Statistica Sinica Preprint No: SS-2017-0243	
Title	Efficient Estimation of Partially Linear Models for Data on Complicated Domains by Bivariate Penalized Splines over Triangulations
Manuscript ID	SS-2017-0243
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0243
Complete List of Authors	Lily Wang Guannan Wang Ming-Jun Lai and Lei Gao
Corresponding Author	Lily Wang
E-mail	lilywang@iastate.edu

Statistica Sinica **30** (2020), 1-23
doi:https://doi.org/10.5705/ss.202017.0243

EFFICIENT ESTIMATION OF PARTIALLY LINEAR MODELS FOR DATA ON COMPLICATED DOMAINS BY BIVARIATE PENALIZED SPLINES OVER TRIANGULATIONS

Li Wang[1], Guannan Wang[2], Ming-Jun Lai[3] and Lei Gao[1]

[1]*Iowa State University,* [2]*College of William & Mary and* [3]*The University of Georgia*

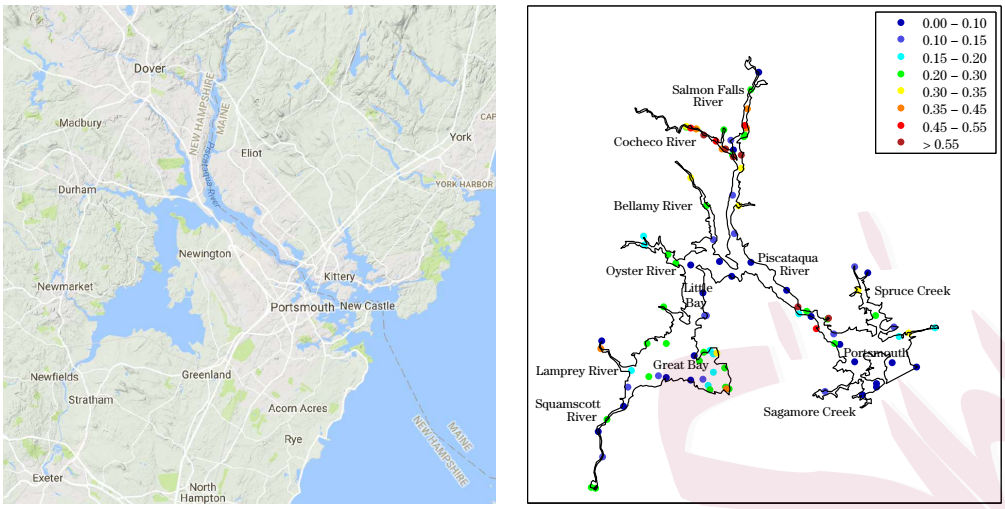
Abstract: In this study, we consider the estimation of partially linear models for spatial data distributed over complex domains. We use bivariate splines over triangulations to represent the nonparametric component on an irregular two-dimensional domain. The proposed method is formulated as a constrained minimization problem that does not require constructing finite elements or locally supported basis functions. Thus, it allows an easier implementation of piecewise polynomial representations of various degrees and various smoothness over an arbitrary triangulation. Moreover, the constrained minimization problem is converted to an unconstrained minimization using a QR decomposition of the smoothness constraints, enabling us to develop a fast and efficient penalized least squares algorithm for fitting the model. The estimators of the parameters are proved to be asymptotically normal under some regularity conditions. The estimator of the bivariate function is consistent, and its rate of convergence is also established. The proposed method enables us to construct confidence intervals and permits inferences for the parameters. The performance of the estimators is evaluated using two simulation examples and a real-data analysis.

Key words and phrases: Bivariate splines, penalty, semiparametric regression, spatial data, triangulation.

1. Introduction

In many geospatial studies, spatially distributed covariate information is available. For example, geographic information systems may contain measurements obtained from satellite images at some locations. These spatially explicit data can be useful in the construction and estimation of regression models. However, the domain over which the variables of interest are defined is often complicated, such as stream networks, islands, and mountains. For example, Figure 1 (a) and (b) show the largest estuary in New Hampshire, together with the loca-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243



(a) (b)

Figure 1. Regional map of estuaries. Dots in (b) represent sample locations.

tion of 97 sites where the mercury in sediment concentrations was surveyed in 2000, 2001, and 2003; see Wang and Ranalli (2007). It is well known that many conventional smoothing tools with respect to the Euclidean distance between observations suffer from the problem of "leakage" across complex domains. This refers to poor estimations over difficult regions as a result of the inappropriate linking of the parts of the domain separated by physical barriers; for more excellent discussions, see Ramsay (2002) and Wood, Bravington and Hedley (2008). Here, we propose using bivariate splines (smooth piecewise polynomial functions over a triangulation of the domain of interest) to model spatially explicit data sets, enabling us to overcome the "leakage" problem and provide a more accurate estimation and prediction.

Here, we focus on the partially linear model (Speckman (1988); He and Shi (1996); Mammen and van de Geer (1997); Liang, Härdle and Carroll (1999); Härdle, Liang and Gao (2000); Ma, Chiou and Wang (2006); Liang and Li (2009)), referred to as the PLM, for data randomly distributed over 2D domains. Specifically, let $\mathbf{X}_i = (X_{i1}, X_{i2})^{\mathrm{T}}$ be the location of the ith point, for $i = 1, \ldots, n$, that ranges over a bounded domain $\Omega \subseteq \mathbb{R}^2$ of arbitrary shape, for example, the domain of the estuaries in New Hampshire, shown in Figure 1. Let Y_i be the response variable and $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^{\mathrm{T}}$ be the predictors at location $\mathbf{X}_i$. Suppose that $\{(\mathbf{Z}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ satisfies the following model:

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

$$Y_i = \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta} + g\left(\mathbf{X}_i\right) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ are unknown parameters, $g(\cdot)$ is some unknown but smooth bivariate function, and ϵ_i, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) random noises with $E(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$. Each ϵ_i is independent of $\mathbf{X}_i$ and $\mathbf{Z}_i$. In many situations, our main interest is in estimating and making inferences for the regression parameters $\boldsymbol{\beta}$, which measures the effect of the covariate $\mathbf{Z}$ after adjusting for the location effect of $\mathbf{X}$.

If $g(\cdot)$ is a univariate function, model (1.1) becomes a typical PLM. In the past three decades, flexible and parsimonious PLMs have been studied extensively and used widely in many statistical applications, from biostatistics, econometrics, engineering to social science; see Chen, Liang and Wang (2011), Huang, Zhang and Zhou (2007), Liu, Wang and Liang (2011), Wang et al. (2011), Ma, Song and Wang (2013), Wang et al. (2014), Zhang, Cheng and Liu (2011) as examples of recent works on PLMs. When $g(\cdot)$ is a bivariate function, there are two popular estimation tools: bivariate P-splines (Marx and Eilers (2005)), and thin plate splines (Wood (2003)). Later, Xiao, Li and Ruppert (2013) proposed a sandwich smoother, which has a tensor product structure that simplifies an asymptotic analysis and can be computed efficiently. However, the application to spatial data analysis over complex domains has been hampered, owing to the scarcity of bivariate smoothing tools that are computationally efficient and theoretically reliable when solving the problem of "leakage" across the domain. Traditional smoothing methods in practical data analyses, such as kernel smoothing, wavelet-based smoothing, tensor product splines, and thin plate splines, usually perform poorly for such data, because they do not consider the shape of the domain and they smooth across concave boundary regions.

There are several challenges when going from rectangular domains to irregular domains with complex boundaries or holes. Recently, studies have examined smoothing over irregular domains, and significant progress has been made. To deal with irregular domains, Wang and Ranalli (2007) proposed replacing the Euclidean distance with the geodesic distance in the low-rank thin-plate spline smoothing method. To calculate the geodesic distances, a graph is constructed where each vertex is the location of an observation and is connected only to its k nearest neighbors. Floyd's algorithm is then used to find the shortest path through the graph. This algorithm has a computing complexity of $O(n^3)$, without considering the selection of the optimal k, which makes the approach costly for large data sets. In addition, their method involves computing the square

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

roots of matrices that are not guaranteed to be positive semi-definite. Ramsay (2002) suggested a penalized least squares approach with a Laplacian penalty, and transformed the problem to one of solving a system of partial differential equations (PDEs). Recently, Sangalli, Ramsay and Ramsay (2013) extended the method in Ramsay (2002) to PLMs, which allows for spatially distributed covariate information to be included in the models. The data smoothing problem in Sangalli, Ramsay and Ramsay (2013) is solved using the finite element method (FEM), a method mainly developed and used to solve PDEs. Although their method is useful in many practical applications, they did not investigate the theoretical properties of the estimation. In addition, our case study in Section 5 and simulation study in the Supplementary Material reveal that the FEM is not sufficiently flexible to estimate the functional part of the model well. Furthermore, Wood, Bravington and Hedley (2008) pointed out that the FEM method requires a very fine triangulation in order to reach a certain approximation power when the underlying function is complicated.

In this study, we tackle the estimation problem using the bivariate splines defined on triangulations (Awanou, Lai and Wenston (2005); Lai and Schumaker (2007)). Our approach is superior to the finite element method (Sangalli, Ramsay and Ramsay (2013)) in that we use spline functions with a more flexible degree and smoothness, enabling us to better approximate the bivariate function $g(\cdot)$. Another important feature of this approach is that it does not require the construction of locally supported splines or finite elements with degree greater than one.

To the best of our knowledge, the statistical aspects of smoothing for PLMs using bivariate splines have not been discussed in the literature. This study presents the first attempt at investigating the asymptotic properties of PLMs for data distributed on complex regions. We study the asymptotic properties of the least squares estimators of $\boldsymbol{\beta}$ and $g(\cdot)$ using bivariate splines defined on triangulations with a penalty term. We show that our estimator of $\boldsymbol{\beta}$ is root-n consistent and asymptotically normal, although the convergence rate of the estimator of the nonparametric component $g(\cdot)$ is slower than root n. A standard error formula for the estimated coefficients is provided and tested to be accurate for practical scenarios. Hence, the proposed method enables us to construct confidence intervals for the regression parameters. We also obtain the convergence rate for the estimator of $g(\cdot)$.

The rest of the paper is organized as follows. In Section 2, we give a brief review of the triangulations and propose our estimation method based on penal-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

ized bivariate splines. We also discuss how to choose the penalty parameters. Section 3 presents our the asymptotic analysis of the proposed estimators. Section 4 provides a detailed numerical study that compares several methods in two different scenarios and explores the estimation and prediction accuracy. In Section 5, we apply the proposed method to the mercury concentration study, where the variables of interest are defined over the estuary in New Hampshire depicted in Figure 1. Concluding remarks are provided in Section 6. All technical details are provided in the Supplementary Material.

2. Triangulations and Penalized Spline Estimators

Our estimation method is based on penalized bivariate splines on triangulations. The idea is to approximate the function $g(\cdot)$ using bivariate splines that are piecewise polynomial functions over a 2D triangulated domain, enabling us to fit $g(\cdot)$ more flexibly. We use this approximation to construct least squares estimators of the linear and nonlinear components of the model with a penalization term. In the remainder of this section, we describe the background of triangulations and B-form bivariate splines, and introduce the penalized spline estimators.

2.1. Triangulations

Triangulation is an effective strategy to handle data distributed over irregular regions with complex boundaries and/or interior holes, and has recently attracted substantial attention in many applied areas, such as geospatial studies, numerical solutions of PDEs, image enhancements, and computer aided geometric design. Many triangulation software packages are available. Section S1 of the Supplementary Material explains how to choose a triangulation for a given data set.

We use τ to denote a triangle that is a convex hull of three points that are not located in one line. A collection $\triangle = \{\tau_1, \ldots, \tau_N\}$ of N triangles is called a triangulation of $\Omega = \cup_{i=1}^{N} \tau_i$, provided that if a pair of triangles in $\triangle$ intersect, then their intersection is either a common vertex or a common edge. Although any kind of polygon shapes can be used for the partition of Ω, we use triangulations because any polygonal domain of arbitrary shape can be partitioned into finitely many triangles to form a triangulation $\triangle$. Given a triangle $\tau \in \triangle$, let $|\tau|$ be its longest edge length, and denote the size of $\triangle$ by $|\triangle| = \max\{|\tau|, \tau \in \triangle\}$, , that is, the length of the longest edge of $\triangle$.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

2.2. B-form bivariate splines

In this section, we briefly introduce bivariate splines. A more in-depth description can be found in Lai and Schumaker (2007), Lai (2008), and Zhou and Pan (2014); the details of the implementation are provided in Awanou, Lai and Wenston (2005). Let $\tau = \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \rangle$ be a nondegenerate (i.e., with nonzero area) triangle with vertices $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Then, for any point $\mathbf{v} \in \mathbb{R}^2$, there is a unique representation of the form $\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3$, with $b_1 + b_2 + b_3 = 1$, where b_1, b_2, and b_3 are the barycentric coordinates of the point $\mathbf{v}$ relative to the triangle τ. The Bernstein polynomials of degree d relative to triangle τ are defined as $B^{\tau,d}_{ijk}(\mathbf{v}) = (d!/i!j!k!)b_1^i b_2^j b_3^k$, for $i + j + k = d$. Then, for any $\tau \in \triangle$, we can write the polynomial piece of spline s restricted on $\tau \in \triangle$ as $s|_\tau = \sum_{i+j+k=d} \gamma^\tau_{ijk} B^{\tau,d}_{ijk}$, where $\boldsymbol{\gamma}_\tau = \{\gamma^\tau_{ijk}, i + j + k = d\}$ are called B-coefficients of s.

For a nonnegative integer r, let $\mathbb{C}^r(\Omega)$ be the collection of all rth continuously differentiable functions over Ω. Given a triangulation $\triangle$, let $\mathbb{S}^r_d(\triangle) = \{s \in \mathbb{C}^r(\Omega) : s|_\tau \in \mathbb{P}_d(\tau), \tau \in \triangle\}$ be a spline space of degree d and smoothness r over triangulation $\triangle$, where $\mathbb{P}_d$ is the space of all polynomials of degree less than or equal to d. Let $\mathbb{S} = \mathbb{S}^r_{3r+2}(\triangle)$ for a fixed smoothness $r \geq 1$. We know that such a spline space has the optimal approximation order (rate of convergence) for noise-free data sets; see Lai and Schumaker (1998) and Lai and Schumaker (2007).

For notational simplicity, let $\{B_\xi\}_{\xi \in \mathcal{K}}$ be the set of degree-d bivariate Bernstein basis polynomials for $\mathbb{S}$, where $\mathcal{K}$ denotes an index set of all Bernstein basis polynomials. Then, we can represent any function $s \in \mathbb{S}$ using the following basis expansion:

$$s(\mathbf{x}) = \sum_{\xi \in \mathcal{K}} B_\xi(\mathbf{x})\gamma_\xi = \mathbf{B}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\gamma}, \tag{2.1}$$

where $\boldsymbol{\gamma}^{\mathrm{T}} = (\gamma_\xi, \xi \in \mathcal{K})$ is the spline coefficient vector. To meet the smoothness requirement of the splines, we need to impose linear constraints on the spline coefficients $\boldsymbol{\gamma}$ in (2.1). We require that $\boldsymbol{\gamma}$ satisfies $\mathbf{H}\boldsymbol{\gamma} = 0$ with $\mathbf{H}$ being the matrix for all smoothness conditions across the shared edges of the triangles, which depends on the smoothness parameter r and the structure of the triangulation. See Zhou and Pan (2014) for examples of $\mathbf{H}$.

2.3. Penalized spline estimators

To define the penalized spline method, for any direction x_j, $j = 1, 2$, let $D^q_{x_j} f(\mathbf{x})$ denote the qth-order derivative in the direction x_j at the point $\mathbf{x} = (x_1, x_2)$. Let

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

$$\mathcal{E}_\upsilon(f) = \sum_{\tau\in\triangle}\int_\tau \sum_{i+j=\upsilon}\binom{\upsilon}{i}(D^i_{x_1}D^j_{x_2}f)^2dx_1dx_2 \tag{2.2}$$

be the energy functional for a fixed integer, $\upsilon \geq 1$ (Lai (2008)). Although all partial derivatives up to the chosen order υ can be included in (2.2), for simplicity, in the remaining part of the paper, we use $\upsilon = 2$. A similar problem can be studied for the more general case of $\upsilon \geq 2$. When $\upsilon = 2$,

$$\mathcal{E}_2(f) = \int_\Omega \left\{(D^2_{x_1}f)^2 + 2(D_{x_1}D_{x_2}f)^2 + (D^2_{x_2}f)^2\right\}dx_1dx_2, \tag{2.3}$$

which is similar to the thin-plate spline penalty (Green and Silverman (1994)), except that the latter is integrated over the entire plane $\mathbb{R}^2$. Sangalli, Ramsay and Ramsay (2013) used a different roughness penalty to that in (2.3). Specifically, they use the integral of the square of the Laplacian of f, that is, $\int_\Omega(D^2_{x_1}f + D^2_{x_2}f)^2dx_1dx_2$. Both forms of penalties are invariant with respect to the Euclidean transformations of the spatial coordinates, thus, the bivariate smoothing does not depend on the choice of the coordinate system.

Given $\lambda > 0$ and the data set $\{(\mathbf{Z}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, we consider the following minimization problem:

$$\min_{\boldsymbol{\beta}}\min_{s\in\mathbb{S}}\sum_{i=1}^n\left\{Y_i - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta} - s\left(\mathbf{X}_i\right)\right\}^2 + \lambda\mathcal{E}_\upsilon(s), \tag{2.4}$$

where $\mathbb{S}$ is a spline space over the triangulation $\triangle$ of Ω.

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ be a vector of n observations of the response variable, $\mathbf{X}_{n\times 2} = \{(X_{i1}, X_{i2})\}_{i=1}^n$ be a location design matrix, and $\mathbf{Z}_{n\times p} = \{(Z_{i1}, \ldots, Z_{ip})\}_{i=1}^n$ be a collection of all covariates. Denote by $\mathbf{B}$ the $n\times K$ evaluation matrix of Bernstein basis polynomials, with the ith row given by $\mathbf{B}_i^{\mathrm{T}} = \{B_\xi(\mathbf{X}_i), \xi \in \mathcal{K}\}$. Then, according to (2.1), $\{s(\mathbf{X}_i)\}_{i=1}^n$ can be written by $\mathbf{B}\boldsymbol{\gamma}$. Thus, the minimization in (2.4) reduces to

$$\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} L(\boldsymbol{\beta},\boldsymbol{\gamma}) = \min_{\boldsymbol{\beta},\boldsymbol{\gamma}}\left\{\|\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\gamma}\|^2 + \lambda\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{P}\boldsymbol{\gamma}\right\} \text{ subject to } \mathbf{H}\boldsymbol{\gamma} = \mathbf{0}, \tag{2.5}$$

where $\mathbf{P}$ is the block diagonal penalty matrix satisfying $\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{P}\boldsymbol{\gamma} = \mathcal{E}_\upsilon(\mathbf{B}\boldsymbol{\gamma})$.

To solve the constrained minimization problem (2.5), we first remove the constraint using a QR decomposition of the transpose of the constraint matrix $\mathbf{H}$. Specifically,

$$\mathbf{H}^{\mathrm{T}} = \mathbf{Q}\mathbf{R} = (\mathbf{Q}_1\ \mathbf{Q}_2)\begin{pmatrix}\mathbf{R}_1\\ \mathbf{0}\end{pmatrix} = \mathbf{Q}_1\mathbf{R}_1, \tag{2.6}$$

where $\mathbf{Q}$ is an orthogonal matrix, $\mathbf{R}_1$ is an upper triangle matrix, and the submatrix $\mathbf{Q}_1$ is the first r columns of $\mathbf{Q}$, where r is the rank of matrix $\mathbf{H}$. The

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

proof of the following lemma is provided in the Supplementary Material.

Lemma 1. *Let* $\mathbf{Q}_1, \mathbf{Q}_2$ *be submatrices, as in* (2.6)*. Let* $\boldsymbol{\gamma} = \mathbf{Q}_2\boldsymbol{\theta}$ *for a vector* $\boldsymbol{\theta}$ *of appropriate size. Then,* $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$*. On the other hand, if* $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$*, then there exists a vector* $\boldsymbol{\theta}$*, such that* $\boldsymbol{\gamma} = \mathbf{Q}_2\boldsymbol{\theta}$*.*

The problem in (2.5), now becomes a conventional penalized regression problem without any constraints:

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}} \left\{ \|\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta} - \mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}\|^2 + \lambda(\mathbf{Q}_2\boldsymbol{\theta})^{\mathrm{T}}\mathbf{P}(\mathbf{Q}_2\boldsymbol{\theta}) \right\}.$$

For a fixed penalty parameter λ, we have

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\theta}} \end{pmatrix} = \left\{ \begin{pmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{Z} & \mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{Q}_2 \\ \mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z} & \mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{Q}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \\ & \lambda\mathbf{Q}_2^{\mathrm{T}}\mathbf{P}\mathbf{Q}_2 \end{pmatrix} \right\}^{-1} \begin{pmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{Y} \\ \mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Y} \end{pmatrix}.$$

Letting

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{Z} & \mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{Q}_2 \\ \mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z} & \mathbf{Q}_2^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{B} + \lambda\mathbf{P})\mathbf{Q}_2 \end{pmatrix}, \tag{2.7}$$

we have

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\theta}} \end{pmatrix} = \mathbf{V}^{-1} \begin{pmatrix} \mathbf{Z}^{\mathrm{T}}\mathbf{Y} \\ \mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Y} \end{pmatrix}.$$

Next, we write

$$\mathbf{V}^{-1} \equiv \mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_{11} & -\mathbf{U}_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1} \\ -\mathbf{U}_{22}\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & \mathbf{U}_{22} \end{pmatrix}, \tag{2.8}$$

where

$$\begin{aligned} \mathbf{U}_{11}^{-1} &= \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} = \mathbf{Z}^{\mathrm{T}}\left[\mathbf{I} - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{B} + \lambda\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right]\mathbf{Z}, \\ \mathbf{U}_{22}^{-1} &= \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = \mathbf{Q}_2^{\mathrm{T}}\left[\mathbf{B}^{\mathrm{T}}\left\{\mathbf{I} - \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\right\}\mathbf{B} + \lambda\mathbf{P}\right]\mathbf{Q}_2. \end{aligned} \tag{2.9}$$

Then, the minimizers of (2.7) can be given precisely, as follows:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right)\mathbf{Y} \\ &= \mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left\{\mathbf{I} - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{B} + \lambda\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right\}\mathbf{Y}, \\ \widehat{\boldsymbol{\theta}} &= \mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}\mathbf{V}_{11}^{-1}\mathbf{Z}^{\mathrm{T}}\right)\mathbf{Y} = \mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\left\{\mathbf{I} - \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\right\}\mathbf{Y}. \end{aligned}$$

Therefore, we obtain the following estimators for $\boldsymbol{\gamma}$ and $g(\cdot)$, respectively:

$$\begin{aligned} \widehat{\boldsymbol{\gamma}} &= \mathbf{Q}_2\widehat{\boldsymbol{\theta}} = \mathbf{Q}_2\mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\left\{\mathbf{I} - \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\right\}\mathbf{Y}, \\ \widehat{g}(\mathbf{x}) &= \mathbf{B}(\mathbf{x})^{\mathrm{T}}\widehat{\boldsymbol{\gamma}} = \sum_{\xi\in\mathcal{K}} B_\xi(\mathbf{x})\widehat{\gamma}_\xi. \end{aligned} \tag{2.10}$$

The fitted values at the n data points are $\widehat{\mathbf{Y}} = \mathbf{Z}\widehat{\boldsymbol{\beta}} + \mathbf{B}\widehat{\boldsymbol{\gamma}} = \mathbf{S}(\lambda)\mathbf{Y}$, where the

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

hat matrix is

$$\mathbf{S}(\lambda) = \mathbf{Z}\mathbf{U}_{11}\mathbf{Z}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right) + \mathbf{B}\mathbf{Q}_2\mathbf{U}_{22}\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{Z}\mathbf{V}_{11}^{-1}\mathbf{Z}^{\mathrm{T}}\right).$$

In a nonparametric regression, the trace of the smoothing matrix, $\mathrm{tr}(\mathbf{S}(\lambda))$, is often referred to as the degrees of freedom of the model fit (Green and Silverman (1993)). The trace can be roughly interpreted as the equivalent number of parameters, and can be thought as a generalization of the definition in a linear regression. Finally, we estimate the variance of the error term, σ^2 by

$$\widehat{\sigma}^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n - \mathrm{tr}(\mathbf{S}(\lambda))}. \tag{2.11}$$

2.4. Penalty parameter selection

Selecting a suitable value for the smoothing parameter λ is critical to the model fitting. A large value of λ enforces a smoother fitted function, with potentially larger fitting errors. A small value yields a rougher fitted function and potentially smaller fitting errors, with sufficiently many data locations. Because the in-sample fitting errors cannot be used to gauge the prediction property of the fitted function, one should target a criterion function that mimics the out-of-sample performance of the fitted model. The generalized cross-validation (GCV) is such a criterion, and is widely used for choosing the penalty parameter. We choose the smoothing parameter λ by minimizing the following GCV criterion:

$$\mathrm{GCV}(\lambda) = \frac{n\|\mathbf{Y} - \mathbf{S}(\lambda)\mathbf{Y}\|^2}{\{n - \mathrm{tr}(\mathbf{S}(\lambda))\}^2},$$

over a grid of values of λ. We use a 10-point grid, where the values of $\log_{10}(\lambda)$ are equally spaced between -6 and 7 in our numerical experiments.

3. Asymptotic Results

This section studies the asymptotic properties for the proposed estimators. To discuss these properties, we first introduce some notation. For any function f over the closure of domain Ω, denote $\|f\|_\infty = \sup_{\mathbf{x}\in\Omega}|f(\mathbf{x})|$ as the supremum norm of function f, and $|f|_{v,\infty} = \max_{i+j=v}\|D_{x_1}^i D_{x_2}^j f(\mathbf{x})\|_\infty$ as the maximum norms of all vth-order derivatives of f over Ω. Let

$$W^{\ell,\infty}(\Omega) = \{f \text{ on } \Omega : |f|_{k,\infty} < \infty, 0 \le k \le \ell\} \tag{3.1}$$

be the standard Sobolev space. For any $j = 1, \ldots, p$, let z_j be the coordinate mapping that maps $\mathbf{z}$ to its jth component such that $z_j(\mathbf{Z}_i) = Z_{ij}$, and let

$$h_j = \mathrm{argmin}_{h\in L^2}\|z_j - h\|_{L^2}^2 = \mathrm{argmin}_{h\in L^2}E\{(Z_{ij} - h(\mathbf{X}_i))^2\} \tag{3.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

be the orthogonal projection of z_j onto L^2.

Before we state the results, we make the following assumptions:

(A1) The random variables Z_{ij} are bounded, uniformly in $i = 1, \ldots, n$, $j = 1, \ldots, p$.

(A2) The eigenvalues of $E\left\{(1\ \mathbf{Z}_i^{\mathrm{T}})^{\mathrm{T}}(1\ \mathbf{Z}_i^{\mathrm{T}})|\mathbf{X}_i\right\}$ are bounded away from zero.

(A3) The noise ϵ satisfies that $\lim_{\eta\to\infty} E\left[\epsilon^2 I(\epsilon > \eta)\right] = 0$.

Assumptions (A1)–(A3) are typical in the semi-parametric smoothing literature; for instance, see Huang, Zhang and Zhou (2007) and Wang et al. (2011). The purpose of Assumption (A2) is to ensure that the vector $(1, \mathbf{Z}_i^{\mathrm{T}})$ is not multicolinear.

We next introduce some assumptions on the properties of the true bivariate function in model (1.1) and the data locations related to the triangulation $\triangle$.

(C1) The bivariate functions $h_j(\cdot)$, for $j = 1, \ldots, p$, and the true function in model (1.1) $g(\cdot) \in W^{\ell+1,\infty}(\Omega)$ in (3.1) for an integer $\ell \geq 1$.

(C2) For every $s \in \mathbb{S}$ and every $\tau \in \triangle$, there exists a positive constant F_1, independent of s and τ, such that

$$F_1\|s\|_{\infty,\tau} \leq \left\{\sum_{\mathbf{X}_i\in\tau,\ i=1,\ldots,n} s\left(\mathbf{X}_i\right)^2\right\}^{1/2}, \quad \text{for all } \tau \in \triangle, \tag{3.3}$$

where $\|s\|_{\infty,\tau}$ denotes the supremum norm of s on triangle τ.

(C3) Let F_2 be the largest number of observations in triangles $\tau \in \triangle$. That is, $F_2 > 0$ is a constant

$$\left\{\sum_{\mathbf{X}_i\in\tau,\ i=1,\ldots,n} s\left(\mathbf{X}_i\right)^2\right\}^{1/2} \leq F_2\|s\|_{\infty,\tau}, \quad \text{for all } \tau \in \triangle. \tag{3.4}$$

We further assume that F_1 and F_2 in (3.3) and (3.4) satisfy $F_2/F_1 = O(1)$.

(C4) The number of triangles N and the sample size n satisfy that $N = Cn^{\gamma}$, for some constant $C > 0$ and $1/(\ell+1) \leq \gamma \leq 1/3$.

(C5) The penalized parameter λ satisfies $\lambda = o(n^{1/2}N^{-1})$.

(C6) Let $\delta_{\triangle} = \max_{\tau\in\triangle} |\tau|/\rho_{\tau}$, where ρ_{τ} is the radius of the largest circle inscribed in τ. The triangulation $\triangle$ is δ-quasi-uniform; that is, there exists a positive constant δ such that the triangulation $\triangle$ satisfies $\delta_{\triangle} \leq \delta$.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

Condition (C1) describes the requirement for the true bivariate function usually used in the literature on nonparametric or semiparametric estimations. Condition (C2) ensures the existence of a discrete least squares spline (von Golitschek and Schumaker (2002)), that is, an unpenalized spline with $\lambda = 0$. Although we can obtain a decent penalized least squares spline fitting without this condition, we need (C2) to study the convergence of the bivariate penalized least squares splines. Heuristically, if a triangle $\tau \in \triangle$ near the boundary of $\triangle$ does not contain sufficient observations, the penalized least square spline will not fit the function well over the triangle τ. Condition (C3) suggests that we should not put too many observations in one triangle. Similar conditions to (C2) and (C3) are used in von Golitschek and Schumaker (2002) and Huang (2003). Condition (C4) requires that the number of triangles is above some minimum, depending upon the degree of the spline, which is similar to the requirement of Li and Ruppert (2008) in the univariate case. It also ensures the asymptotic equivalence of the theoretical and empirical inner products/norms defined at the beginning of Section 3. Condition (C5) is required to reduce the bias of the spline approximation through "under smoothing" and "choosing smaller λ". The study of Lai and Schumaker (2007) shows that the approximation of a bivariate spline space over $\triangle$ is dependent on $\delta_\triangle$, that is, the larger the $\delta_\triangle$ is, the worse the spline approximation is. Condition (C6) suggests using triangulations that are more uniform with a reasonably small $\delta_\triangle$. By choosing a set of appropriate vertices, we have a desired triangulation where $\delta_\triangle$ is sufficiently small, say $\delta_\triangle < 10$.

To avoid confusion, we let $\boldsymbol{\beta}_0$ and g_0 be the true parameter value and function, respectively, in model (1.1). The following theorem states that the rate convergence of $\widehat{\boldsymbol{\beta}}$ is root-n and $\widehat{\boldsymbol{\beta}}$ is asymptotically normal.

Theorem 1. *Suppose Assumptions* (A1)–(A3) *and* (C1)–(C6) *hold. Then the estimator* $\widehat{\boldsymbol{\beta}}$ *is asymptotically normal; that is,* $(n\boldsymbol{\Sigma})^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(\mathbf{0}, \mathbf{I})$, *where* $\mathbf{I}$ *is a* $p \times p$ *identity matrix,*

$$\boldsymbol{\Sigma} = \sigma^{-2} E\{(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\}, \tag{3.5}$$

with $\widetilde{\mathbf{Z}}_i = \{h_1(\mathbf{X}_i), \ldots, h_p(\mathbf{X}_i)\}^{\mathrm{T}}$, *for* $h_j(\cdot)$ *defined in* (3.2), $j = 1, \ldots, p$. *In addition,* $\boldsymbol{\Sigma}$ *can be consistently estimated by*

$$\boldsymbol{\Sigma}_n = \frac{1}{n\widehat{\sigma}^2} \sum_{i=1}^{n} (\mathbf{Z}_i - \widehat{\mathbf{Z}}_i)(\mathbf{Z}_i - \widehat{\mathbf{Z}}_i)^{\mathrm{T}} = \frac{1}{n\widehat{\sigma}^2} (\mathbf{Z} - \widehat{\mathbf{Z}})^{\mathrm{T}} (\mathbf{Z} - \widehat{\mathbf{Z}}), \tag{3.6}$$

where $\widehat{\mathbf{Z}}_i$ *is the* i*th column of* $\widehat{\mathbf{Z}}^{\mathrm{T}} = \mathbf{Z}^{\mathrm{T}} \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^{\mathrm{T}} \mathbf{B}^{\mathrm{T}}$ *and* $\widehat{\sigma}^2$ *is given by* (2.11).

The results in Theorem 1 enable us to construct confidence intervals for the

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

parameters. The next theorem provides the global convergence of the nonparametric estimator $\widehat{g}(\cdot)$.

Theorem 2. *Suppose Assumptions* (A1)–(A3) *and* (C1)–(C6) *hold. Then the bivariate penalized estimator* $\widehat{g}(\cdot)$ *in* (2.10) *is consistent with the true function* g_0*, and satisfies that*

$$\|\widehat{g} - g_0\|_{L^2} = O_P\left(\frac{\lambda}{n\,|\triangle|^3}|g_0|_{2,\infty} + \left(1 + \frac{\lambda}{n\,|\triangle|^5}\right)\frac{F_2}{F_1}|\triangle|^{\ell+1}|g_0|_{\ell,\infty} + \frac{1}{\sqrt{n}|\triangle|}\right).$$

The proofs of the above two theorems are given in the Supplemental Material. Note that the rate of convergence given in Theorem 2 is the same as those for nonparametric spline regression, without including the covariate information obtained in Lai and Wang (2013).

4. Simulation

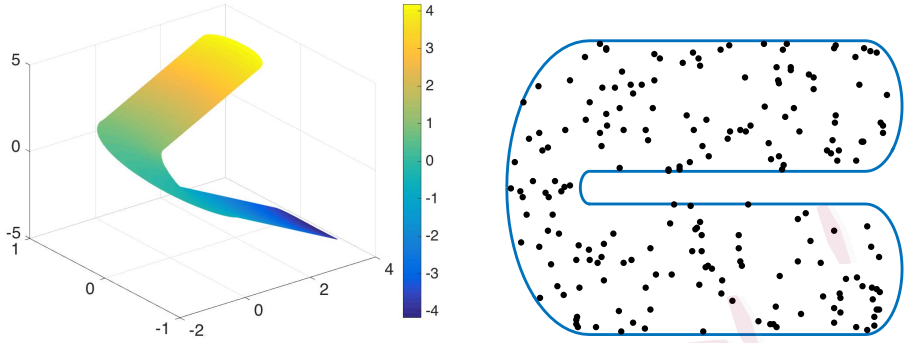
In this section, we carry out a numerical study to assess the performance of the proposed estimators using bivariate penalized splines over triangulations (BPST) over a horseshoe domain. We compare the BPST with filtered kriging (KRIG), thin-plate splines (TPS), the linear finite-elements method (FEM) of Sangalli, Ramsay and Ramsay (2013), and the geodesic low-rank thin-plate splines (GLTPS) of Wang and Ranalli (2007). Additional simulation studies can be found in the Supplementary Material.

For 50×20 grid points on the domain, we simulate data as follows. The response variable Y is generated from the following PLM:

$$Y = \beta_1 Z_1 + \beta_2 Z_2 + g(X_1, X_2) + \epsilon.$$

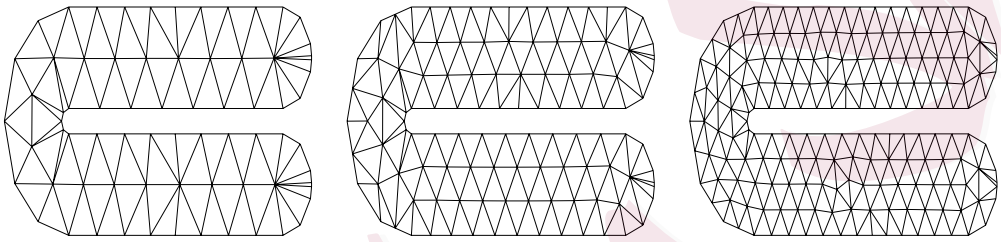
Figure 2 (a) shows the surface of the true function $g(\cdot)$, as used in Wood, Bravington and Hedley (2008) and Sangalli, Ramsay and Ramsay (2013). The random error, ϵ, is generated from an $N(0, \sigma_\epsilon^2)$ distribution with $\sigma_\epsilon = 0.5$. In addition, we set the parameters as $\beta_1 = -1$ and $\beta_2 = 1$. For the design of the explanatory variables, Z_1 and Z_2, two scenarios are considered, based on the relationship between the location variables (X_1, X_2) and the covariates (Z_1, Z_2). Under both scenarios, $Z_1 \sim uniform[-1, 1]$. On the other hand, the variable $Z_2 = \cos[4\pi(\rho(X_1^2 + X_2^2) + (1-\rho)U)]$, where $U \sim uniform[-1, 1]$ and is independent of (X_1, X_2) and Z_1. We consider both an independent design $\rho = 0.0$ and a dependent design $\rho = 0.7$ in this example. Under both scenarios, 100 Monte Carlo replicates are generated. Figure 2 (b) demonstrates the sampled location points of replicate 1. For each replication, we randomly sample $n = 200$ locations

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243



(a) (b)

Figure 2. (a) True function of $g(\cdot)$; (b) Sampled location points of replicate 1.



(a) triangulation $\triangle_1$ (b) triangulation $\triangle_2$ (c) triangulation $\triangle_3$

Figure 3. Three triangulations on the horseshoe domain.

uniformly from the grid points inside the horseshoe domain.

Figure 3 (a)–(c) illustrate three different triangulations used in the BPST method. In the first triangulation ($\triangle_1$), we use 89 triangles (73 vertices). There are 158 triangles (114 vertices) and 286 triangles (186 vertices) in $\triangle_2$ and $\triangle_3$, respectively. To implement the TPS and KRIG methods, we use the R package *fields* under the standard implementation setting of (Furrer, Nychka and Sainand (2011)). For KRIG, we try different covariance structures, and choose the Matérn covariance with smoothness parameter $\nu = 1$, which gives the best prediction. For the GLTPS, following Wang and Ranalli (2007), we use 40 knots with locations selected using the "`cover.design`" method in the package *fields*. For all methods requiring a smoothing or roughness parameter, the GCV is used to choose the values of the parameter.

To assess the accuracy of the estimators, we compute the root mean squared error (RMSE) for each of the components based on 100 Monte Carlo samples. Table 1 shows the RMSEs of the estimates of the parameters β_1, β_2, and σ_ε. The RMSE for the nonlinear function $g(\cdot)$ is computed as the average of

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

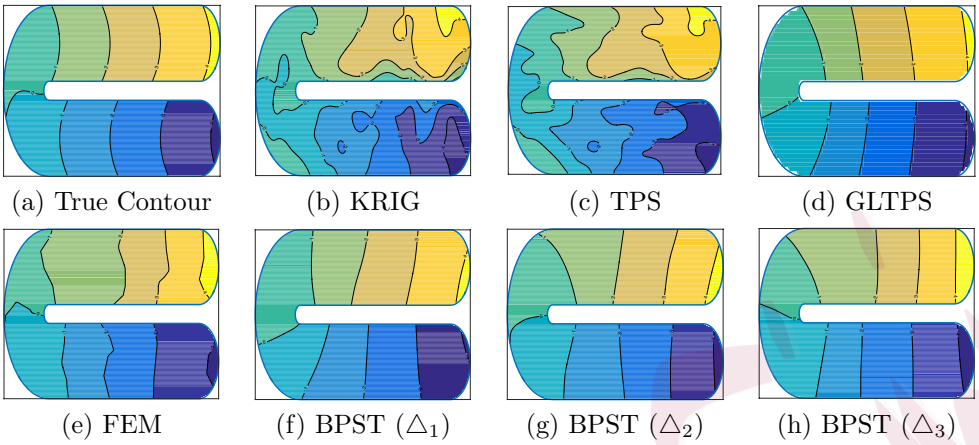
Table 1. Root mean squared errors of the estimates.

ρ	Method	RMSE				CV-RMSPE
		β_1	β_2	σ_ε	$g(\cdot)$	Y
0.0	KRIG	0.0582	0.0433	0.0455	0.3972	0.6728
	TPS	0.0543	0.0426	0.0365	0.3013	0.6037
	GLTPS	0.0625	0.0544	0.0233	0.1565	0.5326
	FEM	0.0560	0.0480	0.0348	0.1558	0.5333
	BPST ($\triangle_1$)	0.0526	0.0498	0.0209	0.1473	0.5299
	BPST ($\triangle_2$)	0.0483	0.0489	0.0220	0.1483	0.5210
	BPST ($\triangle_3$)	0.0544	0.0544	0.0222	0.1458	0.5248
0.7	KRIG	0.0586	0.0440	0.0460	0.3973	0.6728
	TPS	0.0547	0.0402	0.0363	0.3010	0.6038
	GLTPS	0.0612	0.0411	0.0220	0.1553	0.5326
	FEM	0.0562	0.0597	0.0352	0.1567	0.5336
	BPST ($\triangle_1$)	0.0521	0.0563	0.0209	0.1473	0.5294
	BPST ($\triangle_2$)	0.0481	0.0502	0.0222	0.1479	0.5209
	BPST ($\triangle_3$)	0.0543	0.0479	0.0220	0.1457	0.5251

$[1{,}000^{-1}\sum_{i=1}^{1{,}000}\{\widehat{g}(\mathbf{X}_i) - g(\mathbf{X}_i)\}^2]^{1/2}$, based on $1{,}000 = 50 \times 20$ grid points over the 100 Monte Carlo replications. Table 1 shows that BPST produces the best estimation of the nonlinear function $g(\cdot)$, followed by the GLTPS and FEM. The RMSE is nearly constant for all three triangulations, which shows that $\triangle_1$ might be sufficiently fine to capture the feature in the data set. It also suggests that, when this minimum number of triangles is reached, further refining of the triangulation will have little effect on the fitting process, but will make the computational burden unnecessarily heavy. Table 1 also provides the 10-fold cross-validation root mean squared prediction error (CV-RMSPE) for the response variable, defined as $\{n^{-1}\sum_{m=1}^{10}\sum_{i\in\kappa_m}(\widehat{Y}_i - Y_i)^2\}^{1/2}$ over 100 Monte Carlo replications, where $\kappa_1, \ldots, \kappa_{10}$ comprise a random partition of the data set into 10 disjoint subsets of equal size. The CV-RMSPE also shows the superior performance of the BPST method, because it provides the most accurate predictions.

Figure 4 shows the estimated functions over a grid of 500×200 points using different methods for replicate 1 for $\rho = 0.0$. Because such a high-resolution prediction is computationally too expensive for the GLTPS, the prediction map for the GLTPS is based on 100×40 grid points. The plots show that the BPST and GLTPS estimates look visually better than the other four estimates do. In addition, there is a "leakage effect" in the KRIG and TPS estimates. This poor performance is because KRIG and TPS do not take the complex boundary into

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243



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

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

Figure 4. Contour maps for the true function and its estimators ($\rho = 0.0$).

Table 2. Standard error estimates of the coefficients via BPST ($\triangle_2$).

ρ	Parameter	SE_{mc}	SE_{mean}	SE_{median}	SE_{mad}
0.0	β_1	0.0479	0.0651	0.0654	0.0031
	β_2	0.0446	0.0532	0.0530	0.0028
0.7	β_1	0.0477	0.0651	0.0653	0.0029
	β_2	0.0420	0.0518	0.0522	0.0024

account and smooth across the gap inappropriately. Finally, the BPST estimators based on the three triangulations are very similar, supporting our findings for penalized splines that the number of triangles is not critical to the fitting, as long as it is sufficiently large to capture the pattern and features of the data. Similar estimation results are obtained for the case of $\rho = 0.7$. Sample estimated functions are presented in Figure 1 in the Supplementary Material to save space.

Next, we test the accuracy of the standard error (SE) formula in (3.6) for $\hat{\beta}_1$ and $\hat{\beta}_2$; the results are listed in Table 2. The standard deviations of the estimated parameters are computed based on 100 replications, which can be regarded as the true standard errors (column labeled "SE_{mc}") and compared with the mean and median of the 100 estimated standard errors calculated using (3.6) (columns labeled "SE_{mean}" and "SE_{median}", respectively). The column labeled "SE_{mad}" is the interquartile range of the 100 estimated standard errors, divided by 1.349, which is a robust estimate of the standard deviation. Table 2 shows that the averages or medians of the SEs calculated using the formula are very close to the true standard deviations, which confirms the accuracy of the proposed SE formula.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

In terms of the computational complexity, because the GLTPS technique is largely based on Floyd's algorithm, it has cubic time complexity (Miller and Wood (2014)), as does ordinary kriging. In contrast, the TPS, FEM, and BPST can be formulated as a single least squares problem, making them fast and easy to compute. Taking the prediction as an example, we find that as the prediction size increases (sample size is fixed), the computation time for GLTPS and KRIG increases dramatically, whereas BPST provides an almost linear complexity of the prediction size. On a standard PC with a Core i5 @2.9GHz CPU and 16.00GB RAM, the BPST($\triangle_1$) prediction over $2{,}500 \times 1{,}000$ grid points needs only 10 seconds of computing. Then, BPST($\triangle_2$) and BPST($\triangle_3$), with finer triangulations, take just a few seconds longer than BPST($\triangle_1$). However, the GLTPS usually takes hours to complete one estimation and prediction at the 100×40 resolution level. In addition, in our numerical study, KRIG requires a large amount of memory. When the prediction resolution becomes finer than $2{,}500 \times 1{,}000$, KRIG will crash on a standard PC owing to lack of memory.

5. Application to Mercury Concentration Studies

In this section, we apply the proposed method to map the mercury in the sediment concentration over the estuary in New Hampshire; see Figure 1 (a) for a regional map of the estuary. Mercury contamination is a significant public health and environmental problem. When released into the environment, mercury accumulates in water-laid sediments, is ingested by fish, and passed along the food chain to humans. Several rivers flowing into the Great Bay are contaminated with mercury, according to the new Environment New Hampshire report. Estuaries such as Great Bay are ideal locations for the accumulation of contaminants such as mercury that settle from the surrounding watershed (Brown et al. (2015)). The coastal monitoring program, National Coastal Assessment, run by the US Environmental Protection Agency (EPA) and the New Hampshire Department of Environmental Services has developed surveys that can reveal useful information on the status and trends of contaminants.

The spatial data set in our study consists of the mercury concentrations surveyed in 2000/2001 and 2003 at 97 locations in the largest estuary in New Hampshire; see Figure 1 (b) for the measurements of mercury concentrations at different sampled locations. To assist decision-makers to develop effective environmental protection strategies, it is critical to provide measurements of mercury at spatial scales much finer than those at which the mercury was monitored.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

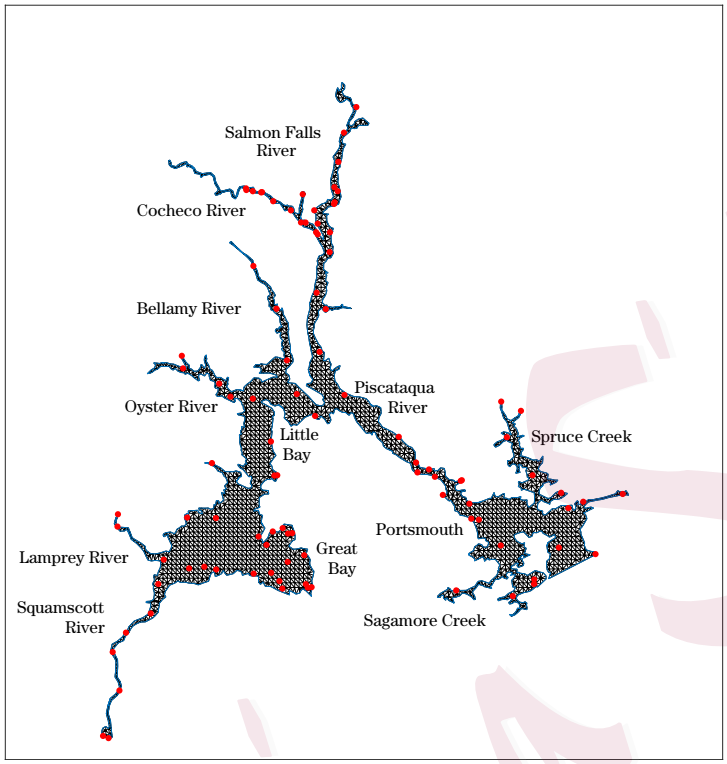


Figure 5. Domain triangulation for estuaries in New Hampshire.

This data set has been studied by Wang and Ranalli (2007) using the GLTPS. Following Wang and Ranalli (2007), we consider a PLM with a linear term for the year effect (Year = 0 if the survey was conducted in 2000/2001; and Year = 1 if the survey was conducted in 2003):

$$\text{Mercury Concentration} = \beta \text{Year} + g(\text{Latitude, Longitude}). \tag{5.1}$$

To fit model (5.1), we use five methods: KRIG, TPS, GLTPS, FEM, and BPST. For KRIG, we choose the Matérn covariance structure to fit the model. The GLTPS is calculated using the setting $k = 5$, as in Wang and Ranalli (2007). For BPST and FEM, the smoothing or roughness parameter is selected using the GCV. Figure 5 shows the triangulation adopted by the BPST. Table 3 summarizes the coefficient estimation results based on the various methods.

The Great Bay estuary is a tidally dominated system and is the drainage confluence of the Lamprey River and Squamscott River. Four additional rivers that flow into the system are the Cocheco, Salmon Falls, Bellamy, and Oyster Rivers. Mercury deposited in the estuaries in New Hampshire is both emitted

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

Table 3. Estimated coefficients with standard errors (SE).

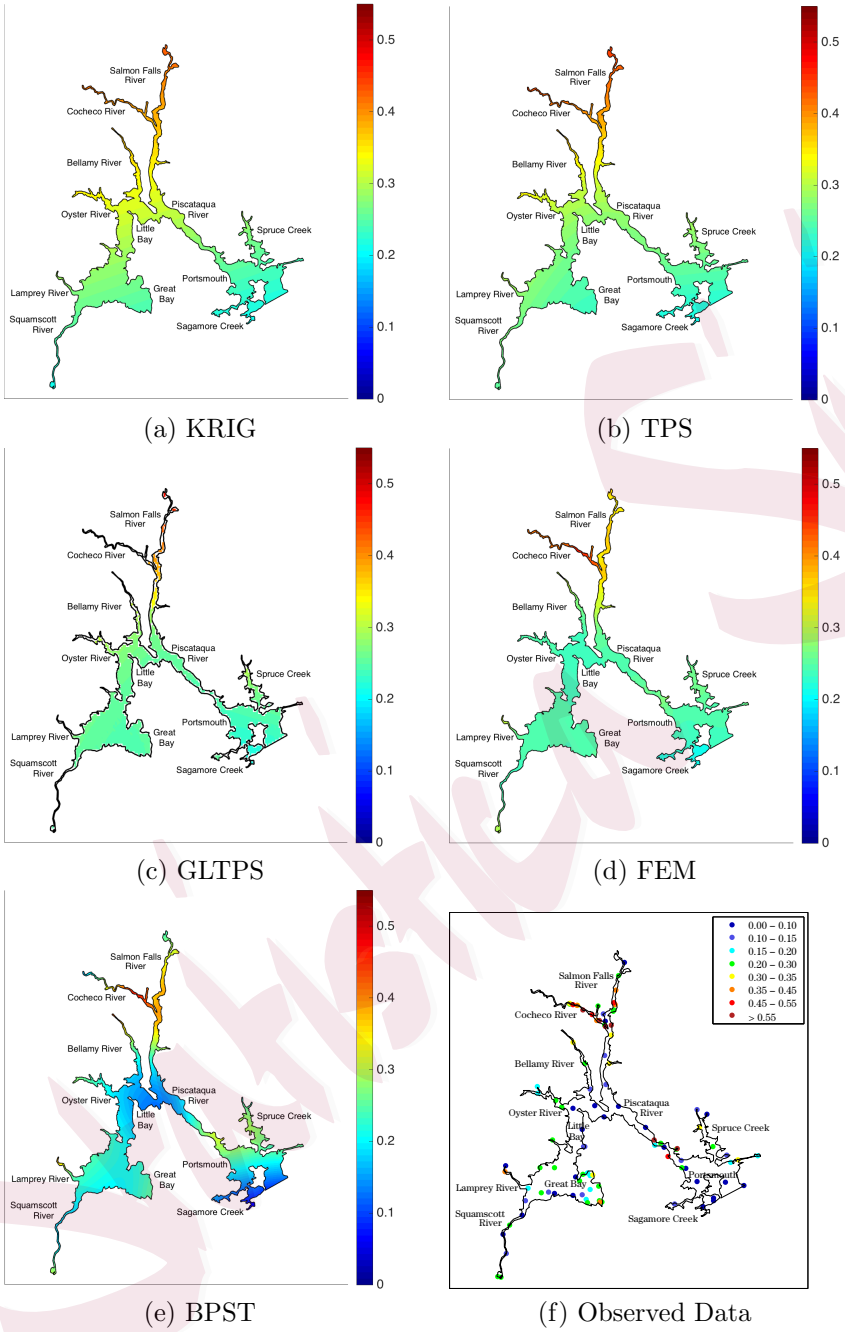
	KRIG	TPS	GLTPS	FEM	BPST
Year	0.096	0.095	0.051	0.076	0.044
SE	0.03	0.03	0.02	0.04	0.04

from in-state sources and carried here from sources upwind. Emissions upwind of New Hampshire are primarily attributable to coal-fired utilities and municipal and medical waste incinerators in the Northeast and Midwest (Abbott et al. (2008)). In general, the spatial distribution in Figure 1 (b) shows higher values in the Salmon Falls River and Cocheco River, lower values in the Piscataqua River and the Portsmouth area, and some localized low spots in the Great Bay estuary.

Prediction maps at the 20×20 m resolution level using different methods are shown in Figure 6. The computation-intensive GLTPS procedure finds it difficult to make such a high-resolution prediction, so we decrease its resolution to 150×150 m. All methods in Figure 6 identify relatively high mercury contamination in the Salmon Falls River and Cocheco River, which is consistent with known historical pollution sources (Abbott et al. (2008)). Figure 6 also illustrates the overspill from the Northern part to the middle area when ordinary spatial smoothing (e.g., KRIG or TPS) is used, because it smoothes across the Salmon Falls River and Cocheco River, with high concentration levels in the northern part. This problem is mitigated for GLTPS and FEM. The BPST smoother does not show signs of leakage in the Piscataqua River and the Portsmouth area of the estuaries, as other methods do. Note the way in which the KRIG and TPS smooth, inappropriately, across the east coast of the Great Bay, so that relatively high mercury concentrations are estimated for the Portsmouth area in the southeastern part of the estuaries. The poor prediction performance of KRIG and TPS suggests that we should not assume that densities in geographically neighboring areas are similar if they are separated by physical barriers.

To evaluate the different methods, we report both the in-sample root mean squared errors (RMSE): $\{n^{-1}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2\}^{1/2}$, and the cross-validation root mean squared prediction errors (RMSPE) of the mercury concentrations. Because there are only 97 observations in this data set, we consider the leave-one-out cross-validation (LOOCV) prediction error instead of the 10-fold cross-validation, as conducted in the simulation studies. Specifically, for each $i = 1, \ldots, 97$, we train the model on every point except i, and then obtain the prediction error on the held-out point. Table 4 summarizes the RMSE and the LOOCV-RMSPE us-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243



(a) KRIG (b) TPS

(c) GLTPS (d) FEM

(e) BPST (f) Observed Data

Figure 6. Prediction maps of mercury concentrations over the estuaries in New Hampshire.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

Table 4. In-sample RMSEs and LOOCV-RMSPE of mercury concentrations.

Method	KRIG	TPS	GLTPS	FEM	BPST
RMSE	0.1397	0.1381	0.1366	0.1263	0.1197
RMSPE	0.1480	0.1473	0.1459	0.1467	0.1402

ing different methods. As expected, when the shape of the boundary is complex, smoothers that respect the complicated boundary shape appropriately are able to reduce the prediction errors. The LOOCV-RMSPE favors the model with the BPST smoother, which not only gives the best model fit, but also provides the most accurate prediction of the concentration values.

6. Conclusion

In this study, we considered PLMs for modeling spatial data with complicated domain boundaries. We introduce a framework of bivariate penalized splines defined on triangulations in a semi-parametric estimation. Our BPST method demonstrates competitive performance compared with existing methods, while providing a number of possible advantages.

First, the proposed method greatly enhances the application of non/semi-parametric methods to spatial data analyses. It solves the problem of "leakage" across complex domains, which many conventional smoothing tools suffer from. The numerical results from the simulation studies and the application show that our method is effective to account for complex domain boundaries. Our method does not require the data to be evenly distributed or on regular-spaced grids, as in the tensor product-smoothing methods. When we have regions of sparse data, bivariate penalized splines provide a more convenient tool for fitting the data than unpenalized splines do, because the roughness penalty helps regularize the estimation. Relative to the conventional FEM, our method provides a more flexible way to use piecewise polynomials of various degrees and smoothness over an arbitrary triangulation for spatial data analyses.

Second, we provide new statistical theories for estimating the PLM for data distributed on complex spatial domains. We show that our estimates of both the parametric and the nonparametric parts of the model enjoy excellent asymptotic properties. In particular, we showed that our estimates of the coefficients in the parametric part are asymptotically normal, and then derived the convergence rate of the nonparametric component under regularity conditions. We also provided a standard error formula for the estimated parameters, and our simulation studies

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

show that the standard errors are estimated with good accuracy. The theoretical results provide measures of the effects of covariates after adjusting for the location effect. In addition, they give valuable insights into the accuracy of our estimate of the PLM and permit a joint inference for the parameters.

Finally, the computation of our proposed method is much more efficient than other approaches, such as kriging and GLTPS. Specifically, for model fitting with n locations, the computational complexity of the ordinary kriging and GLTPS is $O(n^3)$, whereas the computational complexity of our method is only $O(nN^2)$. Here N is the number of triangles in the triangulation, and is usually much smaller than n, as suggested in Condition (C4).

Supplementary Material

The online Supplementary Material Wang et al. (2020) explains how to implement the proposed methods, as well as providing additional simulation and application results and the proofs of Lemma 1 and Theorems 1 and 2.

Acknowledgments

The first author's research was supported, in part, by National Science Foundation grants DMS-1106816 and DMS-1542332, the second author's research was supported, in part, by College of William & Mary Faculty Summer Research Grant and the third author's research was supported, in part, by National Science Foundation grant DMS-1521537 and Simons collaboration grant #280646. The authors would like to thank Haonan Wang and M. Giovanna Ranalli for providing the New Hampshire estuary data. This paper has not been formally reviewed by the EPA. The views expressed here are solely those of the authors. The EPA does not endorse any products or commercial services mentioned in this report. Finally, the authors would like to thank the editor, the associate editor, and reviewers for their valuable comments and suggestions.

References

Abbott, M. L., Lin, C.-J., Martian, P. and Einerson, J. J. (2008). Atmospheric mercury near salmon falls creek reservoir in southern idaho. *Applied Geochemistry* **23**, 438–453.

Awanou, G., Lai, M. J. and Wenston, P. (2005). The multivariate spline method for scattered data fitting and numerical solutions of partial differential equations. *Wavelets and Splines: Athens 2005*, 24–74.

Brown, L. E., Chen, C. Y., Voytek, M. A. and Amirbahman, A. (2015). The effect of sediment mixing on mercury dynamics in two intertidal mudflats at great bay estuary, new

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

hampshire, usa. *Marine Chemistry* **177**, 731–741.

Chen, R., Liang, H. and Wang, J. (2011). On determination of linear components in additive models. *Journal of Nonparametric Statistics* **23**, 367–383.

Furrer, R., Nychka, D. and Sainand, S. (2011). *Package 'fields'. R Package Version* 6.6.1. [online] Available at `http://cran.r-project.org/web/packages/fields/fields.pdf`.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. CRC Press.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London.

Härdle, W., Liang, H. and Gao, J. T. (2000). *Partially Linear Models*. Heidelberg: Springer Physica-Verlag.

He, X. and Shi, P. (1996). Bivariate tensor-troduct b-splines in a partly linear model. *Journal of Multivariate Analysis* **58**, 162–181.

Huang, J. (2003). Asymptotics for polynomial spline regression under weak conditions. *Statistics & Probability Letters* **65**, 207–216.

Huang, J. Z., Zhang, L. and Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scandinavian Journal of Statistics* **34**, 451–477.

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. (1998). Approximation power of bivariate splines. *Advances in Computational Mathematics* **9**, 251–279.

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.

Liang, H., Härdle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **27**, 1519–1535.

Liang, H. and Li, R. (2009). Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association* **104**, 234–248.

Liu, X., Wang, L. and Liang, H. (2011). Estimation and variable selection for semiparametric additive partial linear models. *Statistica Sinica* **21**, 1225–1248.

Ma, S., Song, Q. and Wang, L. (2013). Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data. *Bernoulli* **19**, 252–274.

Ma, Y., Chiou, J.-M. and Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika* **93**, 75–84.

Mammen, E. and van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *The Annals of Statistics* **25**, 1014–1035.

Marx, B. and Eilers, P. (2005). Multidimensional penalized signal regression. *Technometrics* **47**, 13–22.

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

Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical*

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0243

Society, Series B (Statistical Methodology) **64**, 307–319.

Sangalli, L., Ramsay, J. and Ramsay, T. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **75**, 681–703.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Statistical Methodological)* **50**, 413–436.

von Golitschek, M. and Schumaker, L. L. (2002). Bounds on projections onto bivariate polynomial spline spaces with stable local bases. *Constructive Approximation* **18**, 241–254.

Wang, H. and Ranalli, M. G. (2007). Low-rank smoothing splines on complicated domains. *Biometrics* **63**, 209–217.

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

Wang, L., Wang, G., Lai, M. and Gao, L. (2020). Efficient estimation of partially linear models for data on complicated domains via bivariate penalized splines over triangulations. *Statistica Sinica* Supplementary Materials. `http://www3.stat.sinica.edu.tw/statistica/`

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

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **65**, 95–114.

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

Xiao, L., Li, Y. and Ruppert, D. (2013). Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **75**, 577–599.

Zhang, H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? automatic structure discovery for partially linear models. *Journal of American Statistical Association* **106**, 1099–1112.

Zhou, L. and Pan, H. (2014). Smoothing noisy data for irregular regions using penalized bivariate splines on triangulations. *Computational Statistics* **29**, 263–281.

Department of Statistics and the Statistical Laboratory, Iowa State University, Ames, IA 50011, USA.

E-mail: lilywang@iastate.edu

Department of Mathematics, College of William & Mary, Williamsburg, VA 23187, USA.

E-mail: gwang01@wm.edu

Department of Mathematics, University of Georgia, Athens, GA 30602, USA.

E-mail: mjlai@math.uga.edu

Department of Finance, Iowa State University, Ames, IA 50011, USA.

E-mail: lgao@iastate.edu

(Received May 2017; accepted April 2018)