Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Estimation and Inference for Generalized Geoadditive Models

Shan Yu, Guannan Wang, Li Wang, Chenhui Liu & Lijian Yang

To cite this article: Shan Yu, Guannan Wang, Li Wang, Chenhui Liu & Lijian Yang (2019): Estimation and Inference for Generalized Geoadditive Models, Journal of the American Statistical Association, DOI: 10.1080/01621459.2019.1574584

To link to this article: https://doi.org/10.1080/01621459.2019.1574584

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 00, NO. 0, 1–14: Theory and Methods
https://doi.org/10.1080/01621459.2019.1574584

Taylor & Francis
Taylor & Francis Group

Estimation and Inference for Generalized Geoadditive Models

Shan Yu[a*], Guannan Wang[b*], Li Wang[a], Chenhui Liu[c], and Lijian Yang[d]

[a]Department of Statistics, Iowa State University, Ames, IA; [b]Department of Mathematics, College of William & Mary, Williamsburg, VA; [c]Turner-Fairbank Highway Research Center, Federal Highway Administration, McLean, VA; [d]Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing, China

ABSTRACT

In many application areas, data are collected on a count or binary response with spatial covariate information. In this article, we introduce a new class of generalized geoadditive models (GGAMs) for spatial data distributed over complex domains. Through a link function, the proposed GGAM assumes that the mean of the discrete response variable depends on additive univariate functions of explanatory variables and a bivariate function to adjust for the spatial effect. We propose a two-stage approach for estimating and making inferences of the components in the GGAM. In the first stage, the univariate components and the geographical component in the model are approximated via univariate polynomial splines and bivariate penalized splines over triangulation, respectively. In the second stage, local polynomial smoothing is applied to the cleaned univariate data to average out the variation of the first-stage estimators. We investigate the consistency of the proposed estimators and the asymptotic normality of the univariate components. We also establish the simultaneous confidence band for each of the univariate components. The performance of the proposed method is evaluated by two simulation studies. We apply the proposed method to analyze the crash counts data in the Tampa-St. Petersburg urbanized area in Florida. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received September 2018
Accepted January 2019

KEYWORDS
Bivariate splines; Count data; Local polynomial; Penalty; Polynomial splines; Triangulation.

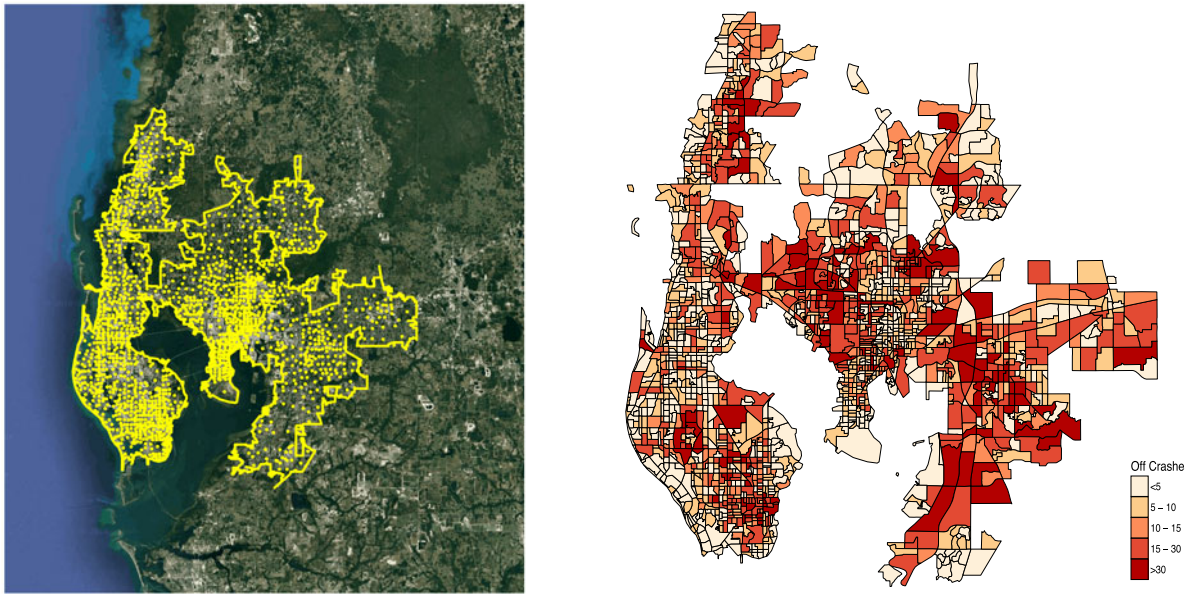
1. Introduction

Regression methods are commonly used in statistics to examine associations between the response and the set of explanatory variables. When the relationship between the variables is complex and cannot be easily modeled by specific linear or nonlinear functions, a generalized additive model (GAM) provides a flexible regression relationship. GAMs were first introduced by Hastie and Tibshirani (1987, 1990), which assume that the mean of the discrete response variable depends on an additive set of predictors through a link function. Since then GAMs have been widely used in many areas. The focus of this work is on GAM for spatial data randomly distributed over a particular geographical region.

To motivate the study, consider the analysis of road traffic crashes, which have been one of the major sources of fatalities and injuries in the United States. Count data models are often used to identify factors significant to crash frequencies, where the influence of demographical and socioeconomic conditions on the crash occurrence and spatial patterns of crashes can assist decision-makers in implementing appropriate road safety management actions (Miaou, Song, and Mallick 2003; Liu and Sharma 2017, 2018). Figure 1(a) shows the Tampa-St. Petersburg urbanized area in Florida, which is consisted of 1761 census block groups. A census block group is a geographical unit with homogeneous population characteristics, economic status, and living conditions for the population survey. Rural areas were excluded from this study, as their population densities often drop to nearly zero, thus, it is not meaningful to analyze their crash data. The frequency of crashes off the state highway system within each census block group in the year of 2014 were collected from the Florida Department of Transportation; see Figure 1(b). In addition, explanatory variables, including vehicle miles traveled, demographic and commuting variables, were collected from the Florida Department of Transportation and U.S. Census Bureau, and many of them exhibit pronounced nonlinear relationships with the response variable.

To analyze such kind of data, researchers face at least three main challenges. First of all, traditional regression methods often assume the data are distributed on a regular sampling grid over a rectangular domain, however, in this study and many other similar cases in spatial studies, observations are dense at some locations while sparse at others, and the shape of the domains is complicated or even shows gaps and holes (see Ramsay 2002; Wood, Bravington, and Hedley 2008). Conventional smoothing tools suffer from the problem of "leakage" across the complex domains, which refers to the poor estimation over difficult regions by smoothing inappropriately across boundary features, such as peninsulas. The second challenge is how to adjust for the spatial effect and provide measures of the nonlinear effect of covariates and assess the impact of uncertainty. In general,

CONTACT Li Wang lilywang@iastate.edu Department of Statistics, Iowa State University, Ames, IA 50011; Lijian Yang yanglijian@tsinghua.edu.cn Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.
*Shan Yu and Guannan Wang are joint first authors of this article.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.



(a) Tampa-St. Petersburg urbanized area

(b) Crashes within individual census block group

Figure 1. Tampa-St. Petersburg urbanized area and its census block groups.

inference for GAM for spatial data is underdeveloped and often made based on ad hoc methods with little understanding of the statistical properties. Last but not least, the computational issue usually is a big challenge for spatial data analysis, as the sample size tends to be large due to the development of remote sensing technology and automated sensor networks. Thus, there is a great need of methodology that is practical, computationally efficient and theoretically reliable.

To address the above-mentioned challenges, in this article we introduce a class of generalized geoadditive models (GGAMs), a synthesis of geostatistics and GAMs, for spatial data randomly distributed over irregular domains. We aim at developing the corresponding estimation and inference procedure for the GGAMs. In the literature, there are two main streams of spatial regression modeling approaches to include spatial information into the model. The first approach adds spatial weights or spatial correlation into a regression modeling, for example, the spatial autoregressive (SAR) model or conditional autoregressive (CAR) model (Lee 2004; Zhu, Huang, and Reyes 2010). The second approach is based on some smoothing techniques, for example, kernel, wavelet, or spline smoothing, Ramsay (2002) use a deterministic smooth surface function to describe the variations and connections among values at different locations. Kammann and Wand (2003) combine the kriging method with the penalized spline regression, and suggest a mixed model representation for model fitting. In our paper, we take the second approach, where the effect of explanatory variables are modeled with additive univariate functions and the spatial effect is modeled via a bivariate function.

In the proposed GGAM, when the spatial component is ignored, the model becomes the traditional GAM since all the components left in the model are univariate functions. There have been a number of proposals for fitting the GAMs, for instance, the spline method of Stone (1986), Xue and Liang (2010), and Wang et al. (2011); the kernel method of Yu, Park, and Mammen (2008), Yang, Sperlich, and Härdle (2003), and Liang et al. (2008); and the two-stage methods of Horowitz and Mammen (2004), Wang and Yang (2007), and Liu, Yang, and Härdle (2013). For estimating bivariate functions defined over rectangular domains, there are several popular smoothing tools: kernel, wavelet, bivariate P-splines (Marx and Eilers 2005) and thin plate splines (Wood 2003). In the past three decades, there has been a great interest in developing the smoothing techniques which can handle irregular domains with complex boundaries (see, e.g., Ramsay 2002; Wang and Ranalli 2007; Wood, Bravington, and Hedley 2008; Sangalli, Ramsay, and Ramsay 2013; Lai and Wang 2013; Miller and Wood 2014; Wang et al. 2019). In this paper, we approximate the bivariate function of the spatial effect using the bivariate splines (smooth piecewise polynomial functions over a triangulation of the domain of interest) over triangulations in Lai and Schumaker (2007). We prefer the bivariate penalized splines (BPS) due to their (i) convenient representations with flexible degrees and various smoothness, (ii) computational efficiency, and (iii) great ability of handling the sparse designs.

Although the above-mentioned spline smoothing seems to be incredibly useful, it provides only convergence rates of the estimators but no asymptotic distributions unless we assume the errors are iid Gaussian, so it is difficult to assign any measure of confidence to the estimators. The simultaneous confidence bands (SCBs) is a powerful tool to evaluate and visualize the variability of the estimators and to make global inferences (see Wang and Yang 2009; Krivobokova, Kneib, and Claeskens 2010; McKeague and Zhao 2006; Wiesenfarth et al. 2012; Zheng et al. 2016 for some related theory and applications of SCBs). To develop the SCBs for each individual component function of the explanatory variables in GGAMs, we propose a one-step spline backfitted local polynomial estimator, referred to as the SBL estimator. In the first stage, we use spline smoothers to approximate the univariate additive components and the geographical

component. In the second stage, local polynomial smoothing is then applied to the cleaned univariate data to average out the variation of the first-stage estimators and obtain SCBs.

Under some regularity conditions, we obtain the asymptotically normal distribution of the SBL estimators and establish the SCBs for the functions of the covariates. Our approach merges the advantages of three smoothing methods (polynomial spline, bivariate spline, and local polynomial) that balance each other out. It enables the spline smoothing to act as an efficient pilot, quickly guiding the fitting toward good solutions. In addition, it allows us to keep the asymptotically normal distribution of the local polynomial estimator, without its computational burden. Furthermore, it properly accounts for all covariate information and spatially improves the estimation accuracy. The entire fitting and inference procedure can be implemented easily and efficiently using standard methodology and software.

The rest of the article is structured as follows. In Section 2, we describe our model, then we give a brief review of univariate splines and bivariate splines over triangulations and introduce the penalized quasi-likelihood estimation method. Section 3 provides the SCBs for the functions of the covariates. Section 4 introduces how to implement the proposed procedure in practice. In Section 5, we conduct simulation studies to evaluate the finite sample performance of the proposed method. Section 6 illustrates our method using a real dataset. Some concluding remarks are given in Section 7. The regularity assumptions are deferred in Appendix A. Proofs and other technical details are given in the online supplementary materials.

2. Methodology

2.1. Model Setup

In the following, let $\mathbf{S}_i = (S_{i1}, S_{i2})^\top$ be the location of ith point, $i = 1, \ldots, n$, which ranges over a bounded domain $\Omega \subseteq \mathbb{R}^2$ of arbitrary shape, for example, a domain with polygon boundary. Let Y_i be the response variable and $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ be the explanatory variables at location $\mathbf{S}_i$.

We assume that the conditional density of Y given $(\mathbf{X}, \mathbf{S}) = (\mathbf{x}, \mathbf{s})$ belongs to the exponential family $f_{Y|\mathbf{X},\mathbf{S}}\left(y \middle| \mathbf{x}, \mathbf{s}\right) = \exp\left[y\xi\left(\mathbf{x}, \mathbf{s}\right) - \mathcal{B}\left\{\xi\left(\mathbf{x}, \mathbf{s}\right)\right\} + \mathcal{C}\left(y\right)\right]$, for known functions $\mathcal{B}$ and $\mathcal{C}$, where ξ is the so-called natural parameter, and is related to the unknown mean response by $\mu\left(\mathbf{x}, \mathbf{s}\right) = \mathrm{E}\left(Y|\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}\right) = \mathcal{B}'\left\{\xi\left(\mathbf{x}, \mathbf{s}\right)\right\}$. In this article, $\mu\left(\mathbf{x}, \mathbf{s}\right)$ is modeled via a link function g in the following additive form

$$g\left\{\mu\left(\mathbf{x}, \mathbf{s}\right)\right\} = \sum_{k=1}^{p} \beta_k(x_k) + \alpha\left(\mathbf{s}\right), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ are unknown univariate smooth functions, and $\alpha(\cdot)$ is an unknown bivariate smooth function.

If $\mathrm{var}\left(Y|\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}\right) = \sigma^2 V\left\{\mu\left(\mathbf{x}, \mathbf{s}\right)\right\}$ for some known positive function V, then estimation of the mean can be achieved by replacing the conditional log-likelihood function $\log\{f_{Y|\mathbf{X},\mathbf{S}}\left(y \middle| \mathbf{x}, \mathbf{s}\right)\}$ with a quasi-likelihood function $\ell\left(\vartheta, y\right)$, which satisfies $\nabla_\vartheta \ell\left(\vartheta, y\right) = \frac{y-\vartheta}{\sigma^2 V(\vartheta)}$. Our estimation method is based on a nonparametric quasi-likelihood approach as described below.

2.2. Spline Approximation

In the first stage, we approximate each of the univariate additive components in the model via univariate polynomial splines. The geographical component is approximated using bivariate penalized splines over triangulation, which is proved to be efficient to deal with data distributed on irregular domains with complicated boundaries. Below we start with a brief review of univariate splines and bivariate splines.

2.2.1. Univariate Polynomial Spline Approximation

Suppose that the covariate X_k is distributed on a compact interval $[a_k, b_k]$, $k = 1, \ldots, p$. We approximate the univariate components $\{\beta_k(\cdot)\}_{k=1}^p$ in (1) by polynomial splines for their simplicity in computation. For any $k = 1, \ldots, p$, let υ_k be a partition of $[a_k, b_k]$ with J_n interior knots, where $\upsilon_k = \left\{a_k = \upsilon_{k,0} < \upsilon_{k,1} < \cdots < \upsilon_{k,J_n} < \upsilon_{k,J_n+1} = b_k\right\}$. The polynomial splines of order $\varrho+1$ are polynomial functions with ϱ-degree (or less) on intervals $[\upsilon k, j, \upsilon_{k,j+1})$, $j = 0, \ldots, J_n - 1$, and $\left[\upsilon_{k,J_n}, \upsilon_{k,J_n+1}\right]$, and have $\varrho - 1$ continuous derivatives globally. Let $\mathcal{U}_k = \mathcal{U}_k^\varrho([a_k, b_k], \upsilon_k)$ be the space of such polynomial splines, and $\mathcal{U}_k^0 = \{u \in \mathcal{U}_k : \mathrm{E}u(X_k) = 0\}$. This ensures that the spline functions are centered (see Xue and Yang 2006; Wang and Yang 2007; Wang et al. 2014).

2.2.2. Bivariate Spline Over Triangulation

In this article, the spatial domain Ω is a polygon of arbitrary shape, which can be partitioned into finitely many triangles. According to Lai and Schumaker (2007), 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 any nonempty intersection between a pair of triangles in $\triangle$ is either a shared vertex or a shared edge. Given a triangle $\tau \in \triangle$, let R_τ be the radius of the largest disk contained in τ, and let $|\tau|$ be the length of the longest edge. Define the shape parameter of τ as the ratio $\pi_\tau = |\tau|/R_\tau$. Note that when π_τ is small, the triangles are relatively uniform in the sense that all angles of triangles in the triangulation $\triangle$ are relatively the same. Denote the size of $\triangle$ by $|\triangle| := \max\{|\tau|, \tau \in \triangle\}$.

For a triangle with nonzero area $\tau \in \triangle$ and any fixed point $\mathbf{v} \in \mathbb{R}^2$, let b_1, b_2, and b_3 be the barycentric coordinates of $\mathbf{v}$ relative to τ. The Bernstein basis polynomials of degree $d \geq 1$ relative to triangle τ is defined as $B_{ijk}^{\tau,d}(\mathbf{v}) = \frac{d!}{i!j!k!} b_1^i b_2^j b_3^k$, $i + j + k = d$. For any integer $d \geq 1$ and triangle τ, let $\mathbb{P}_d(\tau)$ be the space of all polynomials of degree less than or equal to d on τ. Then, any polynomial $\zeta \in \mathbb{P}_d(\tau)$ can be written as $\zeta|_\tau = \sum_{i+j+k=d} \gamma_{ijk}^\tau B_{ijk}^{\tau,d}$, where the coefficients $\boldsymbol{\gamma}_\tau = \{\gamma_{ijk}^\tau, i+j+k = d\}$ are called B-coefficients of ζ.

For any integer $r \geq 0$, let $\mathbb{C}^r(\Omega)$ be the collection of all rth continuously differentiable functions over Ω. Given a triangulation $\triangle$, we define the spline space of degree d and smoothness r over $\triangle$ as $\mathbb{S}_d^r(\triangle) = \{\zeta \in \mathbb{C}^r(\Omega) : \zeta|_\tau \in \mathbb{P}_d(\tau), \tau \in \triangle\}$. Let $\{B_m\}_{m\in\mathcal{M}}$ be the set of bivariate Bernstein basis polynomials for $\mathbb{S}_d^r(\triangle)$, where $\mathcal{M}$ is an index set of $|\mathcal{M}| = N(d+1)(d+2)/2$ basis functions. Then we can represent any function $\zeta \in \mathbb{S}_d^r(\triangle)$ using the following basis expansion

$$\zeta(\mathbf{s}) = \sum_{m\in\mathcal{M}} B_m(\mathbf{s})\gamma_m = \mathbf{B}(\mathbf{s})^\top \boldsymbol{\gamma}, \tag{2}$$

where $\boldsymbol{\gamma}^\top = (\gamma_m, m \in \mathcal{M})$ is the spline coefficient vector. For smooth join between two polynomials on adjoining triangles, we need to impose some linear constraints on the spline coefficients $\boldsymbol{\gamma}$ in (2). To be more specific, we assume that $\boldsymbol{\gamma}$ satisfies $\boldsymbol{\Psi}\boldsymbol{\gamma} = \mathbf{0}$, where $\boldsymbol{\Psi}$ is the matrix that collects the smoothness conditions across all the shared edges of triangles. An example of $\boldsymbol{\Psi}$ is given in Section B.2.1 in the supplementary materials. The above bivariate spline basis can be easily constructed via the R package `BPST` (Wang et al. 2019).

2.3. Penalized Quasi-Likelihood Estimators

In data-sparse regions, penalized splines provide a more convenient tool for data fitting than the unpenalized splines. As discussed in Lai and Wang (2013), the number of triangles for bivariate penalized smoothing is not essential when the number is above some minimum depending upon the degree of the splines. Another advantage is that when we have regions of sparse data, BPS can be considered as a direct ridge regression or shrinkage type method which can alleviate multicollinearity issue. Thus, to reduce complexity in triangulation selection and enhance the performance of bivariate splines in data fitting, we exploit the BPS in the quasi-likelihood of the model below.

To define the penalized spline method, for any direction s_j, $j = 1, 2$, let $\nabla^q_{s_j} f(\mathbf{s})$ be the qth order derivative in the direction s_j at the point $\mathbf{s} = (s_1, s_2)$. Let

$$\mathcal{E}(f) = \int_\Omega \left\{ (\nabla^2_{s_1} f)^2 + 2(\nabla_{s_1} \nabla_{s_2} f)^2 + (\nabla^2_{s_2} f)^2 \right\} ds_1 ds_2. \quad (3)$$

To fit the GGAM, we seek a bivariate function $\alpha(\cdot)$ and univariate functions $\beta_k(\cdot)$, $k = 1, \ldots, p$, that maximize the following penalized quasi-likelihood function

$$\sum_{i=1}^n \ell \left[g^{-1} \left\{ \sum_{k=1}^p \beta_k(X_{ik}) + \alpha(\mathbf{S}_i) \right\}, Y_i \right] - \frac{\lambda}{2} \mathcal{E}(\alpha). \quad (4)$$

Directly solving this penalized maximization is challenging since it involves unstructured nonparametric functions which are subject to the "curse of dimensionality." To overcome this difficulty, we consider some smoothing method. Under some suitable smoothness conditions, β_k's and α can be well approximated by the univariate spline basis functions and the Bernstein basis polynomials introduced in Sections 2.2.1 and 2.2.2, respectively.

Let $u_{kj}(x_k)$, $j \in \mathcal{J}$ be the original B-spline basis functions for the kth covariate, where $\mathcal{J}$ is the index set of the basis functions. Let $u^0_{kj}(x_k) = u_{kj}(x_k) - \frac{\mathrm{E}u_{kj}(X_k)}{\mathrm{E}u_{k1}(X_k)} u_{k1}(x_k)$, $U_{kj}(x_k) = \frac{u^0_{kj}(x_k)}{\mathrm{SD}\{u^0_{kj}(X_k)\}}$, $j \in \mathcal{J}$, then $\mathrm{E}U_{kj}(X_k) = 0$ and $\mathrm{E}U^2_{kj}(X_k) = 1$. Suppose the nonlinear component can be well approximated by a spline function so that, for all $x_k \in [a_k, b_k]$, $\beta_k(x_k) \approx \sum_{j=1}^{J_n+\varrho+1} \theta_{kj} U_{kj}(x_k) = \mathbf{U}_k^\top(x_k)\boldsymbol{\theta}_k$, where $\mathbf{U}_k(x_k) = \big(U_{k1}(x_k), \ldots, U_{k,J_n+\varrho+1}(x_k)\big)^\top$ and $\boldsymbol{\theta}_k = \big(\theta_{k1}, \ldots, \theta_{k,J_n+\varrho+1}\big)^\top$ are a vector of coefficients. Let $\mathbf{S}_{n\times 2} = \{(S_{i1}, S_{i2})\}_{i=1}^n$ be the location design matrix and $\mathbf{X}_{n\times p} = \{(X_{i1}, \ldots, X_{ip})\}_{i=1}^n$ be the collection of all covariates. For $i = 1, \ldots, n$, let $\mathbf{B}_i^\top = \{B_m(\mathbf{S}_i), m \in \mathcal{M}\}$, and denote $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_n)^\top$ the $n \times |\mathcal{M}|$ evaluation matrix of Bernstein basis polynomials.

Then maximizing the penalized quasi-likelihood function in (4) is equivalent to minimizing

$$-\sum_{i=1}^n \ell \left[g^{-1} \left\{ \sum_{k=1}^p \mathbf{U}_k(X_{ik})^\top \boldsymbol{\theta}_k + \mathbf{B}(\mathbf{S}_i)^\top \boldsymbol{\gamma} \right\}, Y_i \right] + \frac{1}{2}\lambda \boldsymbol{\gamma}^\top \mathbf{P} \boldsymbol{\gamma} \text{ subject to } \boldsymbol{\Psi}\boldsymbol{\gamma} = \mathbf{0}, \quad (5)$$

where $\mathbf{P}$ is the block diagonal penalty matrix satisfying that $\boldsymbol{\gamma}^\top \mathbf{P} \boldsymbol{\gamma} = \mathcal{E}(\mathbf{B}\boldsymbol{\gamma})$. See Section B.2.2 in the supplementary materials for the formula used to construct $\mathbf{P}$. It is worth noting that according to Assumption (A3), the optimization problem in (5) has a unique solution.

To solve the constrained minimization problem (5), we first remove the constraint via the following QR decomposition: $\boldsymbol{\Psi}^\top = \mathbf{Q}\mathbf{R} = (\mathbf{Q}_1\ \mathbf{Q}_2)\binom{\mathbf{R}_1}{\mathbf{R}_2}$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}_1$ is an upper triangle matrix, the submatrix $\mathbf{Q}_1$ is the first r (r is the rank of matrix $\boldsymbol{\Psi}$) columns of $\mathbf{Q}$, and $\mathbf{R}_2$ is a matrix of zeros. We reparametrize using $\boldsymbol{\gamma} = \mathbf{Q}_2 \boldsymbol{\gamma}^*$ for some $\boldsymbol{\gamma}^*$, then it is guaranteed that $\boldsymbol{\Psi}\boldsymbol{\gamma} = \mathbf{0}$. The problem (5) is now converted to a conventional penalized minimization problem without any restriction

$$L^P(\boldsymbol{\theta}, \boldsymbol{\gamma}^*) = -\sum_{i=1}^n \ell \left[g^{-1} \left\{ \sum_{k=1}^p \mathbf{U}_k(X_{ik})^\top \boldsymbol{\theta}_k + \mathbf{B}(\mathbf{S}_i)^\top \mathbf{Q}_2 \boldsymbol{\gamma}^* \right\}, Y_i \right] + \frac{\lambda}{2} \boldsymbol{\gamma}^{*\top} \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\gamma}^*. \quad (6)$$

For practitioners, the construction of $\mathbf{P}$ and $\mathbf{Q}_2$ can be carried out via the R package `BPST` (Wang et al. 2019).

Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\gamma}}^*$ be the minimizer of (6), that is, $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}^*) = \arg\min L^P(\boldsymbol{\theta}, \boldsymbol{\gamma}^*)$, then, the univariate spline estimator of $\beta_k(x_k)$ and bivariate spline estimator of $\alpha(\mathbf{s})$ are

$$\widehat{\beta}_k(x_k) = \mathbf{U}_k(x_k)^\top \widehat{\boldsymbol{\theta}}_k,\ \widehat{\alpha}(\mathbf{s}) = \mathbf{B}(\mathbf{s})^\top \widehat{\boldsymbol{\gamma}} = \sum_{m\in\mathcal{M}} B_m(\mathbf{s}) \widehat{\gamma}_m,$$

where the estimated spline coefficients are $\widehat{\boldsymbol{\gamma}} = \{\widehat{\gamma}_m, m \in \mathcal{M}\}^\top = \mathbf{Q}_2 \widehat{\boldsymbol{\gamma}}^*$.

We first establish the L_2 convergence rate of the spline estimators $\widehat{\beta}_k(x_k)$ and $\widehat{\alpha}(\mathbf{s})$. For any Lebesgue measurable function $\psi(\mathbf{u})$ on a domain $\mathcal{D}$, for example, $\mathcal{D} = [a_k, b_k]$ or $\Omega \subseteq \mathbb{R}^2$, let $\|\psi\|^2_{L_2} = \int_{\mathcal{D}} \psi^2(\mathbf{u}) d\mathbf{u}$ and $H = J_n^{-1}$.

Theorem 1. Under Assumptions (A1)–(A6) in Appendix A, the spline estimators $\widehat{\beta}_k(x_k)$ and $\widehat{\alpha}(\mathbf{s})$ satisfy that

$$\sum_{k=1}^p \|\widehat{\beta}_k - \beta_k\|_{L_2} + \|\widehat{\alpha} - \alpha\|_{L_2} = O_{\mathrm{a.s.}} \left\{ (H^{-1/2} + |\triangle|^{-1}) \left(\frac{\log n}{n} \right)^{1/2} + H^{\varrho+1} + |\triangle|^{d+1} + \frac{\lambda}{n|\triangle|^4} \right\}.$$

The proof is provided in the supplementary materials.

3. Two-Stage Estimator and Simultaneous Confidence Band

It is very difficult to derive the asymptotic distribution for the spline estimators introduced in Section 2, so no measures of confidence can be assigned to these estimators. To represent the uncertainty in the estimate of the nonlinear effect of the covariates, we propose the SCB for β_k's via the backfitting method using spline estimators obtained in Section 2 as pilot followed by local polynomial estimators.

The basic idea is that for every $k = 1, \ldots, p$, we estimate the kth additive function $\beta_k(\cdot)$ in model (1) nonparametrically by assuming that other nonparametric components $\{\beta_{k'}(\cdot) : k' \neq k\}$ are known. The problem turns into a univariate function estimation problem. Denote K a continuous kernel function, and let $K_h(u) = K(u/h)/h$ be a rescaling of K, where h is usually called the bandwidth. This leads to an "oracle" smoother $\widehat{\beta}_k^o$ of β_k, where

$$\begin{aligned}&\{\widehat{\beta}_k^o(x_k), \widehat{\beta}_k^{o\prime}(x_k)\}\\ &= \arg\max_{a_0,a_1} \sum_{i=1}^{n} \ell\left[g^{-1}\left\{a_0 + a_1(X_{i,k} - x_k) + \alpha(\mathbf{S}_i) + \sum_{k'\neq k}^{p} \beta_{k'}(X_{ik'})\right\}, Y_i\right] K_{h_k}(X_{ik} - x_k).\end{aligned}$$

Asymptotic properties of smoothers of $\widehat{\beta}_k^o(x_k)$ can be easily established based on these assumptions. We say $x \in \chi_k = [a_k, b_k]$ is a boundary point if and only if $x = a_k + ch_k$ or $x = b_k - ch_k$ for some $0 \leq c < 1$ and an interior point otherwise. Let χ_{h_k} be the interior of the support χ_k. Let $\nu_2 = \int u^2 K(u)\, du$ and f_k be the probability density function of X_k. For $j = 1, 2$, let $\rho_j(x) = \{dg^{-1}(x)/dx\}^j \big/ \{\sigma^2 V(g^{-1}(x))\} = \left[\{g'(g^{-1}(x))\}^j \sigma^2 V(g^{-1}(x))\right]^{-1}$. For the quasi-likelihood function $\ell\{g^{-1}(x), y\}$, let $q_1(x, y) = \frac{\partial}{\partial x}\ell\{g^{-1}(x), y\}$ and $q_2(x, y) = \frac{\partial^2}{\partial x^2}\ell\{g^{-1}(x), y\}$. It is clear that $q_1(x, y) = \{y - g^{-1}(x)\}\rho_1(x)$ and $q_2(x, y) = \{y - g^{-1}(x)\}\rho_1'(x) - \rho_2(x)$ hold. Let $C(K) = \int K'(u)^2 du / \int K(u)^2 du$, $a_h = \sqrt{-2\log(h)}$, and for any $\alpha \in (0, 1)$, denote the quantile

$$Q_h(\alpha) = a_h + a_h^{-1}[\log\{\sqrt{C(K)}/2\pi\} - \log\{-\log\sqrt{1-\alpha}\}].$$

Theorem 2 shows the pointwise and uniform asymptotically normality of the "oracle" estimator $\widehat{\beta}_k^o(x_k)$.

Theorem 2. Suppose Assumptions (A1)–(A4) and (A7) in Appendix A hold. Then, for any $k = 1, \ldots, p, x_k \in \chi_{h_k}$, if $nh_k^5 = O(1)$, we have, as $n \to \infty$,

$$\sigma_{n,k}^{-1}(x_k)\left[\widehat{\beta}_k^o(x_k) - \beta_k(x_k) - \frac{\beta_k''(x_k)}{2}\nu_2 h_k^2\right] \to N(0, 1),$$

and if Assumptions (A1)–(A4), (A7), and (A8)(ii) are satisfied, then

$$\lim_{n\to\infty} \mathrm{P}\left\{\sup_{x_k\in\chi_{h_k}} \sigma_{n,k}^{-1}(x_k)|\widehat{\beta}_k^o(x_k) - \beta_k(x_k)| \leq Q_{h_k}(\alpha)\right\} = 1 - \alpha,$$

where $\sigma_{n,k}^2(x_k) = n^{-1}h_k^{-1}\mathrm{E}\left[\rho_2\{\sum_{k=1}^p \beta_k(X_k) + \alpha(\mathbf{S})\}|X_k = x_k\right]^{-1} f_k(x_k)^{-1} \int K^2(u)du$.

Since the true functions β_k's and α are unknown, we replace those with the pilot estimators $\widehat{\beta}_k$ and $\widehat{\alpha}$ obtained in Section 2 and apply the above backfitting idea to obtain the spline-backfitted local polynomial (SBL) estimator

$$\begin{aligned}&\{\widehat{\beta}_k^{\mathrm{SBL}}(x_k), \widehat{\beta}_k^{\mathrm{SBL}\prime}(x_k)\}\\ &= \arg\max_{a_0,a_1} \sum_{i=1}^{n} \ell\left[g^{-1}\left\{a_0 + a_1(X_{i,k} - x_k) + \widehat{\alpha}(\mathbf{S}_i) + \sum_{k'\neq k}^{p} \widehat{\beta}_{k'}(X_{ik'})\right\}, Y_i\right] K_{h_k}(X_{ik} - x_k).\end{aligned}$$

Theorem 3. Under Assumptions (A1)–(A5), (A6′), (A7), and (A8)(i) in Appendix A, the SBL estimator satisfies

$$\begin{aligned}&\sup_{x_k\in\chi_{h_k}} |\widehat{\beta}_k^{\mathrm{SBL}}(x_k) - \widehat{\beta}_k^o(x_k)| = O_{a.s.}(n^{-1/2}\log n),\\ &\sup_{x_k\in\chi_{h_k}} h_k|\widehat{\beta}_k^{\mathrm{SBL}\prime}(x_k) - \widehat{\beta}_k^{o\prime}(x_k)| = O_{a.s.}(n^{-1/2}\log n).\end{aligned}$$

The proofs of Theorems 2 and 3 are given in the supplementary materials. Theorem 3 implies that the difference between the SBL estimator and the "oracle" estimator is negligible compared to the difference between the "oracle" estimator $\widehat{\beta}_k^o(x_k)$ and the true function $\beta_k(x_k)$, that is, the SBL estimator $\widehat{\beta}_k^{\mathrm{SBL}}(x_k)$ is oracle-efficient. Consequently, the SBL estimator inherits the same asymptotic normality (both pointwise and uniform asymptotically normality) from the "oracle" estimator, as stated in Corollary 1 below, which follows directly from Theorems 2 and 3, thus, the proof is omitted.

Corollary 1. Under Assumptions (A1)–(A5), (A6′), (A7), and (A8)(i) in Appendix A, for any $x_k \in \chi_{h_k}$, we have, as $n \to \infty$,

$$\sigma_{n,k}^{-1}(x_k)\left\{\widehat{\beta}_k^{\mathrm{SBL}}(x_k) - \beta_k(x_k) - \beta_k''(x_k)\nu_2 h_k^2/2\right\} \to N(0, 1).$$

If Assumptions (A1)–(A5), (A6′), (A7), and (A8)(ii) are satisfied, then, for any $\alpha \in (0, 1)$ and $k = 1, \ldots, p$,

$$\lim_{n\to\infty} \mathrm{P}\left\{\sup_{x_k\in\chi_{h_k}} \sigma_{n,k}^{-1}(x_k)|\widehat{\beta}_k^{\mathrm{SBL}}(x_k) - \beta_k(x_k)| \leq Q_{h_k}(\alpha)\right\} = 1 - \alpha.$$

Corollary 1 yields the following $100(1-\alpha)\%$ SCB for $\beta_k(x_k)$

$$\widehat{\beta}_k^{\mathrm{SBL}}(x_k) \pm \sigma_{n,k}(x_k)Q_{h_k}(\alpha),\ x_k \in \chi_{h_k}. \tag{7}$$

4. Implementation

4.1. First Stage

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ be the vector of n observations of the response variable. Denote $\mathbf{U}_i = \left\{\mathbf{U}_1(\mathbf{X}_{i1})^\top, \ldots, \mathbf{U}_p(\mathbf{X}_{ip})^\top\right\}^\top$ and $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_n)^\top$. We use the following iteratively reweighted least square (IRLS) to minimize the objective function in (6), which leads to enormous reductions in computational complexity. Let $\mathbf{G}^{(k)}$, $\mathbf{V}^{(k)}$, and $\boldsymbol{\rho}_2^{(k)}$ be diagonal matrices with elements $g'(\mu_i^{(k)})$, $V(\mu_i^{(k)})$, and $\rho_2(\mu_i^{(k)})$, respectively. At

the $(k+1)$th iteration, we consider the following objective function

$$R^{(k+1)} = \left\| \{\mathbf{V}^{(k)}\}^{-1/2} \left\{ \mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{\gamma}^*) \right\} \right\|^2 + \frac{\lambda}{2} \boldsymbol{\gamma}^{*\top} \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\gamma}^*,$$

where $\boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{\gamma}^*) = g^{-1}(\mathbf{U}^\top \boldsymbol{\theta} + \mathbf{B}^\top \mathbf{Q}_2 \boldsymbol{\gamma}^*)$, and it can be approximated by its first-order Taylor expansion around $(\boldsymbol{\theta}^{(k)}, \boldsymbol{\gamma}^{*(k)})$

$$\boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{\gamma}^*) \approx \boldsymbol{\mu}^{(k)} - \mathbf{J}^{(k)} \left\{ \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\gamma}^* \end{pmatrix} - \begin{pmatrix} \boldsymbol{\theta}^{(k)} \\ \boldsymbol{\gamma}^{*(k)} \end{pmatrix} \right\}$$

with $\boldsymbol{\mu}^{(k)} = \boldsymbol{\mu}(\boldsymbol{\theta}^{(k)}, \boldsymbol{\gamma}^{*(k)}) = \{\mu_i^{(k)}\}_{1 \le i \le n}^\top$ and $\mathbf{J}^{(k)} = \{\mathbf{G}^{(k)}\}^{-1}(\mathbf{U} \ \mathbf{B}\mathbf{Q}_2)$ being the "Jacobian" matrix. Therefore,

$$R^{(k+1)} \approx \left\| \{\mathbf{V}^{(k)}\}^{-1/2} \left[\mathbf{Y} - \boldsymbol{\mu}^{(k)} - \mathbf{J}^{(k)} \left\{ \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\gamma}^* \end{pmatrix} - \begin{pmatrix} \boldsymbol{\theta}^{(k)} \\ \boldsymbol{\gamma}^{*(k)} \end{pmatrix} \right\} \right] \right\|^2 + \frac{\lambda}{2} \boldsymbol{\gamma}^{*\top} \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\gamma}^*.$$

Then, we have

$$\begin{aligned} R^{(k+1)} &\approx \left\| \{\mathbf{V}^{(k)}\}^{-1/2} \{\mathbf{G}^{(k)}\}^{-1} \left[\mathbf{G}^{(k)} (\mathbf{Y} - \boldsymbol{\mu}^{(k)}) + \boldsymbol{\eta}^{(k)} \right. \right. \\ &\quad \left. \left. - \mathbf{U}^\top \boldsymbol{\theta} - \mathbf{B}^\top \mathbf{Q}_2 \boldsymbol{\gamma}^* \right] \right\|^2 + \frac{\lambda}{2} \boldsymbol{\gamma}^{*\top} \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\gamma}^* \\ &= \left\| (\boldsymbol{\rho}_2^{(k)})^{1/2} (\widetilde{\mathbf{Y}}^{(k)} - \mathbf{U}^\top \boldsymbol{\theta} - \mathbf{B}^\top \mathbf{Q}_2 \boldsymbol{\gamma}^*) \right\|^2 \qquad (8) \\ &\quad + \frac{\lambda}{2} \boldsymbol{\gamma}^{*\top} \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\gamma}^*, \qquad (9) \end{aligned}$$

where $\boldsymbol{\eta}^{(k)} = (\eta_1^{(k)}, \ldots, \eta_n^{(k)})^\top$ with $\eta_i^{(k)} = \mathbf{U}_i^\top \boldsymbol{\theta}^{(k)} + \mathbf{B}_i^\top \mathbf{Q}_2 \boldsymbol{\gamma}^{*(k)}$, $\widetilde{\mathbf{Y}}^{(k)} = (\widetilde{Y}_1^{(k)}, \ldots, \widetilde{Y}_n^{(k)})^\top$ with $\widetilde{Y}_i^{(k)} = g'(\mu_i^{(k)})(Y_i - \mu_i^{(k)}) + \eta_i^{(k)}$. By minimizing (9) with respect to $(\boldsymbol{\theta}, \boldsymbol{\gamma}^*)$, we obtain $(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\gamma}^{*(k+1)})$. Notice that to start the iteration, we only need $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\eta}^{(0)}$, not $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\gamma}^{*(0)})$, which simplifies the procedure of choosing initial value. In the Poisson case, we set $\mu_i^{(0)} = y_i + 0.01$ and $\eta_i^{(0)} = g(\mu_i^{(0)})$.

4.1.1. Knots Selection

For univariate spline smoothing, we suggest placing knots on a grid of evenly spaced sample quantiles. Assumption (A6) in Appendix A suggests that the number of knots J_n needs to satisfy: $n^{1/(2\varrho+2)} (\log n)^{-1/(2\varrho+2)} \ll J_n \ll n^{2/5}$, where $\varrho \ge 1$ is the degree of the polynomial spline basis functions. The widely used quadratic/cubic splines both satisfy this condition. In practice we suggest taking the following rule-of-thumb number of interior knots: $J_n = \min\{\lfloor c_1 n^{1/(2\varrho+2)} \rfloor, \lfloor n/(4p) \rfloor\} + 1$, where c_1 is a tuning parameter (typically, $c_1 \in [1, 5]$), and the term $n/(4p)$ is to guarantee that there are at least four observations in each subinterval between two adjacent knots to avoid getting (near) singular design matrices in smoothing.

4.1.2. Triangulation Selection

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, a "good" triangulation usually refers to those with well-shaped triangles, no small angles or/and no obtuse angles. Other criteria include the density control (adaptivity) and optimal size (number of triangles), etc. For a fixed number of triangles, Lai and Schumaker (2007) and Lindgren, Rue, and Lindström (2011) recommend selecting the triangulation according to "max-min" criterion which maximizes the minimum angle of all the angles of the triangles in the triangulation. Monte Carlo experiments in our simulation studies show that there must be enough triangles to capture the features of the surface, but once this minimum necessary number of triangles has been reached, further increasing the number of triangles usually has little effect on the fitting process. Therefore, one needs 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. In practice, if the boundary of the spatial domain is not too complicated, we suggest taking the number of triangles as the following: $N_n = \min\{\lfloor c_2 n^{1/(d+1)} \rfloor, n/4\} + 1$, for some tuning parameter c_2 (typically, $c_2 \in [1, 20]$). However, when the spatial domain does look complicated, N_n can be set (much) larger than n so that the domain can be very well approximated by the triangulation, and the penalty automatically and gracefully handles the situation. Once N_n is chosen, one can build the triangulated meshes using typical triangulation construction methods such as Delaunay triangulation.

4.1.3. Roughness Penalty Selection

The roughness penalty parameter λ can be selected using data-driven approaches, for example, the generalized cross-validation (GCV; Craven and Wahba 1979; Wahba 1990) is such a criterion and is widely used for choosing the penalty parameter. Let $\widehat{\boldsymbol{\rho}}_2$ be the diagonal matrix with elements $\rho_2(\widehat{\mu}_i)$, and

$$\mathbf{C}(\lambda) = \begin{pmatrix} \mathbf{U}^\top \widehat{\boldsymbol{\rho}}_2 \mathbf{U} & \mathbf{U}^\top \widehat{\boldsymbol{\rho}}_2 \mathbf{B} \mathbf{Q}_2 \\ \mathbf{Q}_2^\top \mathbf{B}^\top \widehat{\boldsymbol{\rho}}_2 \mathbf{U} & \mathbf{Q}_2^\top (\mathbf{B}^\top \widehat{\boldsymbol{\rho}}_2 \mathbf{B} + \lambda \mathbf{P}) \mathbf{Q}_2 \end{pmatrix}.$$

Denote the smoothing or hat matrix as $\mathbf{S}(\lambda) = (\mathbf{U} \ \mathbf{B}\mathbf{Q}_2) \mathbf{C}(\lambda)^{-1} (\mathbf{U} \ \mathbf{B}\mathbf{Q}_2)^\top \widehat{\boldsymbol{\rho}}_2$. Let $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)^\top$ with $\widetilde{Y}_i = g'(\widehat{\mu}_i)(Y_i - \widehat{\mu}_i) + \widehat{\eta}_i$. We minimize the following

$$\mathrm{GCV}(\lambda) = \frac{n \| \widehat{\boldsymbol{\rho}}_2^{1/2} \{\widetilde{\mathbf{Y}} - \mathbf{S}(\lambda) \widetilde{\mathbf{Y}}\} \|^2}{\{n - \mathrm{tr}(\mathbf{S}(\lambda))\}^2}$$

over a grid of values to select λ.

4.2. Second Stage

The performance of the SBL estimator is also dependent upon the bandwidth selection and other estimators of the parameters.

4.2.1. Bandwidth Selection

Note that Assumption (A8) in Appendix A requires that the bandwidths in the backfitting are of order around $n^{-1/5}$. Thus, the bandwidth selection can be done using a standard routine in the literature, see discussions in Fan, Heckman, and Wand (1995) and Fan and Gijbels (1996). By minimizing the asymptotic mean integrated squared errors (AMISE), we can obtain the optimal bandwidth

$$h_{k,\mathrm{AMISE}} = n^{-1/5} C_K^{1/5} \left[\frac{\int \rho_{2,k}(x_k) dx_k}{\int \beta_k''(x_k)^2 f_k(x_k) dx_k} \right]^{1/5},$$

where $\rho_{2,k}(x_k) = \mathrm{E}\left[\rho_2\{\sum_{k=1}^{p}\beta_k(X_k)+\alpha(\mathbf{S})\}|X_k = x_k\right]$. The optimal bandwidth could be approximated by

$$n^{-1/5}C_K^{1/5}\left[\frac{n^{-1}\sum_{i=1}^{n}\{\widehat{\rho}_{2,k}(X_{ik})\}^{-1}\widehat{f}_k^{-1}(X_{ik})}{n^{-1}\sum_{i=1}^{n}\widehat{\beta}_k''(X_{ik})^2}\right]^{1/5},$$

where $\widehat{f}_k$ is estimated by kernel density estimation, $\widehat{\beta}_k''$ is obtained from the spline estimator $\widehat{\beta}_k$ and $\widehat{\rho}_{2,k}(X_{ik})$ is an estimator of $\rho_{2,k}(X_{ik})$. Let $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - g^{-1}(\widehat{\eta}_i)\}^2/V(g^{-1}(\widehat{\eta}_i))$, $\widehat{\rho}_{2,k}(x_k) = n^{-1}\sum_{i=1}^{n}\left\{(g'(g^{-1}(\widehat{\eta}_i)))^2\widehat{\sigma}^2V(g^{-1}(\widehat{\eta}_i))\right\}^{-1}K_h(X_{ik}-x_k)$ with a rule-of-the-thumb bandwidth h. To construct pointwise confidence intervals or SCBs of the univariate functions, we suggest taking $h_k = h_{k,\mathrm{AMISE}}(\log n)^{-1/4}$ to satisfy the requirement of Assumption (A8).

4.2.2. Estimating $\sigma_{n,k}(x_k)$

For any $k = 1,\ldots,p$, we obtain the estimator of $\sigma_{n,k}(x_k)$ by $\widehat{\sigma}_{n,k}(x_k) = n^{-1/2}h_k^{-1/2}\widehat{\rho}_{2,k}^{-1}(x_k)\widehat{f}_k^{-1}(x_k)\int K^2(u)\,du$.

5. Simulation

In this section, we conduct two simulation studies to evaluate the finite sample performance of the proposed method.

5.1. Example 1

We consider a modified horseshoe domain in Sangalli, Ramsay, and Ramsay (2013) and randomly sample $n = 1000$ and 2000 locations on the domain. The response variable Y_i is generated from Poisson distribution and negative binomial distribution with log link functions as described in the following two cases.

Case I: Poisson distribution. $Y_i \sim \mathrm{Poisson}(\mu_i)$, where $\log(\mu_i) = \sum_{k=1}^{3}\beta_k(X_{ik}) + \alpha(\mathbf{S}_i)$, $i = 1,\ldots,n$.

Case II: Negative binomial distribution. $Y_i \sim \mathrm{NB}(\theta, \mu_i/(\theta+\mu_i))$, where $\theta = 5$ and $\log(\mu_i)$ is the same as Case I and $\mathrm{E}(Y_i) = \mu_i$ and $\mathrm{var}(Y_i) = \mu_i + \mu_i^2/\theta$.

We conduct 1000 Monte Carlo replications. In each replication, X_{i1}, X_{i2}, and X_{i3} are uniformly generated from $[0,1]$, and $\mathbf{S}_i$ is uniformly generated from the horseshoe domain, and X_{i1}, X_{i2}, and X_{i3} and $\mathbf{S}_i$ are independent of each other. The true univariate functions $\beta_1(x)$, $\beta_2(x)$, $\beta_3(x)$ are illustrated in Figure 2(a)–(c), respectively, where $\beta_1(x) = 1/2\sin(2\pi x) - \cos(\pi x) - \mathrm{E}\{1/2\sin(2\pi X_1) - \cos(\pi X_1)\}$, $\beta_2(x) = 4(x-0.5)^2 - \mathrm{E}\{4(X_2-0.5)^2\}$, and $\beta_3(x) = x - \mathrm{E}(X_3)$. Figure 5(a) shows the contour map of the true bivariate function $\alpha(\cdot)$, which is modified based on a similar function given in Wood, Bravington, and Hedley (2008) and Sangalli, Ramsay, and Ramsay (2013).

To implement the proposed procedure, one needs to select the knots for a univariate spline, the triangulation for a bivariate spline, the bandwidth for the local linear method, as well as the smoothness penalty parameter. We have conducted extensive simulations to examine whether and how sensitive the choice of knots, triangulation, and bandwidth on the performance of the proposed method. For all the univariate splines, we use cubic B-splines with the number of interior knots $J_n = 2, 4, 6, 8$ equally spaced on the sample quantiles. For the bivariate spline smoothing, we consider $d = 2$, $r = 1$ with three different triangulations shown in Figure 3(a)–(c). There are 109 triangles (95 vertices), 163 triangles (124 vertices), and 237 triangles (165 vertices) in $\triangle_1$, $\triangle_2$, and $\triangle_3$, respectively.

We compare our method with the thin plate spline method (TPS) and the soap film smoother (SOAP), which are commonly used for fitting GAMs. The TPS and the SOAP estimators are obtained from the *mgcv* package in R (Wood 2017). For all three methods, GCV is used to choose the values of the penalty parameter. In Case II, the estimator of θ is chosen to ensure that the Pearson estimate of the scale parameter is as close as possible to 1.

To evaluate the accuracy of the estimators, we calculate the mean integrated squared error (MISE) for each of the components based on 1000 Monte Carlo samples. Also, to illustrate the prediction capability, we conduct the 10-fold cross-validation for each Monte Carlo sample and compare the cross-validated mean squared prediction error (MSPE). Table 1 presents the MISE of β_1, β_2, β_3, α and the 10-fold cross-validated MSPE, where the SBL results are based on using four interior knots for univariate splines and $\triangle_1$ for bivariate splines. Table 1 shows that the performance of our method and the TPS method. All three methods are similar in terms of the estimation of $\beta_k(\cdot)$'s, however, our method significantly outperforms the TPS and SOAP when estimating the bivariate function $\alpha(\cdot)$, which results in a big improvement of the cross-validated MSPE.

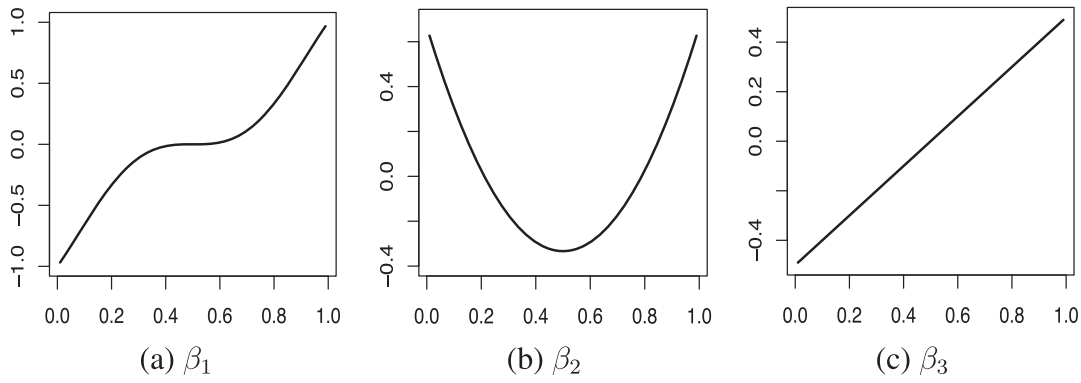
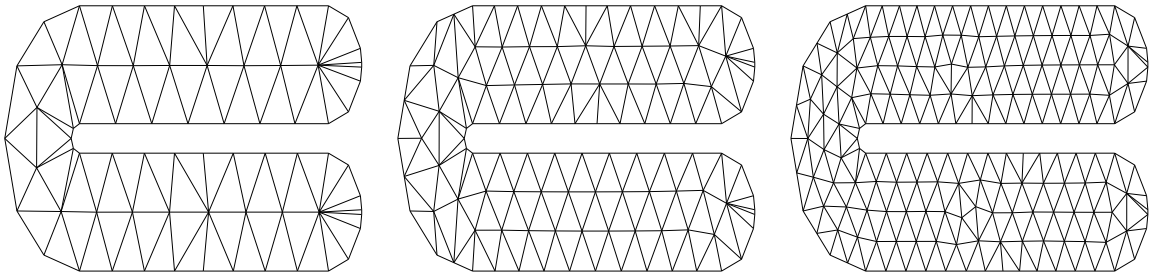


Figure 2. The true univariate component functions in Example 1.



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

Figure 3. Triangulations on the horseshoe domain.

Table 1. MISE of the functional component estimators and 10-fold cross-validated MSPE.

Family	n	Method	MISE				MSPE
			β_1	β_2	β_3	α	
Poisson	1000	SBL	0.0051	0.0037	0.0035	0.8057	1.6186
		TPS	0.0059	0.0045	0.0030	4.1705	2.1645
		SOAP	0.0057	0.0042	0.0034	1.0398	2.0512
	2000	SBL	0.0027	0.0019	0.0016	0.4465	1.4832
		TPS	0.0031	0.0023	0.0017	3.6656	2.0955
		SOAP	0.0029	0.0022	0.0017	0.6361	1.9065
Negative binomial	1000	SBL	0.0079	0.0063	0.0064	0.8490	3.1769
		TPS	0.0079	0.0050	0.0025	1.6768	3.8489
		SOAP	0.0074	0.0046	0.0023	0.8542	3.5008
	2000	SBL	0.0040	0.0034	0.0033	0.4662	2.9303
		TPS	0.0034	0.0030	0.0013	1.4669	3.7571
		SOAP	0.0033	0.0026	0.0012	0.6580	3.3372

Figure 4 depicts the true univariate function β_k (dotted curve). It also shows the corresponding estimator $\widehat{\beta}_k^{\text{SBL}}$ (solid curve) and the 95% SCB for β_k (gray bands) from a typical run generated from Case I or Case II with sample size $n = 2000$, where the estimation and SCB construction are based on four interior quantile knots for the univariate splines and triangulation $\triangle_1$ for the bivariate splines. Figure 5(b)–(g) shows the contour maps of the estimated bivariate functions $\widehat{\alpha}$ over a grid of 500×200 points using our method, TPS and SOAP.

We evaluate the coverage of the proposed SCBs over 20 equally spaced points on $[0, 1]$ and test whether the true functions are covered by the SCBs at these points. Table 2 summarizes the empirical coverage rate of the 95% SCBs from 1000 Monte Carlo experiments. The results clearly show a very good coverage rate of the SCBs.

Figures B.2 and B.3 in the supplementary materials present the MISEs of the estimators based on different combinations of knots and triangulations for Case I and Case II. For the univariate components, one sees that the MISEs are very similar regardless to the choice of knots and triangulations. For the bivariate function α, we have found in our simulation studies that there is a minimum adequate value of the number of triangles in the fitting. Fits using fewer than this minimum number of triangles have low statistical accuracy. We have also found that, when this minimum number of triangles is reached, further refining the triangulation will have little effect on the fitting process, but make the computational burden unnecessarily heavy.

Finally, our proposed method is very user-friendly and computationally efficient. Take the case of fitting GGAMs with the Poisson distribution as an example. Remarkably, it takes only 3.1 sec to estimate all the components in the GGAM with 2000 observations on a standard PC with processor Core i5 @2.7GHz CPU and 8.00GB RAM. This is extremely fast considering that the entire nonparametric regression is done without WARPing.

5.2. Example 2

We conduct another simulation study using the covariates and domain of the crash data analyzed in Section 6. We consider the following negative binomial setting: $Y_i \sim \text{NB}\left(\theta, \frac{\mu_i}{\theta+\mu_i}\right)$, where $\theta = 2.7$, $\log(\mu_i) = \sum_{k=1}^{12} \beta_k(X_{ik}) + \alpha(\mathbf{S}_i)$, $i = 1, \ldots, 1761$, and X_{ik}, $k = 1, \ldots, 12$, are the same covariates as in the crash dataset described in Section 6. The significant univariate functions and bivariate function are set to be the same as the estimates obtained in the crash data analysis, and insignificant univariate functions are set as zero.

The average MISE of each functional component from 1000 Monte Carlo experiments is reported in Table 3. From Table 3, one sees that the MISE are all very small which indicates our method performs very well. We also examine the behavior of the proposed SCBs for the univariate functions. The "coverage" rows in Table 3 summarize the empirical coverage rate of the 95% SCBs from 1000 Monte Carlo experiments, and the results seem to be very reasonable. See Figures B.4 and B.5 in the supplementary materials for the estimators and the SCBs from a typical simulation trial.

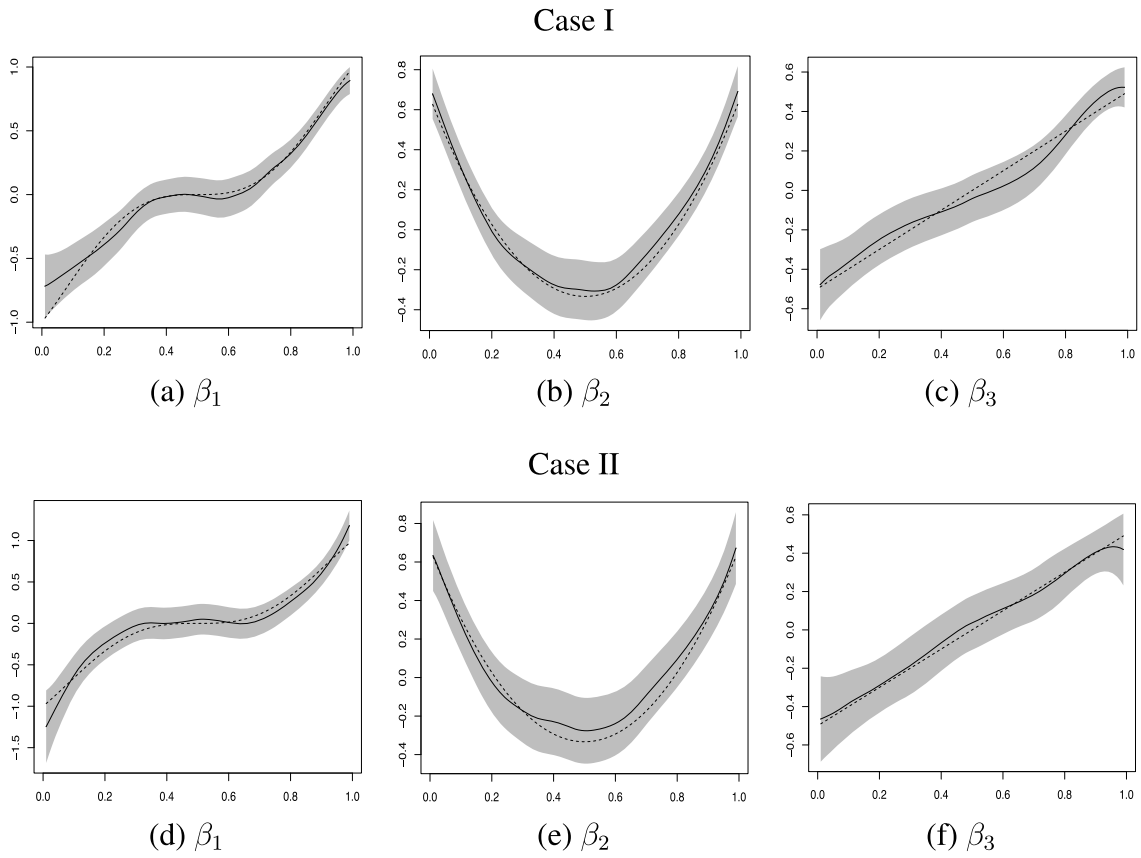


Figure 4. Plots of $\beta_k(x_k)$ (dotted curve), the SBL estimator (solid curve), and the 95% SCBs (gray bands), $k = 1, 2, 3$, based on $n = 2000$ observations.

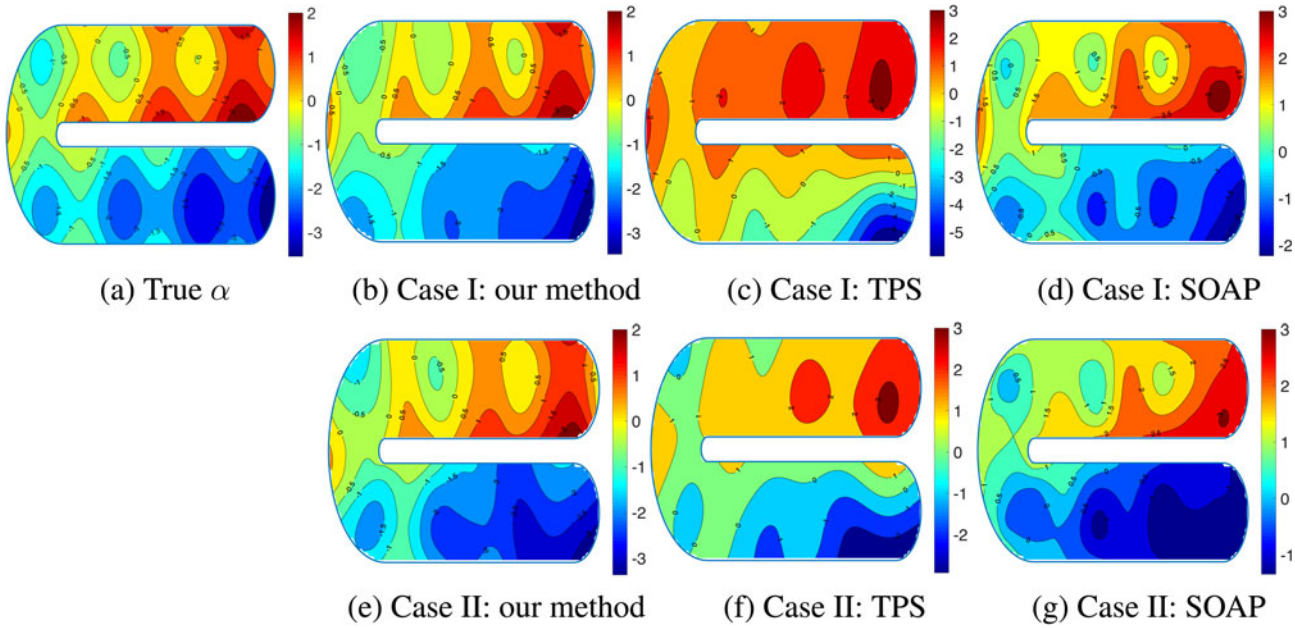


Figure 5. Contour maps for the true bivariate function and its estimators.

Table 2. The coverage rate of the 95% SCBs for univariate functions.

n	Poisson			Negative binomial		
	β_1	β_2	β_3	β_1	β_2	β_3
1000	0.945	0.949	0.967	0.949	0.967	0.960
2000	0.947	0.965	0.979	0.980	0.980	0.970

6. Application to Crash Data

6.1. Domain of Interest

Traffic crashes have been one of the major sources of fatalities and injuries in the United States. Crash frequency analysis is critical for developing and implementing effective safety

Table 3. The MISE of the estimators of the component functions and the coverage rate of the 95% SCB for the univariate functions.

Measurement	Component						
	β_1	β_2	β_3	β_4	β_5	β_6	
MISE	0.0025	0.0017	0.0015	0.0019	0.0033	0.0017	
Coverage	0.786	0.959	0.958	0.954	0.890	0.964	
	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	α
MISE	0.0021	0.0018	0.0016	0.0028	0.0030	0.0032	0.0318
Coverage	0.962	0.963	0.967	0.973	0.965	0.991	–

improvement programs. In this study, we are interested in identifying the spatial pattern of crashes and investigating how demographic, economic, and commuting factors influence crash frequency after the spatial effect is adjusted.

This study focuses on the Tampa-St. Petersburg urbanized area, including Tampa, Clearwater, and St. Petersburg, is a major populated area surrounding Tampa Bay on the west coast of Florida, United States. The area consists of 1761 census block groups, as shown in Figure 1(a) and (b). A census block group is a geographical unit designed by the U.S. Census Bureau to be "relatively homogeneous units with respect to population characteristics, economic status, and living conditions" (Song et al. 2006). Rural areas were excluded from this study as their population densities are often too small to have meaningful crash frequency data. The Tampa Bay and Wilderness Preserve in the north-east are also excluded since they have no traffic. Crash frequency data occurred on a roadway owned or operated by a nonstate entity for each census tract in the year of 2014 were collected from the Florida Department of Transportation.

In addition, 12 covariates involving in demography, economy, and transportation characteristics were collected from the Florida Department of Transportation and U.S. Census Bureau, respectively. See Table 4 for details. Note that the covariates with "$*$" are transformed from the original value by $f(x) = \log(x + 0.01)$. For example, $\text{VMT}^* = \log(\text{VMT} + 0.01)$. The whole dataset contains 1761 observations with 12 covariates.

Table 4. Covariates used in the crash dataset and their corresponding p-value.

Variable	Description	p-value
VMT*	Vehicle miles traveled	<0.01
Population*	Total population	<0.01
Rmale	Proportion of males	0.03
Rhispanic	Proportion of Hispanics in population	<0.01
Rold	Proportion of people age 65 and older	<0.01
Runemployed	Unemployed rate	<0.01
Income*	Median household income	0.19
Rcover	Health insurance cover rate	0.08
MTravelTime	Mean travel time that people take to go to work	0.13
Rwalking*	Proportion of commuting by public transportation	0.57
RpublicTrans*	Proportion of commuting by walking	0.82
Rother*	Proportion of commuting by bicycle, motorcycle, and other means	0.38

6.2. Analysis and Findings

We analyze the crash frequency data using the following GGAM: $Y_i \sim \text{NB}\left(\theta, \frac{\mu_i}{\theta+\mu_i}\right)$, where

$$\begin{aligned}\log(\mu_i) = {} & \beta_1(\text{VMT}_i^*) + \beta_2(\text{Population}_i^*) + \beta_3(\text{Rmale}_i) \\ & + \beta_4(\text{Rhispanic}_i) + \beta_5(\text{Rold}_i) \\ & + \beta_6(\text{Runemployed}_i) + \beta_7(\text{Income}_i^*) + \beta_8(\text{Rcover}_i) \\ & + \beta_9(\text{MTravelTime}_i) + \beta_{10}(\text{Rwalking}_i^*) \\ & + \beta_{11}(\text{RpublicTrans}_i^*) + \beta_{12}(\text{Rother}_i^*) + \alpha(\mathbf{S}_i),\end{aligned}$$

where $\beta_j(\cdot)$'s are unknown univariate functions, $\alpha(\cdot)$ is an unknown bivariate function used to control for the effect of the spatial structure of the observations. For both TPS and our method, the roughness parameter is selected by the GCV. Unlike other traditional methods, the BPS smoother does not smooth over the Tampa Bay area and the Wilderness Preserve, which makes more sense as there are no traffic across these uninhabited areas and our estimate does not link data points on either side of these areas. Figure 6(a) shows the triangulation adopted by our method, which contains 1869 triangles with 1098 vertices. Figure 6(b) shows the estimate of the spatial component, $\alpha(\cdot)$, in

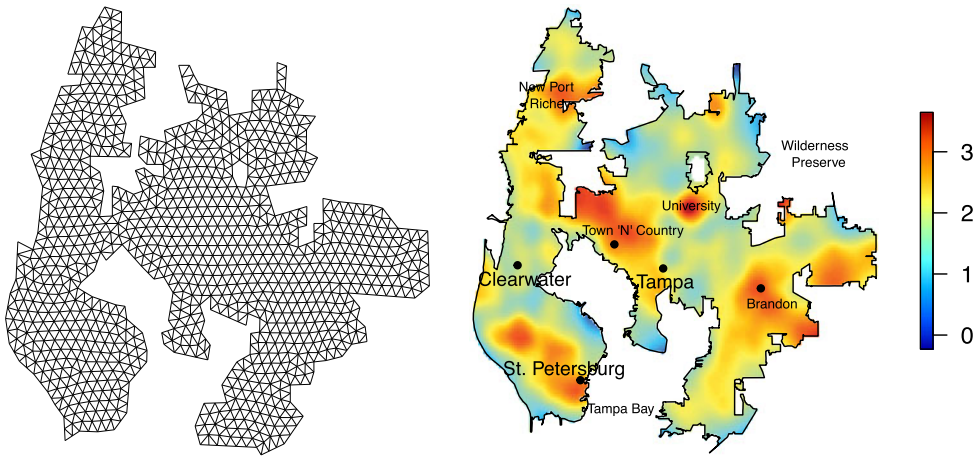


Figure 6. (a) Triangulation of Tampa-St. Petersburg urbanized area; (b) estimated $\alpha(\cdot)$ with cities labeled.

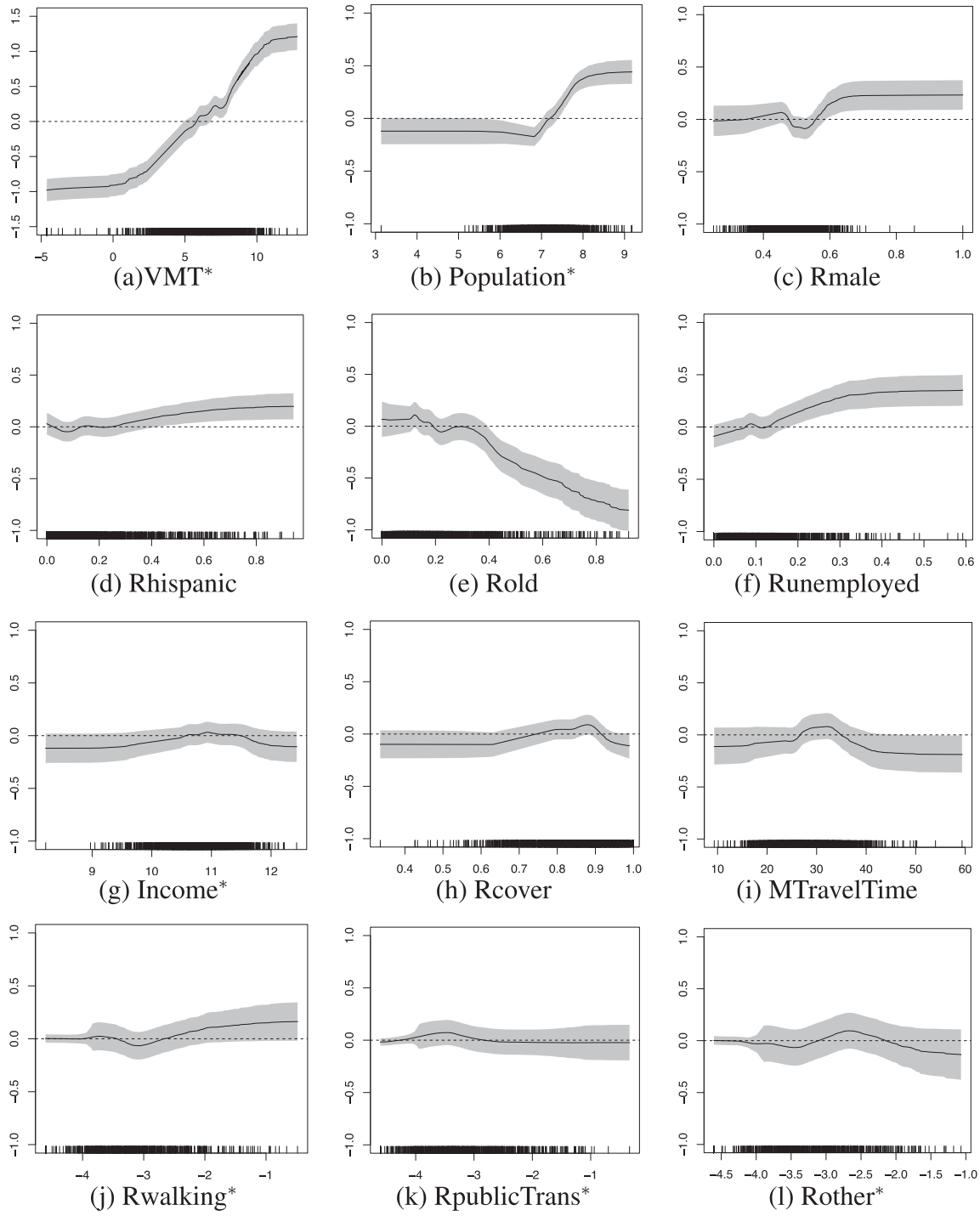


Figure 7. Plots of the SBL estimate (solid line), the 95% SCB (gray band) and the zero line (dashed line).

the model using our method. It identifies relatively high crash frequency in the cities, such as Brandon, St. Petersburg, New Port Richey, and the college town area.

Next we examine the effect of the predictors and test the following hypothesis of the individual functions $H_0 : \beta_k(\cdot) = 0$, $k = 1, \ldots, 12$. In Figure 7, the central black line shows the SBL fit. The gray bands represent the 95% SCBs constructed according to (7). The corresponding p-values of the tests are given in Table 4, which are calculated as the biggest value of α such that the $100(1-\alpha)$ % SCB covers the zero horizontal line.

Clearly, the null hypothesis that $\beta_1(\text{VMT}^*) = 0$, $\beta_2(\text{Population}^*) = 0$, $\beta_3(\text{Rmale}^*) = 0$, $\beta_4(\text{Rhispanic}) = 0$, $\beta_5(\text{Rold}^*) = 0$ and $\beta_6(\text{Runemployed}) = 0$ are rejected at significance level 0.05, since they are not totally covered by the 95% SCBs. For VMT and population, the fitted spline curves present a strong deviation from strict linearity. VMT and population are often considered to have a linear relationship with crash frequency in most studies, however, recently a few studies have shown that these relationships might be heterogeneous or nonlinear (Mannering, Shankar, and Bhat

Table 5. Estimation and prediction results.

MSE		MSPE	
SBL	TPS	SBL	TPS
102.79	133.98	140.81	144.32

2016; Anastasopoulos 2016). From Figure 7, one sees clearly the evidence of threshold effects of the logarithm of VMT and population on crash frequencies. For example, there is a constant effect for VMT (population) below −1 and above 10, while a linear effect of some significance exists in between. Our findings are reasonable: when VMT or population are either very small or very big, crash frequency should be constant, whereas in-between values, crash frequency should increase with VMT or population.

The proportion of unemployment may have mixed effects on crash frequency (Leigh and Waldon 1991), as unemployment increase may lead to less driving, but also more aggressive driving due to the mental stress of job loss or fear of job loss. The net effects are found to be negative in Michigan (Wagenaar 1983) and Iowa (Liu and Sharma 2018), but positive in the Tamp Bay area, which may be attributed to different travel modes and cultures. The proportion of males shows significantly positive. It might be attributed to that males travel more than females for business and work (Collins and Tisdell 2002), and they are also more likely to drive aggressively than female (Shinar and Compton 2004). The proportion of Hispanics is positively significant, which implies that Hispanics tend to have more crashes. Several studies have found that Hispanic drivers have higher rates of safety belt nonuse, speeding, invalid licensure, and alcohol involvement in Colorado (Harper et al. 2000), and Hispanics have more pedestrian crashes in New York City (Ukkusuri, Hasan, and Aziz 2011) and Los Angeles (Loukaitou-Sideris, Liggett, and Sung 2007) as they walk more due to low income. Thus, we believe the same situations may also exist in the study area. The proportion of old people is significant for crash frequency in our study, that is, the region with a higher proportion of old drivers lead to fewer crashes. One interpretation is that old people tend to travel less, which could reduce the traffic crashes within a specific area.

Income is statistically insignificant, which is consistent with another study (Liu and Sharma 2018), where they found travel expenses may only occupy small proportions of incomes for most census tracts, thus, whether the income is lower or higher does not impact people's travel decisions. The health insurance coverage rate is also insignificant. In terms of commuting characteristics, travel time is insignificant at significance level 0.05 and the proportions of travel modes other than driving are also insignificant.

To compare the SBL and the TPS, we adopt two criteria. We use mean squared error (MSE) to measure the model fit and the MSPE based on 10-fold cv to measure the predictive performance. Table 5 provides the MSE and the 10-fold cv MSPE of SBL and TPS. Both of them are in favor of the SBL, which not only gives the best model fit, but also provides the most accurate out-of-sample prediction results.

Finally, we conduct the spatial autoregression test based on the deviance residuals. The deviance residuals are calculated as following

$$d_i \equiv \text{sign}(y_i - \widehat{\mu}_i) \times \left[2 \left\{ y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) - (y_i + \widehat{\theta}) \log\left(\frac{y_i + \widehat{\theta}}{\widehat{\mu}_i + \widehat{\theta}}\right) \right\} \right]^{1/2}.$$

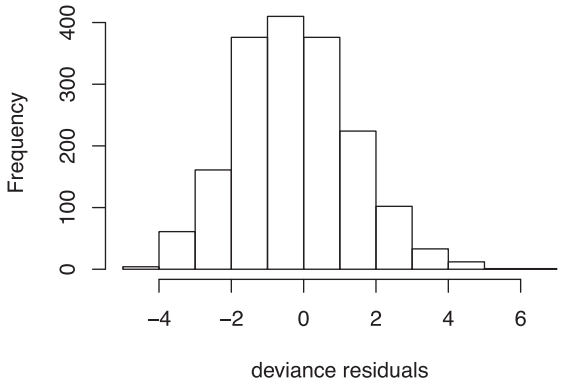


Figure 8. Histogram of the deviance residuals

Figure 8 shows the histogram of these deviance residuals of our method. Then we use the Moran's I to test the spatial autoregression. The Moran's I statistic is calculated as

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(d_i - \bar{d})(d_j - \bar{d})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} (d_i - \bar{d})^2},$$

where w_{ij} is the spatial weights and we use the basic binary coding. If ith observation and jth observation are neighborhood, then $w_{ij} = 1$, otherwise, $w_{ij} = 0$. The test statistic is 0.0098, and the p-value for the Moran's I test is 0.45. Thus, we cannot reject the null hypothesis. It is quite possible that the spatial distribution of feature values is the result of independent random spatial processes.

7. Concluding Remarks

In this paper, we couple the ideas of geostatistics with additive modeling under the framework of penalized quasi-likelihood. We have developed a two-stage procedure accompanied by efficient computational algorithms to carry out estimation and statistical inference for GGAMs. Our approach is balanced in terms of theory, computation, and interpretation. It greatly enhances the application of GAMs to spatial data analysis. We do not require the data to be evenly distributed or on a regular-spaced grid. Our estimation is computationally fast and efficient since our first stage estimation can be formulated as a penalized regression problem. The proposed SCBs provide measures of the effect of covariates after adjusting for the spatial effect, thus, the users can gain valuable insights into the accuracy of their estimation of the GGAM.

This study can be further extended to study the following aspects. First, we have only analyzed the crashes off the state highway system here, whereas crashes on the state highway system should also be analyzed in future studies, thus, they could construct the full picture of traffic safety. When both crash types are considered, the multivariate models may be considered (Zhao et al. 2017; Ma and Kockelman 2006). Second, it may

be also interesting to explore the long-term crash trends with multiple years' data, where spatiotemporal correlations should be considered in modeling (Miaou, Song, and Mallick 2003; Liu and Sharma 2017, 2018; Boulieri et al. 2017).

Appendix A: Regularity Assumptions

Without loss of generality, let $x_k \in [a_k, b_k] = [0,1]$, for $k = 1, \ldots, p$, and the area of Ω be 1. For the univariate splines, we consider equally spaced knots in our theoretical derivation, and denote H as the length of the equally spaced subintervals, then it is clear $H \asymp J_n^{-1}$.

For a univariate function $\psi(\cdot)$, denote $\psi'(\cdot)$, $\psi''(\cdot)$, and $\psi^{(\nu)}(\cdot)$ be its first, second, and νth order derivative, respectively. For any bivariate function $g : \Omega \to R$, denote $|g|_{\upsilon,\infty,\Omega} = \max_{i+j=\upsilon} \|\nabla_{s_1}^i \nabla_{s_2}^j g_\ell(\mathbf{s})\|_{\infty,\Omega}$. Let ν be a nonnegative integer, and $\delta \in (0,1]$ such that $\varrho = \delta + \nu \geq 1$. Let $\mathcal{H}^{(\varrho)}([0,1])$ be the class of functions ψ on $[0,1]$ whose νth derivative exists and satisfies a Lipschitz condition of order δ: $|\psi^{(\nu)}(x) - \psi^{(\nu)}(x')| \leq C_\nu |x - x'|^\delta$, for $x, x' \in [0,1]$. Let

$$\mathcal{D}_k^0([0,1]) = \{g : \mathrm{E}g(X_k) = 0, \mathrm{E}g^2(X_k) < \infty\} \tag{A.1}$$

be the functional space defined on $[0,1]$ and

$$\mathcal{W}^{d+1,\infty}(\Omega) = \{g : |g|_{k,\infty,\Omega} < \infty, 0 \leq k \leq d+1\} \tag{A.2}$$

be the standard Sobolev space.

For the quasi-likelihood function $\ell\{g^{-1}(x), y\}$, let $q_j(x,y) = \frac{\partial^j}{\partial x^j} \ell\{g^{-1}(x), y\}$, $j = 1, 2$. Moreover, let $\varepsilon_i = Y_i - g^{-1}\{\sum_{k=1}^p \beta_k(X_{ik}) + \alpha(\mathbf{S}_i)\}$ be the error term.

The following are the technical assumptions needed to facilitate the technical details, though they may not be the weakest conditions.

(A1) For $k = 1, \ldots, p$, $\beta_k \in \mathcal{H}^{(\varrho)} \cap \mathcal{D}_k^0$, and $\alpha \in \mathcal{W}^{d+1,\infty}(\Omega)$.

(A2) The density function $f(\mathbf{x}, \mathbf{s})$ of $(X_1, \cdots, X_p, \mathbf{S})$ satisfies

$$0 < c_f \leq \inf_{(\mathbf{x},\mathbf{s}) \in [0,1]^p \times \Omega} f(\mathbf{x}, \mathbf{s}) \leq \sup_{(\mathbf{x},\mathbf{s}) \in [0,1]^p \times \Omega} f(\mathbf{x}, \mathbf{s}) \leq C_f < \infty.$$

The marginal densities $f_k(\cdot)$ of X_k have continuous derivatives on $[0,1]$ as well as uniform lower bound c_k and upper bound C_k. The density function $f_{\mathbf{S}}(\cdot)$ of $\mathbf{S}$ is bounded away from zero and infinity on Ω.

(A3) The function $q_2(x,y) < 0$ for $x \in \mathbb{R}$ and y in the range of the response variable. The functions $f_k'(\cdot)$, $\eta^{(3)}(\cdot)$, $V''(\cdot)$, and $g^{(3)}(\cdot)$ are continuous functions, and $\rho_2(\cdot) > 0$. For each $(\mathbf{x}, \mathbf{s}) \in \mathrm{supp}(f)$, $\mathrm{var}(Y|\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s})$ and $g'(\mu(\mathbf{x}, \mathbf{s}))$ are nonzero.

(A4) The errors satisfy $\mathrm{E}\{\varepsilon_i | (\mathbf{X}_i = \mathbf{x}, \mathbf{S}_i = \mathbf{s})\} = 0$ and $\sup_{(\mathbf{x},\mathbf{s}) \in [0,1]^p \times \Omega} \mathrm{E}\{|\varepsilon_i|^{2+\iota} | (\mathbf{X}_i = \mathbf{x}, \mathbf{S}_i = \mathbf{s})\} < \infty$ for some $\iota \in (3, \infty)$; $(\mathbf{X}_i, \mathbf{S}_i, \varepsilon_i)$ are iid.

(A5) The triangulation $\triangle$ is π-quasi-uniform, that is, $(\min_{\tau \in \triangle} R_\tau)^{-1} |\triangle| \leq \pi$ for some positive constant π.

(A6) The number of knots J_n for the univariate splines and the triangulation size $|\triangle|$ satisfy that $J_n \to \infty$, $|\triangle| \to 0$, and $|\triangle|^{-2} J_n n^{-1} \log n \to 0$; and the smoothness penalty parameter satisfies that $\lambda n^{-1} |\triangle|^{-4} \to 0$.

(A6′) For some constant C_1, C_2, $\varrho \geq 1$ and $d \geq 2$, $n^{-2/5} \ll H \ll n^{-1/(2\varrho+2)} (\log n)^{1/(2\varrho+2)}$, $n^{-1/5} \ll |\triangle| \ll n^{-1/(2d+2)} (\log n)^{1/(2d+2)}$ and $\lambda \ll |\triangle|^4 n^{1/2} \log^{1/2} n$.

(A7) The kernel function K is a symmetric probability density with support $[-1,1]$. K is a Lipschitz continuous function for constant C, that is, $|K(x) - K(x')| \leq C|x - x'|$, $\forall x, x' \in [-1,1]$.

(A8) For each $k = 1, \ldots, p$, the bandwidth h_k satisfies either (i) $n h_k^5 = O(1)$, and $h_k^{-1} = O\{n^\delta (\log n)^\upsilon\}$, where $\delta = 1/5$, $\upsilon > 0$ or $1/5 < \delta < (3\iota - 4)/(10 + 5\iota)$ for ι given in Assumption (A4); or (ii) $n^{-1/5} (\log n)^{-\upsilon} \ll h_k \ll n^{-1/5} (\log n)^{-1/5}$ for some $\upsilon > 1/5$.

The above assumptions are mild conditions that can be satisfied in many practical situations. Assumption (A1) describes the requirement on the additive component functions, which are frequently used in the literature of nonparametric estimation. Assumptions (A1) and (A2) are similar to Assumptions (A1) and (A3) in Liu, Yang, and Härdle (2013). Assumption (A3) is analog to Conditions 1–3 in Fan, Heckman, and Wand (1995), which are common assumptions used under the quasi-likelihood frame work. Assumption (A4) is the same as the Assumptions (A3) and (A5) in Zheng et al. (2016). Assumption (A5) is widely used in the triangulation based literature (see Lai and Wang 2013; Wang et al. 2019). Assumptions (A6) and (A6′) show the requirement of the number of interior knots and the size of triangulation to ensure the consistency property of spline estimator and to obtain the SCBs, respectively. Assumptions (A7)–(A8) are regular assumptions in the local polynomial regression literature (see Fan, Heckman, and Wand 1995).

Supplementary Materials

The online supplementary materials contain technical proofs of the theoretical findings, additional simulation results, as well as a detailed explanation on how to implement the proposed bivariate spline smoothing over triangulation.

Funding

This research is supported by the Iowa State University Plant Sciences Institute Scholars Program (Shan Yu), National Science Foundation award DMS-1542332 (Li Wang), and National Natural Science Foundation of China award NSFC 11771240 (Lijian Yang).

References

Anastasopoulos, P. C. (2016), "Random Parameters Multivariate Tobit and Zero-Inflated Count Data Models: Addressing Unobserved and Zero-State Heterogeneity in Accident Injury-Severity Rate and Frequency Analysis," *Analytic Methods in Accident Research*, 11, 17–32. [12]

Boulieri, A., Liverani, S., Hoogh, K., and Blangiardo, M. (2017), "A Space–Time Multivariate Bayesian Model to Analyse Road Traffic Accidents by Severity," *Journal of the Royal Statistical Society*, Series A, 180, 119–139. [13]

Collins, D., and Tisdell, C. (2002), "Gender and Differences in Travel Life Cycles," *Journal of Travel Research*, 41, 133–143. [12]

Craven, P., and Wahba, G. (1979), "Bivariate Splines for Functional Regression Models With Application to Ozone Forecasting." *Numerische Mathematik*, 31, 377–403. [6]

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall. [6]

Fan, J., Heckman, N. E., and Wand, M. P. (1995), "Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions," *Journal of the American Statistical Association*, 90, 141–150. [6,13]

Harper, J. S., Marine, W. M., Garrett, C. J., Lezotte, D., and Lowenstein, S. R. (2000), "Motor Vehicle Crash Fatalities: A Comparison of Hispanic and Non-Hispanic Motorists in Colorado," *The Annals of Emergency Medicine*, 36, 589–596. [12]

Hastie, T., and Tibshirani, R. (1987), "Generalized Additive Models: Some Applications," *Journal of the American Statistical Association*, 82, 371–386. [1]

——— (1990), *Generalized Additive Models*, Hoboken, NJ: Wiley Online Library. [1]

Horowitz, J. L., and Mammen, E. (2004), "Nonparametric Estimation of an Additive Model With a Link Function," *The Annals of Statistics*, 32, 2412–2443. [2]

Kammann, E., and Wand, M. P. (2003), "Geoadditive Models," *Journal of the Royal Statistical Society*, Series C, 52, 1–18. [2]

Krivobokova, T., Kneib, T., and Claeskens, G. (2010), "Simultaneous Confidence Bands for Penalized Spline Estimators," *Journal of the American Statistical Association*, 105, 852–863. [2]

Lai, M. J., and Schumaker, L. L. (2007), *Spline Functions on Triangulations*, New York: Cambridge University Press. [2,3,6]

Lai, M. J., and Wang, L. (2013), "Bivariate Penalized Splines for Regression." *Statistica Sinica*, 23, 1399–1417. [2,4,13]

Lee, L.-F. (2004), "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72, 1899–1925. [2]

Leigh, J. P., and Waldon, H. M. (1991), "Unemployment and Highway Fatalities," *Journal of Health Politics, Policy and Law*, 16, 135–156. [12]

Liang, H., Thurston, S. W., Ruppert, D., Apanasovich, T., and Hauser, R. (2008), "Additive Partial Linear Models With Measurement Errors," *Biometrika*, 95, 667–678. [2]

Lindgren, F., Rue, H., and Lindström, J. (2011), "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach," *Journal of the Royal Statistical Society*, Series B, 73, 423–498. [6]

Liu, C., and Sharma, A. (2017), "Exploring Spatio-Temporal Effects in Traffic Crash Trend Analysis," *Analytic Methods in Accident Research*, 16, 104–116. [1,13]

——— (2018), "Using the Multivariate Spatio-Temporal Bayesian Model to Analyze Traffic Crashes by Severity," *Analytic Methods in Accident Research*, 17, 14–31. [1,12,13]

Liu, R., Yang, L., and Härdle, W. K. (2013), "Oracally Efficient Two-Step Estimation of Generalized Additive Model," *Journal of the American Statistical Association*, 108, 619–631. [2,13]

Loukaitou-Sideris, A., Liggett, R., and Sung, H.-G. (2007), "Death on the Crosswalk: A Study of Pedestrian-Automobile Collisions in Los Angeles," *Journal of Planning Education and Research*, 26, 338–351. [12]

Ma, J., and Kockelman, K. (2006), "Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity," *Transportation Research Record: Journal of the Transportation Research Board*, 1950, 24–34. [12]

Mannering, F. L., Shankar, V., and Bhat, C. R. (2016), "Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data," *Analytic Methods in Accident Research*, 11, 1–16. [12]

Marx, B., and Eilers, P. (2005), "Multidimensional Penalized Signal Regression," *Technometrics*, 47, 13–22. [2]

McKeague, I. W., and Zhao, Y. (2006), "Width-Scaled Confidence Bands for Survival Functions," *Statistics & Probability Letters*, 76, 327–339. [2]

Miaou, S.-P., Song, J. J., and Mallick, B. K. (2003), "Roadway Traffic Crash Mapping: A Space-Time Modeling Approach," *Journal of Transportation and Statistics*, 6, 33–57. [1,13]

Miller, D. L., and Wood, S. N. (2014), "Finite Area Smoothing With Generalized Distance Splines," *Environmental and Ecological Statistics*, 21, 715–731. [2]

Ramsay, T. (2002), "Spline Smoothing Over Difficult Regions," *Journal of the Royal Statistical Society*, Series B, 64, 307–319. [1,2]

Sangalli, L., Ramsay, J., and Ramsay, T. (2013), "Spatial Spline Regression Models," *Journal of the Royal Statistical Society*, Series B, 75, 681–703. [2,7]

Shinar, D., and Compton, R. (2004), "Aggressive Driving: An Observational Study of Driver, Vehicle, and Situational Variables," *Accident Analysis and Prevention*, 36, 429–437. [12]

Song, J. J., Ghosh, M., Miaou, S., and Mallick, B. (2006), "Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping," *Journal of Multivariate Analysis*, 97, 246–273. [10]

Stone, C. J. (1986), "The Dimensionality Reduction Principle for Generalized Additive Models," *The Annals of Statistics*, 14, 590–606. [2]

Ukkusuri, S., Hasan, S., and Aziz, H. (2011), "Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency," *Transportation Research Record: Journal of the Transportation Research Board*, 2237, 98–106. [12]

Wagenaar, A. C. (1983), "Unemployment and Motor Vehicle Accidents in Michigan," Technical Report. [12]

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia: SIAM Publications. [6]

Wang, H., and Ranalli, M. G. (2007), "Low-Rank Smoothing Splines on Complicated Domains," *Biometrics*, 63, 209–217. [2]

Wang, J., and Yang, L. (2009), "Polynomial Spline Confidence Bands for Regression Curves," *Statistica Sinica*, 19, 325–342. [2]

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. [2]

Wang, L., Wang, G., Lai, M. J., and Gao, L. (2019), "Efficient Estimation of Partially Linear Models for Data on Complicated Domains by Bivariate Penalized Splines Over Triangulations," *Statistica Sinica* (in press). [2,13]

Wang, G., Wang, L., and Lai, M. J. (2019), "Package 'BPST'," R Package Version 1.0, available at *http://people.wm.edu/~gwang01/softwares.html*. [4]

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. [3]

Wang, L., and Yang, L. (2007), "Spline-Backfitted Kernel Smoothing of Nonlinear Additive Autoregression Model," *The Annals of Statistics*, 35, 2474–2503. [2,3]

Wiesenfarth, M., Krivobokova, T., Klasen, S., and Sperlich, S. (2012), "Direct Simultaneous Inference in Additive Models and Its Application to Model Undernutrition," *Journal of the American Statistical Association*, 107, 1286–1296. [2]

Wood, S. N. (2003), "Thin Plate Regression Splines," *Journal of the Royal Statistical Society*, Series B, 65, 95–114. [2]

——— (2017), "Package 'mgcv'," R Package Version 1.8-23, available at *https://cran.r-project.org/web/packages/mgcv/index.html*. [7]

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. [1,2,7]

Xue, L., and Liang, H. (2010), "Polynomial Spline Estimation for a Generalized Additive Coefficient Model," *Scandinavian Journal of Statistics*, 37, 26–46. [2]

Xue, L., and Yang, L. (2006), "Additive Coefficient Modeling via Polynomial Spline," *Statistica Sinica*, 16, 1423–1446. [3]

Yang, L., Sperlich, S., and Härdle, W. (2003), "Derivative Estimation and Testing in Generalized Additive Models," *Journal of Statistical Planning and Inference*, 115, 521–542. [2]

Yu, K., Park, B. U., and Mammen, E. (2008), "Smooth Backfitting in Generalized Additive Models," *The Annals of Statistics*, 36, 228–260. [2]

Zhao, M., Liu, C., Li, W., and Sharma, A. (2017), "Multivariate Poisson-Lognormal Model for Analysis of Crashes on Urban Signalized Intersections Approach," *Journal of Transportation Safety and Security*, 10, 1–15. [12]

Zheng, S., Liu, R., Yang, L., and Härdle, W. K. (2016), "Statistical Inference for Generalized Additive Models: Simultaneous Confidence Corridors and Variable Selection," *Test*, 25, 607–626. [2,13]

Zhu, J., Huang, H. C., and Reyes, P. E. (2010), "On Selection of Spatial Linear Models for Lattice Data," *Journal of the Royal Statistical Society*, Series B, 72, 289–402. [2]