Statistica Sinica Preprint No: SS-2024-0118	
Title	Semi-nonparametric Varying Coefficients Models for Imaging Genetics
Manuscript ID	SS-2024-0118
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0118
Complete List of Authors	Ting Li, Yang Yu, Xiao Wang, J.S. Marron and Hongtu Zhu
Corresponding Authors	Hongtu Zhu
E-mails	htzhu@email.unc.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Semi-nonparametric Varying Coefficients Models for Imaging Genetics

Ting Li[1], Yang Yu[2], Xiao Wang[3], J.S. Marron[2] and Hongtu Zhu[4]

[1]*School of Statistics and Data Science, Shanghai University of Finance and Economics.*

[2]*Department of Statistics, University of North Carolina at Chapel Hill.*

[3]*Department of Statistics, Purdue University.*

[4]*Department of Biostatistics, University of North Carolina at Chapel Hill.*

Abstract: Motivated by imaging genetics, this paper introduces a semi-nonparametric varying coefficients modeling framework to reveal varying associations between genetic markers and imaging responses. We aim to conduct a comprehensive theoretical analysis of estimation and inference procedures applicable to these models. By employing the kernel machine method, we estimate unknown varying coefficient functions and derive their representer theorem. We also establish the theoretical properties of these estimated functions, including their rate of convergence, Bahadur representation, point-wise limit distributions, and confidence intervals. Additionally, we propose test statistics under a linear mixed effects model framework to assess the significance of all varying coefficients, taking into account within-subject dependence. The efficacy of our proposed methodology is demonstrated through simulation studies and an application to data from the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Alzheimer's Disease Neuroimaging Initiative study.

Key words and phrases: Imaging Genetics; Kernel ridge regression; Linear mixed effects model; Reproducing kernel Hilbert space.

1. Introduction

Advancements in modern imaging and genetic techniques have facilitated large-scale neuroimaging genetics studies aimed at understanding the genetic foundations of human brain structure and function. Prominent examples include the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Mueller et al., 2005) and the UK Biobank (UKB) study (Miller et al., 2016). Imaging traits are increasingly used as biomarkers for diagnosis and prognosis, as well as endophenotypes to identify genetic markers linked to various brain-related disorders (Elliott et al., 2018; Zhao et al., 2021). These traits provide critical insights into the biological pathways that connect genetics with imaging characteristics and brain-related disorders which are confounded with health factors such as diet and alcohol. Nonetheless, the combined analysis of imaging and genetic data poses significant challenges to current statistical methods due to their high dimensionality and intricate spatio-temporal structures (Zhu et al., 2023).

A compelling data example that motivates our proposed method is derived from ADNI, to identify relevant genetic markers for various brain-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

related diseases (e.g., Alzheimer's disease) in genome-wide imaging genetics (Nathoo et al., 2019; Le and Stein, 2019) using imaging endophenotypes. The hippocampus, crucial for learning and memory, often exhibits significant tissue loss at the early stages of Alzheimer's disease (AD), resulting in a functional disconnection from other brain regions (Rao et al., 2022). Traits linked to the hippocampus have become pivotal biomarkers for grasping the nuances of aging and for diagnostic purposes. Analyzing hippocampus-related imaging phenotypes alongside genetic data is key to unraveling the genetic underpinnings that dictate brain structure and function, and it is instrumental in pinpointing relevant genetic markers. Our focus is to understand the influence of genetic pathways on hippocampal imaging traits after adjusting for clinical and demographic variables. Given the complex and less understood relationships between genes and imaging phenotypes, we propose a versatile framework designed to capture the influence of genetic pathways effectively.

We introduce a semi-nonparametric varying coefficient (SVC) modeling framework to correlate imaging traits with genetic markers while controlling for clinical and demographic covariates. Specifically, we analyze data from n independent subjects, represented as $(\mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i)_{i=1}^n$, where $\mathbf{x}_i$ is a $q \times 1$ vector of gene expressions within a pathway, $\mathbf{z}_i$ is a $p \times 1$ vector of clinical

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

covariates, and $\mathbf{y}_i = \{y_i(s) : s \in \mathcal{S}\}$ are the imaging traits observed in a common compact space $\mathcal{S}$. Drawing inspiration from Liu et al. (2007), our model posits that clinical covariates exert linear effects, while genetic covariates may have non-linear or linear effects using the least squares kernel machine. We consider the following SVC model for imaging genetics

$$y_i(s) = h(\mathbf{x}_i, s) + \mathbf{z}_i^T \boldsymbol{\gamma}(s) + \epsilon_i(s) \quad \text{for} \quad i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\gamma}(s) = (\gamma_1(s), \ldots, \gamma_p(s))^T$ is a $p \times 1$ varying-coefficient function, $h(\mathbf{x}, s)$ is an unknown multivariate smooth function, and $\{\epsilon_i(s) : s \in \mathcal{S}\}$ is a measurement error process and independent of $(\mathbf{x}_i, \mathbf{z}_i)$. Model (1.1) captures the spatial-varying effects of genetic markers $\mathbf{x}$ and clinical variables $\mathbf{z}$ on neuroimaging data. In practice, $\mathbf{y}_i$ can be a 1-dimensional curve, a 2-dimensional surface or matrix, or a 3-dimensional volume extracted from various neuroimaging modalities. For simplicity, we set $\mathcal{S} = [0, 1]$ throughout the paper, although our results are extensible to higher dimensional spaces. Each regression function is assumed to reside in a Reproducing Kernel Hilbert Space (RKHS), with unknown parameters estimated via a least squares kernel machine technique. In our ADNI dataset, $y_i(s)$ measures morphometry along either the left or right hippocampus, with $\mathbf{x}_i$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

comprising genetic data, and $\mathbf{z}_i$ incorporating clinical variables.

The construction of the SVC model is due to the specific characteristics of the ADNI dataset. Our primary objective is to investigate the genetic pathway effects on the hippocampus, while controlling for the parametric effects of clinical and demographic covariates. Given the complex relationship between genetic factors and hippocampal structure, we introduce a flexible nonparametric framework for the genetic data to capture these intricate effects more effectively. In contrast, the clinical variables are modeled linearly, as linear adjustments suffice to account for their influence. This also helps to balance model flexibility with computational feasibility, avoiding the substantial computational burden that would arise from a fully nonparametric approach across all covariates. This strategy is consistent previous studies such as Liu et al. (2007); Kwee et al. (2008); Wu et al. (2011), which utilized general nonparametric models for genetic effects while adjusting for linear effects of clinical covariates.

Model (1.1) reduces to popular function-on-scalar regression models under the functional data analysis (FDA) framework when $h(\cdot, \cdot) = 0$. Various advanced approaches have been proposed to study the effects of scalar covariates on functional responses, such as basis expansion (Reiss et al., 2010; Krafty et al., 2008; Li et al., 2021), kernel smoothing (Zhang and Chen,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

2007; Li et al., 2011; Zhu et al., 2014), and methods within the Bayesian framework (Lindquist et al., 2010; Yang et al., 2020). Comprehensive reviews of FDA models can be found in Morris (2015) and Wang et al. (2016). However, these approaches do not account for the nonlinear varying coefficient effect $h(\mathbf{x}_i, s)$ of the scalar covariates.

Meanwhile, over the past decade, there have been significant advancements in functional data analysis, particularly tailored for imaging data. Notably, varying coefficient models (Zhu et al., 2012; Yuan et al., 2013; Zhu et al., 2014) have been employed to treat imaging data as functional responses, which aids in identifying causal clinical variables and examining their explanatory capabilities. Furthermore, functional principal component analysis (Goodlett et al., 2009; Zhu et al., 2011) is utilized to isolate factor functions that capture the variability in brain structures; while functional (linear) regression analysis (Zhu et al., 2010; Kong et al., 2018; Yu et al., 2021) explores the predictive capacity of imaging data in forecasting certain neurological or clinical outcomes. Despite these methodologies' success in analyzing complex imaging data, they exhibit inherent limitations when applied to massive imaging genetics studies.

In this paper, we introduce the SVC model to study the linear or nonlinear varying coefficient effects of clinical and genetic covariates on the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

functional response. We make several contributions compared to existing literature: (i) We propose an estimation procedure utilizing the kernel machine technique within the RKHS framework. We establish a representer theorem, showing that the solution can be found in a finite-dimensional subspace. (ii) We explore the theoretical properties of our estimators, including convergence rates, the Bahadur representation, and point-wise limiting distribution of the estimators. Additionally, we observe a phase transition phenomenon in the SVC model, where the rate of convergence remains unaffected by the number of design points once they exceed a certain threshold, thereby extending the findings of Cai and Yuan (2011) from the mean function to the SVC model. (iii) We derive point-wise confidence bands for the coefficient functions $\gamma_\nu(\cdot)$ and the multivariate function h, and develop test statistics for assessing the clinical and genetic effects within a linear mixed effects model framework (Zhang and Lin, 2003). We estimate the covariance function from the data nonparametrically, which is non-trivial and more practical. (iv) Our analysis of the ADNI dataset validates the effectiveness and advantages of the proposed SVC model.

2. Estimation Procedure

For $\nu = 1, \ldots, p$, each function $\gamma_\nu(s)$ is assumed to reside in a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}_s$. We also assume $h(\cdot, \cdot)$ belongs to a tensor

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

product space $\mathcal{H}_x \otimes \mathcal{H}_s$ with $\mathcal{H}_x$ being another RKHS. To ensure the model's identification, it is assumed that the expected value of the covariate $\mathbf{z}$ is zero, and the expected value of $h(\cdot,\cdot)$ is zero. For background information and examples of RKHS, refer to references (Wahba, 1990; Hofmann et al., 2008; Gu, 2013). We further assume that $y_i(s)$ is sampled at a sequence of m design points, $0 = s_1 \leq s_2 \leq \ldots \leq s_m = 1$ for all i. Define

$$\mu(\mathbf{x}, \mathbf{z}, s) \quad = h(\mathbf{x}, s) + \mathbf{z}^T \boldsymbol{\gamma}(s) = \textstyle\sum_{\nu=1}^{p} z_\nu \gamma_\nu(s) + h(\mathbf{x}, s).$$

Subsequently, μ is characterized within an RKHS $\mathcal{H}$ as $\mathcal{H} = \bigoplus_{\nu=1}^{p} \big[[z_\nu] \otimes \mathcal{H}_s\big] \oplus \big[\mathcal{H}_x \otimes \mathcal{H}_s\big]$, where $[z_\nu]$ represents the subspace spanned by the basis $\{z_\nu\}$. We denote $\mathcal{H}_\nu = [z_\nu] \otimes \mathcal{H}_s$ for $\nu = 1, \ldots, p$ and $\mathcal{H}_{p+1} = \mathcal{H}_x \otimes \mathcal{H}_s$. Given that $\mathcal{H}_s$ and $\mathcal{H}_x$ are constructed by the reproducing kernels K_s and K_x, the reproducing kernels K_ν for the subspace $\mathcal{H}_\nu$ are defined as

$$K_\nu\big((z_\nu, s), (\tilde{z}_\nu, \tilde{s})\big) \quad = \quad z_\nu \tilde{z}_\nu K_s\big(s, \tilde{s}\big) \text{ and } K_{p+1}\big((\mathbf{x}, s), (\tilde{\mathbf{x}}, \tilde{s})\big) = K_z\big(\mathbf{x}, \tilde{\mathbf{x}}\big) K_s\big(s, \tilde{s}\big).$$

The reproducing kernel for the Hilbert space $\mathcal{H}$ can be defined as

$$\begin{aligned} K\big((\boldsymbol{x}, \mathbf{z}, s), (\tilde{\boldsymbol{x}}, \tilde{\mathbf{z}}, \tilde{s})\big) &= \sum_{\nu=1}^{p} \theta_\nu K_\nu\big((z_\nu, s), (\tilde{z}_\nu, \tilde{s})\big) + \theta_{p+1} K_{p+1}\big((\mathbf{x}, s), (\tilde{\mathbf{x}}, \tilde{s})\big) \\ &= \left\{ \sum_{\nu=1}^{p} \theta_\nu z_\nu \tilde{z}_\nu + \theta_{p+1} K_z(\mathbf{x}, \tilde{\mathbf{x}}) \right\} K_s(s, \tilde{s}), \end{aligned} \tag{2.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{p+1})^T$ is a vector of subsidiary regularization parameters. Estimation in (1.1) is derived by minimizing the following loss function with roughness penalties such that $\widehat{\mu} = \arg\min_{\mu \in \mathcal{H}} \ell_{n,m,\lambda}(\mu)$, where

$$\ell_{n,m,\lambda}(\mu) = (2nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \big[y_i(s_j) - \mu(\mathbf{x}_i, \mathbf{z}_i, s_j)\big]^2 + \frac{\lambda}{2} \sum_{\nu=1}^{p+1} \theta_\nu^{-1} \|P^\nu \mu\|_{\mathcal{H}}^2, \tag{2.2}$$

in which P^ν is the orthogonal projector in $\mathcal{H}$ onto $\mathcal{H}_\nu$ and λ is the primary regularization parameter. Specifically, the unknown function lies in the space $\mathcal{H}$ such that $\mu(\mathbf{x}, \mathbf{z}, s) = \sum_{\nu}^{p+1} \mu_\nu(\mathbf{x}, \mathbf{z}, s)$, where $\mu_\nu(\cdot) \in \mathcal{H}_\nu$. The penalty term can be written as $\|P^\nu \mu\|_{\mathcal{H}}^2 = \theta_\nu^2 \|\mu_\nu\|_{\mathcal{H}_\nu}^2$ with $\|\cdot\|_{\mathcal{H}_\nu}^2$ being the norm in the RKHS $\mathcal{H}_\nu$. This model incorporates $p+2$ regularization parameters in total, however, any configurations of $(\lambda, \boldsymbol{\theta})$ that maintain consistent ratios $\lambda_\nu = \lambda/\theta_\nu$ for $\nu = 1, \ldots, p$ are considered equivalent. The parameter λ serves two main purposes. It first aligns the model with the one-dimensional least squares kernel machine framework, facilitating the application of established analytical results. It also acts as a stabilizing factor in the optimization algorithm. This stabilization ensures that the search for λ within the algorithm remains closely aligned with the optimal solution. The minimization problem in equation (2.2) is solved by applying the representer theorem below.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Theorem 1. *There exists a matrix* $\mathbf{C} = [c_{ij}]_{i=1}^{n}{}_{j=1}^{m} \in \mathbb{R}^{n\times m}$ *such that*

$$\widehat{\mu}(\cdot) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} K\big((\mathbf{x}_i, \mathbf{z}_i, s_j), \cdot\big). \tag{2.3}$$

The representer theorem (Theorem 1) shows that the solution to the optimization problem in (2.2) resides within a finite-dimensional subspace. This subspace is spanned by the kernel function K, evaluated at the design points $\{(\mathbf{x}_i, \mathbf{z}_i, s_j)|i = 1, \ldots, n; j = 1, \ldots, m\}$. Consequently, the solution is expressible as a linear combination of these kernel evaluations, $\boldsymbol{C}$ is the matrix of the coefficients $\{c_{ij}\}_{i=1}^{n}{}_{j=1}^{m}$, making the implementation of the method simple and efficient.

Let $\mathbf{K}_s = [K_s(s_{j_1}, s_{j_2})]_{j_1=1}^{m}{}_{j_2=1}^{m}$ and $\mathbf{K}_x = [K_x(\mathbf{x}_{i_1}, \mathbf{x}_{i_2})]_{i_1=1}^{n}{}_{i_2=1}^{n}$ be the Gram kernel matrices for the kernel functions K_s and K_x, respectively. The Gram kernel matrices of the kernel functions K_ν for $\nu = 1, \ldots, p+1$ are defined as $\mathbf{K}_\nu = \mathbf{Z}_\nu \mathbf{Z}_\nu^T \otimes \mathbf{K}_s$ for $\nu = 1, \ldots, p$ and $\mathbf{K}_{p+1} = \mathbf{K}_x \otimes \mathbf{K}_s$. where $\mathbf{Z}_\nu$ denotes the ν-th column of the matrix $\mathbf{Z}$ and $\otimes$ denotes the Kronecker product. As $\widehat{\mu}(\cdot) = \boldsymbol{c}^\top \sum_{\nu=1}^{p+1} \theta_\nu K_\nu((\boldsymbol{x}_i, \boldsymbol{z}_i, s), \cdot)$ due to the formulation of K in (2.1) and (2.3), then we can have $\|P^\nu \widehat{\mu}\|_{\mathcal{H}}^2 = \theta_\nu^2 \mathbf{c}^T \mathbf{K}_\nu \mathbf{c}$, where $\mathbf{c} = \mathrm{vec}(\mathbf{C}^T) = (c_{11}, \ldots, c_{1m}, \ldots, c_{n1}, \ldots, c_{nm})^T$. Defining $\mathbf{K} = \sum_{\nu=1}^{p+1} \theta_\nu \mathbf{K}_\nu$ and

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

$\mathbf{Y} = (y_{11}, \ldots, y_{1m}, \ldots, y_{n1}, \ldots, y_{nm})^T$, then

$$\widehat{\mu} = \arg\min_{\mathbf{c}} \left\{ \frac{1}{nm} \|\mathbf{Y} - \mathbf{K}\mathbf{c}\|^2 + \lambda \mathbf{c}^T \mathbf{K} \mathbf{c} \right\}, \tag{2.4}$$

resulting in $\widehat{\mathbf{c}} = (\mathbf{K} + nm\lambda\mathbf{I})^{-1}\mathbf{Y}$. Upon reconverting $\widehat{\mathbf{c}}$ into matrix form, we obtain the estimator for the parameter matrix $\mathbf{C}$, denoted as $\widehat{\mathbf{C}}$. Let $\mathbf{K}_s(s) = [K_s(s, s_1), \ldots, K_s(s, s_m)]^T$ and $\mathbf{K}_x(\mathbf{x}) = [K_x(\mathbf{x}, \mathbf{x}_1), \ldots, K_x(\mathbf{x}, \mathbf{x}_n)]^T$, the estimators for $\gamma_\nu(s)$ for $\nu = 1, \ldots, p$ and $h(\mathbf{x}, s)$ are expressed as

$$\widehat{\gamma}_\nu(s) \quad = \theta_\nu \mathbf{Z}_\nu^T \widehat{\mathbf{C}} \mathbf{K}_s(s) \quad \text{and} \quad \widehat{h}(\mathbf{x}, s) \quad = \theta_{p+1} \mathbf{K}_x(\mathbf{x})^T \widehat{\mathbf{C}} \mathbf{K}_s(s). \tag{2.5}$$

The selection of appropriate kernel functions is required for our proposed SVC model. Among the common choices are the Gaussian, polynomial, and sigmoid kernels. The decision on which kernel to employ typically hinges on prior insights into the nature of the regression function being modeled. Notably, the Gaussian kernel is frequently favored in practice due to its advantageous properties and robust performance (Poggio and Girosi, 1990). Detailed discussions on tuning parameter selection are provided in Section S2 of the supplementary material. We have refined the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Broyden, 1970) for optimizing multiple smoothing parameters using the Generalized Cross-Validation

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

(GCV) criterion. Additionally, we have developed a novel algorithm for selecting kernel parameters, which is inspired by scale space theory.

In practice, the kernel based method may be affected by the curse of dimensionality. To address this, several strategies can be employed to mitigate these challenges. The Gaussian kernel works by computing the similarity between data points based on their Euclidean distance in the feature space. It emphasizes local relationships, which can help to mitigate some effects of high dimensionality by focusing on local patterns rather than global ones (Schölkopf and Smola, 2002). Meanwhile, prior to applying the Kernel Machine method, we can consider dimensionality reduction techniques such as Principal Component Analysis (PCA) to reduce the number of features while retaining the most informative components of the data, thus alleviating the burden of high dimensionality. Furthermore, feature selection algorithms can be employed to identify and retain only the most relevant features for the model, reducing the dimensionality of the input space.

3. Linear Operators

In this section, we introduce a novel inner product and some linear operators to derive the Fréchet derivatives of the loss function, which greatly facilitates theoretical studies. Let $\mathbf{u} = (\mathbf{x}, \mathbf{z}, s)$ be an element of the space $\mathcal{U} = \mathcal{X} \times \mathcal{Z} \times \mathcal{S}$, and define $g(\mathbf{z}, s) = \mathbf{z}^T \boldsymbol{\gamma}(s)$. Consequently, the regression

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

function is given by $\mu(\mathbf{u}) = g(\mathbf{z}, s) + h(\mathbf{x}, s)$, which belongs to the RKHS $\mathcal{H} = [\mathcal{H}_z \otimes \mathcal{H}_s] \oplus [\mathcal{H}_x \otimes \mathcal{H}_s]$, where $\mathcal{H}_z, \mathcal{H}_x$ and $\mathcal{H}_s$ are individual RKHSs generated by theirs kernel respectively, and $\otimes$ denotes the tensor product of these spaces. Here, $\mathcal{H}_z$ is the RKHS associated with $\boldsymbol{z}$, which is generated by by the first-order polynomial kernel $K_z(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{z}_1^T \mathbf{z}_2$. This kernel corresponds to the space of linear functions in $\mathbf{z}$. Therefore, any g belongs to the tensor product $\mathcal{H}_z \otimes \mathcal{H}_s$ can be expressed $\boldsymbol{z}^\top \beta(s)$. The inner product in $\mathcal{H}$ is defined as $\langle \mu_1, \mu_2 \rangle_{\mathcal{H}} = \sum_{\nu=1}^{p+1} \theta_\nu^{-1} \langle P^\nu \mu_1, P^\nu \mu_2 \rangle_{\mathcal{H}}$ for any $\mu_1, \mu_2 \in \mathcal{H}$, which is equivalent to $\sum_{\nu=1}^{p+1} \langle P^\nu \mu_1, P^\nu \mu_2 \rangle_{\mathcal{H}}$ for $\theta_\nu > 0$ (Gu, 2013). For simplicity, we assume $\{\theta_\nu = 1\}_\nu^{p+1}$ through our theoretical investigation.

We consider a convenient representation of our error function in (2.2). First, we equip $\mathcal{H}$ with a new inner product defined by

$$\langle \mu_1, \mu_2 \rangle_{\widetilde{\mathcal{H}}} = \langle \mu_1, \mu_2 \rangle_{\mathcal{L}_2} + \lambda \langle \mu_1, \mu_2 \rangle_{\mathcal{H}}, \tag{3.1}$$

where $\langle \mu_1, \mu_2 \rangle_{\mathcal{L}_2} = E_{\mathbf{u}}\{\mu_1(\mathbf{u})\mu_2(\mathbf{u})\}$. The corresponding reproducing kernel is denoted as $\widetilde{K}$, such that for any $f \in \mathcal{H}$, $\langle \widetilde{K}_{\mathbf{u}}, f \rangle_{\widetilde{\mathcal{H}}} = f(\mathbf{u})$ with $\widetilde{K}_{\mathbf{u}} = \widetilde{K}(\mathbf{u}, \cdot)$. We further introduce a positive definite operator $W_\lambda : \mathcal{H} \mapsto \mathcal{H}$ such

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

that $\langle W_\lambda \mu, \tilde{\mu} \rangle_{\widetilde{\mathcal{H}}} := \lambda \langle \mu_1, \mu_2 \rangle_{\mathcal{H}}$, the loss function in (2.2) can be rewritten as

$$\ell_{n,m,\lambda}(\mu) = \frac{1}{2nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \big[y_i(s_j) - \langle \widetilde{K}_{\mathbf{u_{ij}}}, \mu \rangle_{\widetilde{\mathcal{H}}} \big]^2 + \frac{1}{2} \langle W_\lambda \mu, \tilde{\mu} \rangle_{\widetilde{\mathcal{H}}}.$$

Define $S_{n,m,\lambda}(\mu) = -(nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} [y_i(s_j) - \mu(u_{ij})] \widetilde{K}_{\mathbf{u_{ij}}} + W_\lambda \mu$. We have $S_{n,m,\lambda}(\widehat{\mu}) = 0$ and $S_{n,m,\lambda}(\mu_0)$ can be expressed as $S_{n,m,\lambda}(\mu_0) = -(nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \widetilde{K}_{\mathbf{u_{ij}}} \epsilon_i(s_j) + W_\lambda \mu_0$, which plays an important role in deriving the convergence rate and Bahadur representation.

We further explore the series expansion of the two operators $\widetilde{K}_{\mathbf{u}}$ and $W_\lambda \mu$. According to Mercer's theorem, the kernel functions K_z, K_x, and K_s can be decomposed as

$$K_z(\mathbf{z}_1, \mathbf{z}_2) = \sum_{k=1}^{\infty} \varphi_k^{(z)}(\mathbf{z}_1) \varphi_k^{(z)}(\mathbf{z}_2) = \sum_{k=1}^{p} z_{1k} z_{2k}, \quad (3.2)$$

$$K_x(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k=1}^{\infty} \tau_k^{(x)} \varphi_k^{(x)}(\mathbf{x}_1) \varphi_k^{(x)}(\mathbf{x}_2), \quad K_s(s_1, s_2) = \sum_{k=1}^{\infty} \tau_k^{(s)} \varphi_k^{(s)}(s_1) \varphi_k^{(s)}(s_2) \quad (3.3)$$

for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{H}_z$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{H}_x$ and $s_1, s_2 \in \mathcal{H}_s$. Here, $\{\tau_k^{(x)}\}_{k=1}^{\infty}$ and $\{\tau_k^{(s)}\}_{k=1}^{\infty}$ represent the eigenvalues associated with K_x and K_s, while $\{\varphi_k^{(z)}\}_{k=1}^{\infty}$, $\{\varphi_k^{(x)}\}_{k=1}^{\infty}$ and $\{\varphi_k^{(s)}\}_{k=1}^{\infty}$ are eigenfunctions for K_z, K_x and K_s. The eigenfunctions in (3.2) can be specified as $\varphi_k^{(z)}(\mathbf{z}) = z_k$ for $k = 1, \ldots, p$ and $\varphi_k^{(z)}(\mathbf{z}) = 0$ for $k > p$, where $\mathbf{z} = (z_1, \ldots, z_p)^T \in \mathbb{R}^p$. Leveraging (3.2)–(3.3), the eigen-decomposition of the kernel function $K_{\mathbf{u}} = K_z(\mathbf{z}, \cdot) K_s(s, \cdot) +$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

$K_x(\mathbf{x},\cdot)K_s(s,\cdot)$ can be expressed as

$$K_{\mathbf{u}} = \{\sum_{k=1}^{\infty}\varphi_k^{(z)}(\mathbf{z})\varphi_k^{(z)} + \sum_{k=1}^{\infty}\tau_k^{(x)}\varphi_k^{(x)}(\mathbf{x})\varphi_k^{(x)}\}\{\sum_{k'=1}^{\infty}\tau_{k'}^{(s)}\varphi_{k'}^{(s)}(s)\} =: \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\tau_{kk'}\varphi_{kk'}(\mathbf{u})\varphi_{kk'}$$

for $\mathbf{u} \in \mathcal{H}$, where $\phi_{kk'}$ denotes a function $\phi_{kk'}(\cdot)$ and $\tau_{kk'} := (1+\tau_k^{(x)})\tau_{k'}^{(s)}$ if $k \le p$, $\tau_{kk'} := \tau_k^{(x)}\tau_{k'}^{(s)}$ if $k > p$, and $\varphi_{kk'}(\mathbf{u}) := \{\varphi_k^{(z)}(\mathbf{z}) + \varphi_k^{(x)}(\mathbf{x})\}\varphi_{k'}^{(s)}(s)$.

4. Theoretical Properties

In this section, we investigate the asymptotic properties of $\hat{\mu}$. In the smoothing spline framework, rates of convergence can be obtained through either the quadratic approximation approaches (Gu, 2013) or the empirical process techniques (Shang and Cheng, 2013). In our setting, we tackle the problem following a similar spirit as the empirical process techniques.

We need the following assumptions. For positive sequences a_n and b_n, let $a_n \lesssim b_n$ ($a_n \gtrsim b_n$) to indicate that there exists a universal constant $c > 0$ ($c' > 0$) independent of n such that $a_n \le cb_n$ ($a_n \ge c'b_n$) for all $n \in \mathbb{N}$.

Assumption 1. *There are constants $C_{\varphi_x}, C_{\varphi_s} \in (0,\infty)$ and $C_{\tau_x}, C_{\tau_s} \in (0,\infty)$ ensuring that $\sup_k \|\varphi_k^{(x)}\|_{\sup} \le C_{\varphi_x}$, $\sup_k \|\varphi_k^{(s)}\|_{\sup} \le C_{\varphi_s}$, $\sup_{\mathbf{x}} K_x(\mathbf{x},\mathbf{x}) \le C_{\tau_x}$, and $\sup_s K_s(s,s) \le C_{\tau_s}$. Here, $\|\cdot\|_{\sup}$ denotes the supremum norm, defined by $\|f\|_{\sup} := \sup_x |f(x)|$. Additionally, $\mathbf{z}_i$ are uniformly bounded by a constant $C_z \in (0,\infty)$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Assumption 2. *The errors* ϵ_i *are identically independent and satisfy that* $E\{\epsilon_i(s)\} = 0$, $\inf_s E\{\epsilon_i^2(s)\} < \sigma_\epsilon^2 < \infty$, *and* $Cov(\epsilon_i(s), \epsilon_i(s')) = r(s, s')$ *for* $s, s' \in [0, 1]$. *Also,* $\sum_{k,k'} E\Big\{ \int \varphi_{kk'}(\mathbf{u})\epsilon_i(s)ds \int \varphi_{kk'}(\mathbf{u}')\epsilon_i(s')ds' \Big\} < \infty$.

Assumption 1 is widely accepted in the literature (Zhao et al., 2016; Cheng and Shang, 2015). Specifically, for kernels that decay polynomially, it has been established that eigenfunctions derived from the ν-th order Sobolev space are uniformly bounded, provided certain mild smoothness conditions are met (Cheng and Shang, 2015). Similarly, for exponentially decaying kernels, Zhao et al. (2016) has demonstrated that eigenfunctions can be uniformly bounded by 1.336. Assumption 2 imposes expectation conditions on the error function. The proposition below gives the connection between the norms $\|\cdot\|_{\sup}$ and $\|\cdot\|_{\widetilde{\mathcal{H}}}$.

Proposition 1. *For any* $\mu \in \mathcal{H}$, *we have* $\|\mu\|_{\sup} \leq C_\varphi d(\lambda)^{1/2}\|\mu\|_{\widetilde{\mathcal{H}}}$ *where* $C_\varphi = (C_z + C_{\varphi_x})C_{\varphi_s}$ *and* $d(\lambda) := \sum_{kk'}\{1 + \lambda/\tau_{kk'}\}^{-1}$.

The term $d(\lambda)$ can be viewed as the effective dimension of $\mathcal{H}$ (Zhang, 2005). Putting $\tau_{kk'}$ in an increasing order as $\tilde{\tau}_k$ leads to $d(\lambda) = \sum_k 1/(1 + \lambda/\tilde{\tau}_k)$. Using different kernel functions can result in different effective dimensions. We give some specific cases in Examples 1–3.

We present the convergence rate of the SVC estimator. The following theorem details the convergence properties of $\widehat{\mu}$ to its true value μ_0.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Theorem 2. *Suppose that Assumptions 1 and 2 are satisfied, and s_j are independent and identically distributed with a density function such that $\inf_s P(s) \geq c_0 > 0$. Furthermore, if $d(\lambda)^{-1} = o(1)$, $(nm)^{-1}d(\lambda) = o(1)$ and $\sqrt{\log\log\big(nmJ(\mathcal{Q},1)\big)}J(\mathcal{Q},1) = o_p((nm)^{1/2}d(\lambda)^{-1})$ hold, where $J(\mathcal{Q},1)$ is a function of covering number defined in (S1.2) of the supplementary material, then we have $\|\widehat{\mu} - \mu_0\|^2_{\widetilde{\mathcal{H}}} = O_p(d(\lambda)(nm)^{-1} + n^{-1} + \lambda)$.*

Theorem 2 presents the rate of convergence for $\widehat{\mu}$ in $\|\cdot\|^2_{\widetilde{\mathcal{H}}}$, which depends on λ, n, and m. The optimal choice of λ depends on the type of kernels and (n, m). When the number of locations m is large, it has no effect on the rate of convergence and the rate of convergence would be dominated by the term n^{-1}. A phase transition phenomenon happens, and the transition orders for m are different for different kernels. Hence, we extend the results in Cai and Yuan (2011) for the mean function with polynomial kernels to the SVC model with a broad class of reproducing kernels. Furthermore, we consider a substantially different structure which involves not only the observation points of the functional response, but also the scalar covariates. This immediately causes a difference in building the reproducing kernel of the mean function. Implicitly, the term $d(\lambda)$ in the rate of convergence is also related to the dimension of $\boldsymbol{x}$, which will also influence the rate. Following are some commonly encountered examples.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Example 1. If $K_x(\cdot)$ is the finite rank kernel, then the decay rate of $\tilde{\tau}_k = q\tau_k^{(s)}$. For the polynomial decay kernel of order r with $\tilde{\tau}_k \asymp qk^{-2r}$ and $d(\lambda) \asymp (\lambda/q)^{-1/(2r)}$, the optimal choice of λ is $\lambda \asymp (nm)^{-2r/(2r+1)}q^{1/(2r+1)}$, which yields the optimal convergence rate $O_p(q^{1/(2r+1)}(nm)^{-2r/(2r+1)} + n^{-1})$. The optimal rate is of order $(nm)^{-2r/(2r+1)}q^{1/(2r+1)}$ if m is below the order $n^{1/2r}$, and it is of order n^{-1} if $m > n^{1/2r}$. For the exponential decay kernel of order r, we have $\tilde{\tau}_k \asymp q\exp(-\alpha k^r)$ for a constant $\alpha > 0$ and $d(\lambda) \asymp (\log \lambda^{-1}q)^{1/r}$ (Zhao et al., 2016). In this case, the optimal choice of λ is $\lambda \asymp q(nm)^{-1}$ and the corresponding optimal convergence rate is $O_p((nm)^{-1}(\log(nm))^{1/r} + (nm)^{-1}q + n^{-1})$. The phase transition happens when m is of order $(\log n)^{1/r}$.

Example 2. If $K_x(\cdot)$ and $K_s(\cdot)$ are all polynomial decay kernels of order r, then we have $\tilde{\tau}_k \asymp k^{-2r/(q+1)}$ by Wahba (1990) due to the $(p+1)$ elements in h and $d(\lambda) \asymp \lambda^{-(q+1)/(2r)}$ by explicit calculations. Then the optimal choice of λ is $\lambda \asymp (nm)^{-2r/(2r+(q+1))}$ and the optimal convergence rate is $O_p((nm)^{-2r/(2r+q+1)} + n^{-1})$. In this case, the phase transition happens when m is of order $n^{(q+1)/2r}$, which is larger than the order $n^{1/2r}$ of the single polynomial decay kernel case in Example 1. When $m < n^{(q+1)/2r}$, the rate of convergence is the same as the optimal convergence rate in multivariate function estimation (Stone, 1994).

Example 3. If $K_x(\cdot)$ and $K_s(\cdot)$ are all exponential decay kernel of order

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

r with $\tau_k^{(s)} \asymp \tau_k^{(x)} \asymp \exp(-\alpha k^r)$ of each element for a constant $\alpha > 0$, then $\tilde{\tau}_k \asymp \exp(-\alpha k^{r/(q+1)})$ and $d(\lambda) \asymp (\log \lambda^{-1})^{(q+1)/r}$ by direct calculations. Therefore, the optimal choice of λ is $\lambda \asymp (nm)^{-1}$ and the optimal convergence rate is $O_p((nm)^{-1}(\log(nm))^{(q+1)/r} + n^{-1})$. The phase transition happens when m is of order $(\log n)^{(q+1)/r}$.

We present the Bahadur representation in the following theorem to characterize the leading term of our estimator, which is the first-order Fréchet derivative of the loss function. The Bahadur representation is a precise approximation of an estimator, and provides an approximation that facilitates the analysis of the asymptotic properties of the estimator.

Theorem 3. *Suppose that the conditions in Theorem 2 hold, then we have*

$$\|\widehat{\mu} - \mu_0 + S_{n,m,\lambda}(\mu_0)\|_{\widetilde{\mathcal{H}}} = O_p(a_n),$$

where $a_n = (nm)^{-1/2} d(\lambda) \sqrt{\log\log\big(nmJ(\mathcal{Q},1)\big)} J(\mathcal{Q},1)(d(\lambda)/(nm) + n^{-1} + \lambda)^{1/2}$.

Theorem 3 has two important implications. First, it provides a higher order approximation of $\widehat{\mu}$. Different from the result that targets the nonparametric regression for scalar response Shang and Cheng (2013), the functional response and different types of scalar covariates lead to a more complex theoretical investigation. Second, the Bahadur representation greatly

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

facilitates the study of point-wise limit distribution of the estimator $\hat{\mu}(\mathbf{u}_0)$.

We define $d_2(\lambda) = \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}(1+\lambda/\tau_{kk'})^{-2}$. For any $\mathbf{u}_0 \in \mathcal{U}$, denote

$$\sigma^2_{\mathbf{u}_0} = \sigma^2_\epsilon \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\frac{\varphi_{kk'}(\mathbf{u}_0)^2}{(1+\lambda/\tau_{kk'})^2} \quad \text{and} \quad r^2_{\mathbf{u}_0} = \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\frac{r_{kk'}\varphi_{kk'}(\mathbf{u}_0)^2}{(1+\lambda/\tau_{kk'})^2},$$

where $r_{kk'} = E(\int\int \varphi_{kk'}(\mathbf{u})r(s,s')\varphi_{kk'}(\mathbf{u}')dsds')$ and the expectation is taken over $\mathbf{u}$ and $\mathbf{u}'$.

Theorem 4. *If the conditions in Theorem 3 are satisfied and* $a_n^2 d(\lambda) = o_p((nm)^{-1}(d_2(\lambda)+m))$ *, recall that* μ_0 *admits the expansion that* $\mu_0 = \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\mu_{kk'}\varphi_{kk'}$*, if* $\sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\mu_{kk'}/(\tau_{kk'})^{1/2} < \infty$ *and* $(nm\lambda)/(d_2(\lambda)+m) = O_p(1)$*, then we have*

$$\sqrt{\frac{nm}{\sigma^2_{\mathbf{u}_0} + mr^2_{\mathbf{u}_0}}}(\hat{\mu}(\mathbf{u}_0) - \mu_0(\mathbf{u}_0)) \xrightarrow{d} N(0,1). \tag{4.1}$$

Theorem 4 can be used to construct confidence and prediction intervals for the estimated mean function $\hat{\mu}(\mathbf{u}_0)$, as well as point-wise intervals for the estimated coefficients $\widehat{\boldsymbol{\gamma}}$ and $\widehat{h}$. Specifically, we can carefully choose x_0 such that $h(x_0, s) = 0$. For example, it can be achieved by setting $x_0 = \infty$ for the Gaussian kernel. Then, the asymptotic properties of $\widehat{\gamma}_j(s)$ are the same as those of $\hat{\mu}(x_0, \mathbf{z}_0, s)$ with $\mathbf{z}_{0_j} = 1$ and $\mathbf{z}_{0_{j'}} = 0$ for $j \neq j'$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Similarly, the asymptotic properties of $\widehat{h}(x,s)$ are the same as those of $\hat{\mu}(x,0,s)$ corresponding to $\mathbf{z}=0$.

5. Inference procedures

In this section, we derive the covariance functions of the proposed estimators in order to construct point-wise confidence bands of $\widehat{\gamma}_\nu(s)$ and $\widehat{h}(\mathbf{x},s)$ and investigate the nullity of $\gamma_\nu(s)$ and $h(\mathbf{x},s)$ based on a score test approach.

There are two methods for calculating the covariance functions. Both are based on an equivalent formulation of (2.5), which is given by

$$\widehat{h}(\mathbf{x},s)=\theta_{p+1}(\mathbf{K}_x(\mathbf{x})^T\otimes\mathbf{K}_s(s)^T)\widehat{\mathbf{c}} \text{ and } \widehat{\gamma}_\nu(s)=\theta_\nu(\mathbf{Z}_\nu^T\otimes\mathbf{K}_s(s)^T)\widehat{\mathbf{c}} \quad (5.1)$$

for $\nu=1,\ldots,p$. The first method is based on a frequentist statistical approach. Specifically, we treat γ_ν and h as fixed unknown functions and directly calculate the variance functions of $\widehat{h}(\mathbf{x},s)$ and $\widehat{\gamma}_\nu(s)$ based on (5.1) and $\widehat{\mathbf{c}}=\widetilde{\mathbf{K}}^{-1}\mathbf{Y}$, where $\widetilde{\mathbf{K}}=\mathbf{K}+nm\lambda\mathbf{I}$. Based on a frequentist approach, We have for $\nu,\omega=1,\ldots,p$,

$$\begin{aligned}\mathbb{C}\mathrm{ov}_F(\widehat{\gamma}_\nu(s),\widehat{\gamma}_\omega(\tilde{s})) &= \theta_\nu\theta_\omega(\mathbf{Z}_\nu^T\otimes\mathbf{K}_s(s)^T)\widetilde{\mathbf{K}}^{-1}\boldsymbol{\Sigma}_\epsilon\widetilde{\mathbf{K}}^{-1}(\mathbf{Z}_\omega\otimes\mathbf{K}_s(\tilde{s})),\\ \mathbb{C}\mathrm{ov}_F(\widehat{h}(\mathbf{x},s),\widehat{h}(\tilde{\mathbf{x}},\tilde{s})) &= \theta_{p+1}^2(\mathbf{K}_x(\mathbf{x})^T\otimes\mathbf{K}_s(s)^T)\widetilde{\mathbf{K}}^{-1}\boldsymbol{\Sigma}_\epsilon\widetilde{\mathbf{K}}^{-1}(\mathbf{K}_x(\tilde{\mathbf{x}})\otimes\mathbf{K}_s(\tilde{s})),\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

where $\mathbf{\Sigma}_\epsilon = \text{diag}(\Sigma, \ldots, \Sigma)$ and $\Sigma = (\Sigma_{ij})$ with $\Sigma_{ii} = \sigma_\epsilon$ for $i = 1, \ldots, m$ and $\Sigma_{ij} = r(s_i, s_j)$ for $i \neq j$.

Utilizing a Bayesian statistical framework, the true underlying functions $\gamma_\nu(\cdot)$ and $h(\cdot, \cdot)$ are treated as random functions, which are assumed to follow prior Gaussian processes with zero mean and covariance functions $\tau_\nu K_\nu(\cdot, \cdot)$ and $\tau_{p+1} K_{p+1}(\cdot, \cdot)$. This approach aligns with the methodologies used in Zhang and Lin (2003) and Liu et al. (2007), where it is posited that $\mathbf{y}|(\boldsymbol{\gamma}(s), h(\mathbf{x}, s))$ follows a normal distribution $N(\mathbf{z}^T\boldsymbol{\gamma}(s) + h(\mathbf{x}, s), \mathbf{\Sigma}_\epsilon)$. Therefore, model (1.1) can be reformulated as the linear mixed effects model, for $\nu = 1, \ldots, p+1$, $\tau_\nu = (nm\lambda)^{-1}\sigma_\epsilon^2\theta_\nu$,

$$\mathbf{Y} = \sum_{\nu=1}^{p+1} \boldsymbol{\zeta}_\nu + \boldsymbol{\epsilon} \quad \text{and} \quad \boldsymbol{\zeta}_\nu \sim N(0, \tau_\nu \mathbf{K}_\nu).$$

Denote $\tau = (nm\lambda)^{-1}\sigma_\epsilon^2$ and $V = \tau\mathbf{K} + \mathbf{\Sigma}_\epsilon$, The covariances of the random effects ζ_ν, for $\nu = 1, \ldots, p$ and $\omega = 1, \ldots, p$, can be computed as

$$\begin{aligned}
\mathbb{C}\text{ov}_B(\widehat{\gamma}_\nu(s), \widehat{\gamma}_\omega(\tilde{s})) &= \tau_\nu K_s(s, \tilde{s})\mathbb{1}(\nu = \omega) - \tau_\nu\tau_\omega(\mathbf{Z}_\nu^T \otimes \mathbf{K}_s(s)^T)V^{-1}(\mathbf{Z}_\omega \otimes \mathbf{K}(\tilde{s})), \\
\mathbb{C}\text{ov}_B(\widehat{h}(\mathbf{x}, s), \widehat{h}(\tilde{\mathbf{x}}, \tilde{s})) &= \tau_{p+1} K_x(\mathbf{x}, \tilde{\mathbf{z}}) K_s(s, \tilde{s}) - \tau_{p+1}^2(\mathbf{K}_x(\mathbf{x})^T \otimes \mathbf{K}_s(s)^T)V^{-1}(\mathbf{K}_x(\tilde{\mathbf{x}}) \otimes \mathbf{K}_s(\tilde{s})),
\end{aligned}$$

where $\mathbb{1}(\cdot)$ represents the indicator function. These covariances are described as Bayesian posterior covariances within the smoothing spline ANOVA

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

framework as discussed in Gu and Wahba (1993).

We examine two types of hypothesis testing problems,

$$H_0 : \gamma_\nu = 0 \quad \text{vs.} \quad H_1 : \gamma_\nu \neq 0 \in \mathcal{H}_s \quad \text{for } \nu = 1, \ldots, p, \tag{5.2}$$

$$H_0 : h = 0 \quad \text{vs.} \quad H_1 : h \neq 0 \in \mathcal{H}_x \otimes \mathcal{H}_s. \tag{5.3}$$

These hypothesis testing problems are addressed using a score test method derived from the mixed effects framework of the SVC model. The test problems in (5.2) and (5.3) correspond to the following equivalent hypotheses,

$$H_0 : \tau_\nu = 0 \quad \text{vs.} \quad H_1 : \tau_\nu > 0 \quad \text{for } \nu = 1, \ldots, p+1. \tag{5.4}$$

Following Liu et al. (2007), we apply the score test method to address the hypothesis testing problem defined in (5.4). This method involves fixing the kernel parameters initially and subsequently varying them to evaluate the sensitivity of the score test outcomes relative to these parameters. Let $\phi = (\tau_1, \ldots, \tau_p, \tau_{p+1})^\top$ represent the vector of parameters. The score test statistic for τ_ν is defined as:

$$S_\nu(\phi, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\rho}_\nu) = \frac{1}{2}\mathbf{Y}^T V^{-1} \mathbf{K}_\nu(\boldsymbol{\rho_\nu}) V^{-1} \mathbf{Y}, \tag{5.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

where $\boldsymbol{\rho}_\nu$ are the kernel parameters, taking the form of ρ_s for $\nu = 1, \ldots, p$ or (ρ_x, ρ_s) for $\nu = p+1$. Assuming that the true covariance matrix $\boldsymbol{\Sigma}_\epsilon$ is known, the quadratic form in (5.5) implies that $S_\nu(\phi, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\rho}_\nu)$ approximately follows a mixture of chi-square distributions when $\boldsymbol{\rho}_\nu$ is fixed. In practice, however, $\boldsymbol{\Sigma}_\epsilon$ is often unknown, necessitating its estimation through a consistent estimator, $\widehat{\boldsymbol{\Sigma}}_\epsilon$. This estimated matrix is then substituted into $S_\nu(\phi, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\rho}_\nu)$.

Theorem 5. *Suppose that* $H_0 : \tau_\nu = 0$ *is true and* $\phi^0 = (\tau_1^0, \ldots, \tau_{\nu-1}^0, 0, \tau_{\nu+1}^0, \ldots, \tau_{p+1}^0)^\top$ *is the true value of* ϕ*, then*

(i) $S_\nu(\phi^0, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\rho}_\nu) \xrightarrow{d} \sum_\ell \lambda_\ell x_\ell^2$*, where* x_ℓ*s independently follow* $N(0,1)$ *and* $\{\lambda_\ell\}$ *are eigenvalues of* $V^{-1}\mathbf{K}_\nu(\boldsymbol{\rho}_\nu)/2$*.*

(ii) If $H_1 : \tau_\nu = \tau_n$ *hold, then for any sequence* $c_n \to \infty$ *and* $\tau_n \geq c_n \sum_\ell \lambda_\ell / \sum_\ell \lambda_\ell^2$*, the proposed test can reject* H_0 *with probability approaching one.*

(iii) If $\hat{\phi}$ *is a* $\sqrt{n}$ *consistent estimator of* ϕ^0 *under null and if* $\widehat{\boldsymbol{\Sigma}_\epsilon}$ *is a consistent estimator of* Σ *in terms of spectral norm such that* $\|\widehat{\boldsymbol{\Sigma}}_\epsilon^{-1} - \boldsymbol{\Sigma}_\epsilon^{-1}\|_s = o_p(1)$*, then we have* $S_\nu(\hat{\phi}, \widehat{\boldsymbol{\Sigma}}_\epsilon, \boldsymbol{\rho}_\nu) \xrightarrow{d} \sum_\ell \lambda_\ell x_\ell^2$*.*

In Theorem 5 (i), computing λ_ℓs and probability of $\sum_\ell \lambda_\ell x_\ell^2$ is computationally difficult, so we approximate the null distribution of $S_\nu(\hat{\phi}, \widehat{\boldsymbol{\Sigma}}_\epsilon, \boldsymbol{\rho}_\nu)$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

for fixed $\boldsymbol{\rho}_\nu$ by a scaled chi-square $\kappa_\nu \chi^2_{\zeta_\nu}$ distribution using the Satterthwaite method. Specifically, let $\widehat{\phi}_\nu = (\widehat{\tau}_1, \ldots, \widehat{\tau}_{\nu-1}, \widehat{\tau}_{\nu+1}, \ldots, \widehat{\tau}_{p+1})^T$ denote the estimators under the null model. We derive that $\kappa_\nu = \widetilde{I}_{\tau_\nu\tau_\nu}/2\widetilde{e}_\nu$ and $\zeta_\nu = 2\widetilde{e}_\nu^2/\widetilde{I}_{\tau_\nu\tau_\nu}$, where $\widetilde{e}_\nu = \mathrm{tr}(\mathbf{V}^{-1}\mathbf{K}_\nu)/2$ and $\widetilde{I}_{\tau_\nu\tau_\nu} = I_{\tau_\nu\tau_\nu} - \boldsymbol{\mathcal{I}}_{\tau_\nu\phi_\nu}\boldsymbol{\mathcal{I}}^{-1}_{\phi_\nu\phi_\nu}\boldsymbol{\mathcal{I}}^T_{\tau_\nu\phi_\nu}$ with $I_{\tau_\nu\tau_\nu} = 0.5\mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{K}_\nu\mathbf{V}^{-1}\mathbf{K}_\nu\right)$, $[\boldsymbol{\mathcal{I}}_{\tau_\nu\phi_\nu}]_j = 0.5\mathrm{tr}(\mathbf{V}^{-1}\mathbf{K}_\nu\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\phi_{\nu,j}})$, and

$$[\boldsymbol{\mathcal{I}}_{\phi_\nu\phi_\nu}]_{jj'} = 0.5\mathrm{tr}(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\phi_{\nu,j}}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\phi_{\nu,j'}}) \quad \text{for } j, j' = 1, \ldots, p+1.$$

In contrast to the existing literature, which mainly focuses on the null limit distribution of score test under a linear mixed effect model, we explore the separation rate under the alternative hypothesis in Theorem 5 (ii). Theorem 5 (iii) confirms that the null distribution can be approximated using plug-in estimates. To estimate the covariance matrix, we first use model (1.1) to obtain residuals $\hat{\epsilon}_{ij} = y_{ij} - \hat{y}_{ij}$ and then adopt functional principal component analysis (Yao et al., 2005; Zhang and Chen, 2007) to obtain $\widehat{\boldsymbol{\Sigma}}_\epsilon$.

6. Simulation Studies

In this section, we present simulation studies to evaluate the effectiveness of the proposed estimation and inference methods.

Example 4. In this example, we present a study based on real data. To

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

mimic the characteristics of genetic data, we generate the genetic vector $\mathbf{x}_i$ using data from the first LD block on the 15th chromosome in the ADNI dataset. This block consists of 72 SNPs from 606 subjects. Specifically, each $\mathbf{x}_i$ is randomly sampled with replacement from these 72 SNPs across the 606 subjects. We define the function $h(\mathbf{x}, s) = 0.05 \cdot (\sum_{j=1}^{20} \cos(2\pi(x_j - x_{j+20})/3) + s \cdot \sum_{j=41}^{5} 5\sin(\pi(x_j + x_{j+15})/3) + x_{71}x_{72})$, where $s_j = \frac{j-1}{m}$ is an equally spaced design. The covariate $\mathbf{z}$ is generated such that $z_{i1} \sim N(0,1)$ and $z_{i2} \sim N(0,1)$, with the true functions specified as: $\gamma_1(s) = 10s^3 - 15s^2 + 5s + 1$ and $\gamma_2(s) = 3 \cdot (10s^6 - 30s^5 + 25s^4 - 5s^2 + 5/21 + \sin(6\pi s))$. The response $y_i(s_j)$ is given by: $y_i(s_j) = \mathbf{z}_i^T \boldsymbol{\gamma}(s_j) + h(\mathbf{x}_i, s_j) + \epsilon_i(s_j)$, where $\epsilon_i(s_j) \sim N(0, \sigma_\epsilon^2)$.

We explore eight settings by varying $n \in \{30, 50\}$, $m \in \{10, 20\}$, and $\sigma_\epsilon^2 \in \{0.5, 1\}$, with 100 replicates for each setting. In each replicate, the estimation accuracy is assessed using the mean squared errors (MSE) defined as $\|\widehat{f} - f\|_{\mathcal{L}_2}^2 = \int_{\mathcal{D}} (\widehat{f}(\delta) - f(\delta))^2 \mathrm{d}\delta / \Lambda(\mathcal{D})$, where f represents one of the component functions $\gamma_1(\cdot)$, $\gamma_2(\cdot)$, or $h(\cdot, \cdot)$, and $\Lambda(\mathcal{D})$ is the Lebesgue measure of $\mathcal{D}$, the domain of f. Detailed calculations can be found in (S3.1)-(S3.3) in the supplementary material.

Figure 1 displays the average MSE of the estimates across the eight settings. The number of design points m in domain $\mathcal{S}$ significantly impacts

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

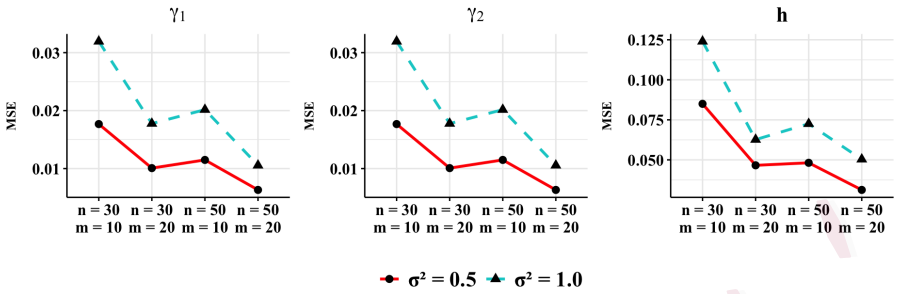

Figure 1: Average MSE in Example 4.

the estimation accuracy. As m increases, the estimation errors for $\widehat{\boldsymbol{\gamma}}(s)$ and $\widehat{h}(\mathbf{x}, s)$ diminish, owing to enhanced resolution in capturing the functional shapes. Similarly, the sample size n positively influences estimation performance. This effect is intuitive for $\widehat{h}(\mathbf{x}, s)$ as it involves $\mathbf{x}$ and more observations of $h(\mathbf{x}_i, \cdot)$ lead to improved estimates. For $\widehat{\gamma}_1(s)$ and $\widehat{\gamma}_2(s)$, increased n provides more information through z_1 and z_2, enhancing the SVC model's accuracy. This improvement is attributed to estimating $\boldsymbol{\gamma}(s)$ based on n repeated measurements across m grid points, facilitating better statistical inference as either n or m increases.

Example 5. This example assesses the efficacy of the proposed test for the null hypothesis $H_0 : \gamma_1(s) = 0$. In this example, the underlying true functions are defined using Bernoulli polynomials $\{B_k(z)\}_{k\geq 1}$. Specifically, the functions $\gamma_1(s)$ and $\gamma_2(s)$ are given by $\gamma_1(s) = a(10B_3(s)+\sin(2\pi s))$ and $\gamma_2(s) = 10B_6(s) + \sin(6\pi s)$. The true h function is defined as $h(x_1, x_2, s) = a[B_2(x_1)B_2(x_2)B_1(s)+10B_1(x_1)B_2(x_2)\cos(2\pi s)]$. We set $s_j = j-1/(m-1)$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

for $j = 1, \ldots, m$, and generate training data with $(z_{i1}, z_{i2})^T \sim N((0,0)^T, \mathbf{I}_2)$ and $(x_{i1}, x_{i2})^T \sim U[0,1]^2$. The error term $\epsilon_{ij} \sim N(0,1)$ and $n = 50, m = 20$.

We evaluate the test size at $a = 0$ and analyze the test power by incrementally increasing a. For both size and power assessments, 2000 datasets are simulated. The same datasets are used to evaluate the test's sensitivity to kernel parameter ρ_s variations, ranging from 0.0001 to 0.2. Figure 2(a) illustrates the power curves of the score test for $\gamma_1(s)$, indicating that the empirical test size approximates the nominal value of 0.05 and remains robust across variations in ρ_s. As a increases, the power of the test rapidly approaches one, irrespective of the ρ_s values. Nonetheless, an optimal choice of ρ_s, such as $\rho_s = 0.02$, can enhance performance.

A similar simulation is conducted to assess the effectiveness of the score test for the hypothesis $H_0 : h(\mathbf{x}, s) = 0$. The kernel parameters ρ_s are varied from 0.1 to 10, and ρ_x from 0.0001 to 0.2 to examine their effects on test performance. The power curves, depicted in Figures 2(b) to 2(d), show that the empirical size of the test closely approximates the nominal level of 0.05 across different (ρ_x, ρ_s) combinations. Notably, the test power ascends rapidly to one and exhibits robustness against variations in the kernel parameters (ρ_x, ρ_s).

We carry out additional simulations to examine the sensitivity of the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

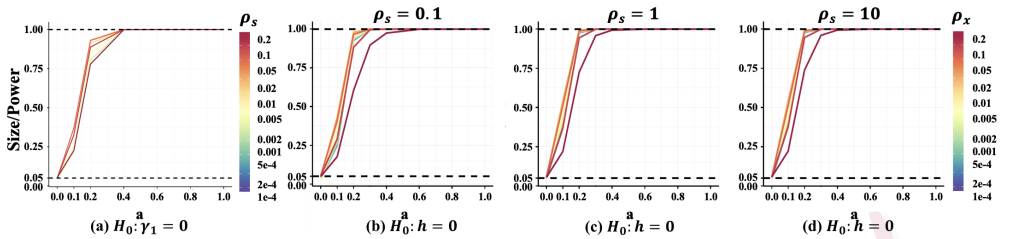

Figure 2: Power curves for the score test under various settings of the kernel parameters ρ_s and ρ_x. Panel (a) shows the effect of different ρ_s values under $H_0 : \gamma_1 = 0$. Panels (b), (c), and (d) illustrate the power curves for $H_0 : h = 0$ with ρ_s set to 0.1, 1, and 10, respectively.

kernel and spread parameters and present the results in Section S3 of the supplementary material. It is observed that when the tuning parameters are within a certain range, the estimates are similar.

7. ADNI Data Analysis

We analyze a dataset extracted from ADNI to investigate the effects of genetic markers and clinical variables on the human hippocampus. The ADNI dataset consists of 606 subjects and includes demographic variables such as Age, Gender (0=Male; 1=Female), Handedness (0=Right; 1=left), Retirement (0=No; 1=Yes), and Years of education. The mean age of the participants is 75.6 years with a standard deviation of 6.6 years, and the average years of education is 15.7 with a standard deviation of 2.9. The sample composition is as follows: 361 males and 245 females; 562 right-handed and 44 left-handed; 497 retired and 109 not retired. Marital status is represented through three dummy variables—Widowed, Divorced, and

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Never Married—with the reference category (all three dummies at zero) indicating Married status. At baseline, 482 participants were married, 75 were widowed, 32 were divorced, and 18 were never married.

For each subject, we extracted the hippocampal morphometry surface measure along the left and right hippocampi and obtained the corresponding density values. We applied the log quantile density transformation proposed in Petersen and Müller (2016) as density functions do not live in a linear space. These morphometry curves along the left or right hippocampus are the functional responses. The hippocampus, a critical brain structure located deep within the temporal lobe, plays vital roles in learning, memory, and spatial navigation. It is particularly susceptible to pathological changes and is associated with various neurodegenerative and neuropsychiatric disorders, including Alzheimer's disease (AD) (Dubois et al., 2016).

We extracted ultra-high dimensional genetic markers by considering linkage disequilibrium (LD) blocks for the genotyped and imputed single-nucleotide polymorphisms (SNPs) across all 22 chromosomes. Linkage disequilibrium is a common biological phenomenon where genetic variants exhibit strong blockwise correlations, as described in Wall and Pritchard (2003). This correlation may cause significant SNPs within a specific LD block to be overlooked if analyzed individually due to their relatively weak

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

signals. To leverage the structural information of LD blocks effectively, we utilized the SVC model to assess the impact of SNPs within each LD block on the hippocampus separately. We implemented the method proposed by Berisa and Pickrell (2016) to identify approximately independent LD blocks, resulting in a total of 1703 LD blocks. The hippocampus surface measure $y_i(s)$ has been centered for each point s.

Given the established asymmetry between the two parts of the hippocampus (Pedraza et al., 2004), we applied our SVC model separately to the left and right hippocampi. To facilitate comparison, we standardized all SNPs and continuous variables. We treated either the left or right hippocampus morphometry curves as the response $y_i(s)$ and considered the following demographic covariates: Age, Gender, Age^2, Age·Gender, Age^2·Gender. These variables has been demonstrated to be important variables in the literature (Lupton et al., 2010; Nebel et al., 2018; Li et al., 2024). We also included the top 10 principal components (PCs) of the whole genome data to correct for population stratification(Price et al., 2006). The $\mathbf{x}_i$ is the SNPs in one of the 1703 LD blocks, leading to 1703 SVC models for either the left or right hippocampus.

Table 1 presents the p-values for demographic covariates associated with the bilateral hippocampus (left and right), which are nearly identical across

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

the 1703 models. This is likely because genetic data typically explain only a small proportion of the variation in hippocampal volume (Stein et al., 2012). Moreover, genetic and clinical variables influence brain structure through distinct pathways. It shows that Age, Gender, Age^2, and their interaction terms Age·Gender and Age^2·Gender are significant for both the bilateral hippocampus. These findings are consistent with literature showing that age and gender are strongly associated with the hippocampus (Guerreiro and Bras, 2015). Figure S4.1 in the supplementary material displays the estimated effects of Age, Gender, Age^2, Age·Gender, and Age^2·Gender for the bilateral hippocampus. The figure reveals strong evidence of symmetry in these estimated covariates in the left and right hippocampi. However, Handedness is found to be significant for the left hippocampus, while Never Married and Education are significant for the right hippocampus, demonstrating the asymmetric structure of the bilateral hippocampus. Figure S4.2 in the supplementary material shows the corresponding estimates and reveals that the three covariates exhibit positive and negative effects on quantile densities of the hippocampus morphometry measures at different quantile levels. This left-right hemispheric asymmetry is an important phenomenon of brain organization (Sha et al., 2021).

In this study, we obtained 1703 p-values by testing the nullity of SNPs

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Table 1: The p-values of all demographic covariates for the left and right hippocampi.

	Age	Gender	Handedness	Widowed	Divorced	Never Married
Left	<1e-16	<1e-16	0.030	0.401	0.326	0.999
Right	<1e-16	<1e-16	0.992	0.166	0.441	0.009
	Retirement	Education	Age·Gender	Age^2	Age^2·Gender	
Left	0.784	0.784	<1e-16	<1e-16	<1e-16	
Right	0.934	0.006	<1e-16	<1e-16	<1e-16	

within one block for either the left or right hippocampus. The Bonferroni correction method, with a commonly used level of 0.05, was adopted to identify important blocks from the 1703 blocks tested. Figure 3 shows the ideogram and Manhattan plots of the significant blocks for the bilateral hippocampus. Specifically, 109 and 245 blocks were declared to be significant for the left and the right hippocampus, respectively. 27 blocks were found to be in common. The figure also reveals polygenic effects on the bilateral hippocampus and different genetic architectures of the left and right hippocampi. Among the 27 common blocks, the well-known block 19q13.32 region on the 19th chromosome is identified to be important for both the left and right hippocampi. This region contains the well-known APOE, a major genetic risk factor for AD (Kim et al., 2009).

Furthermore, we focused on the top 10 significant blocks for the left and right hippocampi. According to the NHGRI-EBI GWAS catalog(Sollis et al., 2023), Figure 4 presents indications of associations between the top 10 blocks and some selected traits. The most significant block for the left hippocampus is located on chromosome 17 and is associated with traits

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

such as education attainment, amyloid-beta, brain measure, reaction time, mathematical ability, and cognitive decline rate in late mild cognitive impairment. The most significant block for the right hippocampus is located on chromosome 5 and is associated with traits such as amyloid beta and neurofibrillary tangles, language functional connectivity, brain morphology, reaction time, educational attainment and mathematical ability. In addition, Figure S4.3 in the supplementary material presents the median positions and p-values of the common 27 blocks and the range and p-values of the top 10 significant blocks for the left and right hippocampi. It shows that the well-known block 19q13.32 region on the 19th chromosome is identified to be important for both the left and right hippocampi.

8. Discussion

We used a semi-nonparametric varying coefficients modeling framework to study the relationship between genetic markers and imaging responses in imaging genetics. We developed an estimation and inference procedure for SVC using the kernel machine method and derived a representer theorem to simplify computation. We also established theoretical properties of the estimated varying coefficient functions. Our analysis of the ADNI study illustrates that our proposed method effectively quantifies the relationship between genetic markers and imaging responses.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

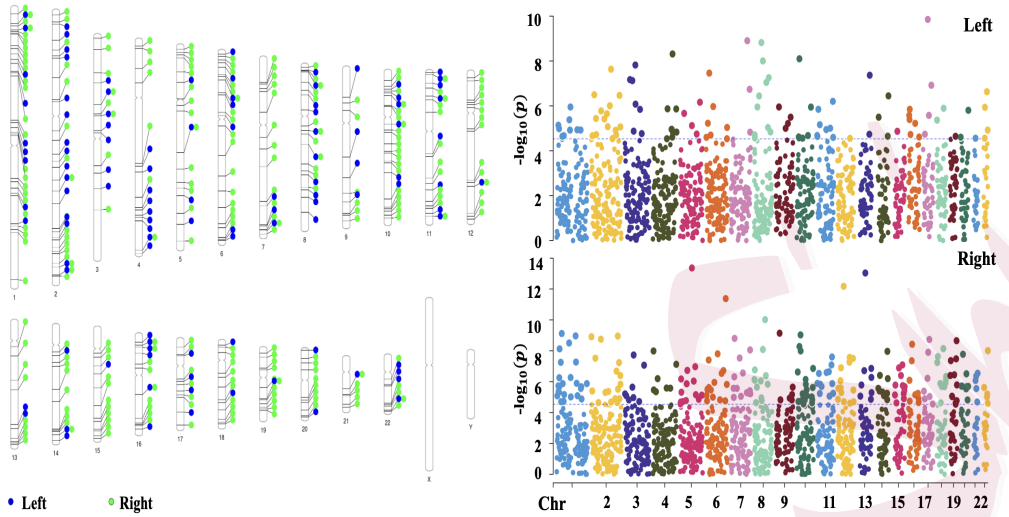

Figure 3: The left panel displays the ideogram plot of significant blocks, with blue and green points indicating the positions of significant blocks for the left and right hippocampi, respectively. The right panel presents the Manhattan plot of significant blocks, where the blue line represents the threshold p-value $= 0.05/1703$.

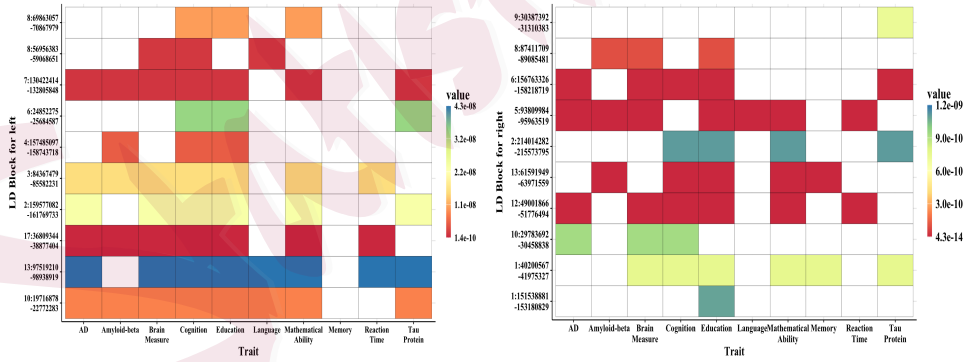

Figure 4: Associations between the SNPs in the top LD blocks for left and right hippocampi with some selected traits. The color indicates the p-value of the LD blocks to the left and right hippocampi.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

Supplementary Material

Additional simulation results, additional real data analysis, and details of all the proofs can be found in the supplementary material.

Acknowledgments

Address for correspondence: Hongtu Zhu, Ph.D., Email: htzhu@email.unc.edu.

Dr. Zhu's work was partially supported by the Gillings Innovation Laboratory on generative AI and by grants from the National Institute on Aging (NIA) of the National Institutes of Health (NIH), including 1R01AG085581, and RF1AG082938, the National Institute of Mental Health (NIMH) grant 1R01MH136055, and the NIH grants R01AR082684, and 1OT2OD038045-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Berisa, T. and J. K. Pickrell (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics 32*(2), 283–285.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 2. the new algorithm. *Journal of Applied Mathematics 6*(3), 222–231.

Cai, T. T. and M. Yuan (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics 39*(5), 2330–2355.

Cheng, G. and Z. Shang (2015). Joint asymptotics for semi-nonparametric regression models with partially linear structure. *The Annals of Statistics 43*(3), 1351–1390.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

REFERENCES

Dubois, B., H. Hampel, H. H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, and K. Blennow (2016). Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia 12*(3), 292–323.

Elliott, L. T., K. Sharp, F. Alfaro-Almagro, S. Shi, K. L. Miller, G. Douaud, J. Marchini, and S. M. Smith (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature 562*(7726), 210–216.

Goodlett, C. B., P. T. Fletcher, J. H. Gilmore, and G. Gerig (2009). Group analysis of DTI fiber tract statistics with application to neurodevelopment. *Neuroimage 45*(1), S133–S142.

Gu, C. (2013). *Smoothing spline ANOVA models*. Springer Science & Business Media.

Gu, C. and G. Wahba (1993). Smoothing spline ANOVA with component-wise bayesian "confidence intervals". *Journal of Computational and Graphical Statistics 2*(1), 97–117.

Guerreiro, R. and J. Bras (2015). The age factor in Alzheimer's disease. *Genome medicine 7*(1), 1–3.

Hofmann, T., B. Schölkopf, and A. J. Smola (2008). Kernel methods in machine learning. *The Annals of Statistics 36*(3), 1171–1220.

Kim, J., J. M. Basak, and D. M. Holtzman (2009). The role of apolipoprotein E in Alzheimer's disease. *Neuron 63*(3), 287–303.

Kong, D., J. G. Ibrahim, E. Lee, and H. Zhu (2018). Flcrm: Functional linear cox regression model. *Biometrics 74*(1), 109–117.

Krafty, R. T., P. A. Gimotty, D. Holtz, G. Coukos, and W. Guo (2008). Varying coefficient model with unknown within-subject covariance for analysis of tumor growth curves. *Biometrics 64*(4), 1023–1031.

Kwee, L. C., D. Liu, X. Lin, D. Ghosh, and M. P. Epstein (2008). A powerful and flexible multilocus association test for quantitative traits. *The American Journal of Human Genetics 82*(2), 386–397.

Le, B. D. and J. L. Stein (2019). Mapping causal pathways from genetics to neuropsychiatric disorders using genome-wide imaging genetics: Current status and future directions. *Psychiatry and Clinical Neurosciences 73*(7), 357–369.

Li, T., Y. Yu, J. Marron, and H. Zhu (2024). A partially functional linear regression framework for

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

integrating genetic, imaging, and clinical data. *The Annals of Applied Statistics 18*(1), 704–728.

Li, X., L. Wang, and H. J. Wang (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association 116*(536), 1994–2008.

Li, Y., H. Zhu, D. Shen, W. Lin, J. H. Gilmore, and J. G. Ibrahim (2011). Multiscale adaptive regression models for neuroimaging data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(4), 559–578.

Lindquist, M. A., J. M. Loh, and Y. R. Yue (2010). Adaptive spatial smoothing of fMRI images. *Statistics and its Interface 3*(1), 3–13.

Liu, D., X. Lin, and D. Ghosh (2007). Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics 63*(4), 1079–1088.

Lupton, M. K., D. Stahl, N. Archer, C. Foy, M. Poppe, S. Lovestone, P. Hollingworth, J. Williams, M. J. Owen, K. Dowzell, et al. (2010). Education, occupation and retirement age effects on the age of onset of alzheimer's disease. *International journal of geriatric psychiatry 25*(1), 30–36.

Miller, K. L., F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, and J. L. Andersson (2016). Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience 19*(11), 1523–1536.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application 2*, 321–359.

Mueller, S. G., M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett (2005). The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics 15*(4), 869–877.

Nathoo, F., L. Kong, and H. Zhu (2019). A review of statistical methods in imaging genetics. *The Canadian Journal of Statistics 47*(1), 108–131.

Nebel, R. A., N. T. Aggarwal, L. L. Barnes, A. Gallagher, J. M. Goldstein, K. Kantarci, M. P. Mallampalli, E. C. Mormino, L. Scott, W. H. Yu, et al. (2018). Understanding the impact of sex and gender in alzheimer's disease: a call to action. *Alzheimer's & Dementia 14*(9), 1171–1183.

Pedraza, O., D. Bowers, and R. Gilmore (2004). Asymmetry of the hippocampus and amygdala in

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

mri volumetric measurements of normal adults. *Journal of the International Neuropsychological Society 10*(5), 664–678.

Petersen, A. and H.-G. Müller (2016). Functional data analysis for density functions by transformation to a hilbert space. *The Annals of Statistics 44*(1), 183–218.

Poggio, T. and F. Girosi (1990). Networks for approximation and learning. *Proceedings of the IEEE 78*(9), 1481–1497.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics 38*(8), 904–909.

Rao, Y. L., B. Ganaraja, B. Murlimanju, T. Joy, A. Krishnamurthy, and A. Agrawal (2022). Hippocampus and its involvement in alzheimer's disease: a review. *3 Biotech 12*(2), 55.

Reiss, P. T., L. Huang, and M. Mennes (2010). Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics 6*(1), Article 28.

Schölkopf, B. and A. J. Smola (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Sha, Z., D. Schijven, A. Carrion-Castillo, M. Joliot, B. Mazoyer, S. E. Fisher, F. Crivello, and C. Francks (2021). The genetic architecture of structural left–right asymmetry of the human brain. *Nature Human Behaviour 5*(9), 1226–1239.

Shang, Z. and G. Cheng (2013). Local and global asymptotic inference in smoothing spline models. *The Annals of Statistics 41*(5), 2608–2638.

Sollis, E., A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. Güneş, P. Hall, J. Hayhurst, et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research 51*(D1), D977–D985.

Stein, J. L., S. E. Medland, A. A. Vasquez, D. P. Hibar, R. E. Senstad, A. M. Winkler, R. Toro, K. Appel, R. Bartecek, Ø. Bergmann, et al. (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nature genetics 44*(5), 552–561.

Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

estimation. *The annals of statistics 22*(1), 118–171.

Wahba, G. (1990). *Spline models for observational data*. Society for Industrial and Applied Mathematics.

Wall, J. D. and J. K. Pritchard (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics 4*(8), 587–597.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*, 257–295.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics 89*(1), 82–93.

Yang, H., V. Baladandayuthapani, A. U. Rao, and J. S. Morris (2020). Quantile function on scalar regression analysis for distributional data. *Journal of the American Statistical Association 115*(529), 90–106.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

Yu, S., G. Wang, L. Wang, and L. Yang (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica 31*, 1463–1487.

Yuan, Y., J. H. Gilmore, X. Geng, M. A. Styner, K. Chen, J.-l. Wang, and H. Zhu (2013). A longitudinal functional analysis framework for analysis of white matter tract statistics. In *International Conference on Information Processing in Medical Imaging*, pp. 220–231. Springer.

Zhang, D. and X. Lin (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics 4*(1), 57–74.

Zhang, J.-T. and J. Chen (2007). Statistical inferences for functional data. *The Annals of Statistics 35*(3), 1052–1079.

Zhang, T. (2005). Learning bounds for kernel regression using effective data dimensionality. *Neural Computation 17*(9), 2077–2098.

Zhao, B., T. Li, Y. Yang, X. Wang, T. Luo, Y. Shan, Z. Zhu, D. Xiong, M. Hauberg, J. Bendl, J. Fullard, P. Roussos, Y. Li, J. Stein, and H. Zhu (2021). Common genetic variation influencing human white

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0118

matter microstructure. *Science 372*(6548), eabf3736.

Zhao, T., G. Cheng, and H. Liu (2016). A partially linear framework for massive heterogeneous data. *The Annals of Statistics 44*(4), 1400–1437.

Zhu, H., J. Fan, and L. Kong (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association 109*(507), 1084–1098.

Zhu, H., L. Kong, R. Li, M. Styner, G. Gerig, W. Lin, and J. H. Gilmore (2011). FADTTS: functional analysis of diffusion tensor tract statistics. *NeuroImage 56*(3), 1412–1425.

Zhu, H., R. Li, and L. Kong (2012). Multivariate varying coefficient model for functional responses. *The Annals of Statistics 40*(5), 2634—2666.

Zhu, H., T. Li, and B. Zhao (2023). Statistical learning methods for neuroimaging data analysis with applications. *Annual Review of Biomedical Data Science 6*, 73–104.

Zhu, H., M. Styner, N. Tang, Z. Liu, W. Lin, and J. H. Gilmore (2010). FRATS: Functional regression analysis of DTI tract statistics. *IEEE transactions on Medical Imaging 29*(4), 1039–1049.