Statistica Sinica Preprint N[illegible]S-2021-0233	
Title	Empirical Likelihoo[illegible] Ratio Tests for V[illegible] Co[illegible]ficient Geo Models
Manuscript ID	SS-2021-0233
URL	http://www.stat.sinica.edu.tw/statisti[illegible]
DOI	10.5705/ss.202021.0233
Complete List of Authors	Shuoyang Wan[illegible] Honglang Wang, Yichuan [illegible]ao, [illegible]uan[illegible]o a[illegible] Y[illegible] L[illegible]
Corresponding Author	[illegible]anq[illegible] Cao
E-m[illegible]	[illegible]9@auburn.edu
Notice: Accepted versio[illegible]oje[illegible]o [illegible]glish editing.	

Statistica Sinica

Empirical Likelihood Ratio Tests for Varying Coefficient Geo Models

Shuoyang Wang[a], Honglang Wang[b], Yichuan Zhao[c], Guanqun Cao[a], Yingru Li[d]

[a]Auburn University, [b]Indiana University–Purdue University Indianapolis,

[c]Georgia State University, [d]University of C[illegible]

Abstract:

In this paper, we investigate the varying coefficient model for spatial data distributed over two-dimensional domains. First, the univariate component and the geographical component in the model are approximated via univariate polynomial splines and bivariate penalized splines over triangulation, respectively. The estimators of the univariate and bivariate functions are consistent, and their convergence rates are also established. Second, we propose the empirical likelihood based test procedures to conduct both pointwise and simultaneous inferences for the varying coefficient functions. The asymptotic distributions of the test statistics are derived under the null and local alternative hypotheses. The proposed methods also perform favorably in finite sample applications, as we demonstrate them in simulation studies and an application to an adult obesity prevalence data in the United States.

Key words and phrases: B-spline, Bivariate spline, Empirical likelihood, Geo data, Nonparametric hypothesis testing

1. Introduction

Varying coefficient models (VCMs) introduced by Hastie and Tibshirani (1993) are commonly applied regression models to examine the interactive associations between the response and predictors. The appeal of these models is that the regression coefficients are allowed to vary as a smooth function of some variables of interest to detect non-linear interactions. Because of the flexibility, VCMs have been widely applied to many scientific areas during the past three decades. See Fan and Zhang (2008) for a selective overview on the major methodological and theoretical developments of VCMs. The focus of this work is on VCMs for spatial data randomly distributed over an arbitrary geographical region.

Our work is motivated by inferential problems of examining the effects of county-level food retail environment on obesity rates in U.S. with the effect varying over median household income. County food retail environments are measured by availability and healthfulness of food retail stores. More detailed information of the dataset is provided in Section 6. Based on this dataset, socioeconomists want to disentangle how county-level associations between food environment and obesity rates change with median household income levels. This leads to model the effect of food retail environment as functions of household income levels. However, considering the geographic dependence, the classical VCM is not sufficient.

In this work, we propose the varying coefficient geo model (VCGM) to solve the above motivating application. To be more specific, assume $\boldsymbol{S}_i = (S_{i1}, S_{i2})^\top$ be location for i-th subject, $i = 1, \ldots, n$. 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. $\{(Y_i, Z_i, \mathbf{X}_i)\}$ is 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:

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

where $\boldsymbol{\beta}(Z) = (\beta_1(Z), \ldots, \beta_p(Z))^\top$, with each $\beta_k(\cdot)$ is an unknown varying-coefficient function, $\alpha(\boldsymbol{S}_i)$ is an unknown smoothing bivariate function representing the spatial component, and ε_i's are independent and identically distributed random noises, with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$ are independent of $(Z_i, \mathbf{X}_i, \boldsymbol{S}_i)$. Our primary interest is to estimate and make 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. There have been a plenty number of proposals for fitting the VCM, for example, the local linear method Fan and Zhang (1999), the spline method Huang et al. (2002) and the two-stage methods Wang and Yang (2007); Liu et al. (2013). There are also several methods for estimating bivariate functions defined over 2D domains. Within the nonparametric framework, it includes 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 it can handle irregular 2D domains with complex boundaries and is computationally efficient.

The focus of this paper is 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 due 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 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 non-linear constraints. See Ashby (2011) for some explicit examples of non-linear hypotheses.

In contrast with estimation, less work has been done for the inference of varying coefficient functions, with a few exceptions. For example, Huang et al.

(2002) proposed a goodness-of-fit test based on the comparison of the weighted residual sum of squares. It is a specific incidence of generalized likelihood ratio studied by Fan et al. (2001). More recently, Yu et al. (2020) proposed spline backfitted local polynomial to estimate and make simultaneous inferences of the univariate components in the geo-additive model. Although the above-mentioned methods seems to be incredibly 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 the unstable asymptotic distribution of the test statistics.

In this paper, we propose both pointwise and simultaneous tests for the hypothesis (1.2) based on empirical likelihood (EL). EL is a non-parametric likelihood which was introduced by Owen (1988, 1990). In spite of its nonparametric construction based on observed data points, the EL shares some convenient properties of parametric likelihood and has many desirable advantages in deriving confidence sets for unknown parameters. Owen (2001); Chen and Van Keilegom (2009) gave an overview of EL method. Among those previous works, the EL method has been extended to VCMs for various data types; see, for example Xue and Zhu (2007); Xue and Wang (2012); Yang et al. (2014); Liu and Zhao (2020). 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 for the Google flu trend data, the spatial information contained in the dataset has been ignored. Bandyopadhyay et al. (2015); Van Hala et al. (2015) considered EL method for inference over a broad class of spatial data exhibiting stochastic spatial patterns. But they neither considered the flexible VCGM, nor the spatial information.

Different from existing VCMs, our proposed VCGM takes into account 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. While due to the necessity of estimating the standard errors, which is a typical challenging in nonparametric analysis, the Wald-type simultaneous inference are not stable in Liu and Yang (2020). Second, as DiCiccio et al. (1991) proved, the EL is Bartlett correctable and, thus, it has an advantage over the bootstrap method. To the best of our knowledge, this is the first work of proposing VCGMs and introducing EL ratio test for spatial data, which is a nontrivial extension.

The rest of the paper is organized as follows. We propose the spline estimators for both univariate and bivariate functions and develop their asymptotic consistency in Section 2. The pointwise and simultaneous EL tests are studied 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, number of univariate spline knots and kernel bandwidth selection. Simulation studies are presented in Section 5, followed by analysis of the real data example in Section 6. We summarize the proposed methodology and discuss the future work in Section 7. Major technical details are included in the supplementary material.

2. Univariate and bivariate splines estimations

In the estimation stage, we approximate each varying coefficient by univariate polynomial splines. The geographical function $\alpha(\cdot)$ is approximated via bivariate penalized splines over triangulation. Here we first introduce some notations for univariate spline and bivariate spline.

2.1 Setup

Suppose that the covariate Z is distributed on a compact interval $[a, b]$. Due to the simplicity of computation, we approximate the univariate components $\beta_k(z)$ in (1.1) by polynomial splines. Define a partition of $[a, b]$ with J_n interior knots as $v = \{a = v_0 \leq v_1 \leq \ldots \leq v_{J_{n+1}} = b\}$. For some $\varrho \geq 1$, the polynomial splines of order $\varrho + 1$ are polynomial functions with ϱ-degree on intervals $[v_j, v_{j+1})$, $j = 0, \ldots J_n - 1$, and $[v_{J_n}, v_{J_{n+1}}]$, and have $\varrho - 1$ continuous derivatives globally and let $\mathcal{U} = \mathcal{U}([a,b])$ be the space of such polynomial splines. Let $U_j(z)$, $j = 1, \ldots, J_n + \varrho + 1$, 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 bivariate penalized splines method is efficient to deal with data distributed on irregular domains with complicated boundaries (Yu et al., 2020; Wang et al., 2020). In the following, we briefly introduce the techniques of triangulations and describe the bivariate penalized spline smoothing method for VCGM. We refer to Lai and Schumaker (2007); Wang et al. (2020) for a detailed introduction of the triangulation technique and how to construct the bivariate spline basis functions on the triangulation.

According to Lai and Schumaker (2007), let $\tau = \langle \mathrm{s}_1, \mathrm{s}_2, \mathrm{s}_3 \rangle$ be a nonempty-triangle with three vertices, s_1, s_2, and s_3. There is a unique representation in the form for every point $\boldsymbol{s} \in \mathbb{R}^2$, $\boldsymbol{s} = b_1\mathrm{s}_1 + b_2\mathrm{s}_2 + b_3\mathrm{s}_3$ with $b_1 + b_2 + b_3 = 1$, where b_1, b_2, 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 $B_{ijk}^{\tau,d}(\boldsymbol{s}) = \frac{d}{i!j!k!} b_1^i b_2^j b_3^k$. 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 $\zeta|_\tau = \sum_{i+j+k=d} \gamma^\tau_{ijk} B^{\tau,d}_{ijk}$, where the coefficients $\boldsymbol{\gamma}_\tau = \{\gamma^\tau_{ijk}, i+j+k = d\}$ are called B-coefficients of ζ. For any integer $r \geq 0$, let $\mathbb{C}^r(\Omega)$ be the collection of all r-th continuously differentiable functions over Ω. Given a triangulation $\triangle$, define the spline space of degree d and smoothness r over $\triangle$ as $\mathbb{S}^r_d(\triangle) = \{\zeta \in \mathbb{C}^r(\Omega) : \zeta|_\tau \in \mathbb{P}_d(\tau), \tau \in \triangle\}$. Let $\{B_m\}_{m\in\mathcal{M}}$ be the set of bivariate Bernstein basis polynomials for $\mathbb{S}^r_d(\triangle)$, where $\mathcal{M}$ is an index set with cardinality $|\mathcal{M}| = N(d+1)(d+2)/2$. Then we rewrite any function $\zeta \in \mathbb{S}^r_d(\triangle)$ using the following 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

Generally there are three approaches to conduct spline estimation: smoothing splines, regression splines, and penalized splines. Smoothing splines request as many parameters as the number of observations. Regression splines only need a small number of knots placed judiciously, but appropriate algorithms are needed for knots selection. Penalized splines combine the features of smoothing splines and regression splines. A roughness penalty is incorporated with a relative large number of knots. In terms of bivariate spline smoothing, Wang et al. (2020); Yu et al. (2020) have discussed 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 the 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, whilst 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:

$$\sum_{i=1}^{n}\left\{Y_i-\sum_{k=1}^{p}\sum_{j=1}^{J_n+\varrho+1}\eta_{jk}b_j(Z_i)X_{ik}-\sum_{m\in\mathcal{M}}B_m(\boldsymbol{s})\gamma_m\right\}^2+\frac{\lambda_n}{2}\mathcal{E}(\alpha) \tag{2.3}$$

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

For 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 $\mathbf{\Psi}$ can be found in Yu et al. (2020). Thus, the penalized least-squares problem (2.3) becomes

$$\sum_{i=1}^{n}\left\{Y_i - \sum_{k=1}^{p}\sum_{j=1}^{J_n+\varrho+1}\eta_{j,k}U_j(Z_i)X_{ik} - \sum_{m\in\mathcal{M}}B_m(\boldsymbol{s})\gamma_m\right\}^2 + \frac{1}{2}\lambda_n\boldsymbol{\gamma}^\top\mathbf{P}\boldsymbol{\gamma} \quad (2.4)$$

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

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

a $n\times p(J_n+\varrho+1)$ matrix. To solve the constrained minimization problem (2.4), we first remove the constraint via QR decomposition of the transpose of the constraint matrix $\mathbf{\Psi}$. Specifically, we have $\mathbf{\Psi}^\top = \mathbf{Q}\mathbf{R} = \begin{pmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{pmatrix}\begin{pmatrix}\mathbf{R}_1 \\ \mathbf{0}\end{pmatrix}$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangle matrix, the submatrix $\mathbf{Q}_1$ is the first r columns of $\mathbf{Q}$, where r is the rank of matrix $\mathbf{\Psi}$, and $\mathbf{0}$ is a matrix of zeros. According to Lemma 1 in Wang et al. (2020), the problem (2.4), is now converted to a 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 matrix inverse that

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

where

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

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

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

$$\widehat{\beta}_k(z) = \mathbf{U}(z)^\top\widehat{\boldsymbol{\eta}}_k \quad \text{and} \quad \widehat{\alpha}(\boldsymbol{s}) = \mathbf{B}(\boldsymbol{s})^\top\widehat{\boldsymbol{\gamma}}, \text{ where } \widehat{\boldsymbol{\gamma}} = \boldsymbol{Q}_2\widehat{\boldsymbol{\theta}}. \tag{2.5}$$

We now investigate the asymptotic properties of the spline estimates $\widehat{\beta}_k(z)$ and $\widehat{\alpha}(\boldsymbol{s})$. To avoid the confusion, let $\beta_{0,k}(\cdot)$ and $\alpha_0(\cdot)$ be the true functions of

$\beta_k(\cdot)$ and $\alpha(\cdot)$ in model (2.5). 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, the spline estimators $\widehat{\beta}_k$ and $\widehat{\alpha}$ satisfy that*

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

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) is reduced to the same model in Lai and Wang (2013). When the varying coefficients reduce to linear coefficients, the model (1.1) is reduced to the same model in Wang et al. (2020). In these two reduced models, the convergence rate of $\widehat{\alpha}$ developed above [illegible] the same as the ones given in Lai and Wang (2013) and Wang et al. (2020), i.e., $O_p\left\{n^{-1/2}|\triangle|^{-1} + \frac{\lambda_n}{n|\triangle|^3} + \left(1 + \frac{\lambda_n}{n|\triangle|^5}\right)|\triangle|^{d+1}\right\}$. When the geo function $\alpha(\cdot)$ is [illegible] in the model (1.1), the convergence rate of $\widehat{\beta}_k$ is reduced 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 build hypothesis testing statistics based on these estimators in the next section.

3. Empirical likelihood ratio tests for varying coefficients

There are extraordinarily challenges to derive the asym[illegible] a[illegible] measure of variability for the spline estimators introduced in Section 2. Similar findings have been discussed in Liu et al. (2013); Yu et al. (202[illegible]). To investigate the uncertainty in the estimation of the varying effect of the covariates, we propose the inference for the hypothesis (1.2) via the EL method with bivariate penalized spline estimators plugged in[illegible] function.

To test (1.2) and construct a[illegible] function for $\boldsymbol{\beta}(z)$, we first introduce an auxiliary random ve[illegible]

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

where $K(\cdot)$ is a continuous kernel function and 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 directly used to make 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 $\alpha_0(\cdot)$ by the estimator $\widehat{\alpha}(\boldsymbol{S}_i)$ given in (2.5), i.e.,

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

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

$$\check{\boldsymbol{\beta}}(z) = \left\{\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^\top K_h(Z_i - z)\right\}^{-1}\left\{\sum_{i=1}^n (Y_i - \widehat{\alpha}(\boldsymbol{S}_i))\boldsymbol{X}_i K_h(Z_i - z)\right\}. \tag{3.6}$$

After replacing the true function $\alpha_0(\cdot)$, we show 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 VCM 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 $-2\log R\{\boldsymbol{\beta}(z)\}$ would be similar as the common VCM situations. According to Theorem 1, we established the orders of the first two moments for $g_i\{\boldsymbol{\beta}_0(z)\}$ as in Proposition 1 by bounding $E\left\{\mathbf{X}_i\mathbf{X}_i^\top (\boldsymbol{\beta}_0(Z_i) - \boldsymbol{\beta}_0(z)) K_h(Z_i - z)\right\}$ and $E\{\mathbf{X}_i K_h(Z_i - z)(\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i))\}$ with careful choice of the lower bound of J_n and upper bound of $|\triangle|$. The details can be found in the proof of Proposition 1 in the supplementary material.

With slight abuse of notation, we define the EL function

$$L\{\boldsymbol{\beta}(z)\} = \max_{p_1,\ldots,p_n}\left\{\prod_{i=1}^n p_i : 0 \le p_i \le 1, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_i\{\boldsymbol{\beta}(z)\} = 0\right\}. \tag{3.7}$$

The maximization of (3.7) can be solved by 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

$$\ell(z) := \min_{H\{\boldsymbol{\beta}(z)\}=0} \sum_{i=1}^{n} \log\left\{1 + \boldsymbol{\delta}^\top(z) g_i\{\boldsymbol{\beta}(z)\}\right\}. \qquad (3.[illegible])$$

To investigate the power of the tests, we consider the local alternative H_1 : $H\{\boldsymbol{\beta}_0(z)\} = b_n \boldsymbol{d}(z)$, where b_n is a sequence of num[illegible] $\boldsymbol{d}(z) \neq 0$ is a q-dimensional function. For any fixed non-zero function [illegible], b_n depicts the order of signals that a test can detect. The smallest order of b_n has been discovered in Chen and Zhong (2010), which has shown 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. [illegible] larger than the parametric rate $n^{-1/2}$.

The following theo[illegible] 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, a[illegible] for each $z \in [a, b]$ under the null hypothesis: $H\{\boldsymbol{\beta}_0(z)\} = 0$ we have [illegible] χ_q^2. For each $z \in [a, b]$ and any fixed real vector of function $\mathbf{d}(z)$, under the alternative hypothesis $H_1 : H\{\boldsymbol{\beta}_0(z)\} = (nh)^{-1/2}\mathbf{d}(z)$, we have*

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

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

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 [illegible] confidence interval for $\beta_j(z)$ is given by $\{\beta_j(z) : 2\ell(z) \leq \chi^2_{1,\alpha}\}$.

For simultaneous test on H_0 in (1.2) for all $z \in [a, b]$, we consider the Cramér-von Mises type test statistic. Since $2\ell(z)$ can be viewed as the distance between $H\{\boldsymbol{\beta}(z)\}$ and 0, we propose the following test statistic for H_0

$$D_n = \int_a^b \ell(z)w(z)dz, \tag{3.9}$$

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 $\sigma_0^2 = 2\mu_{20}^{-2}\int_a^b w^2(t)dt\int_{-2}^2\{K^{(2)}(u)\}^2du$. *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 the 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 errors probability in practice, we suggest a bootstrap procedure to generate the empirical quantile and perform the homogeneous testing. The distribution consistency of this method has been discussed in Wang et al. (2018). The proposed bootstrap procedure consists of the following steps.

Step 1. For each subject, calculate residuals $\widetilde{e}_i = Y_i - \check{\boldsymbol{\beta}}(Z_i)^\top \boldsymbol{X}_i - \widehat{\alpha}_i(\boldsymbol{S}_i)$, with local constant estimator $\check{\boldsymbol{\beta}}(z)$ in (2.6). Compute the sample variance of $\widetilde{e}_i$ and denote it as $\widetilde{\sigma}^2$;

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

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

Remark 3. In the 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 spline estimator $\widehat{\boldsymbol{\beta}}(z)$ to generate residuals, as $\check{\beta}(z)$ is maximum empirical likelihood estimator involved in the construction of $\ell(z)$ and D_n.

The following proposition provides the justification of the bootstrap procedure. The proof is similar to Theorem 4 in Wang et al. (2018).

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* $\mathcal{X}_n$, $\mathcal{L}\left(D_n^{(b)}|\mathcal{X}_n\right)$, *converges to* $\mathcal{L}(D_n)$ *almost surely.*

4. Implementation

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

4.1 Tuning parameters selection in univariate and bivariate splines smoothing

In this work, we do not need the spline estimator $\widehat{\boldsymbol{\beta}}(z)$ for the inference of $\boldsymbol{\beta}(z)$ directly. However, $\widehat{\alpha}(\boldsymbol{s})$ is essential for constructing EL ratio tests (3.8) and its estimating procedure involves $\widehat{\boldsymbol{\beta}}(z)$. Hence, we need to make sure that $\widehat{\boldsymbol{\beta}}(z)$ is estimated efficiently. For univariate spline smoothing, [illegible] 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: $|\triangle|^{1/(\varrho+1)}\ n^{2/(5\varrho+5)} \ll J_n \ll |\triangle|^2 n \log^{-1}(n)$. [illegible] the widely used cubic splines, in practice we suggest taking the following rule-of-thumb number of interior knots: J_n [illegible] $\max\{\lfloor c_1 n$[illegible]$\rfloor$ [illegible] where the tuning parameter $c_1 \in [1, 3]$. Similar technique has been considered in Yu et al. (2020). We have also compared the proposed knots selecting method with other data driven methods, i.e., AIC and BIC. The well selected parameters via 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 the practical application.

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 by 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, Delaunay triangulation can be used to build the triangulated meshes. According to our numerical experience, when given the smoothness $r = 1$, comparing with the setting $d = 2$ or 3, setting $d = 5$ requires too much unnecessary computational time, its improvement on accuracy is negligible. We suggest using $r = 1$ and $d = 2$ or 3 in practice, since they can provide enough accuracy for model selection and reduce computational cost simultaneously. Similar settings are also found in Lai and Wang (2013); Yu et al. (2020); Kim et al. (202[illegible]).

The generalized cross-validation (GCV) criterion is one of the efficient methods to select smoothing parameters λ_n, which also has good theoretical properties (Wahba, 1[illegible]). 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

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

We choose the smoothing parameter λ_n by minimizing

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

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

4.2 Bandwidth selection

The performance of the EL pointwise and simultaneous tests depend on the choice of the bandwidth h. We apply the 5-fold cross-validation criterion and choose the bandwidth h by minimizing

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

where $\mathcal{F}_k$ denotes the sample index set for 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 pointwise tests and $h = \lfloor c_3 n^{1/5} \rfloor$ for simultaneous tests, where $c_3 \in \{0.1, 0.2, \ldots, 0.9, 1\}$.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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:

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

where X_{ij}'s and ϵ_i's are independently generated from $N(0,1)$, and Z_i's follows $Unif[0,1]$ independently. In addition, we choose the Epanechnikov kernel $K(x) = 3/4\,(1-x^2)_+$ for local linear estimation, where $(a)_+ = \max(a, 0)$. The sample sizes are chosen to be $n = 500, 1000, 2000$. We consider two different spatial domains for bivariate function $\alpha(\cdot)$: 1) a rectangular domain $[0,1]^2$; 2) a modified horseshoe domain used by Sangalli et al. (2013); Wang et al. (2020). For each Monte Carlo replication, we randomly sample n locations uniformly from the grid points inside the two spatial domains, respectively. 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 [illegible] 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

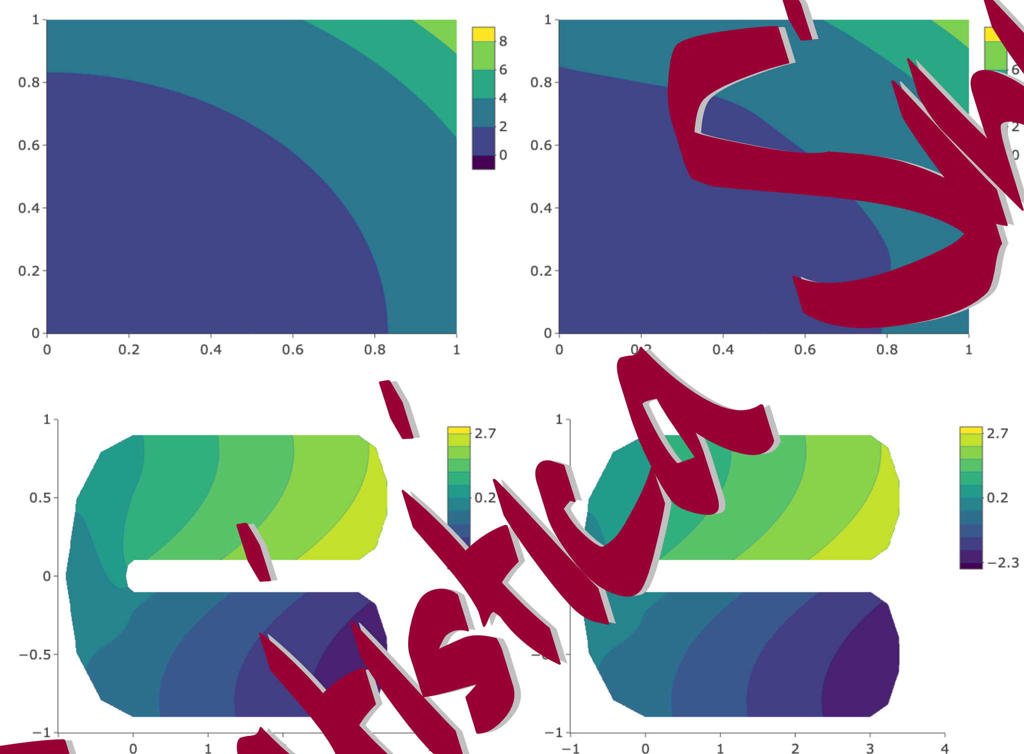

Figure 1: Contour maps of the true function $\alpha_0(\cdot)$ (first column) and the estimators (second column) over square region (first row) and horseshoe region (second row).

sample size $n = 2,000$. The proposed estimates look visually close to the true functions. Figure 2 shows the boxplot of the MSEs of spline estimators for both regions. One can easily find that the MSEs and the corresponding standard deviations are decreasing with the increasing of sample sizes.

We first conduct pointwise hypothesis testings. [illegible] $\{(\beta_1, \beta_2)^\top\}$ [illegible] β[illegible] β_2 to test $H_0 : \beta_1(z) = \beta_2(z)$ versus $H_1 : \beta_1(z)$ [illegible] here, [illegible] $\beta_1(z) = (2 + a)\sin(2\pi z)$ and $\beta_2(z) = 2\sin(2\pi z)$ for some nonnegative [illegible] model (5.10) to evaluate the empirical size (when $a = 0$) an[illegible] (when $a > 0$) at 5% nominal level. Figure 3 shows the empirical sizes and powers with two different domains of $\alpha(\boldsymbol{s})$ and different z [illegible].3, 0.4, 0.6, 0.7\}$. Given each z, empirical size is reasona[illegible] nd nominal level 5% for all different sample sizes, and power[illegible] w[illegible] a until reaching 1. As expected, larger sample size lea[illegible] ver.

[illegible]ext, we set β[illegible] sin[illegible]), $\beta_2(z) = 2\sin(z + 1/2)$ in model (5.10) and apply the [illegible] ection 3 to construct pointwise confidence intervals for β_1[illegible] ominal level. Table 1 summarizes the empirical coverage probability as [illegible] ges 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 are increasing with sample size, and are around 95% when $n = 2,000$. It can also been seen that the length of confidence inter-

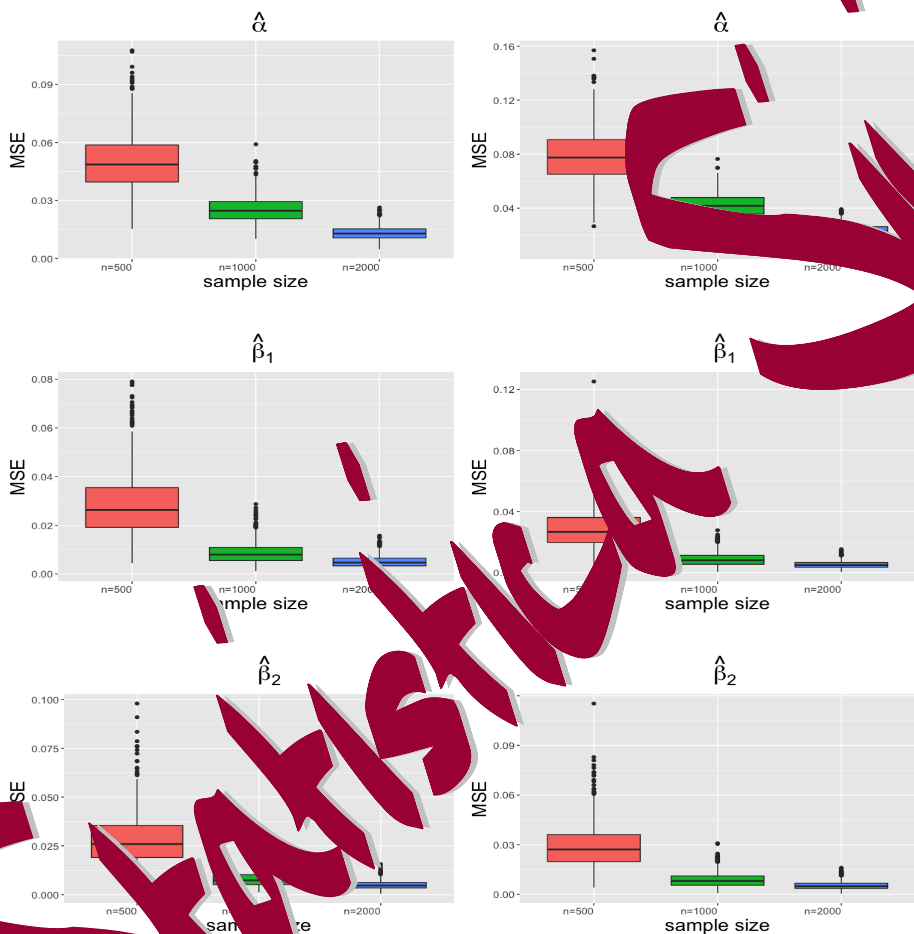

Figure 2: Mean squared error of spline estimators. First column: square region; Second column: horseshoe region.

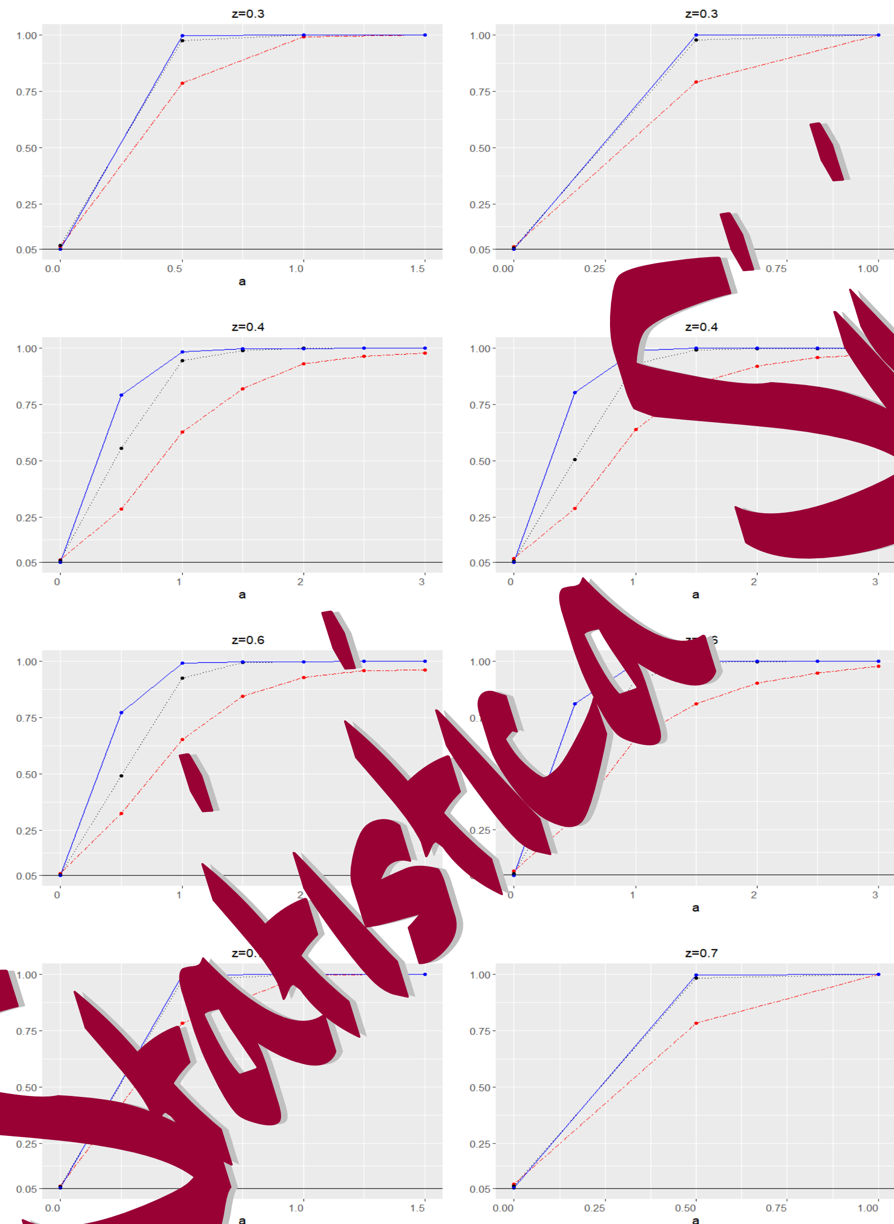

Figure 3: Empirical size and power for the pointwise test $H_0 : \beta_1(z) = \beta_2(z)$ at 5% nominal level. -·-·-: $n = 500$; - - -: $n = 1,000$; —— : $n = 2,000$. First column: square region; Second column: horseshoe region.

vals is decreasing as the sample size is increasing.

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 (5.10). We evaluate the empirical size (when $a = 0$) and power (when $a > 0$) and the results are presented in Table 2. All tests are under two different 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 was estimated by 500 bootstrap samples in each simulation run. From Table 2, we find that the empirical size for each n is around nominal level 5%, and the trends of powers are reasonably controlled.

6. Real Data Analysis

The unequal food retail environment (FRE) has been recognized as a critical contextual factor contributing to geographic disparities in the obesity. However, there is no clear conclusion on the relationship between FRE and obesity due to diverse measures of FRE and socioeconomic disparities. In order to resolve this challenge, this study included multiple types of food stores, restaurants, and Supplemental Nutrition Assistance Program stores to assess FRE from two important perspectives of FRE, X_1, availability and X_2, healthfulness. In partic-

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

	n	$z = 0.3$	$z = 0.4$	$z = 0.6$	$z = 0.7$
	500	0.920 (0.265)	0.935 (0.260)	[illegible]	0.934 (0.262)
Square	1000	0.931 (0.234)	0.947 (0.233)	[illegible] (0.225)	[illegible]
	2000	0.949 (0.135)	0.944 (0.134)	0.950 (0.165)	[illegible]
	500	0.938 (0.278)	0.942 (0.272)	0.948 (0.263)	0.945 (0.263)
Horseshoe	1000	[illegible]	0.951 (0.208)	[illegible]	0.949 (0.199)
	2000	0.944 (0.156)	[illegible]	[illegible] (0.154)	0.949 (0.154)

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

	n	$a = 0$	[illegible]	$a = 0.2$	$a = 0.3$	$a = 0.4$	$a = 0.5$	$a = 0.6$
	500	[illegible]	[illegible]	0.274	0.604	0.868	0.984	1
Square	1000	[illegible]	0.136	0.572	0.927	0.997	1	1
	2000	[illegible]	0.262	0.868	1	1	1	1
	500	0.046	0.078	0.280	0.597	0.879	0.975	1
Horseshoe	1000	0.049	0.140	0.561	0.937	0.999	1	1
	2000	0.052	0.256	0.889	0.999	1	1	1

ular, X_1 is a composite index of densities of food stores, restaurants, and Supplemental Nutrition Assistance Program (SNAP) stores and X_2 is a composite index of 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 taken by their geographical location, and Z_i is taken by median household income. We model the county level obesity rate (Y) as the following VCGM:

$$Y_i = \beta_0(Z_i) + X_{i1}\beta_1(Z_i) + X_{i2}\beta_2(Z_i) + \alpha(\boldsymbol{S}_i) + \epsilon_i, i = 1, \cdots, 3,091, \quad (6.11)$$

To check if two covariates X_1 and X_2 are significant in model (6.11), we first conduct two simultaneous tests $H_{01} : \beta_1(z) = 0$ and $H_{02} : \beta_2(z) = 0$ for all z. For simultaneous test H_{01}, the test statistic is $D_n = 28.888$ and 95% quantile of the bootstrap samples is $\widehat{d}_{0.05} = 11.666$; for simultaneous test H_{02}, the test statistic is $D_n =$ [illegible] and the 95% quantile of the bootstrap samples is $\widehat{d}_{0.05} = 11.69$[illegible]. Hence both null hypotheses are rejected, indicating that at least for some points, $\beta_1(z) \neq 0$ and $\beta_2(z) \neq 0$. Next, we further investigate the pointwise properties for these varying coefficient functions. Figure 4 shows the pointwise confidence bands and empirical maximum likelihood estimators for $\beta_0(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$ and penalized bivariate spline estimator $\widehat{\alpha}(\cdot)$. From the pointwise confidence bands, we can conclude that food availability (X_1) and healthfulness (X_2) have strong nonlinear effects on reducing county

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

obesity rates given the higher household income level, especially when income value is larger than $\$100,000$. Interestingly, the pointwise confidence bands and zero lines together indicate that for those counties with the median household income less than about $\$75,000$, food availability ($X_1$) has no significant impact on the obesity rate. At the mean while, the composite index of healthfulness (X_2) has significant negative impact on the obesity rate for counties with median household income less than about $\$100,000$. This finding suggests that increasing the value of healthfulness can help reducing adult obesity rate for county whose median household income is less than about $\$100,000$. As there are few numbers of counties having household income greater than $\$100,000$, the confidence bands are much wider in that region. Given the relative large variation, food availability has negative effect and index of healthfulness has no significant impact on the obesity rate [illegible]. As expected, Figure 4 also indicates that the traditional deep-south states have large positive value of geo value $\alpha(\cdot)$, suggesting that these states have higher obesity rates than other places with similar FRE [illegible] reflects that besides food retail environment, local food preference, culture [illegible] other factors have also influenced county obesity rates.

As the social scientists doubt the association of FRE and obesity may differ with the county median household income $z_0 = 56,516$. We perform the pointwise hypothesis testing $H_{0P} : \beta_1(z_0) = \beta_2(z_0)$ vs. $H_{1P} : \beta_1(z_0) \neq \beta_2(z_0)$

to test if availability and healthfulness have the same contribution to the obesity rates at z_0. We use cubic B-splines for three univariate splines, and we consider $d = 2$ and $r = 1$ for the bivariate spline smoothing. The corresponding pointwise test statistics based on data is 0.137, which accepts H_{0P}. Thus we conclude that availability and healthfulness does not have significantly different contribution to obesity rate at the median household income part. For availability and healthfulness, we derive the pointwise confidence interval separately, which are $[-0.552, 0.099]$ and $[-0.356, -0.235]$. This indicates that at 95% significance level, we believe at $z_0 = 56,516$, availability has no contribution to obesity rates; nevertheless, healthfulness has negative contribution to obesity rates. The results reflect that, compared to the availability, healthfulness is a more influential factor for shaping the spatial pattern of obesity rates across counties. The associations between obesity rates and both these FRE indicators vary greatly with the change of county median household income and across the space.

In this work, we propose both of 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. Meanwhile, regression coefficients are allowed to vary systematically

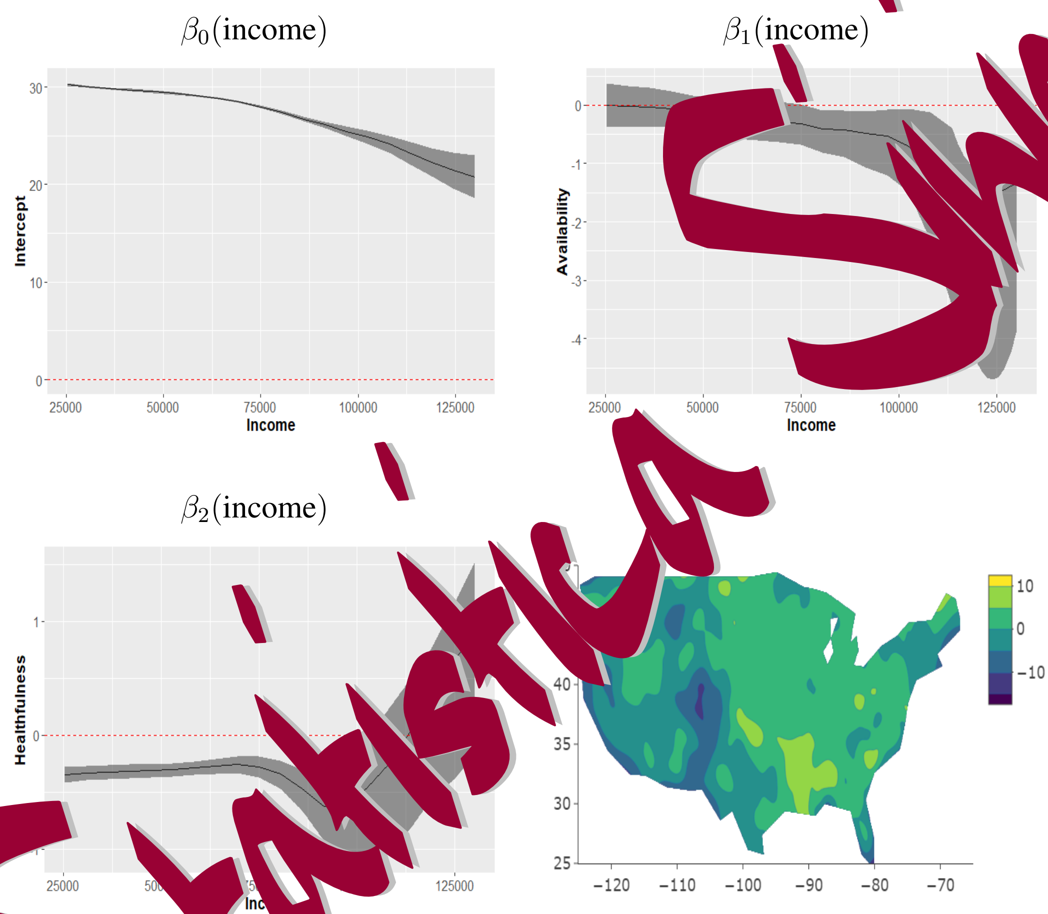

Figure 4: 95% pointwise confidence bands for β_0 (top left), β_1 (top right) and β_1 (bottom left) (—: maximum empirical likelihood estimator $\check{\beta}$; - - -: zero line) and penalized bivariate spline estimator $\widehat{\alpha}$ (bottom right).

and smoothly in some variables. Due to 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 and fascinating properties that have not been investigated.

Supplementary Material

Technical assumptions, proofs of Proposition 1, Theorems 1, 2 and 3 are provided in the supplementary material.

Acknowledgements

We thank the Associate Editor and referees for their helpful and constructive comments, which lead to significant improvement in this paper. S. Wang's and G. Cao's research was partially supported by NSF award DMS 1736470. G. [illegible] research was partially supported by Simons Foundation under Grant #849413. Y. Z[illegible]'s research was supported by NSF award DMS 2006304 and Simons Foundation under Grant #638679.

References

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

Bandyopadhyay, S., S. Lahiri, and D. Nordman (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *The Annals of Statistics 25*(2), 519–545.

Chen, S. and I. Van Keilegom (2009). A review on empirical likelihood methods for regression. *TEST 18*(3), 415—447.

Chen, S. and P.-s. Zhong (2010). Anova for longitudinal data with missing values. *The Annals of Statistics 36*(6), 3630–3659.

DiCiccio, T., P. Hall, and J. Romano (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics 19*(2), 1053–1061.

Fan, J., C. Zhang, and J. Zhang (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics 29*(1), 153–193.

Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics 27*(5), 1491–1518.

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

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B. 55*(4), 757–796.

Huang, J. Z., C. O. Wu, and L. Zhou (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika 89*(1), 111–128.

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

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

Liu, R., L. Yang, and W. K. Härdle (2013). Oracally efficient two-step estimation of generalized additive model. *Journal of the American Statistical Association 108*(502), 619–631.

Liu, R. and Y. Zhao (2020). Empirical likelihood inference for generalized additive partially linear models. *TEST*, in press.

Marx, B. D. and P. H. C. Eilers (2005). Multidimensional penalized signal regression. *Technometrics 47*(1), 13–22.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*(1), 90–120.

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

Sangalli, L. M., J. O. Ramsay, and T. O. Ramsay (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B 75*(4), 681–703.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics 10*(4), 1040–1053.

Van Hala, M., D. Nordman, and Z. Zhu (2015). Empirical likelihood for irregularly located spatial data. *Statistica Sinica 25*(2), 1399–1420.

Wahba, G. (1990). *Spline models for observational data*, Volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Wang, H., P.-S. Zhong, Y. Cui, and Y. Li (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. 80*(2), 343–364.

Wang, L., G. Wang, M. Lai, and L. Gao (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 L. Yang (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics 35*(6), 247[illegible].

Wood, S. N. (2003). Thin plate regression splines. *J. R. Stat. Soc. Ser. B Stat. Methodol. 65*(1), 95–114.

Xue, L. and Q. Wang (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli* [illegible], 8[illegible]–856.

Xue, L. and L. Zhu (20[illegible]). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association* [illegible]-654.

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

Yu, S., G. Wang, L. Wang, C. Liu, and L. Yang (2020). Estimation and inference for generalized geoadditive models. *J. Amer. Statist. Assoc. 115*(530), 761–774.

Shuoyang Wang, Auburn University

E-mail: szw0100@auburn.edu

Honglang Wang, Indiana University–Purdue University Indianapolis

E-mail:hlwang@iupui.edu

Yichuan Zhao, Georgia State University

E-mail: yichuan@gsu.edu

Guanqun Cao, Auburn University

E-mail: gzc0009@auburn.edu

Yinru Li, Central Florida University

E-mail: yingru.li@ucf.edu