Supplementary Material to "Efficient Estimation of Partially Linear Models for Spatial Data over Complex Domains via Penalized Bivariate Splines on Triangulations"

Li WANG[1], Guannan WANG[2], Ming-Jun LAI[3] and Lei GAO[1]
[1]Iowa State University, Ames, IA 50011, USA
[2]College of William & Mary, Williamsburg, VA 23185, USA.
[3]The University of Georgia, Athens, GA 30602, USA.

This note contains detailed explanations of how to choose the triangulation for the proposed method, more simulation studies, and proofs of Lemma 1, Theorems 1, 2 with some technical details.

S.1 Choosing the Triangulation

The triangulation selection is one of the key ingredients for obtaining good performance of the bivariate splines estimation. An optimal triangulation is a partition of the domain which is best according to some criterion that measures the shape, size or number of triangles. For example, one of the well-known criteria used to control the shape with a triangulation is the "max-min" criterion which maximizes the minimum angle of all the angles of the triangles in the triangulation. Based on the "max-min" criterion, the Delaunay triangulation algorithm can be implemented to avoid sliver triangles (a triangle that is almost flat) when a set of appropriate vertices is chosen. In the past few decades, various packages have been developed to realize the Delaunay algorithm; see MATLAB program *delaunay.m* or MATHEMATICA function *DelaunayTriangulation*. "Triangle" (Shewchuk, 1996) is also widely used in many applications, and one can download it for free from `http://www.cs.cmu.edu/~quake/triangle.html`. It is a C++ program for two-dimensional mesh generation and construction of Delaunay triangulations. "DistMesh" is another method to generate unstructured triangular and tetrahedral meshes; see the *DistMesh* generator on `http:`

`//persson.berkeley.edu/distmesh/`. A detailed description of the program is provided by Persson and Strang (2004). Once the shape of triangulations is handled, we can simply focus on how to select the number of triangles, K, for quasi-uniform triangulations in all the numerical studies.

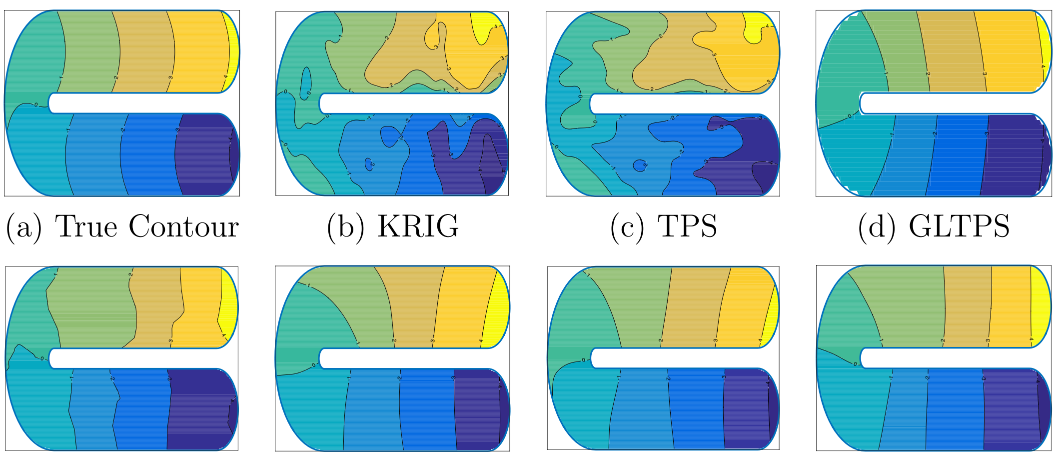
As is usual with the one-dimensional (1-D) penalized least squares (PLS) splines, the number of knots is not important given that it is above some minimum depending upon the degree of the smoothness; see Li and Ruppert (2008). For bivariate PLS splines, Lai and Wang (2013) and Wang et al. (2017) also observed that the number of triangles K is not very critical, provided K is larger than some threshold. In fact, one of the main advantages of using PLS splines over unpenalized splines is the flexibility of choosing knots in the 1-D setting and choosing triangles in the 2-D setting. For unpenalized splines, one has to have large enough sample according to the requirement of the degree of splines on each subinterval in the 1-D case or each triangle in the 2-D case to guarantee that a solution can be found. However, there is no such requirement for PLS splines. When the smoothness $r \geq 1$, the only requirement for bivariate PLS splines is that there is at least one triangle containing three points which are not in one line (Lai, 2008). Also, PLS splines perform similarly to unpenalized splines as long as the penalty parameter λ is very small. So in summary, the proposed bivariate PLS splines are very flexible and convenient for data fitting, even for smoothing sparse and unevenly sampled data over a domain with complicated boundary.

In practice, to form a good triangulation, we need to make certain that the triangulation is sufficiently fine to capture the feature in the dataset and not so large that computational burden is unnecessarily heavy. Wang et al. (2017) proposed to choose the number of triangles by generalized cross-validation (GCV) (Craven and Wahba (1979); Wahba (1990)). As suggested by Wang et al. (2017), we consider a sequence of trial values of the number of vertices of the triangles "equally-spaced" on the domain, and apply the Delaunay triangulation method. The more vertices we insert, the finer the triangulation. For each trial value, the PLS spline is fitted, and the value in that trial sequence that minimizes the GCV is selected. Wang et al. (2017) provides extensive numerical studies to illustrate the practical performance of the GCV triangulation selection scheme.

S.2 More Simulation Results

S.2.1 Additional simulation result from Example 1

Figure S.1 shows the estimated functions over a grid of 500×200 points via different methods for replicate 1 with $\rho = 0.7$. From those plots, it is clear that the BPST and GLTPS estimates perform better than the other four estimates. There seems to be some "leakage effect" in KRIG and TPS estimates, which is likely caused by the fact that KRIG and TPS do not take the complex boundary into any account and smooth across the gap inappropriately. Finally, as what we expected that the BPST estimators based on the three different triangulations are very similar, which confirms that the number of triangles is not very critical for the penalized spline fitting as long as it is sufficiently large enough to capture the pattern and features of the data.



(a) True Contour (b) KRIG (c) TPS (d) GLTPS

(e) FEM (f) BPST ($\triangle_1$) (g) BPST ($\triangle_2$) (h) BPST ($\triangle_3$)

Figure S.1: Contour maps for the true function and its estimators ($\rho = 0.7$).

S.2.2 A new simulation example

In this example, we consider a rectangular domain, $[0, 1]^2$, where there is no irregular shape or complex boundaries problem. In this case, classical methods for spatial data analysis, such as KRIG and TPS, will not encounter any difficulty.

We obtain the true signal and noisy observation for each coordinate pair lying on a 101×101-grid over $[0,1]^2$ using the following model:

$$Y = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(X_1, X_2) + \epsilon,$$

where $\boldsymbol{\beta} = (-1, 1)^{\mathrm{T}}$ and $g(x_1, x_2) = 10\{(x_1-0.5)^2+(x_2-0.5)^2\}$. The random error, ϵ, is generated from an $N(0, \sigma_\epsilon^2)$ distribution with $\sigma_\epsilon = 0.5$. Similar to Example 1, we simulate $Z_1 \sim uniform[-1, 1]$, and $Z_2 = \cos[4\pi(\rho(X_1^2 + X_2^2) + (1 - \rho)U)]$, where $\rho = 0.0$ or 0.7, $U \sim uniform[-1, 1]$ and is independent from (X_1, X_2) and Z_1. Next we take 100 Monte Carlo random samples of size $n = 200$ from the 101×101 points.

Figure S.2 (a) and (b) display the true quadratic surface and the contour map, respectively. We use the triangulation in Figure S.2 (e) and (f), and there are 8 triangles and 9 vertices as well as 18 triangles and 16 vertices, respectively. In addition, the points in Figure S.2 (d) demonstrate the sampled location points of replicate 100.

We compare the proposed BPST estimator with estimators from the KRIG, TPS, LFE methods, which are implemented in the same way as in Section 4. To see the accuracy of the estimators, we compute the RMSEs of the coefficient estimators and the estimator of σ_ϵ. To see the overall prediction accuracy, we make prediction on the 101×101 grid points on the domain for each replication using different methods, and compare the predicted values with the true function of $g(\cdot)$ at these grid points, and we report the average mean squared prediction errors (MSPE) over all replications.

All the results are summarized in Table S.1. As expected, KRIG and TPS work pretty well since the domain is regular in this example. In both scenarios, BPST performs the best. One also notices that, compared with the FEM, our BPST estimator shows much better performance in terms of both estimation and prediction, because BPST provides a more flexible and easier construction of splines with piecewise polynomials of various degrees and smoothness than the FEM method. As pointed out in Wood et al. (2008), the FEM method may require a very fine triangulation in order to reach certain approximation power, however, BPST doesn't need such a strict fineness requirement as it uses piecewise polynomials of higher degree yielding an larger order approximation power.

Figures S.3 and S.4 show the estimated functions via different methods for the last replicate. Compare with the true function in Figure S.2, the BPST estimate looks visually better than the other estimates. In addition, from Figures S.3 and

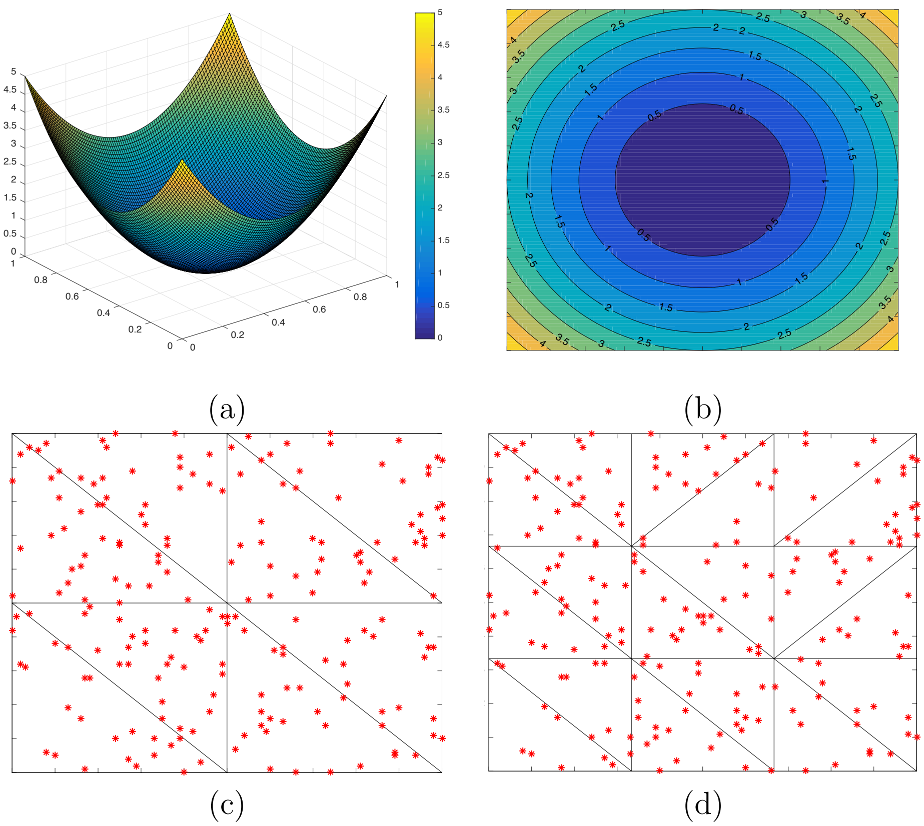


Figure S.2: (a) true function of $g(\cdot)$; (b) contour map of $g(\cdot)$; (c) first triangulation ($\triangle_1$); and (d) second triangulation ($\triangle_2$) on the domain.

S.4, one also sees that the BPST estimators based on $\triangle_1$ and $\triangle_2$ are very similar, which agrees our findings for penalized splines. In summary, Monte Carlo experiment in this study also shows that once the minimum necessary number of triangles has been reached for BPST, further increasing of the number of triangles usually have little effect on the fitting process.

Table S.1: Root mean squared errors of the estimates.

	Method	β_1	β_2	σ_ε	$g(\cdot)$
	KRIG	0.0640	0.0557	0.0369	0.1797
	TPS	0.0647	0.0551	0.0286	0.1640
0.0	LFE	0.0772	0.0604	0.0669	0.2978
	BPST($\triangle_1$)	0.0642	0.0546	0.0266	0.1495
	BPST($\triangle_2$)	0.0640	0.0556	0.0273	0.1395
	KRIG	0.0647	0.0530	0.0365	0.1800
	TPS	0.0653	0.0515	0.0281	0.1640
0.7	LFE	0.0769	0.0607	0.0668	0.2978
	BPST($\triangle_1$)	0.0645	0.0513	0.0263	0.1497
	BPST($\triangle_2$)	0.0644	0.0512	0.0265	0.1476

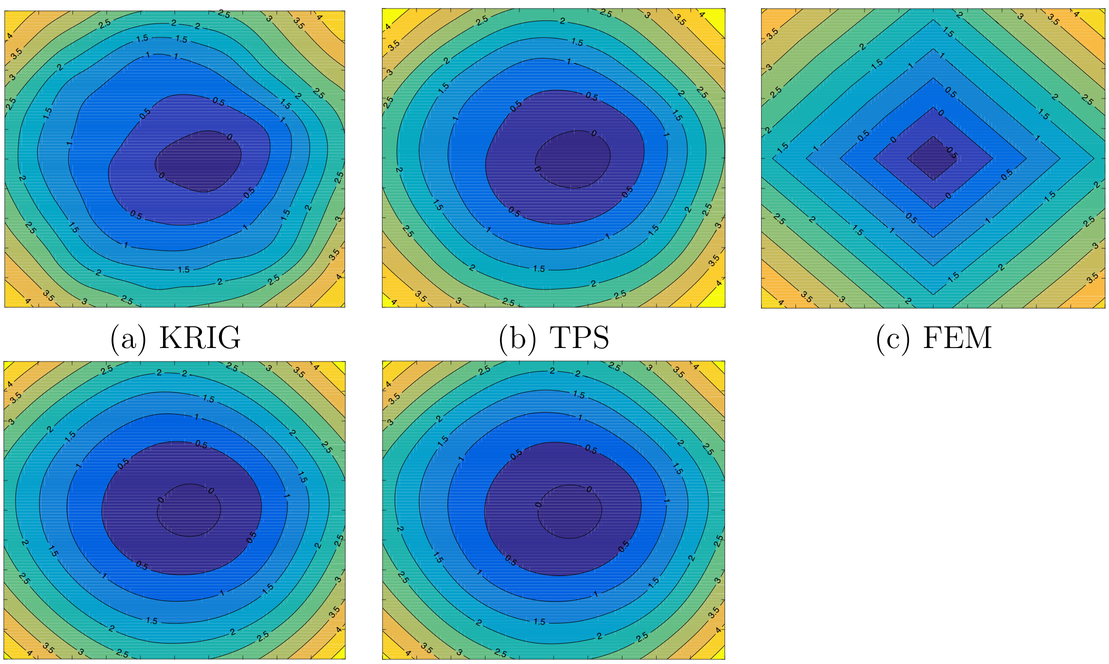
Table S.2 lists the accuracy results of the standard error formula in (18) for $\hat{\beta}_1$ and $\hat{\beta}_2$ using BPST with triangulation $\triangle_1$. From Table S.2, one sees that the estimated standard errors based on sample size $n = 200$ are very accurate.

Table S.2: Standard error estimates of the BPST coefficients.

ρ	Parameter	$\mathrm{SE}_{\mathrm{mc}}$	$\mathrm{SE}_{\mathrm{mean}}$	$\mathrm{SE}_{\mathrm{median}}$	$\mathrm{SE}_{\mathrm{mad}}$
0.0	β_1	0.0643	0.0622	0.0621	0.0032
	β_2	0.0546	0.0517	0.0516	0.0028
0.7	β_1	0.0645	0.0621	0.0622	0.0030
	β_2	0.0515	0.0519	0.0518	0.0026

S.3 Residual Plots from Mercury Concentration Studies

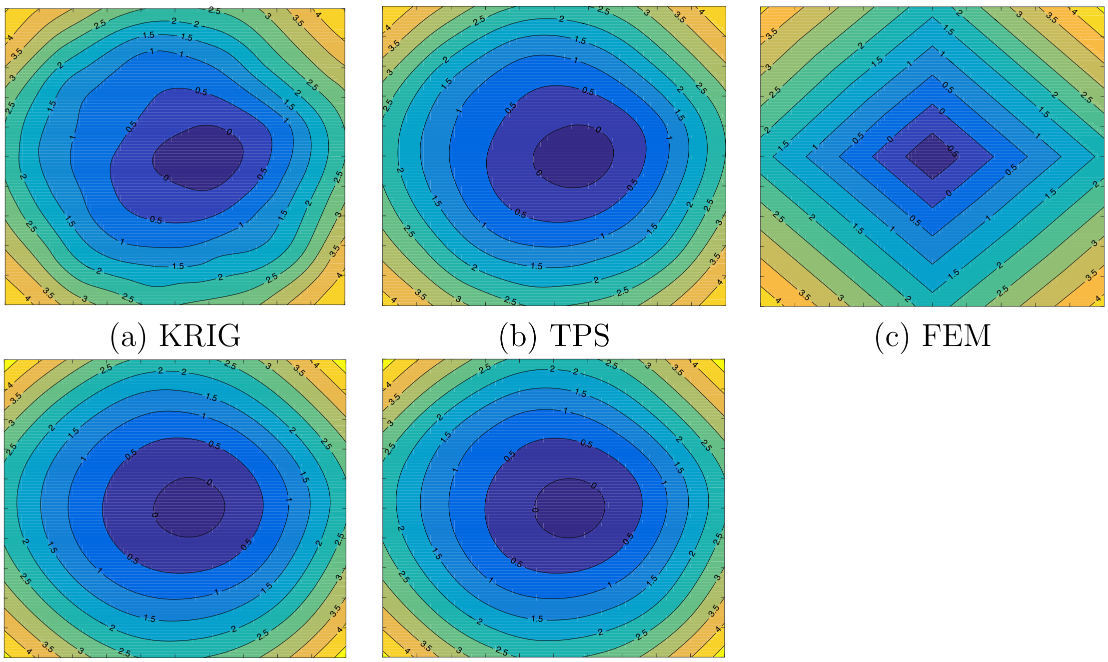
In this section, we provide some diagnosis plots of the residuals. Figure S.5 provides the residuals vs fits plots for five different methods. From Figure S.5, one sees that



(a) KRIG (b) TPS (c) FEM

(d) BPST ($\triangle_1$) (e) BPST ($\triangle_2$)

Figure S.3: Contour maps for the estimators ($\rho = 0.0$).



(a) KRIG (b) TPS (c) FEM

(d) BPST ($\triangle_1$) (e) BPST ($\triangle_2$)

Figure S.4: Contour maps for the estimators ($\rho = 0.7$).

the residuals "bounce randomly" around the zero line, and no residual "stands out" from the basic random pattern of residuals. Figure S.6 further demonstrates the residual scatter plot using five different methods. As seen in Figure S.6, the absolute values of the residuals are relatively higher in the middle of the Piscataqua river for KRIG and TPS compared to that of the BPST. Due to the small sample size and the complex terrain, all methods have some difficulty in the estimation at the confluence of the Salmon Falls River and Cocheco River. According to Steve et al. (2009), the accumulation of mercury in this area is complex and includes aspects of transport from urban point sources, atmospheric deposition from local and distant sources, prevailing currents, equilibrium processes between overlying water and the quality of sediments. Further research is warranted.

S.4 Technical Lemmas

In the following, we use c, C, c_1, c_2, C_1, C_2, etc. as generic constants, which may be different even in the same line. For functions f_1 and f_2 on $\Omega \times \mathbb{R}^p$, we define the empirical inner product and norm as $\langle f_1, f_2 \rangle_n = \frac{1}{n}\sum_{i=1}^n f_1(\mathbf{X}_i, \mathbf{Z}_i) f_2(\mathbf{X}_i, \mathbf{Z}_i)$ and $\|f_1\|_n^2 = \langle f_1, f_1 \rangle_n$. If f_1 and f_2 are L^2-integrable, we define the theoretical inner product and theoretical L^2 norm as $\langle f_1, f_2 \rangle_{L^2} = E\left\{f_1(\mathbf{X}_i, \mathbf{Z}_i) f_2(\mathbf{X}_i, \mathbf{Z}_i)\right\}$ and $\|f_1\|_{L^2} = \langle f_1, f_1 \rangle_{L^2}$. Furthermore, let $\|\cdot\|_{\mathcal{E}_v}$ be the norm introduced by the inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}_v}$, where, for g_1 and g_2 on Ω,

$$\langle g_1, g_2 \rangle_{\mathcal{E}_v} = \int_\Omega \left\{ \sum_{i+j=v} \binom{v}{i} (D_{x_1}^i D_{x_2}^j g_1)^2 \right\}^{1/2} \left\{ \sum_{i+j=v} \binom{v}{i} (D_{x_1}^i D_{x_2}^j g_2)^2 \right\}^{1/2} dx_1 dx_2.$$

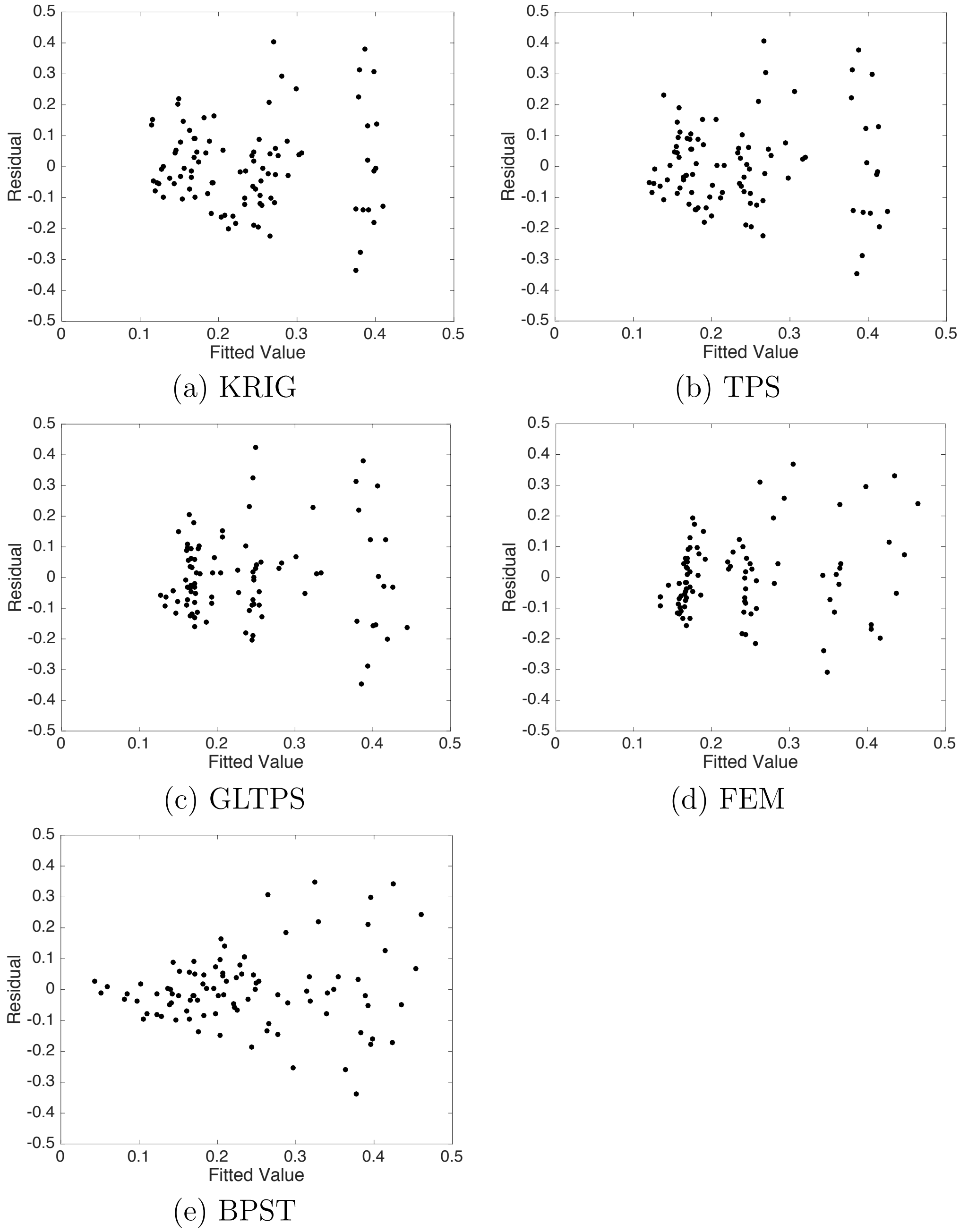
Proof of Lemma 1. By (7), we have $\mathbf{H}^{\mathrm{T}} = \mathbf{Q}_1 \mathbf{R}_1$ since $\mathbf{R}_2 = \mathbf{0}$. That is, $\mathbf{H} = \mathbf{R}_1^{\mathrm{T}} \mathbf{Q}_1^{\mathrm{T}}$. Thus,

$$\mathbf{H}\boldsymbol{\gamma} = \mathbf{H}\mathbf{Q}_2\boldsymbol{\theta} = \mathbf{R}_1^{\mathrm{T}} \mathbf{Q}_1^{\mathrm{T}} \mathbf{Q}_2 \boldsymbol{\theta} = \mathbf{0}$$

since $\mathbf{Q}_1^{\mathrm{T}} \mathbf{Q}_2 = \mathbf{0}$. On the other hand, if

$$\mathbf{0} = \mathbf{H}\boldsymbol{\gamma} = \mathbf{R}_1^{\mathrm{T}} \mathbf{Q}_1^{\mathrm{T}} \boldsymbol{\gamma},$$

we have $\mathbf{Q}_1^{\mathrm{T}} \boldsymbol{\gamma} = 0$ since $\mathbf{R}_1$ is invertible. Thus, $\boldsymbol{\gamma}$ is in the perpendicular subspace of the space spanned by the columns of $\mathbf{Q}_1$. That is, $\boldsymbol{\gamma}$ is in the space spanned by the columns of $\mathbf{Q}_2$. Thus, there exists a vector $\boldsymbol{\theta}$ such that $\boldsymbol{\gamma} = \mathbf{Q}_2 \boldsymbol{\theta}$. These complete the proof. □



(a) KRIG

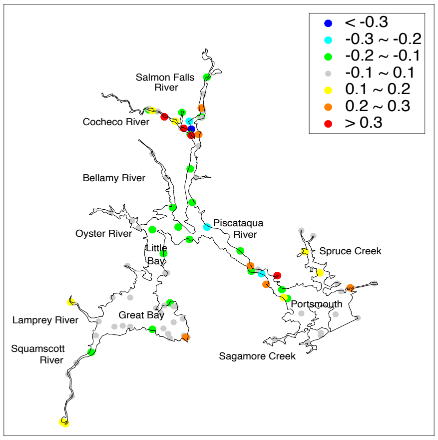
(b) TPS

(c) GLTPS

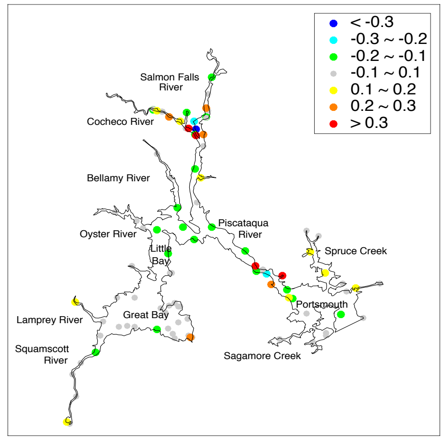
(d) FEM

(e) BPST

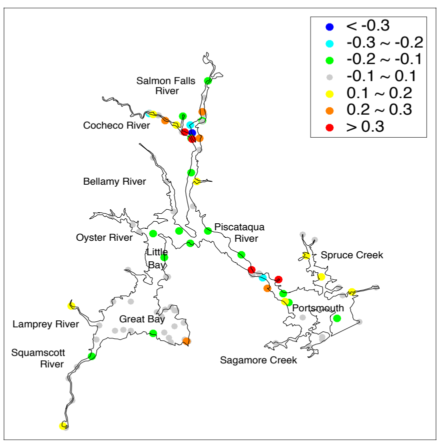
Figure S.5: Plots of the residuals vs fitted values of mercury concentrations.



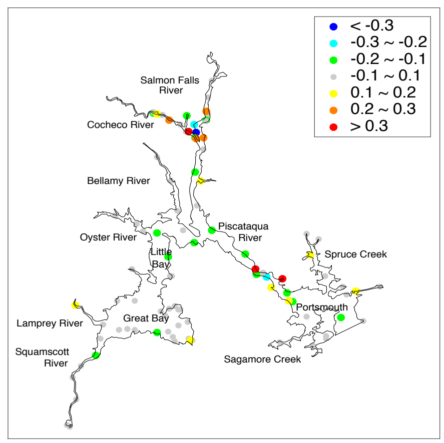
(a) KRIG



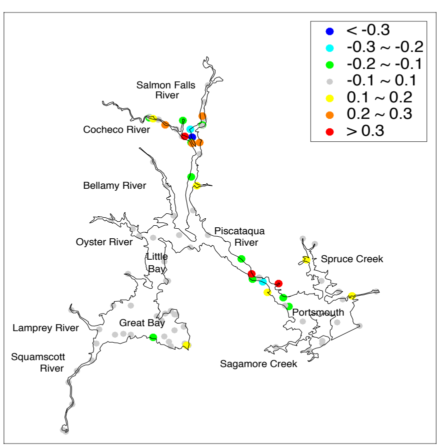
(b) TPS



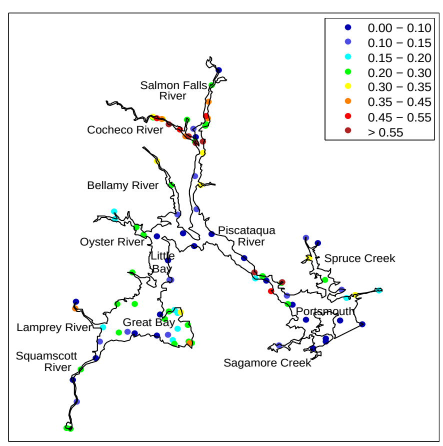
(c) GLTPS



(d) FEM



(e) BPST



(f) Observed Data

Figure S.6: Residual maps of mercury concentrations over the estuaries in New Hampshire.

LEMMA **S.1.** *[Lai and Schumaker (2007)] Let $\{B_\xi\}_{\xi\in\mathcal{K}}$ be the Bernstein polynomial basis for spline space $\mathbb{S}$ with smoothness r, where $\mathcal{K}$ stands for an index set. Then there exist positive constants c, C depending on the smoothness r and the shape parameter δ in Condition (C6) such that*

$$c|\triangle|^2 \sum_{\xi\in\mathcal{K}} |\gamma_\xi|^2 \leq \| \sum_{\xi\in\mathcal{K}} \gamma_\xi B_\xi \|_{L^2}^2 \leq C|\triangle|^2 \sum_{\xi\in\mathcal{K}} |\gamma_\xi|^2$$

for all $\gamma_\xi, \xi \in \mathcal{K}$.

With the above stability condition, Lai and Wang (2013) established the following uniform rate at which the empirical inner product approximates the theoretical inner product.

LEMMA **S.2.** *[Lemma 2 of the Supplement of Lai and Wang (2013)] Let $g_1 = \sum_{\xi\in\mathcal{K}} c_\xi B_\xi$, $g_2 = \sum_{\zeta\in\mathcal{K}} \widetilde{c}_\zeta B_\zeta$ be any spline functions in $\mathbb{S}$. Under Conditions (C4) and (C6),*

$$\sup_{g_1,g_2\in\mathbb{S}} \left| \frac{\langle g_1, g_2\rangle_n - \langle g_1, g_2\rangle_{L^2}}{\|g_1\|_{L^2} \|g_2\|_{L^2}} \right| = O_P\left\{(N\log n)^{1/2}/n^{1/2}\right\}.$$

For any smooth bivariate function $g(\cdot)$ and $\lambda > 0$, define

$$s_{\lambda,g} = \mathrm{argmin}_{s\in\mathbb{S}} \sum_{i=1}^{n} \{g(\mathbf{X}_i) - s(\mathbf{X}_i)\}^2 + \lambda\mathcal{E}_\upsilon(s) \tag{S.1}$$

the penalized least squares splines of $g(\cdot)$. Then the non-penalized solution $s_{0,g}$ is the discrete least squares spline estimator of $g(\cdot)$.

LEMMA **S.3.** *[Corollary of Theorem 6 in Lai (2008)] Assume $g(\cdot)$ is in Sobolev space $W^{\ell+1,\infty}(\Omega)$. For bi-integer (α_1, α_2) with $0 \leq \alpha_1 + \alpha_2 \leq \upsilon$, there exists an absolute constant C depending on r and δ, such that with probability approaching 1,*

$$\|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - s_{0,g})\|_\infty \leq C\frac{F_2}{F_1}|\triangle|^{\ell+1-\alpha_1-\alpha_2}|g|_{\ell+1,\infty},$$

where F_2 appears in Assumption (C3) and $F_1 > 0$ is a constant in a different version of Assumption C2 (Lai, 2008).

We remark that the current version of Assumption (C2) is an improvement of the original Assumption (C2). The improvement requires an extensive study. We leave it to a future publication.

LEMMA **S.4.** *Suppose $g(\cdot)$ is in the Sobolev space $W^{\ell+1,\infty}(\Omega)$, and let $s_{\lambda,g}$ be its penalized spline estimator defined in (S.1). Under Conditions (C2), (C3) and (C6),*

$$\begin{aligned}\|g - s_{\lambda,g}\|_n &= O_P\left\{\frac{F_2}{F_1}|\triangle|^{\ell+1}|g|_{\ell+1,\infty}\right.\\ &\qquad\left.+\frac{\lambda}{n|\triangle|^2}\left(|g|_{v,\infty}+\frac{F_2}{F_1}|\triangle|^{\ell+1-v}|g|_{\ell+1,\infty}\right)\right\}.\end{aligned}$$

Proof. Note that $s_{\lambda,g}$ is characterized by the orthogonality relations

$$n\langle g - s_{\lambda,g}, u\rangle_n = \lambda\langle s_{\lambda,g}, u\rangle_{\mathcal{E}_v}, \quad \text{for all } u \in \mathbb{S}, \tag{S.2}$$

while $s_{0,g}$ is characterized by

$$\langle g - s_{0,g}, u\rangle_n = 0, \quad \text{for all } u \in \mathbb{S}. \tag{S.3}$$

By (S.2) and (S.3), $n\langle s_{0,g} - s_{\lambda,g}, u\rangle_n = \lambda\langle s_{\lambda,g}, u\rangle_{\mathcal{E}_v}$, for all $u \in \mathbb{S}$. Replacing u by $s_{0,g} - s_{\lambda,g}$ yields that

$$n\|s_{0,g} - s_{\lambda,g}\|_n^2 = \lambda\langle s_{\lambda,g}, s_{0,g} - s_{\lambda,g}\rangle_{\mathcal{E}_v}. \tag{S.4}$$

Thus, by Cauchy-Schwarz inequality,

$$\begin{aligned}n\|s_{0,g} - s_{\lambda,g}\|_n^2 &\le \lambda\|s_{\lambda,g}\|_{\mathcal{E}_v}\|s_{0,g} - s_{\lambda,g}\|_{\mathcal{E}_v}\\ &\le \lambda\|s_{\lambda,g}\|_{\mathcal{E}_v}\sup_{f\in\mathbb{S}}\left\{\frac{\|f\|_{\mathcal{E}_v}}{\|f\|_n}, \|f\|_n \ne 0\right\}\|s_{0,g} - s_{\lambda,g}\|_n.\end{aligned}$$

Similarly, using (S.4), we have

$$n\|s_{0,g} - s_{\lambda,g}\|_n^2 = \lambda\left\{\langle s_{\lambda,g}, s_{0,g}\rangle_{\mathcal{E}_v} - \langle s_{\lambda,g}, s_{\lambda,g}\rangle_{\mathcal{E}_v}\right\} \ge 0.$$

Therefore, by Cauchy-Schwarz inequality,

$$\|s_{\lambda,g}\|_{\mathcal{E}_v}^2 \le \langle s_{\lambda,g}, s_{0,g}\rangle_{\mathcal{E}_v} \le \|s_{\lambda,g}\|_{\mathcal{E}_v}\|s_{0,g}\|_{\mathcal{E}_v},$$

which implies that $\|s_{\lambda,g}\|_{\mathcal{E}_v} \le \|s_{0,g}\|_{\mathcal{E}_v}$. Therefore,

$$\|s_{0,g} - s_{\lambda,g}\|_n \le n^{-1}\lambda\|s_{0,g}\|_{\mathcal{E}_v}\sup_{f\in\mathbb{S}}\left\{\frac{\|f\|_{\mathcal{E}_v}}{\|f\|_n}, \|f\|_n \ne 0\right\}.$$

By Lemma S.3, with probability approaching 1,

$$
\begin{aligned}
\|s_{0,g}\|_{\mathcal{E}_v} &\leq C_1 A_\Omega \left\{ |g|_{v,\infty} + \sum_{\alpha_1+\alpha_2=v} \left\| D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - s_{0,g}) \right\|_\infty \right\} \\
&\leq C_2 A_\Omega \left(|g|_{v,\infty} + \frac{F_2}{F_1} |\triangle|^{\ell+1-v} |g|_{\ell+1,\infty} \right), \qquad \text{(S.5)}
\end{aligned}
$$

where A_Ω denotes the area of Ω. By Markov's inequality, for any $f \in \mathbb{S}$, $\|f\|_{\mathcal{E}_v} \leq C |\triangle|^{-2} \|f\|_{L^2}$. Lemma (S.2) implies that

$$
\sup_{f \in \mathcal{S}} \{ \|f\|_n / \|f\|_{L^2} \} \geq 1 - O_P \left\{ (N \log n)^{1/2} / n^{1/2} \right\}.
$$

Thus, we have

$$
\begin{aligned}
\sup_{f \in \mathbb{S}} \left\{ \frac{\|f\|_{\mathcal{E}_v}}{\|f\|_n}, \|f\|_n \neq 0 \right\} &\leq C |\triangle|^{-2} \left[1 - O_P \left\{ (N \log n)^{1/2} / n^{1/2} \right\} \right]^{-1/2} \\
&= O_P \left(|\triangle|^{-2} \right). \qquad \text{(S.6)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\|s_{0,g} - s_{\lambda,g}\|_n &= O_P \left\{ \frac{\lambda}{n |\triangle|^2} \left(|g|_{v,\infty} + \frac{F_2}{F_1} |\triangle|^{\ell+1-v} |g|_{\ell+1,\infty} \right) \right\}, \\
\|g - s_{\lambda,g}\|_n &\leq \|g - s_{0,g}\|_n + \|s_{0,g} - s_{\lambda,g}\|_n .
\end{aligned}
$$

By Lemma S.3,

$$
\|g - s_{0,g}\|_n \leq \|g - s_{0,g}\|_\infty = O_P \left(\frac{F_2}{F_1} |\triangle|^{\ell+1} |g|_{\ell+1,\infty} \right).
$$

Thus, the desired result is established. □

LEMMA **S.5.** *Under Assumptions (A1), (A2), (C4)-(C6), there exist constants* $0 < c_U < C_U < \infty$*, such that with probability approaching 1 as* $n \to \infty$*,* $c_U \mathbf{I}_{p\times p} \leq n\mathbf{U}_{11} \leq C_U \mathbf{I}_{p\times p}$*, where* $\mathbf{U}_{11}$ *is given in (9).*

Proof. Denote by

$$
\mathbf{\Gamma}_\lambda = \frac{1}{n} \left(\mathbf{B}^{\mathrm{T}} \mathbf{B} + \lambda \mathbf{P} \right) = \left[\frac{1}{n} \sum_{i=1}^n B_\xi (\mathbf{X}_i) B_\zeta (\mathbf{X}_i) + \frac{\lambda}{n} \langle B_\xi, B_\zeta \rangle_{\mathcal{E}_v} \right]_{\xi,\zeta \in \mathcal{K}}
$$

a symmetric positive definite matrix. Then for $\mathbf{V}_{22}$ defined in (8), we can rewrite it as $\mathbf{V}_{22} = n\mathbf{Q}_2^{\mathrm{T}}\boldsymbol{\Gamma}_\lambda\mathbf{Q}_2$. Let $\alpha_{\min}(\lambda)$ and $\alpha_{\max}(\lambda)$ be the smallest and largest eigenvalues of $\boldsymbol{\Gamma}_\lambda$. As shown in the proof of Theorem 2 in the Supplement of Lai and Wang (2013), there exist positive constants $0 < c_3 < C_3$ such that under Conditions (C4) and (C5), with probability approaching 1, we have

$$c_3|\triangle|^2 \leq \alpha_{\min}(\lambda) \leq \alpha_{\max}(\lambda) \leq C_3\left(|\triangle|^2 + \frac{\lambda}{n|\triangle|^2}\right).$$

Therefore, we have

$$c_4\left(|\triangle|^2 + \frac{\lambda}{n|\triangle|^2}\right)^{-1}\|\mathbf{a}\|^2 \leq n\mathbf{a}^{\mathrm{T}}\mathbf{V}_{22}^{-1}\mathbf{a} = \mathbf{a}^{\mathrm{T}}(\mathbf{Q}_2^{\mathrm{T}}\boldsymbol{\Gamma}_\lambda\mathbf{Q}_2)^{-1}\mathbf{a} \leq C_4|\triangle|^{-2}\|\mathbf{a}\|^2.$$

Thus, by Assumption (A2), we have with probability approaching 1

$$\begin{aligned} c_5\left(|\triangle|^2 + \frac{\lambda}{n|\triangle|^2}\right)^{-1}|\triangle|^2\|\mathbf{a}\|^2 &\leq \mathbf{a}^{\mathrm{T}}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{a} \\ &= \mathbf{a}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{a} \leq C_5\|\mathbf{a}\|^2. \quad \text{(S.7)} \end{aligned}$$

According to (9) and (10), we have

$$(n\mathbf{U}_{11})^{-1} = n^{-1}(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}) = n^{-1}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}).$$

The desired result follows from Assumptions (A1), (A2) and (S.7). □

S.5 Proof of Theorem 1

Let $\mu_i = \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + g_0(\mathbf{X}_i)$, $\boldsymbol{\mu}^{\mathrm{T}} = (\mu_1, \ldots, \mu_n)$, and let $\boldsymbol{\epsilon}^{\mathrm{T}} = (\epsilon_1, \ldots, \epsilon_n)$. Define

$$\widetilde{\boldsymbol{\beta}}_\mu = \mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right)\boldsymbol{\mu}, \quad \text{(S.8)}$$

$$\widetilde{\boldsymbol{\beta}}_\epsilon = \mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right)\boldsymbol{\epsilon}. \quad \text{(S.9)}$$

Then $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = (\widetilde{\boldsymbol{\beta}}_\mu - \boldsymbol{\beta}_0) + \widetilde{\boldsymbol{\beta}}_\epsilon$.

LEMMA **S.6.** *Under Assumptions (A1), (A2), (C1)-(C5),* $\|\widetilde{\boldsymbol{\beta}}_\mu - \boldsymbol{\beta}_0\| = o_P\left(n^{-1/2}\right)$ *for* $\widetilde{\boldsymbol{\beta}}_\mu$ *in (S.8).*

Proof. Let $\mathbf{g}_0 = (g_0(\mathbf{X}_1), \ldots, g_0(\mathbf{X}_n))^{\mathrm{T}}$. It is clear that

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}_\mu - \boldsymbol{\beta}_0 &= \mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right)\mathbf{g}_0 \\
&= \mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left[\mathbf{g}_0 - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{B} + \lambda\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{g}_0\right] \\
&= n\mathbf{U}_{11}\mathbf{A},
\end{aligned}$$

where $\mathbf{A} = (A_1, \ldots, A_p)^{\mathrm{T}}$, with

$$A_j = n^{-1}\mathbf{Z}_j^{\mathrm{T}}\left[\mathbf{g}_0 - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{B} + \lambda\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{g}_0\right]$$

for $\mathbf{Z}_j^{\mathrm{T}} = (Z_{1j}, ..., Z_{nj})$. Next we derive the order of A_j, $1 \le j \le p$, as follows. For any $g_j \in \mathbb{S}$, by (S.2) we have

$$A_j = \langle z_j, g_0 - s_{\lambda,g_0}\rangle_n = \langle z_j - g_j, g_0 - s_{\lambda,g_0}\rangle_n + \frac{\lambda}{n}\langle s_{\lambda,g_0}, g_j\rangle_{\mathcal{E}_v}.$$

For any $j = 1, \ldots, p$, let $h_j(\cdot)$ be the function $h(\cdot)$ that minimizes $E\{Z_{ij} - h(\mathbf{X}_i)\}^2$ as defined in (14). According to Lemma S.3, there exists a function $\widetilde{h}_j \in \mathbb{S}$ satisfy

$$\|\widetilde{h}_j - h_j\|_\infty \le C\frac{F_2}{F_1}|\triangle|^{\ell+1}|h_j|_{\ell+1,\infty}, \tag{S.10}$$

then

$$A_j = \langle z_j - h_j, g_0 - s_{\lambda,g_0}\rangle_n + \langle h_j - \widetilde{h}_j, g_0 - s_{\lambda,g_0}\rangle_n + \frac{\lambda}{n}\langle s_{\lambda,g_0}, \widetilde{h}_j\rangle_{\mathcal{E}_v} = A_{j,1} + A_{j,2} + A_{j,3}.$$

Since h_j satisfies $\langle z_j - h_j, \psi\rangle_{L_2(\Omega)} = 0$ for any $\psi \in L_2(\Omega)$, $E\left(A_{j,1}\right) = 0$. According to Proposition 1 in Lai and Wang (2013),

$$\begin{aligned}
\|g_0 - s_{\lambda,g_0}\|_\infty &= O_P\left\{\frac{F_2}{F_1}|\triangle|^{\ell+1}|g_0|_{\ell+1,\infty}\right. \\
&\quad \left. + \frac{\lambda}{n|\triangle|^3}\left(|g_0|_{2,\infty} + \frac{F_2}{F_1}|\triangle|^{\ell-1}|g_0|_{\ell+1,\infty}\right)\right\}.
\end{aligned}$$

Next,

$$\begin{aligned}
\mathrm{Var}\left(A_{j,1}\right) &= \frac{1}{n^2}\sum_{i=1}^n E\left[\{Z_{ij} - h_j(\mathbf{X}_i)\}(g_0 - s_{\lambda,g_0})\right]^2 \\
&\le \frac{\|g_0 - s_{\lambda,g_0}\|_\infty^2}{n}\|z_j - h_j\|_{L^2}^2,
\end{aligned}$$

which together with $E\left(A_{j,1}\right)=0$ implies that

$$\begin{aligned}|A_{j,1}| &= O_P\left\{\frac{F_2}{n^{1/2}F_1}\left|\triangle\right|^{\ell+1}|g_0|_{\ell+1,\infty}\right.\\&\left.\quad+\frac{\lambda}{n^{3/2}\left|\triangle\right|^3}\left(|g_0|_{2,\infty}+\frac{F_2}{F_1}\left|\triangle\right|^{\ell-1}|g_0|_{\ell+1,\infty}\right)\right\}.\end{aligned}$$

Cauchy-Schwartz inequality, Lemma S.4 and (S.10) imply that

$$\begin{aligned}|A_{j,2}| &\leq \|h_j-\widetilde{h}_j\|_n\,\|g_0-s_{\lambda,g_0}\|_n\\&= O_P\left(\frac{F_2}{F_1}\left|\triangle\right|^{\ell+1}|h_j|_{\ell+1,\infty}\right)\times O_P\left\{\frac{F_2}{F_1}\left|\triangle\right|^{\ell+1}|g_0|_{\ell+1,\infty}\right.\\&\left.\quad+\frac{\lambda}{n\left|\triangle\right|^2}\left(|g_0|_{2,\infty}+\frac{F_2}{F_1}\left|\triangle\right|^{\ell-1}|g_0|_{\ell+1,\infty}\right)\right\}.\end{aligned}$$

Finally, by (S.5), we have

$$\begin{aligned}|A_{j,3}| &\leq \frac{\lambda}{n}\|s_{\lambda,g_0}\|_{\mathcal{E}_v}\|\widetilde{h}_j\|_{\mathcal{E}_v}\leq\frac{\lambda}{n}\|s_{0,g_0}\|_{\mathcal{E}_v}\|\widetilde{h}_j\|_{\mathcal{E}_v}\\&\leq \frac{\lambda}{n}C_1\left(|g_0|_{2,\infty}+\frac{F_2}{F_1}\left|\triangle\right|^{\ell-1}|g_0|_{\ell+1,\infty}\right)\\&\quad\times\left(|h_j|_{2,\infty}+\frac{F_2}{F_1}\left|\triangle\right|^{\ell-1}|h_j|_{\ell+1,\infty}\right).\end{aligned}$$

Combining all the above results yields that

$$\begin{aligned}|A_j| &= O_P\left[n^{-1/2}\left\{\frac{F_2}{F_1}\left|\triangle\right|^{\ell+1}|g_0|_{\ell+1,\infty}+\frac{\lambda}{n\left|\triangle\right|^3}\left(|g_0|_{2,\infty}\right.\right.\right.\\&\left.\left.\left.\quad+\frac{F_2}{F_1}\left|\triangle\right|^{\ell-1}|g_0|_{\ell+1,\infty}\right)\right\}\right]\end{aligned}$$

for $j=1,\ldots,p$. By Assumptions (C3)-(C5), $|A_j|=o_P(n^{-1/2})$, for $j=1,\ldots,p$. In addition, we have $nU_{11}=O_P(1)$ according to Lemma S.5. Therefore, $\|\widetilde{\boldsymbol{\beta}}_\mu-\boldsymbol{\beta}_0\|=o_P\left(n^{-1/2}\right)$. □

Lemma **S.7.** *Under Assumptions (A1)-(A3) and (C1)-(C6), as $n\rightarrow\infty$,*

$$\left[\mathrm{Var}\left(\widetilde{\boldsymbol{\beta}}_\epsilon\left|\left\{\left(\mathbf{X}_i,\mathbf{Z}_i\right),i=1,\ldots,n\right\}\right.\right)\right]^{-1/2}\widetilde{\boldsymbol{\beta}}_\epsilon\longrightarrow N\left(0,\mathbf{I}_{p\times p}\right),$$

where $\widetilde{\boldsymbol{\beta}}_\epsilon$ is given in (S.9).

Proof. Note that

$$\widetilde{\boldsymbol{\beta}}_\epsilon = \mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right)\boldsymbol{\epsilon}.$$

For any $\mathbf{b} \in \mathbb{R}^p$ with $\|\mathbf{b}\| = 1$, we can write $\mathbf{b}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_\epsilon = \sum_{i=1}^{n} \alpha_i \epsilon_i$, where

$$\alpha_i^2 = n^{-2}\mathbf{b}^{\mathrm{T}}(n\mathbf{U}_{11})\left(\mathbf{Z}_i^{\mathrm{T}} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}_i\right)\left(\mathbf{Z}_i - \mathbf{B}_i^{\mathrm{T}}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\right)(n\mathbf{U}_{11})\mathbf{b},$$

and conditioning on $\{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \ldots, n\}$, $\alpha_i\epsilon_i$'s are independent. By Lemma S.5, we have

$$\max_{1\le i\le n} \alpha_i^2 \le Cn^{-2} \max_{1\le i\le n}\left\{\|\mathbf{Z}_i\|^2 + \left\|\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}_i\right\|^2\right\},$$

where for any $\mathbf{a} \in \mathbb{R}^p$,

$$\mathbf{a}^{\mathrm{T}}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}_i\mathbf{a} = n^{-1}\mathbf{a}^{\mathrm{T}}\mathbf{V}_{12}(\mathbf{Q}_2^{\mathrm{T}}\boldsymbol{\Gamma}_\lambda\mathbf{Q}_2)^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}_i\mathbf{a} \le Cn^{-1}|\triangle|^{-2}\mathbf{a}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}_i\mathbf{a},$$

and the j-th component of $n^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}_i$ is $\frac{1}{n}\sum_{i'=1}^{n} Z_{i'j}\sum_{\xi\in\mathcal{K}} B_\xi(\mathbf{X}_{i'}) B_\xi(\mathbf{X}_i)$. Using Assumptions (A1) and (A2), we have

$$E\left\{\frac{1}{n}\sum_{i'=1}^{n} Z_{i'j}\sum_{\xi\in\mathcal{K}} B_\xi(\mathbf{X}_{i'}) B_\xi(\mathbf{X}_i)\right\}^2 = O(1),$$

for large n, thus with probability approaching 1,

$$\max_{1\le i\le n}\left|\frac{1}{n}\sum_{i'=1}^{n}\sum_{\xi\in\mathcal{K}} Z_{i'j}B_\xi(\mathbf{X}_{i'}) B_\xi(\mathbf{X}_i)\right| = O_P(1),$$
$$\max_{1\le i\le n}\|\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}_i\|^2 = O_P(|\triangle|^{-2}).$$

Therefore, $\max_{1\le i\le n}\alpha_i^2 = O_P\left(n^{-2}|\triangle|^{-2}\right)$. Next, with probability approaching 1,

$$\begin{aligned}
\sum_{i=1}^{n}\alpha_i^2 &= \mathrm{Var}\left[\mathbf{b}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_\epsilon\,|\{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \ldots, n\}\right] \\
&= \mathbf{b}^{\mathrm{T}}\mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right)\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right)\mathbf{Z}\mathbf{U}_{11}\mathbf{b}\sigma^2 \\
&= n^{-1}\mathbf{b}^{\mathrm{T}}(n\mathbf{U}_{11})\left\{n^{-1}\sum_{i=1}^{n}(\mathbf{Z}_i - \widehat{\mathbf{Z}}_i)(\mathbf{Z}_i - \widehat{\mathbf{Z}}_i)^{\mathrm{T}}\right\}(n\mathbf{U}_{11})\mathbf{b}\sigma^2, \quad \text{(S.11)}
\end{aligned}$$

where $\widehat{\mathbf{Z}}_i$ is the i-th column of $\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}$. Using Lemma S.5 again, we have $\sum_{i=1}^n \alpha_i^2 \geq cn^{-1}$. So $\max_{1\leq i\leq n} \alpha_i^2/\sum_{i=1}^n \alpha_i^2 = O_P\left(n^{-1}|\triangle|^{-2}\right) = o_P\left(1\right)$ from Assumption (C4). By Linderberg-Feller CLT, we have

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

Then the desired result follows. □

For any $j = 1,\ldots,p$ and $\lambda > 0$, define

$$s_{\lambda,z_j} = \mathrm{argmin}_{s\in\mathbb{S}} \sum_{i=1}^n \{z_j(\mathbf{X}_i) - s(\mathbf{X}_i)\}^2 + \lambda\mathcal{E}_v(s), \tag{S.12}$$

where z_j is the coordinate mapping that maps $\mathbf{z}$ to its j-th component.

Lemma **S.8.** *Under Assumptions (A1), (A2), (C2), (C3) and (C6), for s_{λ,z_j} defined in (S.12), $\left\|s_{0,z_j} - s_{\lambda,z_j}\right\|_n = O_P(\lambda n^{-1}|\triangle|^{-5})$, $j = 1,\ldots,p$.*

Proof. Note that

$$n\left\langle z_j - s_{\lambda,z_j}, u\right\rangle_n = \lambda\left\langle s_{\lambda,z_j}, u\right\rangle_{\mathcal{E}_v}, \quad \left\langle z_j - s_{0,z_j}, u\right\rangle_n = 0, \quad \text{for all } u\in\mathcal{S},$$

Inserting $u = s_{0,z_j} - s_{\lambda,z_j}$ in the above yields that

$$n\left\|s_{0,z_j} - s_{\lambda,z_j}\right\|_n^2 = \lambda\left\langle s_{\lambda,z_j}, s_{0,z_j} - s_{\lambda,z_j}\right\rangle_{\mathcal{E}_v} = \lambda(\langle s_{\lambda,z_j}, s_{0,z_j}\rangle_{\mathcal{E}_v} - \langle s_{\lambda,z_j}, s_{\lambda,z_j}\rangle_{\mathcal{E}_v}).$$

By Cauchy-Schwarz inequality, $\left\|s_{\lambda,z_j}\right\|_{\mathcal{E}_v}^2 \leq \left\langle s_{\lambda,z_j}, s_{0,z_j}\right\rangle_{\mathcal{E}_v} \leq \left\|s_{\lambda,z_j}\right\|_{\mathcal{E}_v}\left\|s_{0,z_j}\right\|_{\mathcal{E}_v}$, which implies

$$\left\|s_{\lambda,z_j}\right\|_{\mathcal{E}_v} \leq \left\|s_{0,z_j}\right\|_{\mathcal{E}_v}. \tag{S.13}$$

By (S.6), we have for large n

$$n\left\|s_{0,z_j} - s_{\lambda,z_j}\right\|_n^2 \leq \lambda\left\|s_{\lambda,z_j}\right\|_{\mathcal{E}_v}\left\|s_{0,z_j} - s_{\lambda,z_j}\right\|_n \times O_P(|\triangle|^{-2}),$$

thus, $\left\|s_{0,z_j} - s_{\lambda,z_j}\right\|_n \leq \left\|s_{0,z_j}\right\|_{\mathcal{E}_v} \times O_P(\lambda n^{-1}|\triangle|^{-2})$. Markov's inequality implies that

$$\left\|s_{0,z_j}\right\|_{\mathcal{E}_v} \leq \frac{C_1}{|\triangle|^2}\left\|s_{0,z_j}\right\|_\infty. \tag{S.14}$$

Note that $\left\|s_{0,z_j}\right\|_\infty \le C|\triangle|^{-2}\max_{\xi\in\mathcal{K}}|n^{-1}\sum_{i=1}^n B_\xi(\mathbf{X}_i)Z_{ij}|$ with probability approaching one. According to Assumptions (A1) and (A2),

$$\max_{\xi\in\mathcal{K}}\left|n^{-1}\sum_{i=1}^n B_\xi(\mathbf{X}_i)Z_{ij}\right| = O_P(|\triangle|).$$

The desired results follows. □

Lemma **S.9.** *Under Assumptions (A1)-(A3) and (C1)-(C6), for the covariance matrix* $\boldsymbol{\Sigma}$ *defined in (17), we have* $c_\Sigma^*\mathbf{I}_p \le \boldsymbol{\Sigma} \le C_\Sigma^*\mathbf{I}_p$, *and*

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

Proof. According to (S.11),

$$\mathrm{Var}\left(\widetilde{\boldsymbol{\beta}}_\epsilon\,|\{(\mathbf{X}_i,\mathbf{Z}_i)\}\right) = n^{-1}(n\mathbf{U}_{11})\left\{n^{-1}\sum_{i=1}^n(\mathbf{Z}_i-\widehat{\mathbf{Z}}_i)(\mathbf{Z}_i-\widehat{\mathbf{Z}}_i)^{\mathrm{T}}\right\}(n\mathbf{U}_{11})\sigma^2.$$

By the definition of $\mathbf{U}_{11}^{-1}$ in (10), we have

$$(n\mathbf{U}_{11})^{-1} = \frac{1}{n}\sum_{i=1}^n\mathbf{Z}_i(\mathbf{Z}_i-\widehat{\mathbf{Z}}_i)^{\mathrm{T}} = \left(\langle z_j, z_{j'} - s_{\lambda,z_{j'}}\rangle_n\right)_{1\le j,j'\le p}.$$

As in the proof of Lemma S.6, let $\widetilde{h}_j\in\mathbb{S}$ and h_j satisfy (S.10). Then,

$$\langle z_j, z_{j'} - s_{\lambda,z_{j'}}\rangle_n = \langle z_j - \widetilde{h}_j, z_{j'} - s_{\lambda,z_{j'}}\rangle_n + \frac{\lambda}{n}\langle s_{\lambda,z_{j'}},\widetilde{h}_j\rangle_{\mathcal{E}_v}. \tag{S.15}$$

According to (S.5), (S.13) and (S.14), we have

$$\begin{aligned}\left|\langle s_{\lambda,z_{j'}},\widetilde{h}_{j'}\rangle_{\mathcal{E}_v}\right| &\le \|s_{\lambda,z_{j'}}\|_{\mathcal{E}_v}\|\widetilde{h}_{j'}\|_{\mathcal{E}_v} \le \|s_{0,z_{j'}}\|_{\mathcal{E}_v}\|\widetilde{h}_{j'}\|_{\mathcal{E}_v}\\ &\le \frac{C}{|\triangle|^2}\|s_{0,z_{j'}}\|_\infty\|\widetilde{h}_{j'}\|_{\mathcal{E}_v}\\ &\le \frac{CC^*}{|\triangle|^3}\left(\left|h_j'\right|_{2,\infty} + \frac{F_2}{F_1}|\triangle|^{\ell+1-v}\left|h_j'\right|_{\ell+1,\infty}\right).\end{aligned}$$

Note that

$$\langle z_j-\widetilde{h}_j, z_{j'}-s_{\lambda,z_{j'}}\rangle_n = \langle z_j-h_j, z_{j'}-h_{j'}\rangle_n + \langle h_j-\widetilde{h}_j, h_{j'}-\widetilde{h}_{j'}\rangle_n + \langle z_j-h_j, h_{j'}-\widetilde{h}_{j'}\rangle_n$$

$$+\langle h_j - \widetilde{h}_j, z_{j'} - h_{j'}\rangle_n + \langle z_j - h_j, \widetilde{h}_{j'} - s_{\lambda,z_{j'}}\rangle_n + \langle h_j - \widetilde{h}_j, \widetilde{h}_{j'} - s_{\lambda,z_{j'}}\rangle_n. \quad \text{(S.16)}$$

By (S.10), the second term on the right side of (S.16) satisfies that

$$\left|\langle h_j - \widetilde{h}_j, h_{j'} - \widetilde{h}_{j'}\rangle_\infty\right| \le \|h_j - \widetilde{h}_j\|_\infty \|h_{j'} - \widetilde{h}_{j'}\|_\infty = o_P(1).$$

By Lemma S.2 and (S.10), the third term on the right side of (S.16) satisfies that

$$\left|\langle z_j - h_j, h_{j'} - \widetilde{h}_{j'}\rangle_n\right| \le \{\|z_j - h_j\|_{L^2}(1 + o_P(1))\} \|h_{j'} - \widetilde{h}_{j'}\|_\infty = o_P(1).$$

Similarly, $\left|\langle h_j - \widetilde{h}_j, z_{j'} - h_{j'}\rangle_n\right| = o_P(1)$. From the triangle inequality, we have

$$\|\widetilde{h}_j - s_{\lambda,z_j}\|_n \le \|\widetilde{h}_j - h_j\|_n + \|h_j - s_{0,z_j}\|_n + \|s_{0,z_j} - s_{\lambda,z_j}\|_n.$$

According to (S.10) and Lemma S.8, we have

$$\|\widetilde{h}_j - s_{\lambda,z_j}\|_n \le \|h_j - s_{0,z_j}\|_n + o_P(1).$$

Define $h^*_{j,n} = \mathrm{argmin}_{h\in\mathbb{S}}\|z_j - h\|_{L^2}$, then, from the triangle inequality,

$$\|h_j - s_{0,z_j}\|_n \le \|h_j - h^*_{j,n}\|_n + \|h^*_{j,n} - s_{0,z_j}\|_n$$

Note that $\|h_j - h^*_{j,n}\|_{L^2} = o_P(1)$. Lemma S.2 implies that $\|h_j - h^*_{j,n}\|_n = o_P(1)$. Next note that $\|s_{0,z_j} - h^*_{j,n}\|^2_{L^2} = \|z_j - s_{0,z_j}\|^2_{L^2} - \|z_j - h^*_{j,n}\|^2_{L^2}$ and $\|z_j - s_{0,z_j}\|_n \le \|z_j - h^*_{j,n}\|_n$. Using Lemma S.2 again, we have

$$\|s_{0,z_j} - h^*_{j,n}\|^2_{L^2} = o_P(\|z_j - h^*_{j,n}\|^2_{L^2}) + o_P(\|z_j - s_{0,z_j}\|^2_{L^2}).$$

Since there exists a constant C such that $\|z_j - h^*_{j,n}\|_{L^2} \le C$, so we have

$$\|z_j - s_{0,z_j}\|_{L^2} \le \|z_j - h^*_{j,n}\|_{L^2} + \|h^*_{j,n} - s_{0,z_j}\|_{L^2} \le C + \|h^*_{j,n} - s_{0,z_j}\|_{L^2}.$$

Therefore, $\|h^*_{j,n} - s_{0,z_j}\|_{L^2} = o_P(1)$. Lemma S.2 implies that $\|h^*_{j,n} - s_{0,z_j}\|_n = o_P(1)$. As a consequence,

$$\|s_{0,z_j} - h_j\|_n = o_P(1). \quad \text{(S.17)}$$

For the fifth item, by Lemma S.2 and (S.17), we have

$$\begin{aligned}\left|\langle z_j - h_j, \widetilde{h}_{j'} - s_{\lambda,z_{j'}}\rangle_n\right| &\le \{\|z_j - h_j\|_{L^2}(1 + o_P(1))\}\left\{\|h_j - s_{0,z_j}\|_n + o_P(1)\right\} \\ &= o_P(1).\end{aligned}$$

Similarly, for the sixth item, we have

$$\left|\langle h_j - \widetilde{h}_j, \widetilde{h}_{j'} - s_{\lambda,z_{j'}}\rangle_n\right| \leq \|h_j - \widetilde{h}_j\|_n \left\{\|h_j - s_{0,z_j}\|_n + o_P(1)\right\} = o_P(1). \quad \text{(S.18)}$$

Combining the above results from (S.15) to (S.18) gives that

$$\langle z_j, z_{j'} - s_{\lambda,z_{j'}}\rangle_n = \langle z_j - h_j, z_{j'} - h^*_{j'}\rangle_n + o_P(1).$$

Therefore,

$$(n\mathbf{U}_{11})^{-1} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)^{\mathrm{T}} + o_P(1) = E[(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)^{\mathrm{T}}] + o_P(1).$$

Hence,

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

S.6 Proof of Theorem 2

Let $\mathbf{H}_\mathbf{Z} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$, then

$$\widehat{\boldsymbol{\theta}} = \mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{H}_\mathbf{Z}\mathbf{Y} = \mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{H}_\mathbf{Z}\mathbf{g}_0 + \mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{H}_\mathbf{Z}\boldsymbol{\epsilon} = \widetilde{\boldsymbol{\theta}}_\mu + \widetilde{\boldsymbol{\theta}}_\epsilon.$$

According to Lemma S.3, $\|s_{0,g_0} - g_0\|_\infty \leq C\frac{F_2}{F_1}|\triangle|^{\ell+1}|g_0|_{\ell+1,\infty}$. Denote by $\boldsymbol{\gamma}_0 = \mathbf{Q}_2\boldsymbol{\theta}_0$ the spline coefficients of s_{0,g_0}. Then we have the following decomposition: $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \widetilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}_0 + \widetilde{\boldsymbol{\theta}}_\epsilon$. Note that

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}_0 &= \mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{H}_\mathbf{Z}\mathbf{g}_0 - \boldsymbol{\theta}_0 \\ &= \mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{H}_\mathbf{Z}(\mathbf{g}_0 - \mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}_0) - \lambda\mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}_0.\end{aligned}$$

According to (10), for any $\mathbf{a}$

$$\mathbf{a}^{\mathrm{T}}\mathbf{U}_{22}^{-1}\mathbf{a} = \mathbf{a}^{\mathrm{T}}\mathbf{Q}_2^{\mathrm{T}}\left(\mathbf{B}^{\mathrm{T}}\mathbf{H}_\mathbf{Z}\mathbf{B} + \lambda\mathbf{P}\right)\mathbf{Q}_2\mathbf{a}.$$

Since $\mathbf{H}_\mathbf{Z}$ is idempotent, so its eigenvalues π_j is either 0 or 1. Without loss of generality we can arrange the eigenvalues in decreasing order so that $\pi_j = 1$, $j = 1, \ldots, m$ and $\pi_j = 1$, $j = m+1, \ldots, n$. Therefore, we have

$$\mathbf{a}^{\mathrm{T}}(n\mathbf{U}_{22})^{-1}\mathbf{a} = \frac{1}{n}\sum_{j=1}^{m}\pi_j\mathbf{a}^{\mathrm{T}}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\mathbf{B}\mathbf{Q}_2\mathbf{a} + \frac{\lambda}{n}\mathbf{a}^{\mathrm{T}}\mathbf{Q}_2^{\mathrm{T}}\mathbf{P}\mathbf{Q}_2\mathbf{a},$$

where $\mathbf{e}_j$ be the indicator vector which is a zero vector except for an entry of one at position j. Using Markov's inequality, we have

$$\frac{\lambda}{n}\mathcal{E}_v\left(\sum_{\xi\in\mathcal{K}} a_\xi B_\xi\right) \leq \frac{\lambda}{n}\frac{C_1}{|\triangle|^2}C_2\|\mathbf{a}\|^2.$$

Thus, by Conditions (C4) and (C5), $n\mathbf{a}^{\mathrm{T}}\mathbf{U}_{22}\mathbf{a} \leq C|\triangle|^{-2}$. Next

$$\begin{aligned}\|\mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{H}_{\mathbf{Z}}(\mathbf{g}_0-\mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}_0)\| &\leq C^{1/2}|\triangle|^{-1}n^{-1}\|\mathbf{B}^{\mathrm{T}}\mathbf{H}_{\mathbf{Z}}(\mathbf{g}_0-\mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}_0)\| \\ &\leq C^{1/2}|\triangle|^{-1}n^{-1}\left[\sum_{\xi\in\mathcal{K}}\{\mathbf{B}_\xi^{\mathrm{T}}\mathbf{H}_{\mathbf{Z}}(\mathbf{g}_0-\mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}_0)\}^2\right]^{1/2} \\ &= O_P\left(\frac{F_2}{F_1}|\triangle|^{\ell}|g_0|_{\ell+1,\infty}\right),\end{aligned}$$

and

$$\lambda\|\mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}_0\| \leq \frac{C\lambda}{n|\triangle|^4}\|s_{0,g_0}\|_{\mathcal{E}_v} \leq \frac{C\lambda}{n|\triangle|^4}\left(|g_0|_{2,\infty}+\frac{F_2}{F_1}|\triangle|^{\ell-1}|g_0|_{\ell+1,\infty}\right).$$

Thus,

$$\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\| = O_P\left\{\frac{\lambda}{n\left|\triangle\right|^4}|g_0|_{2,\infty}+\left(1+\frac{\lambda}{n\left|\triangle\right|^5}\right)\frac{F_2}{F_1}\left|\triangle\right|^{\ell}\left|g_0\right|_{\ell+1,\infty}\right\}.$$

For any $\boldsymbol{\alpha}$ with $\|\boldsymbol{\alpha}\|=1$, we write $\boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\boldsymbol{\theta}}_\epsilon = \sum_{i=1}^n \alpha_i\epsilon_i$ and

$$\alpha_i^2 = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{H}_{\mathbf{Z}}\mathbf{B}\mathbf{Q}_2\mathbf{U}_{22}\boldsymbol{\alpha}.$$

Following the same arguments as those in Lemma S.7, we have $\max_{1\leq i\leq n}\alpha_i^2 = O_P(n^{-2}|\triangle|^{-4})$. Thus,

$$\|\widetilde{\boldsymbol{\theta}}_\epsilon\| \leq |\triangle|^{-1}|\boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\boldsymbol{\theta}}_\epsilon| = |\triangle|^{-1}|\sum_{i=1}^n\alpha_i\epsilon_i| = O_P(|\triangle|^{-2}n^{-1/2}).$$

Therefore,

$$\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\| = O_P\left\{\frac{\lambda}{n\left|\triangle\right|^4}|g_0|_{2,\infty}+\left(1+\frac{\lambda}{n\left|\triangle\right|^5}\right)\frac{F_2}{F_1}\left|\triangle\right|^{\ell}\left|g_0\right|_{\ell+1,\infty}+\frac{1}{\sqrt{n}|\triangle|^2}\right\}.$$

Observing that $\widehat{g}(\mathbf{x}) = \mathbf{B}(\mathbf{x})\widehat{\boldsymbol{\gamma}} = \mathbf{B}(\mathbf{x})\mathbf{Q}_2\widehat{\boldsymbol{\theta}}$, we have

$$\|\widehat{g} - g_0\|_{L^2} \leq \|\widehat{g} - s_{0,g_0}\|_{L^2} + C\frac{F_2}{F_1}|\triangle|^{\ell+1}|g_0|_{\ell+1,\infty}.$$

According to Lemma S.1, we have.

$$\begin{aligned} \|\widehat{g} - g_0\|_{L^2} \ \leq \ & C\left(|\triangle|\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| + \frac{F_2}{F_1}|\triangle|^{\ell+1}|g_0|_{\ell+1,\infty}\right) = O_P\left\{\frac{\lambda}{n\,|\triangle|^3}|g_0|_{2,\infty}\right. \\ & \left. + \left(1 + \frac{\lambda}{n\,|\triangle|^5}\right)\frac{F_2}{F_1}\,|\triangle|^{\ell+1}\,|g_0|_{\ell+1,\infty} + \frac{1}{\sqrt{n}|\triangle|}\right\}. \end{aligned}$$

The proof is completed.

Bibliography

Craven, P. and Wahba, G. (1979), "Smoothing noisy data with spline functions," *Numerische Mathematik*, 31, 377–403.

Lai, M. J. (2008), "Multivariate splines for data fitting and approximation," *Conference Proceedings of the 12th Approximation Theory,*, 210–228.

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

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

Li, Y. and Ruppert, D. (2008), "On the asymptotics of penalized splines," *Biometrika*, 95, 415–436.

Persson, P. O. and Strang, G. (2004), "A simple mesh generator in MATLAB," *SIAM Review*, 46, 329–345.

Shewchuk, J. R. (1996), "Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator," in *Applied Computational Geometry: Towards Geometric Engineering*, eds. Lin, M. C. and Manocha, D., Springer-Verlag, vol. 1148 of *Lecture Notes in Computer Science*, pp. 203–222, from the First ACM Workshop on Applied Computational Geometry.

Steve, J., Christian, K., and Gareth, H. (2009), "Distribution of mercury and trace metals in shellfish and sediments in the Gulf of Maine," *Symposcience*, online available at http://symposcience.lyon.cemagref.fr/exl–doc/colloque/ART–00002561.pdf.

Wahba, G. (1990), *Spline models for observational data*, SIAM Publications, Philadelphia.

Wang, G., Mu, J., and Wang, L. (2017), "Efficient smoothing parameter and triangulation selection for bivariate penalized spline regression," *Manuscript*, online available at http://people.wm.edu/ gwang01/triangulation.pdf.

Wood, S. N., Bravington, M. V., and Hedley, S. L. (2008), "Soap film smoothing," *Journal of the Royal Statistical Society, Series B*, 70, 931–955.