.6. *Extra Simulation Results*

In this section, we provide additional results from the simulation studies presented in Section 5 of the main paper. Figures S.2–S.3 depict the contour plots for the estimated coefficient functions for a typical simulation run in Simulation Study 1 when $n = 500,\ 1000$, respectively. Figures S.4 shows the contour plots for the estimated coefficient functions for a typical simulation run in Simulation Study 2 when $n = 500$.

References

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

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

Miller, D.L., and Wood, S.N. (2014), 'Finite area smoothing with generalized distance splines', *Environmental and Ecological Statistics*, 21, 715–731.

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

Sun, Y., Yan, H., Zhang, W., and Lu, Z. (2014), 'A semiparametric spatial dynamic model', *The Annals of Statistics*, 42, 700–727.

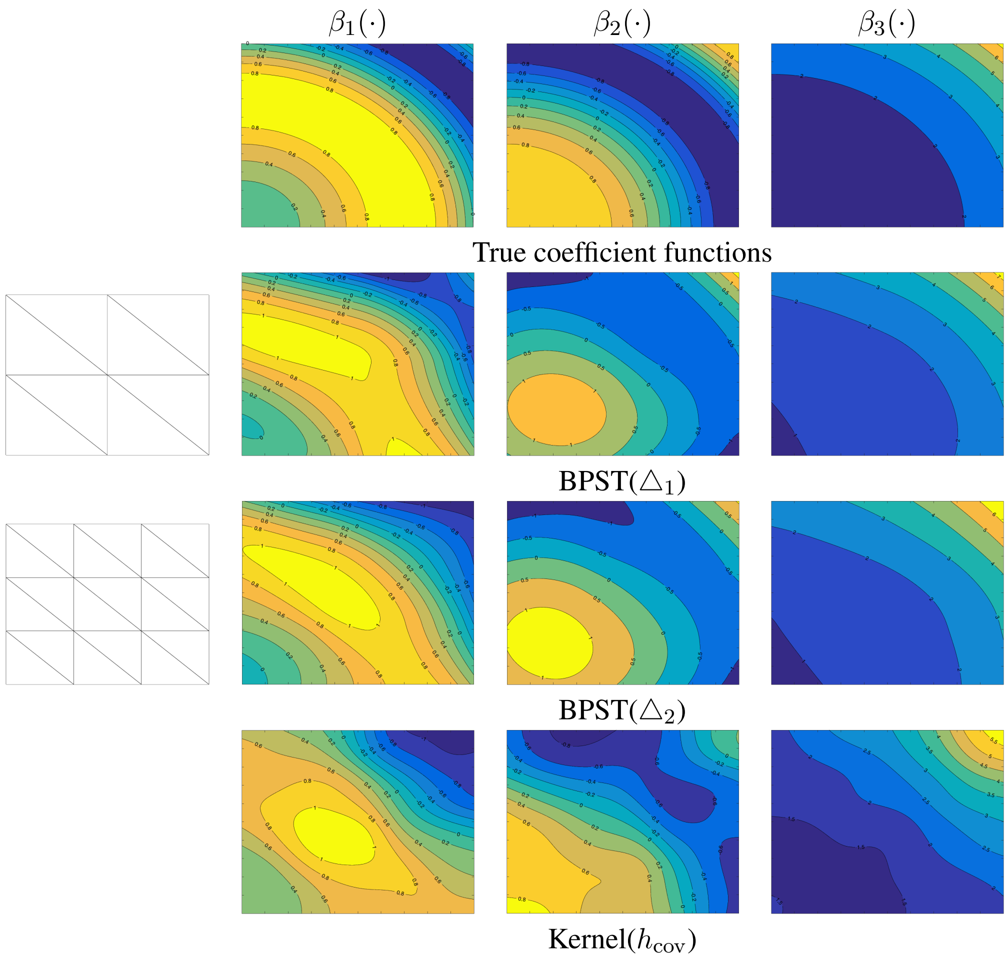


Figure S.2. Simulation Study 1: contour plots for the true and estimated coefficient functions when $n = 500$. BPST($\triangle_j$): bivariate penalized spline estimator based on triangulation $\triangle_j$, $j = 1, 2, 3$; Kernel(h_{cov}): local smoothing method (Sun et al. 2014) with bandwidth h_{cov}.

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

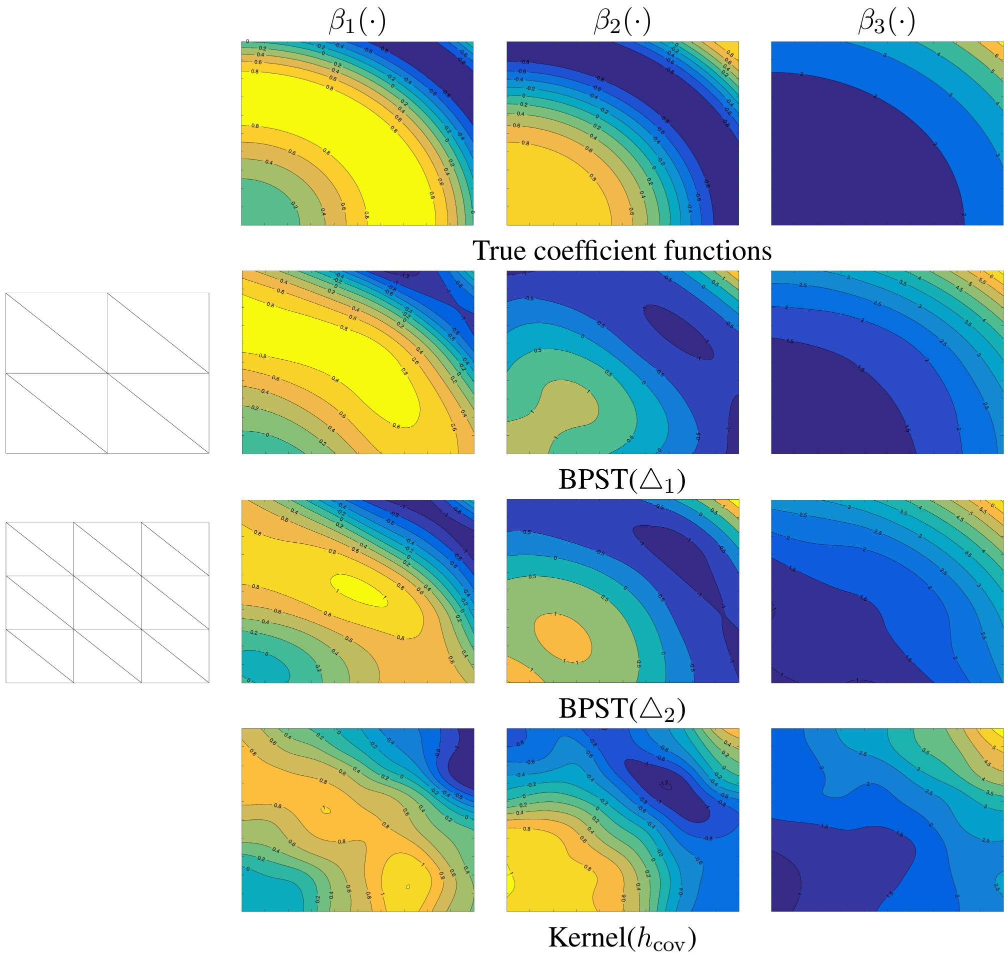


Figure S.3. Simulation Study 1: contour plots for the true and estimated coefficient functions when $n = 1000$. BPST($\triangle_j$): bivariate penalized spline estimator based on triangulation $\triangle_j$, $j = 1, 2, 3$; Kernel(h_{cov}): local smoothing method (Sun et al. 2014) with bandwidth h_{cov}..

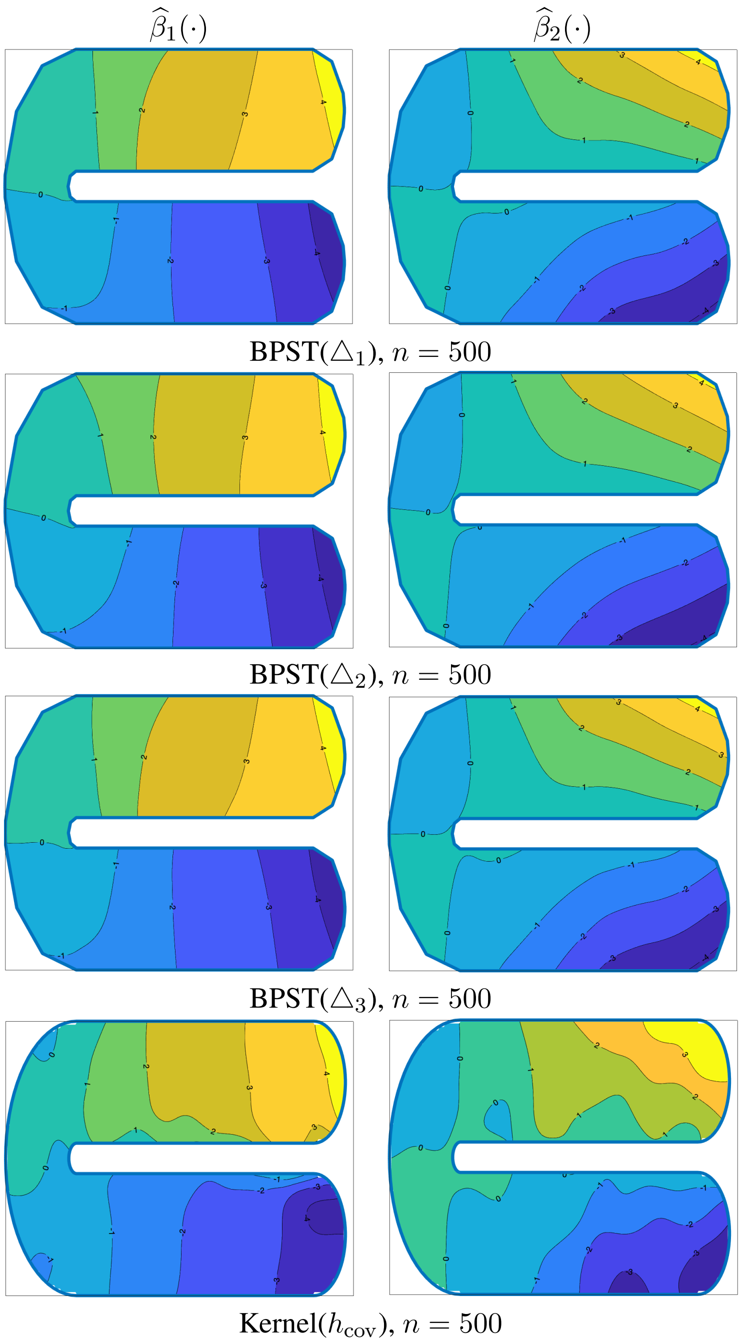


Figure S.4. Simulation Study 2: the estimated contour plots for coefficient functions. BPST($\triangle_j$): bivariate penalized spline estimator based on triangulation $\triangle_j$, $j = 1, 2, 3$; Kernel(h_{cov}): local smoothing method (Sun et al. 2014) with bandwidth h_{cov}.