Statistica Sinica **33** (2023), 1093-1114
doi:https://doi.org/10.5705/ss.202021.0233

EMPIRICAL LIKELIHOOD RATIO TESTS FOR VARYING COEFFICIENT GEO MODELS

Shuoyang Wang[1], Honglang Wang[2], Yichuan Zhao[3], Guanqun Cao[1] and Yingru Li[4]

[1]*Auburn University,* [2]*Indiana University–Purdue University Indianapolis,* [3]*Georgia State University and* [4]*University of Central Florida*

Abstract: In this study, we investigate varying-coefficient models for spatial data distributed over two-dimensional domains. First, we approximate the univariate components and the geographical component in the model using univariate polynomial splines and bivariate penalized splines over triangulation, respectively. The spline estimators of the univariate and bivariate functions are consistent, and their convergence rates are also established. Second, we propose empirical likelihood-based test procedures to conduct both pointwise and simultaneous inferences for the varying-coefficient functions. We derive the asymptotic distributions of the test statistics under the null and local alternative hypotheses. The proposed methods perform favorably in finite-sample applications, as we show in simulations and an application to adult obesity prevalence data in the United States.

Key words and phrases: B-spline, bivariate spline, empirical likelihood, Geo data, non-parametric hypothesis testing

1. Introduction

Varying-coefficient models (VCMs)'s introduced by Hastie and Tibshirani (1993), are regression models commonly applied to examine the interactive associations between a response and predictors. These models are appealing because the regression coefficients are allowed to vary as a smooth function of some variables of interest to detect nonlinear interactions. Because of their flexibility, VCMs have been widely applied in many scientific areas. See Fan and Zhang (2008) for a selective overview of the major methodological and theoretical developments on VCMs. This study focuses on VCMs for spatial data randomly distributed over an arbitrary geographical region.

Our work is motivated by inference problems examining the effects of the county-level food retail environment on obesity rates in United States, with the effect varying over median household income. County food retail environments

Corresponding author: Guanqun Cao, Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA. E-mail: gzc0009@auburn.edu.

are measured by the availability and healthfulness of their food retail stores. More detailed information of this data set is provided in Section 6. Based on this data set, socioeconomists attempt to disentangle how county-level associations between the food environment and obesity rates change with median household income levels. This leads to modeling the effect of food retail environments as functions of household income levels. However, owing to the geographic dependence, the classical VCM is not sufficient.

In this work, we propose the varying-coefficient geo model (VCGM) to solve the motivating application. Specifically, assume $\boldsymbol{S}_i = (S_{i1}, S_{i2})^\top$ is the location of the ith subject, for $i = 1, \ldots, n$. The location $\boldsymbol{S}$ ranges over a two-dimensional bounded domain $\Omega \in \mathbb{R}^2$ of any arbitrary shape. We observe data of the form $\{Y_i, Z_i, \mathbf{X}_i, \boldsymbol{S}_i\}$, where Y_i is a response variable, $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ is a vector of scalar covariates, and Z_i is a scalar predictor. Furthermore, $\{(Y_i, Z_i, \mathbf{X}_i)\}_{i=1}^n$ are observed at location $\boldsymbol{S}_i$. Suppose that $\{(Y_i, Z_i, \mathbf{X}_i, \boldsymbol{S}_i)\}_{i=1}^n$ satisfies the following VCGM:

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta}(Z_i) + \alpha(\boldsymbol{S}_i) + \varepsilon_i, \quad \boldsymbol{S}_i \in \Omega, \quad i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\beta}(Z) = (\beta_1(Z), \ldots, \beta_p(Z))^\top$, with each $\beta_k(\cdot)$ as an unknown varying-coefficient function, $\alpha(\boldsymbol{S}_i)$ is an unknown smoothing bivariate function representing the spatial component, and ε_i denotes independent and identically distributed (i.i.d.) random noise, with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$ independent of $(Z_i, \mathbf{X}_i, \boldsymbol{S}_i)$. Our primary interest is to estimate and conduct an inference for $\boldsymbol{\beta}(\cdot)$ and $\alpha(\cdot)$ based on the given observations $\{(Y_i, Z_i, \mathbf{X}_i, \boldsymbol{S}_i)\}_{i=1}^n$.

In the proposed VCGM, when the spatial component $\alpha(\cdot)$ is ignored, the model becomes the traditional VCM. Numerous studies have proposed methods for fitting the VCM, for example, the local linear method Fan and Zhang (1999), spline method Huang, Wu and Zhou (2002), and two-stage methods Wang and Yang (2007); Liu, Yang and Härdle (2013). There are also several methods for estimating bivariate functions defined over 2D domains. Within the nonparametric framework, these include bivariate P-splines (Marx and Eilers (2005)), thin plate splines (Wood (2003)), and bivariate splines (Wang et al. (2020); Yu et al. (2020)). Here, we apply bivariate splines over triangulations (Lai and Schumaker (2007)), because they can handle irregular 2D domains with complex boundaries and they are computationally efficient.

This study focuses on proposing pointwise (at a specific z) and simultaneous (for all $z \in [a, b]$) testing procedures for the following hypothesis under model (1.1):

$$H_0 : H\{\boldsymbol{\beta}_0(z)\} = 0 \text{ v.s. } H_1 : H\{\boldsymbol{\beta}_0(z)\} \neq 0, \tag{1.2}$$

where $H(\mathbf{b})$ is a q-dimensional function of $\mathbf{b} = (b_1, \ldots, b_p) \in \mathbb{R}^p$, such that $\mathbf{C}(\mathbf{b}) := \partial H(\mathbf{b})/\partial \mathbf{b}^\top$ is a $q \times p$ full-rank matrix ($q \leq p$), for all $\mathbf{b}$. The above hypothesis is very general, owing to the choice flexibility of $H(\mathbf{b})$. It includes many interesting hypotheses as special cases, for instance, $H_0 : \beta_{0,k}(z) = 0$ for all k if $H(\mathbf{b}) = \mathbf{b}$, a test for any arbitrary linear constraints on $\boldsymbol{\beta}_0$ if $H(\mathbf{b}) = \boldsymbol{\Lambda}\mathbf{b} - \mathbf{c}_0$ for a $q \times p$ known matrix $\boldsymbol{\Lambda}$ and a known vector $\mathbf{c}_0$, and even tests with nonlinear constraints. See Ashby (2011) for explicit examples of nonlinear hypotheses.

In contrast to estimation, few studies have examined inferences of varying-coefficient functions. Huang, Wu and Zhou (2002) proposed a goodness-of-fit test based on a comparison of the weighted residual sum of squares. This is a specific example of the generalized likelihood ratio studied by Fan, Zhang and Zhang (2001). More recently, Yu et al. (2020) proposed a spline backfitted local polynomial to estimate and make simultaneous inferences of the univariate components in a geo-additive model. Although the above-mentioned methods seem useful, they are not applicable to the general hypothesis in (1.2). Furthermore, the testing procedure involves a plug-in variance estimate, which leads to an unstable asymptotic distribution of the test statistics.

In this paper, we propose both pointwise and simultaneous tests for the hypothesis in (1.2) based on the empirical likelihood (EL). The EL is a nonparametric likelihood, introduced by Owen (1988, 1990). In spite of its nonparametric construction based on observed data points, the EL shares some convenient merits of the parametric likelihood, and has many desirable advantages in deriving confidence sets for unknown parameters. Owen (2001) and Chen and Van Keilegom (2009) provide an overview of the EL method. The EL method has been extended to VCMs for various data types; see, for example, Xue and Zhu (2007), Xue and Wang (2012),Yang, Li and Peng (2014) and Liu and Zhao (2021). Recently, Wang et al. (2018) considered test procedures based on the EL to conduct inferences for a class of functional concurrent linear models. However, when they applied the method to Google flu trend data, they ignored the spatial information contained in the data set. Bandyopadhyay, Lahiri and Nordman (2015) and Van Hala, Nordman and Zhu (2015) considered the EL method for inference over a broad class of spatial data exhibiting stochastic spatial patterns. However, they did not consider the flexible VCGM, or the spatial information.

In contrast to existing VCMs, our proposed VCGM properly accounts for all covariates and spatial information, which improves the model flexibility. The proposed EL-based inference has many advantages over normal approximation-based methods. First, it does not involve a plug-in estimate for the limiting variance. Owing to the necessity of estimating the standard errors, which is a typical challenge in nonparametric models, the Wald-type simultaneous inference is not stable in Liu and Zhao (2021). Second, as DiCiccio, Hall and Romano (1991) proved, the EL is Bartlett correctable and, thus, has an advantage over the bootstrap method. To the best of our knowledge, this is the first work to propose a VCGM and conduct an EL ratio test for spatial data, which is a nontrivial extension.

The rest of the paper is organized as follows. We propose spline estimators for both univariate and bivariate functions and develop their asymptotic consistency in Section 2. The pointwise and simultaneous EL tests are presented in Section 3, where we investigate the asymptotic distributions of the test statistics under both the null hypothesis and local alternatives. In Section 4, we address implementation issues such as triangulation, the number of univariate spline knots, and the kernel bandwidth selection. Simulation studies are presented in Section 5, followed by an analysis of a real-data example in Section 6. We summarize the proposed methodology and discuss future work in Section 7. Major technical details are included in the Supplementary Material.

2. Univariate and Bivariate Spline Estimations

In the estimation stage, we approximate each varying coefficient using univariate polynomial splines. The geographical function $\alpha(\cdot)$ is approximated using bivariate penalized splines over triangulation. First, we introduce some notation for univariate and bivariate splines.

2.1. Setup

Suppose that the covariate Z is distributed on a compact interval $[a, b]$. Owing to the simplicity of the computation, we approximate the univariate components $\beta_k(z)$ in (1.1) using polynomial splines. Define a partition of $[a, b]$ with J_n interior knots as $v = \{a = v_0 \leq v_1 \leq \cdots \leq v_{J_{n+1}} = b\}$. For some $\varrho \geq 1$, the polynomial splines of order $\varrho + 1$ are polynomial functions with ϱ-degree on intervals $[v_j, v_{j+1})$, for $j = 0, \ldots J_n - 1$, and $[v_{J_n}, v_{J_{n+1}}]$, and have $\varrho - 1$ continuous derivatives globally. Let $\mathcal{U} = \mathcal{U}([a, b])$ be the space of such polynomial splines. Let $U_j(z)$, for $j = 1, \ldots, J_n + \varrho + 1$, be the original B-spline basis functions for the coefficient functions. Suppose for $z \in [a, b]$, $\beta_k(z) \approx \sum_{j=1}^{J_n+\varrho+1} \eta_{kj} U_j(z) = \mathbf{U}(z)^\top \boldsymbol{\eta}_k$, where $\mathbf{U}(z) = (U_1(z), \ldots, U_{J_n+\varrho+1}(z))^\top$ and $\boldsymbol{\eta}_k = (\eta_{1k}, \ldots, \eta_{J_n+\varrho+1,k})^\top$.

It has been proved that the bivariate penalized splines method is efficient in dealing with data distributed on irregular domains with complicated boundaries (Yu et al., 2020; Wang et al., 2020)). In the following, we briefly introduce the triangulation techniques and describe the bivariate penalized spline smoothing method for the VCGM. See Lai and Schumaker (2007) and Wang et al. (2020) for a detailed introduction of the triangulation technique and how to construct the bivariate spline basis functions over triangulation.

According to Lai and Schumaker (2007), let $\tau = \langle \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3 \rangle$ be a nonempty-area triangle with three vertices, $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$. There is a unique representation in the form for any point $\boldsymbol{s} \in \mathbb{R}^2$, $\boldsymbol{s} = b_1\mathbf{s}_1 + b_2\mathbf{s}_2 + b_3\mathbf{s}_3$, with $b_1 + b_2 + b_3 = 1$, and b_1, b_2, and b_3 are the barycentric coordinates of the point $\boldsymbol{s}$ relative to the triangle τ. We define the Bernstein polynomials of degree d relative to triangle τ as $B_{ijk}^{\tau,d}(\boldsymbol{s}) = (d/i!j!k!)b_1^i b_2^j b_3^k$. The spatial domain Ω is a polygon of arbitrary shape, which can be partitioned into finitely many triangles. Let a collection $\triangle = \{\tau_1, \ldots, \tau_N\}$ of N triangles be 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. For any triangle $\tau \in \triangle$, denote T_τ as the radius of the largest disk contained in τ. Let $|\tau|$ be the length of the longest edge. Denote the size of $\triangle$ as $|\triangle| = \max\{|\tau| : \tau \in \triangle\}$. 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 uniquely 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 $\triangle$, 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 with cardinality $|\mathcal{M}| = N(d+1)(d+2)/2$. Then, we rewrite any function $\zeta \in \mathbb{S}_d^r(\triangle)$ using the basis expansion $\zeta(\boldsymbol{s}) = \sum_{m\in\mathcal{M}} B_m(\boldsymbol{s})\gamma_m = \mathbf{B}(\boldsymbol{s})^\top \boldsymbol{\gamma}$, where $\boldsymbol{s} \in \Omega$ and $\boldsymbol{\gamma} = (\gamma_m, m \in \mathcal{M})^\top$ is the bivariate spline coefficient vector.

2.2. Penalized least-squares estimators

In general, there are three approaches to conduct a spline estimation: smoothing splines, regression splines, and penalized splines. Smoothing splines request as many parameters as the number of observations. Regression splines need only a small number of knots, placed judiciously, but appropriate algorithms are needed to select the knots. Penalized splines combine the features of smoothing splines and regression splines. A roughness penalty is incorporated with relatively large number of knots. Wang et al. (2020) and Yu et al. (2020) discuss the advantages and necessity of penalized bivariate spline smoothing. Note that, given some suitable smoothness conditions, $\beta_k(\cdot)$ and $\alpha(\cdot)$ can be well represented by a univariate spline basis expansion and the Bernstein basis polynomials introduced in

Section 2.1. It is well known that increasing the number of triangles may overfit the data and increase the variance, while decreasing the number of triangles may result in a rigid and restrictive function that has more bias. Consequently, to improve the data fitting efficiency, reduce the computation complexity, and avoid over fitting, we consider the following penalized least-squares problem:

$$\sum_{i=1}^{n}\left\{Y_i-\sum_{k=1}^{p}\sum_{j=1}^{J_n+\varrho+1}\eta_{jk}U_j(Z_i)X_{ik}-\sum_{m\in\mathcal{M}}B_m(\boldsymbol{s})\gamma_m\right\}^2+\frac{\lambda_n}{2}\mathcal{E}(\alpha), \tag{2.1}$$

where

$$\mathcal{E}(\alpha)=\sum_{\tau\in\triangle}\int_{\tau}\sum_{i+j=2}\binom{2}{i}(\nabla_{s_1}^{i}\nabla_{s_2}^{j}\alpha)^2ds_1ds_2$$

is the roughness penalty for $\alpha(\cdot)$, λ_n is the roughness penalty parameter, and $\nabla_{s_q}^{v}$ is the vth order derivative in the direction s_q at the point $\boldsymbol{s}$, for $q=1,2$.

For a smooth join between two polynomials on adjoining triangles, we impose some linear constraints on the spline coefficients $\boldsymbol{\gamma}:\boldsymbol{\Psi}\boldsymbol{\gamma}=0$, where $\boldsymbol{\Psi}$ is the matrix that collects the smoothness conditions across all the shared edges of triangles. An example of $\boldsymbol{\Psi}$ can be found in Yu et al. (2020). Thus, the penalized least-squares problem (2.1) becomes

$$\sum_{i=1}^{n}\left\{Y_i-\sum_{k=1}^{p}\sum_{j=1}^{J_n+\varrho+1}\eta_{j,k}U_j(Z_i)X_{ik}-\sum_{m\in\mathcal{M}}B_m(\boldsymbol{s})\gamma_m\right\}^2+\frac{1}{2}\lambda_n\boldsymbol{\gamma}^{\top}\mathbf{P}\boldsymbol{\gamma}, \tag{2.2}$$

subject to $\boldsymbol{\Psi}\boldsymbol{\gamma}=0$, where $\mathbf{P}$ is the block diagonal penalty matrix satisfying $\boldsymbol{\gamma}^{\top}\mathbf{P}\boldsymbol{\gamma}=\mathcal{E}(\mathbf{B}\boldsymbol{\gamma})$. In the following, let $\mathbf{Y}=(Y_1,\ldots,Y_n)^{\top}$ be collections of Y_i. Denote

$$\mathbf{W}=\begin{pmatrix}\mathbf{U}(Z_1)^{\top}(X_{11}) & \cdots & \mathbf{U}(Z_1)^{\top}(X_{1p})\\ \vdots & \ddots & \vdots\\ \mathbf{U}(Z_n)^{\top}(X_{n1}) & \cdots & \mathbf{U}(Z_n)^{\top}(X_{np})\end{pmatrix}$$

as an $n\times p(J_n+\varrho+1)$ matrix. To solve the constrained minimization problem (2.2), we first remove the constraint using a QR decomposition of the transpose of the constraint matrix $\boldsymbol{\Psi}$. Specifically, we have $\boldsymbol{\Psi}^{\top}=\mathbf{QR}=\begin{pmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{pmatrix}\begin{pmatrix}\mathbf{R}_1 & \mathbf{0}\end{pmatrix}^{T}$, where $\mathbf{Q}$ is an orthogonal matrix, $\mathbf{R}$ is an upper-triangle matrix, the submatrix $\mathbf{Q}_1$ is the first r columns of $\mathbf{Q}$, where r is the rank of the matrix $\boldsymbol{\Psi}$, and $\mathbf{0}$ is a matrix of zeros. According to Lemma 1 in Wang et al. (2020), the problem (2.2) is now converted to the following conventional penalized regression problem without any constraints:

$$\min_{\boldsymbol{\eta},\boldsymbol{\theta}} \left\{ \|Y - \mathbf{W}\boldsymbol{\eta} - \mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}\|^2 + \lambda_n(\mathbf{Q}_2\boldsymbol{\theta})^\top \mathbf{P}(\mathbf{Q}_2\boldsymbol{\theta}) \right\},$$

where $\boldsymbol{\eta} = \left(\eta_{11}, \ldots, \eta_{p(J_n+\varrho+1)}\right)$ and $\mathbf{Q}_2\boldsymbol{\theta} = \boldsymbol{\gamma}$. For a fixed penalty parameter λ_n, we have

$$\begin{pmatrix} \widehat{\boldsymbol{\eta}} \\ \widehat{\boldsymbol{\theta}} \end{pmatrix} = \left\{ \begin{pmatrix} \mathbf{W}^\top\mathbf{W} & \mathbf{W}^\top\mathbf{B}\mathbf{Q}_2 \\ \mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{W} & \mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{B}\mathbf{Q}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \\ & \lambda_n\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2 \end{pmatrix} \right\}^{-1} \begin{pmatrix} \mathbf{W}^\top\mathbf{Y} \\ \mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{Y} \end{pmatrix}.$$

Define

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{W}^\top\mathbf{W} & \mathbf{W}^\top\mathbf{B}\mathbf{Q}_2 \\ \mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{W} & \mathbf{Q}_2^\top(\mathbf{B}^\top\mathbf{B} + \lambda_n\mathbf{P})\mathbf{Q}_2 \end{pmatrix}.$$

It follows from well-known block matrix forms of a matrix inverse that

$$\mathbf{V}^{-1} := \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & -\mathbf{A}_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & \mathbf{A}_{22} \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{A}_{11}^{-1} &= \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} = \mathbf{W}^\top[\mathbf{I} - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^\top(\mathbf{B}^\top\mathbf{B} + \lambda_n\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^\top\mathbf{B}^\top]\mathbf{W} \\ \mathbf{A}_{22}^{-1} &= \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = \mathbf{Q}_2^\top[\mathbf{B}^\top\{\mathbf{I} - \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top\}\mathbf{B} + \lambda_n\mathbf{P}]\mathbf{Q}_2. \end{aligned}$$

Hence, $\widehat{\boldsymbol{\eta}} = \mathbf{A}_{11}\mathbf{W}^\top \left\{\mathbf{I} - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^\top(\mathbf{B}^\top\mathbf{B} + \lambda_n\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^\top\mathbf{B}^\top\right\} \mathbf{Y}$ and $\widehat{\boldsymbol{\theta}} = \mathbf{A}_{22}$ $\mathbf{Q}_2^\top\mathbf{B}^\top\{\mathbf{I} - \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top\}\mathbf{Y}$. Thus, the estimators of $\beta_k(\cdot)$ and $\alpha(\cdot)$ are

$$\widehat{\beta}_k(z) = \mathbf{U}(z)^\top\widehat{\boldsymbol{\eta}}_k \quad \text{and} \quad \widehat{\alpha}(\boldsymbol{s}) = \mathbf{B}(\boldsymbol{s})^\top\widehat{\boldsymbol{\gamma}}, \text{ respectively, where } \widehat{\boldsymbol{\gamma}} = \boldsymbol{Q}_2\widehat{\boldsymbol{\theta}}. \quad (2.3)$$

We now investigate the asymptotic properties of the spline estimates $\widehat{\beta}_k(z)$ and $\widehat{\alpha}(\boldsymbol{s})$. To avoid confusion, let $\beta_{0,k}(\cdot)$ and $\alpha_0(\cdot)$ be the true functions of $\beta_k(\cdot)$ and $\alpha(\cdot)$ in model (2.3). For any Lebesgue measurable function $\phi(\boldsymbol{s})$ on a domain $\mathcal{D}$, where $\mathcal{D} = [a, b]$ or $\Omega \subseteq \mathbb{R}^2$, let $\|\phi\|_{L_2}^2 = \int_{\mathcal{D}} \phi^2(\boldsymbol{s})d\boldsymbol{s}$.

Theorem 1 (Rate of Convergence)**.** *Suppose that Assumptions* (A1)–(A6) *in the Supplementary Material hold. Then the spline estimators $\widehat{\beta}_k$ and $\widehat{\alpha}$ satisfy*

$$\begin{aligned} &\|\widehat{\alpha} - \alpha_0\|_{L_2} \\ &= O_p\left\{ J_n^{-\varrho-1}|\triangle| + n^{-1/2}|\triangle|^{-1} + \frac{\lambda_n}{n|\triangle|^3} + \left(1 + \frac{\lambda_n}{n|\triangle|^5}\right)|\triangle|^{d+1} \right\}, \\ &\sum_{k=1}^{p} \|\widehat{\beta}_k - \beta_{0,k}\|_{L_2} = O_p\left(n^{-1/2}J_n^{1/2} + n^{-1}|\triangle|^{-1} + J_n^{-\varrho-1}\right). \end{aligned}$$

Remark 1. This consistency result echoes similar phenomena discovered by other nonparametric regression literature. In fact, when only spatial information is available and no other scale covariates are included, the model (1.1) reduces to the same model in Lai and Wang (2013). When the varying coefficients reduce to linear coefficients, model (1.1) reduces to the same model in Wang et al. (2020). In these two reduced models, the above convergence rate of $\widehat{\alpha}$ is the same as those given in Lai and Wang (2013) and Wang et al. (2020), that is, $O_p\{n^{-1/2}|\triangle|^{-1} + \lambda_n/(n|\triangle|^3) + (1 + \lambda_n/(n|\triangle|^5))|\triangle|^{d+1}\}$. When the geo function $\alpha(\cdot)$ is excluded from model (1.1), the convergence rate of $\widehat{\beta}_k$ reduces to $O_p(n^{-1/2}J_n^{1/2} + J_n^{-\varrho-1})$. If $\beta_{0,k}$ have bounded second-order derivatives ($\varrho = 1$) and $J_n \asymp n^{1/5}$, we have $\|\widehat{\beta}_k - \beta_{0,k}\|_{L_2} = O_p(n^{-2/5})$, achieving the optimal nonparametric rate Stone (1982).

Given these consistency results of the proposed univariate and bivariate spline estimators, we can now build hypothesis testing statistics based on these estimators.

3. Empirical Likelihood Ratio Tests for Varying Coefficients

It is challenging to derive the asymptotic distribution and the measure of variability for the spline estimators introduced in Section 2. Similar findings have been discussed in Liu, Yang and Härdle (2013) and Yu et al. (2020). To investigate the uncertainty in the estimation of the varying effect of the covariates, we propose an inference for hypothesis (1.2) using the EL method, with bivariate penalized spline estimators plugged in for the geo function.

To test (1.2) and construct an EL ratio function for $\boldsymbol{\beta}(z)$, we first introduce an auxiliary random vector

$$g_i\{\boldsymbol{\beta}(z), \alpha_0\} = \left(Y_i - \boldsymbol{\beta}(z)^\top \boldsymbol{X}_i - \alpha_0(\boldsymbol{S}_i)\right) \boldsymbol{X}_i K_h(Z_i - z),$$

where $K(\cdot)$ stands for a continuous kernel function, h is a bandwidth, and $K_h(\cdot) = K(\cdot/h)/h$ is a rescaling of K. Note that $Eg_i\{\boldsymbol{\beta}(z), \alpha_0\}$ is close to zero if $\boldsymbol{\beta}(z) = \boldsymbol{\beta}_0(z)$. Hence, the problem of testing whether $\boldsymbol{\beta}(z)$ is the true function $\boldsymbol{\beta}_0(z)$ is equivalent to testing whether $Eg_i\{\boldsymbol{\beta}(z), \alpha_0\}$ is close to zero, for $i = 1, 2, \ldots, n$. According to Owen (2001), this can be done by using the EL; that is, we can define the profile EL ratio function

$$R\{\boldsymbol{\beta}(z), \alpha_0\} = \max_{p_i:1\le i\le n} \left\{ \prod_{i=1}^n np_i : 0 \le p_i \le 1, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_i\{\boldsymbol{\beta}(z), \alpha_0\} = 0 \right\}.$$

The rich EL literature has shown that $-2\log R\{\boldsymbol{\beta}_0(z),\alpha_0\}$ is asymptotically chi-squared with p degrees freedom. However, $R\{\boldsymbol{\beta}(z),\alpha_0\}$ cannot be used directly to make a statistical inference on $\boldsymbol{\beta}(z)$, because $R\{\boldsymbol{\beta}(z),\alpha_0\}$ contains the unknown function $\alpha_0(\cdot)$. A natural way is to replace $\alpha_0(\cdot)$ with the estimator $\widehat{\alpha}(\boldsymbol{S}_i)$ given in (2.3), that is,

$$g_i\{\boldsymbol{\beta}(z)\} := g_i\{\boldsymbol{\beta}(z),\widehat{\alpha}\} = \Big(Y_i - \boldsymbol{\beta}^\top(z)\boldsymbol{X}_i - \widehat{\alpha}(\boldsymbol{S}_i)\Big)\,\boldsymbol{X}_i K_h(Z_i - z).$$

Note that the solution to $\sum_{i=1}^n g_i\{\boldsymbol{\beta}(z)\} = 0$ corresponds to the local constant estimator

$$\check{\boldsymbol{\beta}}(z) = \left\{\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^\top K_h(Z_i - z)\right\}^{-1}\left\{\sum_{i=1}^n (Y_i - \hat{\alpha}(\boldsymbol{S}_i))\boldsymbol{X}_i K_h(Z_i - z)\right\}. \quad (3.1)$$

After replacing the true function $\alpha_0(\cdot)$, we show that the discrepancy between $g_i\{\boldsymbol{\beta}_0(z)\}$ and $g_i\{\boldsymbol{\beta}_0(z),\alpha_0\}$ is asymptotically negligible in the following proposition. Let $\mu_{jj'} = \int u^{j'}K^j(u)du$ and $\boldsymbol{\Omega}(z) = E(\mathbf{X}_1\mathbf{X}_1^\top|Z=z)$.

Proposition 1. *Under Assumptions* (A1)—(A5), (A6′), (A7), *and* (A8) *in the Supplementary Material, we have*

$$E[g_i\{\boldsymbol{\beta}_0(z)\}] = O(h^2)$$

and

$$Var[g_i\{\boldsymbol{\beta}_0(z)\}] = \sigma^2\boldsymbol{\Omega}(z)f(z)\mu_{20}h^{-1}\{1+o(1)\},$$

where $f(z)$ is the probability density function of Z.

Remark 2. To investigate the EL tests for the geo spatial model, the key point is to check the asymptotic property of $g_i\{\boldsymbol{\beta}_0(z)\}$. More specifically, if the first and second moments of $g_i\{\boldsymbol{\beta}_0(z)\}$ have the same orders as those of $g_i\{\boldsymbol{\beta}_0(z),\alpha_0\}$, the asymptotic distribution of $-2\log R\{\boldsymbol{\beta}(z)\}$ is similar to the common VCM cases. According to Theorem 1, we establish the orders of the first two moments for $g_i\{\boldsymbol{\beta}_0(z)\}$ as in Proposition 1 by bounding $E\left\{\mathbf{X}_i\mathbf{X}_i^\top\left(\boldsymbol{\beta}_0(Z_i)-\boldsymbol{\beta}_0(z)\right)K_h(Z_i-z)\right\}$ and $E\left\{\mathbf{X}_iK_h(Z_i-z)(\alpha_0(\boldsymbol{S}_i)-\widehat{\alpha}(\boldsymbol{S}_i))\right\}$, with a careful choice of the lower bound of J_n and the upper bound of $|\triangle|$. The details can be found in the proof of Proposition 1 in the Supplementary Material.

With a slight abuse of notation, we define the EL function

$$L\{\boldsymbol{\beta}(z)\} = \max_{p_i:1\le i\le n}\left\{\prod_{i=1}^n p_i : 0\le p_i\le 1, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_i\{\boldsymbol{\beta}(z)\} = 0\right\}. \quad (3.2)$$

We can maximize (3.2) using the Lagrange multiplier technique, which leads to the following log-EL:

$$\log L\{\boldsymbol{\beta}(z)\} = -\sum_{i=1}^{n} \log\left\{1 + \boldsymbol{\delta}^{\top}(z) g_i\{\boldsymbol{\beta}(z)\}\right\} - n\log n,$$

where $\boldsymbol{\delta}(z)$ is determined by the equation: $\sum_{i=1}^{n} g_i\{\boldsymbol{\beta}(z)\}[1 + \boldsymbol{\delta}^{\top}(z) g_i\{\boldsymbol{\beta}(z)\}]^{-1} = 0$. Therefore, the negative log-EL ratio statistic for testing $H_0 : H\{\boldsymbol{\beta}_0(z)\} = 0$ is

$$\ell(z) := \min_{H\{\boldsymbol{\beta}(z)\}=0} \sum_{i=1}^{n} \log\left\{1 + \boldsymbol{\delta}^{\top}(z) g_i\{\boldsymbol{\beta}(z)\}\right\}. \tag{3.3}$$

To investigate the power of the tests, we consider the local alternatives $H_1 : H\{\boldsymbol{\beta}_0(z)\} = b_n \boldsymbol{d}(z)$, where b_n is a sequence of numbers converging to zero and $\boldsymbol{d}(z) \neq 0$ is a q-dimensional function. For any fixed nonzero function $\boldsymbol{d}(z)$, b_n depicts the order of signals that a test can detect. The smallest order of b_n is given in Chen and Zhong (2010), who show that the EL method can detect alternatives of order $(nh)^{-1/2}$ for pointwise tests and order $n^{-1/2}h^{-1/4}$ for simultaneous tests. Both orders are larger than the parametric rate $n^{-1/2}$.

The following theorem summarizes the asymptotic distribution of $2\ell(z)$ under both the local alternative and the null hypothesis H_0 for each fixed z.

Theorem 2. *Under Assumptions* (A1)—(A5)*,* (A6′)*,* (A7)*, and* (A8) *in the Supplementary Material, and for each $z \in [a, b]$ under the null hypothesis: $H\{\boldsymbol{\beta}_0(z)\} = 0$, we have $2\ell(z) \xrightarrow{d} \chi^2_q$. For each $z \in [a, b]$ and any fixed real vector of function $\boldsymbol{d}(z)$, under the alternative hypothesis $H_1 : H\{\boldsymbol{\beta}_0(z)\} = (nh)^{-1/2}\boldsymbol{d}(z)$, we have*

$$2\ell(z) \xrightarrow{d} \chi^2_q\left(\boldsymbol{d}^{\top}(z)\boldsymbol{R}(z)\boldsymbol{d}(z)\right),$$

where $\mathbf{R}(z) = \sigma^2\mu_{20} f(z)\left\{\mathbf{C}(z)\boldsymbol{\Omega}(z)\mathbf{C}^{\top}(z)\right\}^{-1}$ *and*

$$\mathbf{C}(z) = \mathbf{C}\left(\boldsymbol{\beta}(z)\right) = \frac{\partial H\left(\boldsymbol{\beta}(z)\right)}{\partial \boldsymbol{\beta}(z)^{\top}}.$$

According to the Theorem 2, we can construct a pointwise confidence interval for each $\beta_j(z)$. The construction of the confidence interval is based on an asymptotic α-level test when $H\{\boldsymbol{\beta}(z)\} = \beta_j(z)$. We reject H_0 at a fixed point z if $2\ell(z) > \chi^2_{1,\alpha}$, where $\chi^2_{1,\alpha}$ is the upper α-quantile of χ^2_1, and a $100(1-\alpha)\%$ confidence interval for $\beta_j(z)$ is given by $\{\beta_j(z) : 2\ell(z) \leq \chi^2_{1,\alpha}\}$.

For the simultaneous test on H_0 in (1.2), for all $z \in [a, b]$, we consider the Cramér—von Mises type test statistic. Because $2\ell(z)$ can be viewed as the distance between $H\{\boldsymbol{\beta}(z)\}$ and zero, we propose the following test statistic for H_0:

$$D_n = \int_a^b 2\ell(z)w(z)dz, \tag{3.4}$$

where $w(z)$ is some probability weight function.

Theorem 3. *Under Assumptions* (A1)—(A5), (A6′), (A7), *and* (A8) *in the Supplementary Material, with the null hypothesis $H_0 : H\{\boldsymbol{\beta}_0(\cdot)\} = 0$, as $n \to \infty$, we have*

$$h^{-1/2}\{D_n - q\} \xrightarrow{d} N\left(0, q\sigma_0^2\right),$$

where $\sigma_0^2 = 2\mu_{20}^{-2}\int_a^b w^2(t)dt \int_{-2}^2\{K^{(2)}(u)\}^2 du$. When the alternative hypothesis $H_1 : H\{\boldsymbol{\beta}_0(z)\} = n^{-1/2}h^{-1/4}\boldsymbol{d}(z)$ holds, we have

$$h^{-1/2}\{D_n - q\} \xrightarrow{d} N\left(\mu_0, q\sigma_0^2\right),$$

where $\mu_0 = \int_a^b \mathbf{d}^\top(z)\mathbf{R}(z)\mathbf{d}(z)w(z)dz$.

Although the above theorem guarantees the asymptotic normality of D_n, the convergence rate is $h^{-1/2}$. According to Assumption (A6′), the rate is $o(n^{1/10})$, which is much slower than the classical nonparametric rate $n^{2/5}$. To obtain accurate type-I and type-II error probabilities in practice, we suggest a bootstrap procedure to generate the empirical quantile and perform the simultaneous testing. The distribution consistency of this method is discussed in Wang et al. (2018). The proposed bootstrap procedure consists of the following steps:

Step 1. For each subject, calculate the residual $\widetilde{e}_i = Y_i - \check{\boldsymbol{\beta}}(Z_i)^\top \boldsymbol{X}_i - \widehat{\alpha}_i(\boldsymbol{S}_i)$, with the local constant estimator $\check{\boldsymbol{\beta}}(z)$ in (3.1). Compute the sample variance of $\widetilde{e}_i$, and denote it as $\widetilde{\sigma}^2$;

Step 2. For the bth bootstrapping, for $b = 1, \ldots, B$, construct observation $Y_i^{(b)} = \check{\boldsymbol{\beta}}(Z_i)^\top \boldsymbol{X}_i + \widehat{\alpha}_i(\boldsymbol{S}_i) + \epsilon_i^{(b)}$, where $\epsilon_i^{(b)}$ are independently generated from a normal distribution satisfying $E(\epsilon_i^{(b)}) = 0$ and $Var(\epsilon_i^{(b)}) = \widetilde{\sigma}^2$. Apply $\{Y_i^{(b)}\}_{i=1}^n$ as new observations, and compute the bootstrapped version of D_n, denoted by $D_n^{(b)}$;

Step 3. Calculate the $100(1-\alpha)\%$ quantile of the bootstrap samples $\{D_n^{(b)}\}_{b=1}^B$, and denote it as $\widehat{d}_\alpha$. Reject the null hypothesis if $D_n > \widehat{d}_\alpha$.

Remark 3. In step 1, $\check{\boldsymbol{\beta}}(z)$ is the solution to $n^{-1}\sum_{i=1}^n g_i\left(\boldsymbol{\beta}(z), \widehat{\alpha}\right) = 0$. We use $\check{\boldsymbol{\beta}}(z)$ instead of the spline estimator $\widehat{\boldsymbol{\beta}}(z)$ to generate residuals, because $\check{\beta}(z)$ is the maximum empirical likelihood estimator involved in the construction of $\ell(z)$ and D_n.

The following proposition justifies the bootstrap procedure. The proof is similar to Theorem 4 in Wang et al. (2018). Thus it is omitted.

Proposition 2. *Let* $\mathcal{X}_n = \{(Y_i, Z_i, \boldsymbol{X}_i, \boldsymbol{S}_i)\}_{i=1}^n$ *be the original data, and* $\mathcal{L}(D_n)$ *be the asymptotic distribution of* D_n *under the null hypothesis. Under Assumptions* (A1)—(A6), (A6′), (A7) *and* (A8), *the conditional distribution of* $D_n^{(b)}$ *given* $\mathcal{X}_n$, $\mathcal{L}(D_n^{(b)}|\mathcal{X}_n)$ *converges to* $\mathcal{L}(D_n)$ *almost surely.*

4. Implementation

In extensive numerical studies, we find that the selections of the knots for the univariate spline, triangulation, and the choice of bandwidth are crucial, especially for simultaneous tests. In the following, we discuss the selection procedures one by one.

4.1. Selection of the tuning parameters in univariate and bivariate spline smoothing

In this work, we do not directly need the spline estimator $\widehat{\boldsymbol{\beta}}(z)$ for the inference of $\boldsymbol{\beta}(z)$. However, $\widehat{\alpha}(\boldsymbol{s})$ is essential for constructing the EL ratio tests (3.3), and its estimating procedure involves $\widehat{\boldsymbol{\beta}}(z)$. Hence, we need to make sure that $\boldsymbol{\beta}(z)$ is estimated efficiently. For univariate spline smoothing, we suggest applying knots on a grid of equally spaced sample quantiles. Assumption (A6′) in the Supplementary Material suggests that the number of knots J_n needs to satisfy $|\triangle|^{1/(\varrho+1)}\ n^{2/(5\varrho+5)} \ll J_n \ll |\triangle|^2 n \log^{-1}(n)$. Given the widely used cubic splines, in practice, we suggest the rule-of-thumb number of interior knots $J_n = \max\left\{\lfloor c_1 n^{2/(5\varrho+5)}\rfloor + 1, 3\right\}$, where the tuning parameter $c_1 \in [1, 3]$. A similar is considered in Yu et al. (2020). We also compared the proposed knot selection method with other data-driven methods, namely, the AIC and BIC. The well-selected parameters using the AIC and BIC are similar to our proposed rule-of-thumb choices. Therefore, for the purpose of efficient computation, we recommend the rule-of-thumb choices for practical applications.

When selecting the number of triangles, we need to balance the computational burden and the approximation accuracy. According to Yu et al. (2020) and Assumption (A6′), in practice, when the boundary of the spatial domain is not extremely complicated, we suggest taking the number of triangles as the following: $N = \min\left\{\lfloor c_2 n^{4/(5d+5)}\rfloor, n/4\right\} + 1$, for some tuning parameter c_2. Typically, $c_2 \in [1, 5]$ and is chosen using cross-validation. When the boundary of the spatial domain looks complicated, we suggest N to be much larger than n, and the triangulation can approximate the complicated domain precisely. Once N is chosen, a typical triangulation method, such as Delaunay triangulation, can be used to build the triangulated meshes. From our numerical experience, when the smoothness $r = 1$, compared with the setting $d = 2$ or 3, using $d = 5$ requires too much unnecessary computational time, because its improvement in terms of accuracy is negligible. We suggest using $r = 1$ and $d = 2$ or 3 in practice, because they provide enough accuracy for smooth functions and reduce the computational cost. Similar settings are also found in Lai and Wang (2013), Yu et al. (2020) and Kim, Wang and Zhou (2021).

The generalized cross-validation (GCV) criterion is an efficient method for selecting the smoothing parameters λ_n, and also has good theoretical properties (Wahba (1990)). The fitted values at the n data points are $\widehat{\mathbf{Y}} = \mathbf{W}\widehat{\boldsymbol{\eta}} + \mathbf{B}\mathbf{Q}_2\widehat{\boldsymbol{\theta}}$, and the smoothing matrix is

$$\mathbf{S}(\lambda_n) = \mathbf{W}\mathbf{A}_{11}\mathbf{W}^\top \left\{\mathbf{I} - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^\top(\mathbf{B}^\top\mathbf{B} + \lambda_n\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^\top\mathbf{B}^\top\right\}$$
$$+\mathbf{B}\mathbf{Q}_2\mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\left\{\mathbf{I} - \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top\right\}.$$

We choose the smoothing parameter λ_n by minimizing

$$GCV(\lambda_n) = \frac{n\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{[n - tr\,\{\mathbf{S}(\lambda_n)\}]^2}$$

over a grid of values of λ_n. We use a 10-point grid, where the values of $\log_{10}(\lambda_n)$ are equally spaced between -6 and 1 in our numerical studies. The aforementioned bivariate spline smoothing methods are all implemented using the R package "BPST" developed by Wang et al. (2020).

4.2. Bandwidth selection

The performance of the EL pointwise and simultaneous tests depends on the choice of the bandwidth h. We apply the five-fold cross-validation criterion and choose the bandwidth h by minimizing

$$CV(h) = 5^{-1}\sum_{k=1}^{5}|\mathcal{F}_k|^{-1}\sum_{i\in\mathcal{F}_k}\left\{Y_i - \check{\boldsymbol{\beta}}^{(-k)}(Z_i)^\top\mathbf{X}_i - \widehat{\alpha}^{(-k)}(\boldsymbol{S}_i)\right\}^2,$$

where $\mathcal{F}_k$ denotes the subject index set for the kth folder and $|\mathcal{F}_k|$ denotes the cardinality of $\mathcal{F}_k$ over a grid of values of h. In our numerical studies, we select the bandwidth $h = \lfloor c_3 n^{1/5}\rfloor + 0.02$ for the pointwise tests, and $h = \lfloor c_3 n^{1/5}\rfloor$ for the simultaneous tests, where $c_3 \in \{0.1, 0.2, \ldots, 0.9, 1\}$.

5. Simulation

In this section, we conduct simulation studies to evaluate the finite-sample performance of the proposed methodology. We generate the data from the following VCGM:

$$Y_i = X_{i1}\beta_1(Z_i) + X_{i2}\beta_2(Z_i) + \alpha(\boldsymbol{S}_i) + \epsilon_i, i = 1, \ldots, n, \tag{5.1}$$

where X_{ij} and ϵ_i are independently generated from $N(0, 1)$, and Z_i follows $Unif[0, 1]$ independently. In addition, we choose the Epanechnikov kernel $K(x) = 3/4\left(1 - x^2\right)_+$ for the local linear estimation, where $(a)_+ = \max(a, 0)$. The sample sizes are chosen to be $n = 500, 1000, 2000$. We consider two spatial domains for the bivariate function $\alpha(\cdot)$: 1) a rectangular domain $[0, 1]^2$; and 2) a modified horseshoe domain used by Sangalli, Ramsay and Ramsay (2013) and Wang et al. (2020). For each Monte Carlo replication, we randomly sample n locations uniformly from the grid points inside the two spatial domains. Under all scenarios, 1,000 Monte Carlo replicates are conducted. For all the univariate splines, we use cubic B-splines with $\varrho = 3$. For the bivariate spline smoothing, we consider $d = 3$ and $r = 1$.

To check the accuracy of the proposed spline estimators, we compute the mean squared error (MSE) for α, β_1, and β_2. Figure 1 shows the surface and the contour map of the true bivariate function $\alpha(\cdot)$ and the estimated one when the sample size $n = 2000$. The proposed estimates look visually close to the true functions. Figure 2 shows the box plot of the MSEs of the spline estimators for both regions, showing that the MSEs and the corresponding standard deviations decrease as the sample size increases.

We first conduct pointwise hypothesis tests. Let $H\{(\beta_1, \beta_2)^\top\} = \beta_1 - \beta_2$ to test $H_0 : \beta_1(z) = \beta_2(z)$ versus $H_1 : \beta_1(z) \neq \beta_2(z)$, where we set $\beta_1(z) = (2 + a)\sin(2\pi z)$ and $\beta_2(z) = 2\sin(2\pi z)$, for some nonnegative a in model (5.1), to evaluate the empirical size (when $a = 0$) and power (when $a > 0$) at the 5% nominal level. Figure 3 shows the empirical size and power with two different domains of $\alpha(\boldsymbol{s})$ and different $z \in \{0.3, 0.4, 0.6, 0.7\}$. Given each z, the empirical size is reasonably controlled around the nominal level of 5% for all sample sizes, and the power increases with a until reaching one. As expected, a larger sample size leads to greater power.

Next, we set $\beta_1(z) = 1/2\sin(z)$, $\beta_2(z) = 2\sin(z + 1/2)$ in model (5.1), and apply the procedure in Section 3 to construct pointwise confidence intervals for

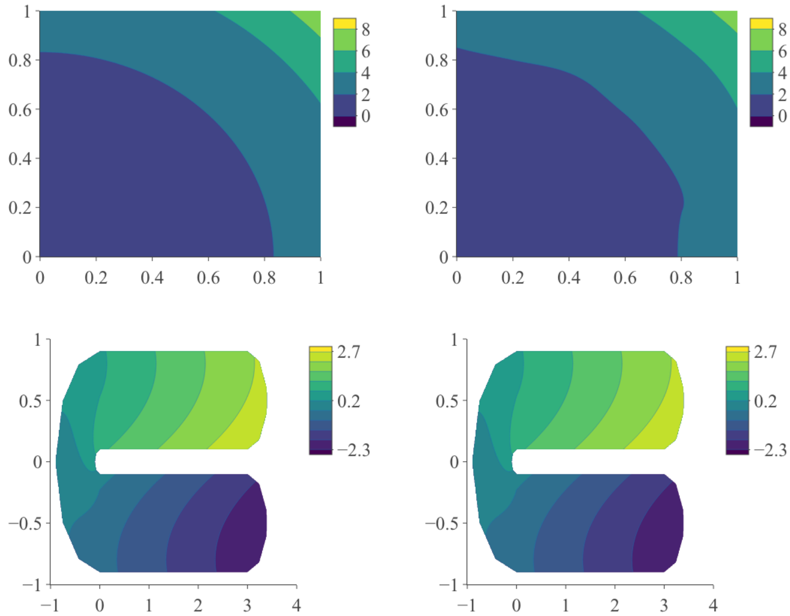

Figure 1. Contour maps of the true function $\alpha_0(\cdot)$ (first column) and the estimators (second colmun) over the square region (first row) and the horseshoe region (second row).

$\beta_1(z)$ at the 95% nominal level. Table 1 summarizes the empirical coverage probability (as percentages) and the average length of the confidence intervals (in parentheses) for $\beta_1(z)$ at $z = 0.3, 0.4, 0.6, 0.7$. From the table, we see that for different z, the coverage rates increase with the sample size, and are around 95% when $n = 2000$. Furthermore, the length of the confidence intervals decreases as the sample size increases.

Finally, we consider simultaneous inference. We test $H_0 : \beta_1(z) = \beta_2(z)$ for all $z \in [0, 1]$ versus $H_1 : \beta_1(z) \neq \beta_2(z)$, for some z, where we set $\beta_1(z) = (2 + a)\sin(2\pi z)$ and $\beta_2(z) = 2\sin(2\pi z)$ for $a \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ in model (5.1). We evaluate the empirical size (when $a = 0$) and power (when $a > 0$); the results are presented in Table 2. All tests are under two scenarios of bivariate function regions. In the construction of the test statistics D_n, we choose the weight function $w(z) = 1$ for $z \in (0, 1)$, and $w(z) = 0$ otherwise. The critical value of the test is estimated using 500 bootstrap samples in each simulation run. From Table 2, we find that the empirical size for each n is around the nominal level of 5%, and the trend of the power is reasonably controlled.

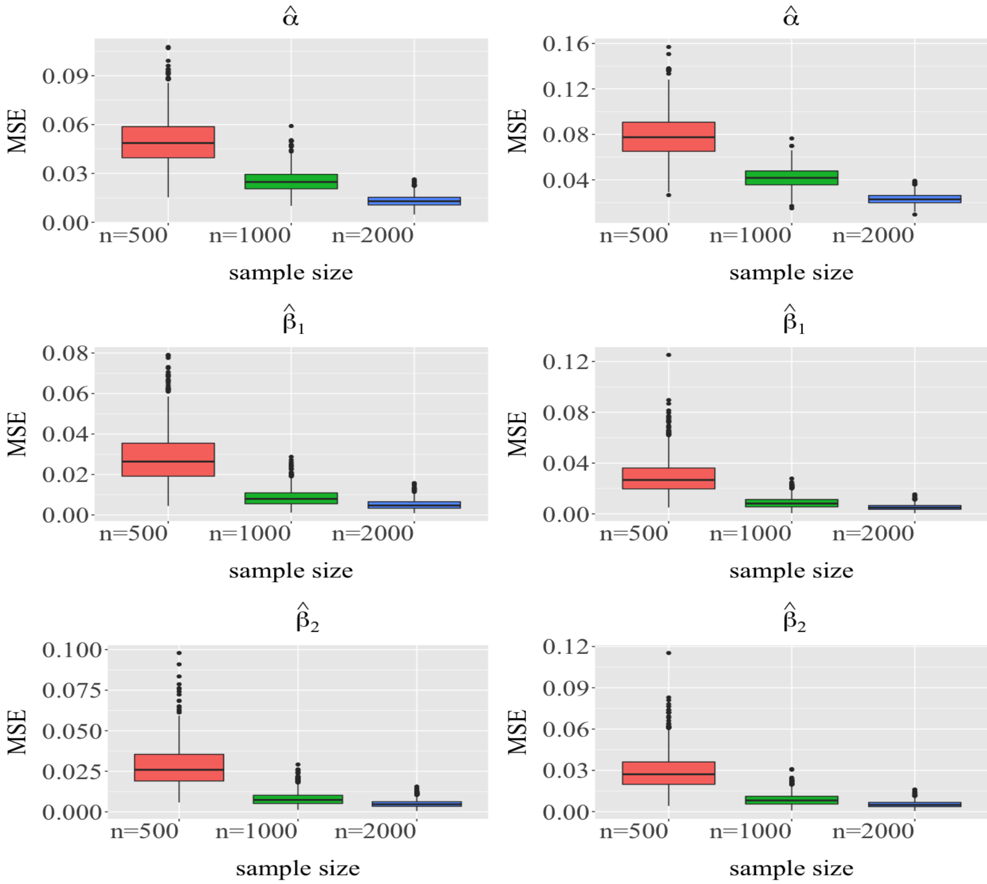

Figure 2. Mean squared error of the spline estimators. First column: the square region; Second column: the horseshoe region.

Table 1. Coverage rate and average length (in parentheses) of confidence intervals.

	n	$z = 0.3$	$z = 0.4$	$z = 0.6$	$z = 0.7$
Square	500	0.920 (0.265)	0.935 (0.260)	0.934 (0.308)	0.934 (0.262)
	1,000	0.931 (0.234)	0.947 (0.233)	0.959 (0.225)	0.947 (0.224)
	2,000	0.949 (0.135)	0.944 (0.134)	0.950 (0.165)	0.959 (0.163)
Horseshoe	500	0.938 (0.278)	0.942 (0.272)	0.948 (0.263)	0.945 (0.263)
	1,000	0.940 (0.207)	0.951 (0.208)	0.948 (0.206)	0.949 (0.199)
	2,000	0.944 (0.156)	0.949 (0.154)	0.951 (0.154)	0.949 (0.154)

6. Real-Data Analysis

The unequal food retail environment (FRE) has been recognized as a critical contextual factor contributing to geographic disparities in obesity. However, there is no clear conclusion on the relationship between the FRE and obesity,

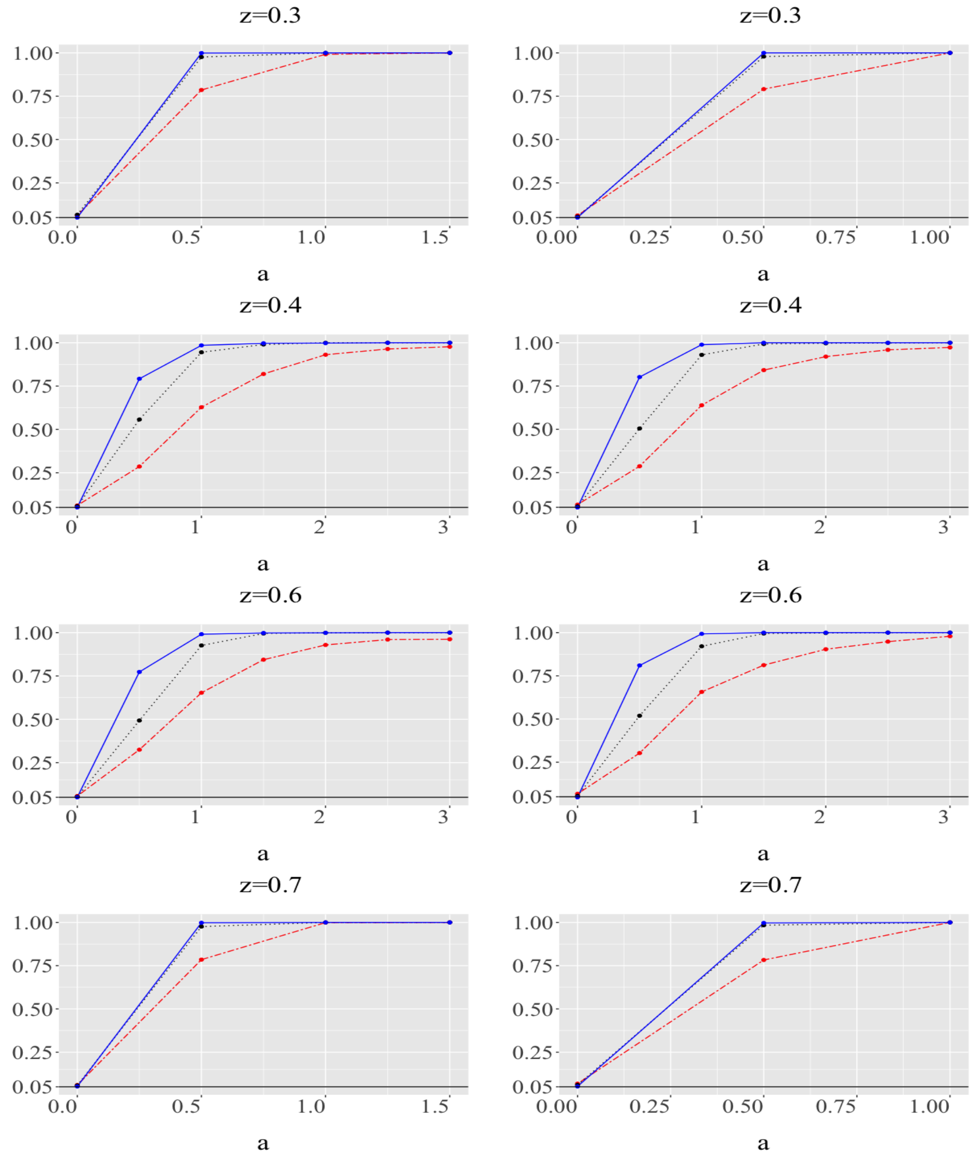

Figure 3. Empirical size and power for the pointwise test $H_0 : \beta_1(z) = \beta_2(z)$ at the 5% nominal level. -··-: $n = 500$; - - -: $n = 1000$; —— : $n = 2000$. First column: square region; Second column: horseshoe region.

owing to diverse measures of the FRE and socioeconomic disparities. In order to resolve this challenge, we include multiple types of food stores, restaurants, and Supplemental Nutrition Assistance Program (SNAP) stores to assess the FRE from two important perspectives: X_1, availability, and X_2, healthfulness. In particular, X_1 is a composite index of the densities of food stores, restaurants, and

Table 2. Empirical size and power for the simultaneous test $H_0 : \beta_1(\cdot) = \beta_2(\cdot)$.

	n	$a = 0$	$a = 0.1$	$a = 0.2$	$a = 0.3$	$a = 0.4$	$a = 0.5$	$a = 0.6$
Square	500	0.045	0.091	0.274	0.604	0.868	0.984	1
	1,000	0.045	0.136	0.572	0.927	0.997	1	1
	2,000	0.050	0.262	0.868	1	1	1	1
Horseshoe	500	0.046	0.078	0.280	0.597	0.879	0.975	1
	1,000	0.049	0.140	0.561	0.937	0.999	1	1
	2,000	0.052	0.256	0.889	0.999	1	1	1

SNAP stores, and X_2 is a composite index of the ratios of healthy to unhealthy food stores, full service restaurants to fast food restaurants, and healthy to unhealthy SNAP stores. Data are collected from 3,091 counties in the United States in 2018. For each county, $\boldsymbol{S}_i = (S_{i1}, S_{i2})^\top$ is their geographical location, and Z_i is their median household income. We model the county-level obesity rate (Y) as the following VCGM:

$$Y_i = \beta_0(Z_i) + X_{i1}\beta_1(Z_i) + X_{i2}\beta_2(Z_i) + \alpha(\boldsymbol{S}_i) + \epsilon_i, \quad i = 1, \ldots, 3091. \tag{6.1}$$

To check whether the two covariates X_1 and X_2 are significant in model (6.1), we first conduct two simultaneous tests $H_{01} : \beta_1(z) = 0$ and $H_{02} : \beta_2(z) = 0$, for all z. For the simultaneous test H_{01}, the test statistic is $D_n = 28.888$, and the 95% quantile of the bootstrap samples is $\widehat{d}_{0.05} = 11.666$; for the simultaneous test H_{02}, the test statistic is $D_n = 85.060$, and the 95% quantile of the bootstrap samples is $\widehat{d}_{0.05} = 11.696$. Hence, both null hypotheses are rejected, indicating that at least for some point z, $\beta_1(z) \neq 0$ and $\beta_2(z) \neq 0$. Next, we investigate the pointwise properties for these varying-coefficient functions. Figure 4 shows the 95% pointwise confidence bands and empirical maximum likelihood estimators for $\beta_0(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$, and the penalized bivariate spline estimator $\widehat{\alpha}(\cdot)$. From the pointwise confidence bands, we conclude that food availability (X_1) and healthfulness (X_2) have strong nonlinear effects on reducing county obesity rates, given the higher household income level, especially when the income value is larger than USD 100,000. Interestingly, the pointwise confidence bands and zero lines together indicate that for those counties with a median household income less than about USD 75,000, food availability (X_1) has no significant impact on the obesity rate. However, the composite index of healthfulness (X_2) has significant negative impact on the obesity rate of counties with a median household income less than about USD 100,000. This finding suggests that increasing the value of healthfulness can help to reduce adult obesity rates in counties with a median household income of less than about USD 100,000. Because there are few counties with a household income greater than USD 100,000, the confidence bands are much wider in that region. Given the relatively large variation, food availability has a negative effect , and the index of healthfulness has no significant impact on the obesity rate. As expected, Figure 4 also indicates that the traditional deep-south states have a large positive geo value $\alpha(\cdot)$, suggesting that these states have higher obesity rates than others with similar FRE values. This reflects that, in addition to the FRE, local food preference, culture, and other factors also influenc county obesity rates.

Because social scientists doubt the association between FRE and obesity may differ with county median household income, $z_0 = 56{,}516$. We perform the pointwise hypothesis test $H_{0P} : \beta_1(z_0) = \beta_2(z_0)$ versus $H_{1P} : \beta_1(z_0) \neq \beta_2(z_0)$ to test whether availability and healthfulness have the same contribution to obesity rates at z_0. We use cubic B-splines for three univariate splines, and consider $d = 2$ and $r = 1$ for the bivariate spline smoothing. The corresponding pointwise test statistic based on the data is 0.137, which accepts H_{0P}. Thus, we conclude that availability and healthfulness do not have significantly different contributions to obesity rates at the median household income point. For availability and healthfulness, we derive the pointwise confidence intervals separately, which are $[-0.552, 0.099]$ and $[-0.356, -0.235]$, respectively. This indicates that at the 95% significance level, we believe that at $z_0 = 56{,}516$, availability has no contribution to obesity rates; however, healthfulness has a negative contribution to obesity rates. The results reflect that, compared with availability, healthfulness is a more influential factor shaping the spatial pattern of obesity rates across counties. The associations between obesity rates and both FRE indicators vary greatly with changes in the county median household income and across space.

7. Conclusion

In this work, we have proposed both pointwise and simultaneous tests for a general hypothesis in a spatial VCM. Compared with classical VCMs, the proposed VCGM is able to handle spatial information in any regular or irregular 2D domains. Furthermore, regression coefficients are allowed to vary systematically and smoothly in some variables. Owing to the advantages over normal approximation-based methods, the EL method is proposed for conducting the inference. We argue that the proposed hypothesis testing method for the VCGM has attractive properties that have not been investigated.

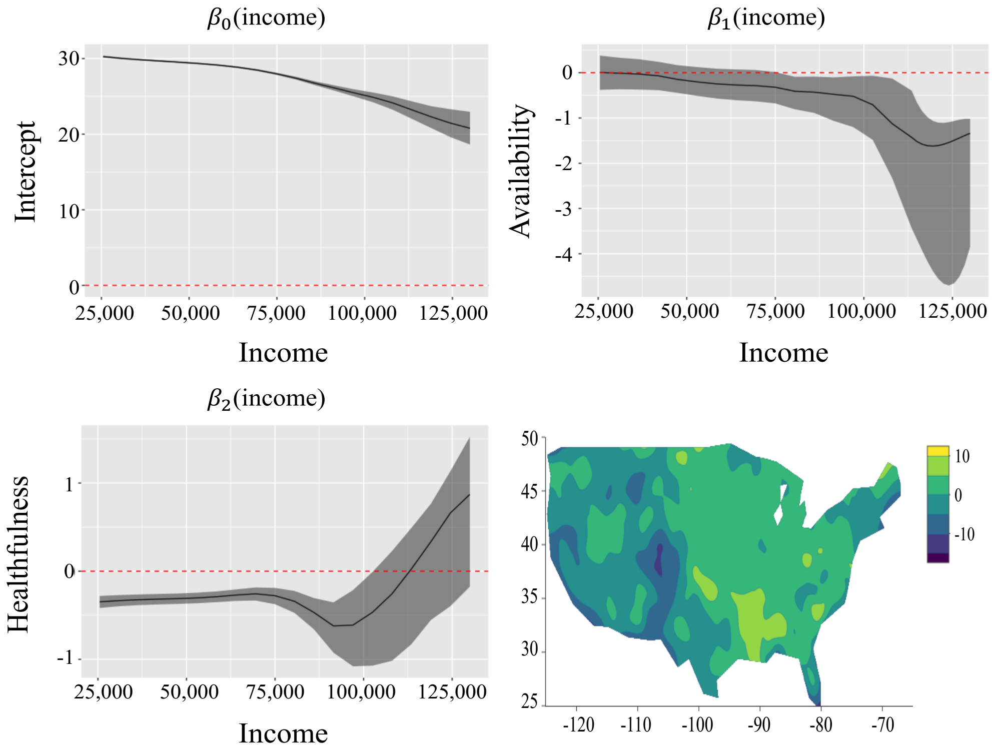

Figure 4. 95% pointwise confidence bands for β_0 (top left), β_1 (top right), and β_1 (bottom left) (——: maximum empirical likelihood estimator $\check{\beta}$; - - -: zero line), and the penalized bivariate spline estimator $\widehat{\alpha}$ (bottom right).

Supplementary Material

Technical assumptions and proofs of Proposition 1, and Theorems 1, 2, and 3 are provided in the online Supplementary Material.

Acknowledgments

We thank the associate editor and two referees for their helpful and constructive comments. S. Wang's and G. Cao's research was partially supported by NSF award DMS 1736470. G. Cao's research was also partially supported by the Simons Foundation under Grant #849413. Y. Zhao's research was supported by NSF award DMS 2006304 and the Simons Foundation under Grant #638679.

References

Ashby, F. G. (2011). *Statistical Analysis of fMRI Data*. MIT Press, Cambridge.

Bandyopadhyay, S., Lahiri, S. and Nordman, D. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *The Annals of Statistics* **25**, 519–545.

Chen, S. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18**, 415–447.

Chen, S. and Zhong, P.-s. (2010). Anova for longitudinal data with missing values. *The Annals of Statistics* **36**, 3630–3659.

DiCiccio, T., Hall, P. and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics* **19**, 1053–1061.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics* **29**, 153–193.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics* **27**, 1491–1518.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface* **1**, 179–195.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B.* **55**, 757–796.

Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128.

Kim, M., Wang, L. and Zhou, Y. (2021). Spatially varying coefficient models with sign preservation of the coefficient functions. *Journal of Agricultural, Biological and Environmental Statistics* **26**, 367–386.

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

Lai, M.-J. and Schumaker, L. L. (2007). *Spline Functions on Triangulations*. Volume 110 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

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.

Liu, R. and Zhao, Y. (2021). Empirical likelihood inference for generalized additive partially linear models. *TEST* **30**, 569–585.

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

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18**, 90–120.

Owen, A. B. (2001). *Empirical Likelihood.* CRC Press, Boca Raton.

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

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics* **10**, 1040–1053.

Van Hala, M., Nordman, D. and Zhu, Z. (2015). Empirical likelihood for irregularly located spatial data. *Statistica Sinica* **25**, 1399–1420.

Wahba, G. (1990). *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Wang, H., Zhong, P.-S., Cui, Y. and Li, Y. (2018). Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **80**, 343–364.

Wang, L., Wang, G., Lai, M. and Gao, L. (2020). Efficient estimation of partially linear models for data on complicated domains by bivariate penalized splines over triangulations. *Statistica Sinica* **30**, 347–369.

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics* **35**, 2474–2503.

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

Xue, L. and Wang, Q. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli* **18**, 836–856.

Xue, L. and Zhu, L. (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association* **102**, 642–654.

Yang, Y., Li, G. and Peng, H. (2014). Empirical likelihood of varying coefficient errors-in-variables models with longitudinal data. *Journal of Multivariate Analysis* **127**, 1–18.

Yu, S., Wang, G., Wang, L., Liu, C. and Yang, L. (2020). Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association* **115**, 761–774.

Shuoyang Wang

Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA.

E-mail: szw0100@auburn.edu

Honglang Wang

Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202, USA.

E-mail: hlwang@iupui.edu

Yichuan Zhao

Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA.

E-mail: yichuan@gsu.edu

Guanqun Cao

Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA.

E-mail: gzc0009@auburn.edu

Yingru Li

Department of Sociology, University of Central Florida, Orlando, FL 32816, USA.

E-mail: yingru.li@ucf.edu

(Received June 2021; accepted October 2021)