Density-on-scalar Single-index Quantile Regression Model

Xingcai Zhou *
School of Statistics and Data Science, Nanjing Audit University, Nanjing, China
Shengxian Ding *
Department of Statistics, Florida State University, Tallahassee, FL
Jiangyan Wang
School of Statistics and Data Science, Nanjing Audit University, Nanjing, China
Rongjie Liu
Department of Statistics, Florida State University, Tallahassee, FL
Linglong Kong
Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada
and
Chao Huang †
Department of Statistics, Florida State University, Tallahassee, FL

July 11, 2022

Abstract

This paper develops a density-on-scalar single-index quantile regression modeling framework to investigate the relationship between imaging responses and covariates of interest while tackling the imaging heterogeneity. Conventional association analysis methods typically assume that the imaging responses are well aligned after some preprocessing steps. However, this assumption is often violated in practice due to imaging heterogeneity. This heterogeneity is primarily caused by the different pathological patterns across subjects. Although some distribution-based approaches are developed to deal with this heterogeneity, major challenges have been posted due to the nonlinear subspace formed by the distributional responses and the unknown nonlinear association structure. Our method can successfully address all the challenges. We establish both estimation and inference procedures for the unknown functions in our model. The asymptotic properties of both estimation and inference procedures are systematically investigated. The finite-sample performance of our proposed method is assessed by using both Monte Carlo simulations and a real data example on brain cancer images from The Cancer Imaging Archive Glioblastoma Multiforme collection.

Keywords: Heterogeneity; Glioblastoma; Density function; Single-index; Quantile regression.

*These two authors contributed equally.

†Corresponding author (Email: chuang7@fsu.edu).

1 Introduction

With the rapid development of modern imaging acquisition technology, more and more imaging studies have been conducted in recent years, e.g., The Cancer Imaging Archive (Clark et al. 2013) and UK Biobank Study (Bycroft et al. 2018). Among these studies, we are usually interested in the association between imaging responses and several covariates of interest, such as age and gender. To establish this association, the pixel-based morphometry was first developed where statistical analysis is performed at each pixel on the basis of the general linear model and the statistical parametric map, comprising the pixel-wise hypothesis testing results, is derived throughout the entire image domain (Mechelli et al. 2005). However, these methods have limitations as the image spatial smoothness and correlation are not well characterized or even ignored. To address this issue, the functional data analysis tools are adopted and the function-on-scalar model has been widely used for different image modalities (Morris 2015, Wang et al. 2016).

In the function-on-scalar model, the imaging responses are typically assumed to be well aligned after some preprocessing steps, e.g., image registration and time warping (Srivastava & Klassen 2016). However, this assumption is often violated in practice due to the imaging heterogeneity. The heterogeneity is primarily caused by the different pathological patterns across subjects. For example, in brain cancer image analysis, the brain tumor is of great interest instead of the whole brain. However, it is impossible to align all the tumors as they can significantly vary across subjects and/or time in terms of their number, size, and location (Roy et al. 2018). As an illustration, we consider the fluid-attenuated inversion recovery (FLAIR) brain image dataset from The Cancer Imaging Archive Glioblastoma Multiforme (TCIA-GBM) collection (Bakas et al. 2017*a*). The left two columns in Figure 1 present FLAIR images for four patients with corresponding tumor segmentation annotation. Different colors represent different tumor sub-regions, i.e., necrosis (blue), edema (yellow), and enhancing tumor (orange). We observe different tumor

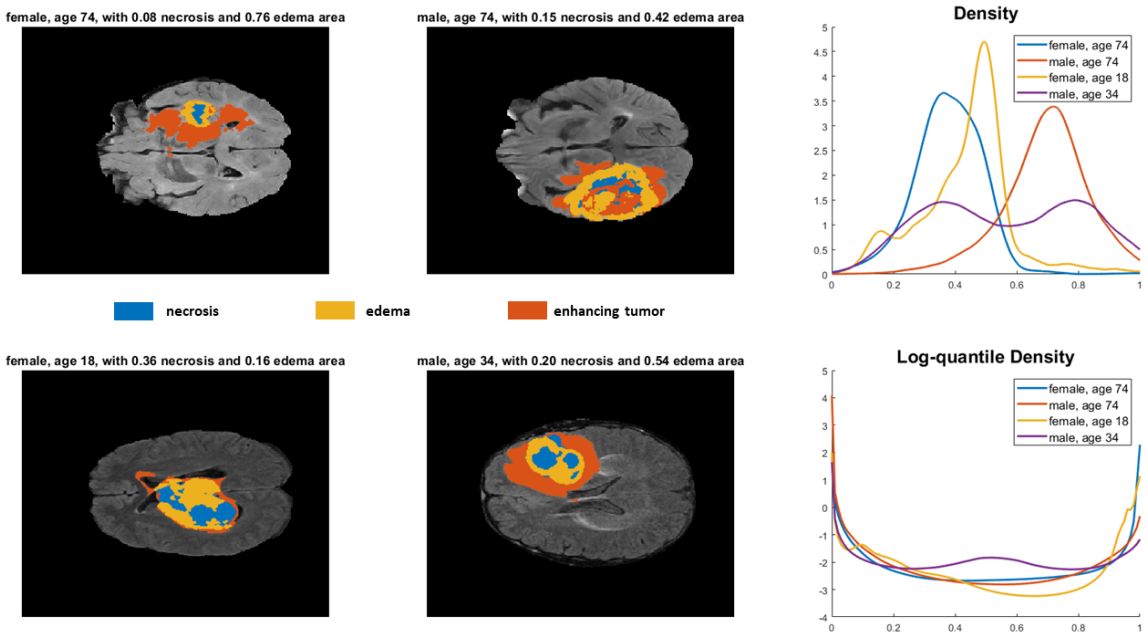

Figure 1: Illustration of imaging heterogeneity. Left two columns show FLAIR brain images of four patients from TCIA-GBM collection with corresponding tumor segmentation annotation. Different tumor sub-regions are highlighted in different colors: necrosis (blue), edema (yellow), and enhancing tumor (orange). The two panels in the right column show the boundary-corrected kernel density estimates and corresponding LQD transformed curves of all four patients.

patterns across patients, which makes the conventional registration approaches not applicable for the tumor regions. Thus, developing such association analysis tools that are resistant to the tumor heterogeneity is important for understanding the role of imaging biomarkers in the pathological progression of glioblastoma.

Currently, there are two kinds of approaches to tackle this imaging heterogeneity in the association studies, i.e., feature-based approaches and distribution-based approaches. In the first kind of approaches, a large number of imaging features, e.g., radiomic features and deep learning based features, are extracted from medical images, and then regression analysis is conducted to investigate the relationship between each feature and clinical variables. While theses features

are resistant to the imaging heterogeneity and have potential to uncover the underlying patterns, the feature-based approaches have drawbacks. Specifically, as some imaging features (e.g., min, mean, and variance) are derived according to the predefined formulas (Lambin et al. 2012), it may fail in recovering the association if the primary effects are not contained in the manually extracted features. Although the deep learning based features have the power to automatically learn advanced abstract representations, they often face a lack of explainability and interpretability in the association analysis (Bai et al. 2021).

In the second kind of approaches, instead of imaging features, some functional representations, such as density curves (Petersen et al. 2019, Mohammed et al. 2021) and quantile functions (Yang et al. 2020), are extracted and treated as the imaging responses regressed on a set of scalar covariates of interest. The functional representations can provide much richer information compared to the manually extracted features and better interpretability compared to the deep learning based features. However, there are couple of challenges in existing methods. First, the extracted functional representations usually don't constitute a linear Hilbert subspace due to some inherent constraints. For example, the density function is always nonnegative and integrated to one (Petersen & Müller 2016). Thus, some popular functional data analysis tools established in the linear function space, such as the functional principal component analysis (FPCA, Yao et al. (2005)), may not be feasible for the functional data from the nonlinear subspace. Second, the linear model is typically used in existing literature to characterize the relationship between extracted imaging responses and scalar covariates (Zhu et al. 2012), which is restrictive in practice and may substantially lose efficiency when the true relationship is nonlinear (Luo et al. 2016). Although there are some recent approaches employing the single-index or additive models to represent the nonlinear relationship between the imaging responses and covariates of interest (Han et al. 2019, Bhattacharjee & Müller 2021, Ghosal et al. 2021), they proposed their methods with the focus on

estimation procedures rather than inference procedures.

The aim of this paper is to develop a density-on-scalar single-index quantile regression modeling framework to investigate the relationship between distribution-based imaging responses and scalar covariates of interest. To address the challenges in existing approaches, we start with the density function of pixel intensities (the top panel of the right column in Figure 1) and consider some transformation functions, e.g., the log quantile density (LQD) transformation in Petersen & Müller (2016) (the bottom panel of the right column in Figure 1), to transfer the density functions to some functional representations lying in the linear Hilbert subspace. Meanwhile, the nonlinear relationship between imaging responses and scalar covariates are established through a unknown link function. We develop a three-step estimation procedure to estimate unknown functions in our proposed model. In addition to the estimation procedure, a global Wald-type test statistic, a simultaneous confidence band, and a simultaneous confidence region are constructed for the varying coefficients, while a simultaneous confidence band is constructed for the link function. We also systematically investigate the asymptotic properties in both estimation and inference procedures. Furthermore, both simulation studies and real data analysis show that our proposed method outperforms other competitors in terms of both estimation accuracy and robustness.

2 Methods

2.1 Density-on-scalar single-index quantile regression model

In this paper, we would like to investigate how the image signals (e.g., image intensities from tumor regions) are affected by the covariates of interest (e.g., age and gender). Suppose that we observe both image data and covariates of interest from n unrelated subjects. For the i-th subject, it is assumed that $\boldsymbol{x}_i$ is a $p \times 1$ vector including covariates of interest and $\nu_i(t)$ is the imaging

measurement derived at $t \in \mathcal{T}$, where $\mathcal{T} \subset \mathbb{R}^d$ is the region of interest (ROI), including m_i pixels $\{t_1, \ldots, t_{m_i}\}$, $i = 1, \ldots, n$. Instead of the pixel-wise imaging measurements, we are interested in their distributional representations, i.e., density functions $\{f_i\}_{i=1}^n$. However, the density function space $\mathcal{F}$ has its inherent constraints of being non-negative and integrated to one. To address this issue, we construct the conditional mean of f_i through the transformation function:

$$\mu(\cdot|\boldsymbol{x}_i) = \Psi^{-1}\left(\mathbb{E}\left(\Psi(f_i)|\boldsymbol{x}_i\right)\right), \tag{1}$$

where the transformation $\Psi(\cdot) : \mathcal{F} \to \mathcal{L}_2([0,1])$ maps the density function into an unrestricted square integrable random function on $[0,1]$. In this paper, we consider adopting the LQD transformation, i.e., $\Psi(f_i) = -\log(f_i(Q_i))$ where $Q_i(\cdot)$ is the corresponding quantile function (Petersen & Müller 2016). Note that, as alternatives to LQD transformation, some other transformations satisfying the required criteria are also available in existing literature, including the log hazard transformation (Petersen & Müller 2016), the square-root velocity function representation (Srivastava & Klassen 2016), and among others.

To investigate the relationship between the transformed density functions and the covariates of interest, we further specify the mean structure $\mu(\cdot|\boldsymbol{x}_i)$ and introduce a density-on-scalar single-index quantile regression model:

$$\Psi(f_i)(s) = g\left(\boldsymbol{x}_i^\top \boldsymbol{\beta}(s)\right) + \epsilon_i(s),\ s \in [0,1], \tag{2}$$

where $\boldsymbol{\beta}(s)$ is a $p \times 1$ vector including unknown varying-coefficient functions, $g(\cdot)$ is an unknown index function, and $\epsilon_i(s)$ is a measurement error process including the individual variation. In practice, the density function f_i is not actually observed, but must be estimated from the pixel-wise imaging measurements $\{\nu_i(t_l)\}_{l=1}^{m_i}$. Through some nonparametric density estimation methods, our model in (2) can be rewritten based on the estimated density functions $\{\hat{f}_i\}_{i=1}^n$:

$$y_i(s) = g\left(\boldsymbol{x}_i^\top \boldsymbol{\beta}(s)\right) + \eta_i(s),\ s \in [0,1], \tag{3}$$

where $y_i(s) = \Psi(\hat{f}_i)(s)$ and $\eta_i(s) = \epsilon_i(s) + \delta_i(s)$ with $\delta_i(s) = \Psi(\hat{f}_i)(s) - \Psi(f_i)(s)$ be the stochastic error resulting from the estimation of f_i. In particular, similar to Liu et al. (2020), we assume that the τ-th quantile of $\eta_i(s)$ conditional on $\boldsymbol{x}_i$ equals to zero, that is, $\mathbb{P}\{\eta_i(s) < 0|\boldsymbol{x}_i\} = \tau$ for any $s \in [0,1]$. Furthermore, $\{\eta_i(s)\}_{i=1}^n$ are assumed to be independent copies of SP$(0, \gamma_\eta)$, i.e., a stochastic process with mean function 0 and covariance function $\gamma_\eta(s, s')$. To ensure identifiability, we impose following constraints on $\boldsymbol{\beta}(s)$: $\|\boldsymbol{\beta}(s)\|_2 = 1$ with its first entry positive for all $s \in [0,1]$. Without loss of generality, the transformed functions are further assumed to be measured at the same m grid points, i.e., $\mathcal{S}_m = \{s_l : 0 < s_1 < \cdots < s_m < 1\}$.

2.2 Estimation procedure

Our estimation procedure mainly consists of three steps: (i) estimating the individual density functions $\{f_i\}_{i=1}^n$; (ii) estimating the coefficient functions $\boldsymbol{\beta}(s)$ and the index function $g(\cdot)$; and (iii) smoothing the individual stochastic functions $\{\eta_i(s)\}_{i=1}^n$. In addition, the detailed bandwidth selection procedures can be found in the Supplementary Material.

In step (i), according to the results in Petersen & Müller (2016), a boundary-corrected kernel estimator of f_i can be constructed as

$$\hat{f}_i(t) = \left\{ \sum_{l=1}^{m_i} \int_0^1 K\left(\frac{\varsigma - \nu_i(t_l)}{h_f}\right) \omega(\varsigma, h_f) d\varsigma \right\}^{-1} \sum_{l=1}^{m_i} K\left(\frac{t - \nu_i(t_l)}{h_f}\right) \omega(t, h_f), \tag{4}$$

where $K(\cdot)$ is the kernel function and the bandwidth $h_f < 1/2$. The weight function $\omega(t, h_f)$ is set to $(\int_{-t/h_f}^1 K(\varsigma) d\varsigma)^{-1}$ as $t \in [0, h_f)$, $(\int_{-1}^{(1-t)/h_f} K(\varsigma) d\varsigma)^{-1}$ as $t \in (1 - h_f, 1]$, and 1 otherwise. Compared to traditional kernel density estimators, the boundary-corrected one in (4) converges to f_i uniformly, i.e., $\sup_{f_i \in \mathcal{F}} ||\hat{f}_i - f_i||_\infty \to 0$ as $m_i \to \infty$ (Petersen & Müller 2016).

In step (ii), we first start with the locally weighted quantile regression loss function:

$$\sum_{i=1}^n \sum_{l=1}^m \rho_\tau \left(y_i(s_l) - g(\boldsymbol{x}_i^\top \boldsymbol{\beta}(s)) \right) K_{h_\beta}(s_l - s), \tag{5}$$

where $\rho_\tau(u) = u\{\tau - \mathbb{I}(u < 0)\}$ is a check function (Koenker & Bassett 1978), $\mathbb{I}(\cdot)$ is an indicator function, and $K_{h_\beta}(\cdot) = h_\beta^{-1} K(\cdot/h_\beta)$. Then, the corresponding gradient of (5) with respect to $\boldsymbol{\beta}(s)$ can be written as $\sum_{i=1}^{n}\sum_{l=1}^{m}\{\boldsymbol{\varphi}_0(\boldsymbol{\beta}(s); y_i(s_l), \boldsymbol{x}_i) + \boldsymbol{\varphi}_1(\boldsymbol{\beta}(s); y_i(s_l), \boldsymbol{x}_i)\}$, in which

$$\boldsymbol{\varphi}_0(\boldsymbol{\beta}(s); y_i(s_l), \boldsymbol{x}_i) = (\boldsymbol{x}_i - \tilde{\boldsymbol{x}}_i)\psi_\tau\left(y_i(s_l) - g\left(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)\right)\right)\dot{g}\left(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)\right)K_{h_\beta}(s_l - s),$$

$$\boldsymbol{\varphi}_1(\boldsymbol{\beta}(s); y_i(s_l), \boldsymbol{x}_i) = \tilde{\boldsymbol{x}}_i\psi_\tau\left(y_i(s_l) - g\left(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)\right)\right)\dot{g}\left(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)\right)K_{h_\beta}(s_l - s),$$

with $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \mathbb{E}\{\boldsymbol{x}_i|\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)\}$, $\psi_\tau(u) = \tau - \mathbb{I}(u < 0)$ is the score function of $\rho_\tau(u)$, and $\dot{g}(\cdot)$ denotes the first order derivative of $g(\cdot)$. According to the estimation efficiency of the central mean subspace (Ma & Zhu 2014), $\boldsymbol{\varphi}_0(\boldsymbol{\beta}(s); y_i(s_l), \boldsymbol{x}_i)$ belongs to the tangent space of model (3) with respect to $\boldsymbol{\beta}(s)$ while $\boldsymbol{\varphi}_1(\boldsymbol{\beta}(s); y_i(s_l), \boldsymbol{x}_i)$ belongs to its orthogonal component.

To construct an efficient empirical log likelihood ratio function for $\boldsymbol{\beta}(s)$, we introduce n independent auxiliary random vectors $\{\boldsymbol{U}_i(\boldsymbol{\beta}(s)) \doteq \frac{1}{m}\sum_{l=1}^{m}\boldsymbol{\varphi}_1(\boldsymbol{\beta}(s); y_i(s_l), \boldsymbol{x}_i)\}_{i=1}^{n}$. Then, the block empirical likelihood method (You et al. 2006) is adopted to accommodate the within-curve correlation structure, and the empirical log likelihood ratio function of $\boldsymbol{\beta}(s)$ is

$$\ell_n(\boldsymbol{\beta}(s)) = -2\sup_{p_i}\left\{\sum_{i=1}^{n}\log(np_i)|p_i \geq 0, \sum_{i=1}^{n}p_i = 1, \sum_{i=1}^{n}p_i\boldsymbol{U}_i(\boldsymbol{\beta}(s)) = 0\right\}, \tag{6}$$

where the unknown quantities in $\boldsymbol{U}_i(\boldsymbol{\beta}(s))$ can be estimated via some nonparametric approaches. In particular, the conditional expectation in $\tilde{\boldsymbol{x}}_i$ is approximated by the Nadaraya-Watson estimator while the estimators $\widehat{g}(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s))$ and $\widehat{\dot{g}}(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s))$ are derived by using the weighted least squares method (Luo et al. 2016). Then the optimal weights $\{p_i\}_{i=1}^{n}$ are obtained from the Lagrange multiplier method, and the function $\ell_n(\boldsymbol{\beta}(s))$ in (6) can be further written as

$$\hat{\ell}_n(\boldsymbol{\beta}(s)) = 2\sum_{i=1}^{n}\log\left(1 + \boldsymbol{\lambda}(\boldsymbol{\beta}(s))^\top\widehat{\boldsymbol{U}}_i(\boldsymbol{\beta}(s))\right), \tag{7}$$

where $\boldsymbol{\lambda}(\boldsymbol{\beta}(s))$ satisfies $\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{U}}_i(\boldsymbol{\beta}(s))/\{1 + \boldsymbol{\lambda}(\boldsymbol{\beta}(s))^\top\widehat{\boldsymbol{U}}_i(\boldsymbol{\beta}(s))\} = 0$.

Then, the estimator $\hat{\boldsymbol{\beta}}(s)$ can be derived via minimizing the empirical log likelihood ratio function in (7), and the estimator $\hat{g}(\boldsymbol{x}^\top\boldsymbol{\beta}(s))$ can be further derived via minimizing the following

loss function with respect to $\boldsymbol{G}(\boldsymbol{x}^\top\boldsymbol{\beta}(s)) \doteq [g(\boldsymbol{x}^\top\boldsymbol{\beta}(s)), h_g\dot{g}(\boldsymbol{x}^\top\boldsymbol{\beta}(s))]^\top$:

$$\sum_{i=1}^{n}\sum_{l=1}^{m}\rho_\tau\left\{y_i(s_l)-\widehat{\boldsymbol{z}}_{il}^\top(s)\boldsymbol{G}(\boldsymbol{x}^\top\boldsymbol{\beta}(s))\right\}K_{h_g}\left(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s_l)-\boldsymbol{x}^\top\boldsymbol{\beta}(s)\right), \tag{8}$$

where $\widehat{\boldsymbol{z}}_{il}(s) = [1, \{\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s_l) - \boldsymbol{x}^\top\boldsymbol{\beta}(s)\}/h_g]^\top$ and h_g is the bandwidth.

In step (iii), by using the local linear regression technique (Fan & Gijbels 1996), we can smooth the updated residual functions, i.e., $\{\tilde{y}_i(s) \doteq y_i(s) - \hat{g}(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s)), s \in [0,1]\}_{i=1}^n$, to obtain the estimators of the individual functions $\eta_i(s)$. In particular, the estimator $\hat{\eta}_i(s)$ is derived as $\hat{\eta}_i(s) = \sum_{l=1}^m \varrho(h_\eta, s)\tilde{y}_i(s_l)$, where $\varrho(h_\eta, s) = (1,0)\{\sum_{l=1}^m K_{h_\eta}(s_l - s)\boldsymbol{\vartheta}_l(s)\boldsymbol{\vartheta}_l^\top(s)\}^{-1}K_{h_\eta}(s_l - s)\boldsymbol{\vartheta}_l(s)$ with $\boldsymbol{\vartheta}_l(s) = [1, (s_l - s)/h_\eta]^\top$ and h_η is the bandwidth. Furthermore, we use the empirical covariance $(n-p)^{-1}\sum_{i=1}^n \widehat{\eta}_i(s)\widehat{\eta}_i(s')$ to estimate $\widehat{\gamma}_\eta(s, s')$. In particular, the FPCA can be applied to the empirical covariance to further capture the within-curve dependence (Yao & Lee 2006). In addition, the estimated density function of $\{\eta_i(s)\}_{i=1}^n$ at each s, i.e., $\hat{f}_\eta(\cdot; s)$, can be derived through the kernel estimation of "residual-based" empirical distribution (Müller et al. 2007).

2.3 Inference procedure

Our inference procedure mainly consists of three components: (i) hypothesis testing of $\boldsymbol{\beta}(s)$; (ii) constructing the simultaneous confidence band and simultaneous confidence region of $\boldsymbol{\beta}(s)$; and (iii) constructing the simultaneous confidence band of $g(\cdot)$.

In component (i), we consider the linear hypotheses on $\boldsymbol{\beta}(s)$ as below:

$$\mathbb{H}_0 : \mathbf{C}\boldsymbol{\beta}(s) = \boldsymbol{c}(s) \text{ for all } s \in [0,1] \text{ vs. } \mathbb{H}_1 : \mathbf{C}\boldsymbol{\beta}(s) \neq \boldsymbol{c}(s) \text{ for some } s \in [0,1], \tag{9}$$

where $\mathbf{C}$ is a $r \times p$ matrix with rank r and $\boldsymbol{c}(s)$ is a $r \times 1$ vector of functions. A Wald-type global test statistic T_n is defined as

$$T_n = \int_0^1 \left\{\mathbf{C}\hat{\boldsymbol{\beta}}_\tau(s) - \boldsymbol{c}(s)\right\}^\top \left\{\mathbf{C}\widehat{\boldsymbol{\Xi}}_n(s,s)\mathbf{C}^\top\right\}^{-1} \left\{\mathbf{C}\hat{\boldsymbol{\beta}}_\tau(s) - \boldsymbol{c}(s)\right\} ds, \tag{10}$$

where $\widehat{\mathbf{\Xi}}_n(s,s) = \tau(1-\tau)\hat{f}_\eta^{-2}(0;s)\{\sum_{i=1}^n \widehat{\mathbf{B}}_i(s)\}^{-1}\{\sum_{i=1}^n \widehat{\mathbf{A}}_i(s)\widehat{\mathbf{A}}_i^\top(s)\}\{\sum_{i=1}^n \widehat{\mathbf{B}}_i^\top(s)\}^{-1}$ with $\widehat{\mathbf{A}}_i(s) = \{\boldsymbol{x}_i - \widehat{\mathbb{E}}(\boldsymbol{x}_i|\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s))\}\dot{\hat{g}}(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s))$ and $\widehat{\mathbf{B}}_i(s) = \widehat{\mathbf{A}}_i(s)\boldsymbol{x}_i^\top\dot{\hat{g}}(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s))$. As the asymptotic null distribution of T_n is a χ^2-type mixture one, which is quite complicated, it is hard to directly obtain the percentiles of T_n under the null hypothesis $\mathbb{H}_0$. To address this issue, a wild bootstrap method is developed consisting of the following four steps:

- Fit the model (3) under $\mathbb{H}_0$ and obtain the estimated functions $\hat{\boldsymbol{\beta}}(s)$, $\hat{g}(\cdot)$, the residuals $\{\hat{e}_i(s) \doteq y_i(s) - \hat{g}(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s))\}_{i=1}^n$, and the global test statistic T_n;

- Generate random samples $\{\varsigma_i^{(r)}(s_l)\}_{l=1}^m$ independently from $N(0,1)$, and then construct

$$y_i^{(r)}(s_l) = \hat{g}(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s_l)) + \varsigma_i^{(r)}(s_l)\hat{e}_i(s_l),\ i = 1, \cdots, n;$$

- Based on $\{\boldsymbol{x}_i, \{y_i^{(r)}(s_l)\}_{l=1}^m\}_{i=1}^n$, recalculate $\hat{\boldsymbol{\beta}}^{(r)}(s)$ and the global test statistic $T_n^{(r)}$;

- Repeat the previous two steps R times to have $\{T_n^{(1)}, \ldots, T_n^{(R)}\}$, which yields the empirical p-value, i.e., $p = \sum_{r=1}^R \mathbb{I}\{T_n^{(r)} \geq T_n\}/R$.

In component (ii), we would like to constructe the simultaneous confidence band and simultaneous confidence region of coefficient functions $\boldsymbol{\beta}(s)$. First, for a given level α, the $1-\alpha$ simultaneous confidence band of the j-th element in $\boldsymbol{\beta}(s)$, i.e., $\beta_j(s)$, is constructed as follows:

$$\left(\hat{\beta}_j(s) - \frac{\boldsymbol{C}_j(\alpha)}{\sqrt{n}}, \hat{\beta}_j(s) + \frac{\boldsymbol{C}_j(\alpha)}{\sqrt{n}}\right),\ j = 1, \cdots, p, \tag{11}$$

where $\boldsymbol{C}_j(\alpha)$ is approximated by the efficient resampling method developed in Zhu et al. (2012), which involves the following three steps:

- Fit model (3) and obtain $\hat{\boldsymbol{\beta}}(s)$, $\hat{g}(\cdot)$, and $\{\hat{\eta}_i(s)\}_{i=1}^n$;

- Generate the samples $\{\varsigma_i^{(r)}\}_{i=1}^n$ from $N(0,1)$ independently and construct the process sample $\zeta_j^{(r)}(s) = \boldsymbol{e}_{j,p}^\top\hat{f}_\eta(0;s)\{\frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{B}}_i(s)\}^{-1}\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \varsigma_i^{(r)}\widehat{\mathbf{A}}_i(s)\psi_\tau(\hat{\eta}_i(s))\}$, where $\boldsymbol{e}_{j,p}$ is a $p \times 1$ vector with the j-th element being 1 and 0 otherwise;

- Repeat the second step for R times to obtain $\{\sup_s |\zeta_j^{(1)}(s)|, \ldots, \sup_s |\zeta_j^{(R)}(s)|\}$, and then adopt its $1-\alpha$ empirical percentile to estimate $C_j(\alpha)$.

Next, the $1-\alpha$ simultaneous confidence region of $\boldsymbol{\beta}(s)$ can be constructed according to

$$\mathbb{P}\{\|\sqrt{n}(\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s))\|^2 \leq C_\beta(\alpha),\ \forall s \in [0,1]\} = 1-\alpha, \tag{12}$$

where the critical value $C_\beta(\alpha)$ can be approximated via utilizing an efficient resampling method:

- Fit model (3) and obtain $\hat{\boldsymbol{\beta}}(s)$, $\hat{g}(\cdot)$, and $\{\hat{\eta}_i(s)\}_{i=1}^n$;
- Generate the samples $\{\varsigma_i^{(r)}\}_{i=1}^n$ from $N(0,1)$ independently and construct the process sample $w^{(r)}(s) = \|\hat{f}_\eta(0;s)\{\frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{B}}_i(s)\}^{-1}\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \varsigma_i^{(r)}\widehat{\mathbf{A}}_i(s)\psi_\tau(\hat{\eta}_i(s))\}\|^2$;
- Repeat the second step for R times to obtain $\{\sup_s w^{(1)}(s), \ldots, \sup_s w^{(R)}(s)\}$, and then adopt its $1-\alpha$ empirical percentile to estimate $C_\beta(\alpha)$.

In component (iii), the $1-\alpha$ simultaneous confidence band of $g(\cdot)$ is proposed as follows:

$$(\hat{g}(z) - C_g(\alpha), \hat{g}(z) + C_g(\alpha))\,, \tag{13}$$

where the crtitical value $C_g(\alpha)$ can be determined via the bootstrap method:

- Fit the model (3) and obtain the estimated functions $\hat{\boldsymbol{\beta}}(s)$, $\hat{g}(\cdot)$, and the residuals $\{\hat{e}_i(s_l) \doteq y_i(s_l) - \hat{g}(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s_l)),\ l = 1, \ldots, m\}_{i=1}^n$;
- Generate random samples $\{\varsigma_i^{(r)}(s_l)\}_{l=1}^m$ independently from $N(0,1)$, and then construct

$$y_i^{(r)}(s_l) = \hat{g}(\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}(s_l)) + \varsigma_i^{(r)}(s_l)\hat{e}_i(s_l),\ i = 1, \cdots, n;$$

- Based on $\{\boldsymbol{x}_i, \{y_i^{(r)}(s_l)\}_{l=1}^m\}_{i=1}^n$, recalculate the estimator $\hat{g}^{(r)}(z)$ and the difference $\Delta_g^{(r)}(z) = \hat{g}(z) - \hat{g}^{(r)}(z)$;
- Repeat the previous two steps R times to have $\{\sup_z |\Delta_g^{(1)}(z)|, \ldots, \sup_z |\Delta_g^{(R)}(z)|\}$, and then adopt its $1-\alpha$ empirical percentile to estimate $C_g(\alpha)$.

3 Asymptotic properties

We systematically investigate the asymptotic properties of all estimators proposed in Section 2.2 and inference procedures in Section 2.3. Assumptions used to facilitate the technical details can be found in the Supplementary Material.

The following theorems tackle the theoretical properties of $\hat{\boldsymbol{\beta}}(s)$, $\hat{g}(\cdot)$, and $\{\hat{\eta}_i(s)\}_{i=1}^n$. The detailed proof can be found in the Supplementary Material.

Theorem 1 *Under Assumptions 1-14, one has:*

(i) Global uniform Bahadur representation of $\hat{\boldsymbol{\beta}}(s)$*:*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}(s))=-f_\eta(0;s)\left\{\frac{1}{n}\sum_{i=1}^n\mathbf{B}_i(s)\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{A}_i(s)\psi_\tau(\eta_i(s))\right\}+o_p(1), \quad (14)$$

where $\mathbf{A}_i(s)=[\boldsymbol{x}_i-\mathbb{E}\{\boldsymbol{x}_i|\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)\}]\dot{g}(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s))$ *and* $\mathbf{B}_i(s)=\mathbf{A}_i(s)\boldsymbol{x}_i^\top\dot{g}(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s))$.

(ii) Weak convergence of $\hat{\boldsymbol{\beta}}(s)$*:*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}(s))\Rightarrow\mathbb{G}_1(s),\ in\ s\in\mathcal{S}, \quad (15)$$

where $\mathbb{G}_1(s)$ *is a centered Gaussian process with covariance matrix*

$$\boldsymbol{\Xi}_1(s,s')=f_\eta^{-1}(0;s)f_\eta^{-1}(0;s')\{F_\eta(0,0;s,s')-\tau^2\}\mathbb{E}[\mathbf{B}_i(s)]^{-1}\mathbb{E}[\mathbf{A}_i(s)\mathbf{A}_i^\top(s')]\mathbb{E}[\mathbf{B}_i^\top(s)]^{-1}$$

with $F_\eta(\cdot,\cdot;s,s')$ *the joint distribution of* $(\eta(s),\eta(s'))$.

Theorem 2 *Under Assumptions 1-14, the following result holds:*

(i) Weak convergence of $\hat{g}(\boldsymbol{x}^\top\hat{\boldsymbol{\beta}}(s))$*:*

$$\sqrt{n}\left\{\hat{g}(\boldsymbol{x}^\top\hat{\boldsymbol{\beta}}(s))-g(\boldsymbol{x}^\top\boldsymbol{\beta}(s))\right\}\Rightarrow\mathbb{G}_2(s), \quad (16)$$

where $\mathbb{G}_2(s)$ *is a centered Gaussian process with covariance matrix*

$$\boldsymbol{\Xi}_2(s,s')=f_\eta^{-1}(0;s)f_\eta^{-1}(0;s')[F_\eta(0,0;s,s')-\tau^2]\boldsymbol{e}_{1,2}^\top\mathbb{E}[\boldsymbol{\Gamma}_i(s)]^{-1}\mathbb{E}[\boldsymbol{\upsilon}_i(s)\boldsymbol{\upsilon}_i^\top(s')]\mathbb{E}[\boldsymbol{\Gamma}_i(s')]^{-1}\boldsymbol{e}_{1,2}$$

with $\mathbf{\Gamma}_i(s) = \int_0^1 \boldsymbol{a}_i(u,s)\boldsymbol{a}_i(u,s)^\top \kappa(u,s)du$ *and* $\boldsymbol{\upsilon}_i(s) = \int_0^1 \boldsymbol{a}_i(u,s)\kappa(u,s)du$. *Here* $\kappa(u,s) = K_{h_\beta}\{\xi_i(u,s)\}\pi(u)$, $\xi_i(u,s) = \boldsymbol{x}_i^\top\boldsymbol{\beta}(u) - \boldsymbol{x}^\top\boldsymbol{\beta}(s)$, *and* $\boldsymbol{a}_i(u,s) = [1, h_g^{-1}\xi_i(u,s)]^\top$.

(ii) Weak convergence of $\hat{g}(z)$*:*

$$\sqrt{n}\,\{\hat{g}(z) - g(z)\} \Rightarrow \mathbb{G}_3(z),\ z \in \mathcal{Z}, \tag{17}$$

where $\mathcal{Z}$ *is the domain of function* $g(z)$ *and* $\mathbb{G}_3(z)$ *is a centered Gaussian process with covariance matrix* $\mathbf{\Xi}_3(z,z')$ *defined similarly as* $\mathbf{\Xi}_2(s,s')$ *via replacing* $\boldsymbol{x}^\top\boldsymbol{\beta}(s)$ *by* z.

Theorem 3 *Under Assumptions 1-14, one obtains*

(a) $\sup_{s\in[0,1]} |\hat{\eta}_i(s) - \eta_i(s)| = O_p(n^{-1/2} + h_\eta^2)$.

(b) $\sup_t |\hat{f}_\eta(t;s) - f_\eta(t;s)| = O_p\left\{n^{-1/2} + h_\eta^2 + h_{f_\eta}^2 + n^{-3/2}h_{f_\eta}^{-4} + h_\eta^6 h_{f_\eta}^{-4} + \sqrt{\frac{\log n}{nh_{f_\eta}}}\right\}$,

for any $s \in [0,1]$.

(c) $\sup_{s,s'\in[0,1]} |\hat{\gamma}_\eta(s,s') - \gamma_\eta(s,s')| = O_p(n^{-1/2} + h_\eta^2)$.

The following theorem derives the asymptotic distribution of global test statistic T_n under the null hypothesis. The detailed proof can be found in the Supplementary Material.

Theorem 4 *Under Assumptions S1-S14, one has*

$$T_n \Rightarrow \int_{s\in\mathcal{S}} \mathbb{G}_C^\top(s)\mathbb{G}_C(s)ds, \tag{18}$$

where $\mathbb{G}_C(s) = \{\mathbf{C}\mathbf{\Xi}_1(s,s)\mathbf{C}^\top\}^{-1/2}\mathbf{C}\mathbb{G}_1(s)$ *is a centered Gaussian process, in which* $\mathbb{G}_1(s)$ *and* $\mathbf{\Xi}_1(s,s)$ *are defined in Theorem 1.*

4 Simulation study

To investigate the performance of our proposed method, we generated synthetic data from model (2) using the LQD transformation Ψ and the index function $g(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)) = \sin(2\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)) +$

$2\cos(2+\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)), i = 1, 2, \ldots, n$, where the varying coefficients $\boldsymbol{\beta}(s)$ are given by $(1+s^2, (1-s)^2)^\top$, and then scaled as $\boldsymbol{\beta}(s)/\|\boldsymbol{\beta}(s)\|$ for model identification. For a given covariate vector $\boldsymbol{x}_i$, inspired by the idea in Han et al. (2019), we implemented random sampling based on error-contaminated random quantile functions

$$\mathcal{Q}(s|\boldsymbol{x}_i) = \theta_\eta(\boldsymbol{x}_i)^{-1}\int_0^s \exp\{g(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)) + \eta_i(s)\}ds,$$

where $\theta_\eta(\boldsymbol{x}_i) = \int_0^1 \exp\{g(\boldsymbol{x}_i^\top\boldsymbol{\beta}(s)) + \eta_i(s)\}ds$ and $\eta_i(s) = \iota_{i1}\xi_1(s) + \iota_{i2}\xi_2(s)$ with $\iota_{i1} \sim N(0, 0.01^2), \iota_{i2} \sim N(0, 0.05^2), \xi_1(s) = \cos(2\pi s)$, and $\xi_2(s) = \sin(\pi s)$. Furthermore, the covariates are generated by $\boldsymbol{x}_i = \Phi(\boldsymbol{v}_i)$, where Φ is the bivariate standard normal CDF and $\boldsymbol{v}_i \sim N(\mathbf{0}_p, \boldsymbol{\Sigma})$, $i = 1, 2, \ldots, n$, where $\boldsymbol{\Sigma}$ is a 2×2 matrix with elements $0.6^{|j'-j|}$ for $j, j' = 1, 2$. Then we generated $u_{i,1}, \ldots, u_{i,N_i} \sim$ Uniform$(0, 1)$ independent of $\boldsymbol{x}_i$, and obtained random samples $\{\mathcal{Q}(u_{i,t}|\boldsymbol{x}_i)\}_{t=1}^{N_i}$ as the simulated observations, which will be used in estimating the response densities $\{f_i\}_{i=1}^n$. Examples of 20 randomly generated densities are illustrated in the Supplementary Material. Without loss of generality, we assume $N_i = N$ for each subject and consider scenarios with $N = 500$ and 1000 to assess how it affects the estimation performance of the proposed method. For the sample size n of simulated subjects, we choose $n = 100$ and 200. Finally, we generated 200 datasets for each simulation scenario. Note that we did not directly generate the density functions as in (2), nor were the quantile assumption of residual errors in (3) used in any way in simulating the data. Therefore, the data generating process does not give any inherent advantage to our approach over others.

We applied our method to the simulated data, where the unknown densities were estimated first and the coefficient functions along with the single index function were estimated for $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$ respectively. As an illustration, for $\tau = 0.5$, $N = 1000$, and $n = 200$, the mean performance of the estimated functions $\boldsymbol{\beta}(s)$ and $g(\cdot)$ based on our proposed method are shown in Figure 2. As expected, the estimated curves (broken lines) are very close to the

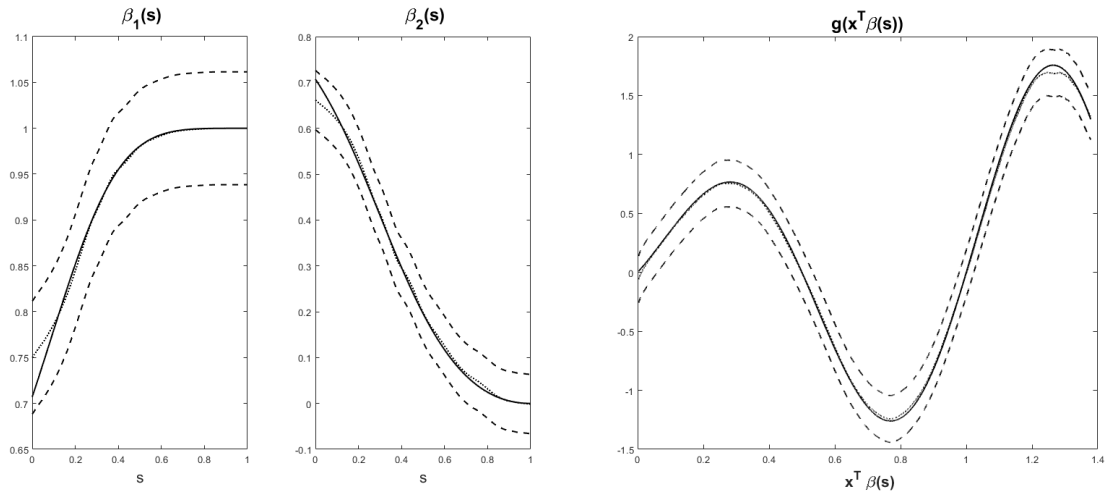

Figure 2: Simulation results based on our proposed method with $\tau = 0.5$, $N = 1000$, and $n = 200$. In each panel, the solid line represents the true function, the broken line represents the estimated function, and dashed lines are the corresponding 95% simultaneous confidence bands.

corresponding true curves (solid lines), and the true curves are all in the range of the corresponding 95% simultaneous confidence bands (dashed lines). In addition to our proposed method, we also considered a few alternative approaches and assessed their performance, including (i) a grid-wised single index model in Ichimura (1993), where the coefficient functions $\boldsymbol{\beta}(s)$ were estimated at each grid without using the kernel weights to borrow information from the neighborhoods, and (ii) the single index varying coefficient model proposed in Luo et al. (2016). Note that the two competitors were originally developed for well registered functional responses, not for the scenarios we are interested in here. To make it comparable, we applied them to the transformed densities instead of the simulated random samples. Therefore, the same density estimation procedure is adopted in the three methods, and the corresponding density estimation performance is evaluated by the mean and standard deviation of integrated squared error (ISE) in Table 1. Furthermore, to compare the estimation performance among all three methods, the mean and standard deviation of ISE are reported for the proposed estimators of functions $\boldsymbol{\beta}(s), g(\cdot)$, and the plug-in estimators of true density functions $\{\Psi^{-1}[g(\boldsymbol{x}_i^\top \boldsymbol{\beta}(s))]\}_{i=1}^n$ in Table 1. Additional es-

timation results from 200 simulated datasets for $N = 1000$ can be found in the Supplementary Material.

Table 1: Estimation results from 200 simulated datasets for $N = 500$ measured by the mean and standard deviation (std) of ISE.

n	$f_i(\cdot)$	Method	τ	$\beta_1(s)$	$\beta_2(s)$	$g(\cdot)$	$\Psi^{-1}[g(\boldsymbol{x}_i^\top \boldsymbol{\beta}(s))]$
			0.1	0.0008 (0.0032)	0.0012 (0.0016)	0.0218 (0.0067)	0.0142 (0.0066)
			0.3	0.0004 (0.0001)	0.0009 (0.0003)	0.0065 (0.0017)	0.0080 (0.0025)
		our method	0.5	0.0004 (0.0001)	0.0008 (0.0003)	0.0038 (0.0012)	0.0055 (0.0016)
100	0.0171 (0.0024)		0.7	0.0003 (0.0001)	0.0007 (0.0003)	0.0069 (0.0018)	0.0046 (0.0017)
			0.9	0.0015 (0.0089)	0.0007 (0.0005)	0.0235 (0.0041)	0.0057 (0.0032)
		grid-wised	-	0.0123 (0.0001)	0.0078 (0.0285)	0.0061 (0.0103)	0.0100 (0.0131)
		Luo et al. (2016)	-	0.0004 (0.0012)	0.0052 (0.0260)	0.0046 (0.0102)	0.0071 (0.0128)
			0.1	0.0004 (0.0006)	0.0009 (0.0004)	0.0200 (0.0041)	0.0123 (0.0040)
			0.3	0.0004 (0.0001)	0.0007 (0.0002)	0.0054 (0.0009)	0.0069 (0.0017)
		our method	0.5	0.0003 (0.0001)	0.0006 (0.0002)	0.0030 (0.0007)	0.0048 (0.0011)
200	0.0174 (0.0018)		0.7	0.0003 (0.0001)	0.0005 (0.0002)	0.0057 (0.0010)	0.0034 (0.0012)
			0.9	0.0022 (0.0147)	0.0005 (0.0009)	0.0216 (0.0023)	0.0035 (0.0013)
		grid-wised	-	0.0123 (0.0001)	0.0038 (0.0007)	0.0037 (0.0004)	0.0072 (0.0010)
		Luo et al. (2016)	-	0.0004 (0.0007)	0.0020 (0.0080)	0.0030 (0.0032)	0.0052 (0.0037)

From Table 1, it can be found that: (i) for the kernel density estimators in all three methods, when the number of random samples, N, is getting larger, the estimation performance is getting better in terms of both the mean and standard deviation of ISE; (ii) except for the grid-wise method, the increasing of N improves the estimation performance of functions $\boldsymbol{\beta}(s), g(\cdot)$ and individual density functions $\{\Psi^{-1}[g(\boldsymbol{x}_i^\top \boldsymbol{\beta}(s))]\}_{i=1}^n$; (iii) all three methods possess better estimation performance when the sample size n increases; (iv) both our proposed method and the method in Luo et al. (2016) outperform the grid-wised method for all different scenarios, which shows the power of local kernel weights in estimating the coefficient functions; and (v) when $\tau = 0.5$ or 0.7,

our proposed method performs better than the method in Luo et al. (2016) for most scenarios in terms of the mean ISE and more robust than the method in Luo et al. (2016) for all the scenarios in terms of the standard deviation of ISE.

Next, for $\tau = 0.5, N = 1000$, and $n = 200$, the empirical coverage probabilities of the simultaneous confidence bands (SCB) of $\beta_1(s), \beta_2(s)$ and $g(\cdot)$ and the simultaneous confidence regions (SCR) of $\boldsymbol{\beta}(s)$ are summarized in Table 2(a) based on 200 replications. As expected, all the empirical coverage probabilities are close to the significance levels.

Table 2: Inference results at $1 - \alpha$ significance level.

(a) Empirical coverage probabilities of SCB and SCR.

	SCB			SCR
α	$\beta_1(\cdot)$	$\beta_2(\cdot)$	$g(\cdot)$	$\boldsymbol{\beta}(\cdot)$
0.05	0.980	0.830	0.960	0.940
0.01	0.985	0.895	0.980	0.970

(b) Rejection rates for hypothesis testing (19).

	c				
α	0	0.1	0.2	0.3	0.4
0.05	0.0052	0.7710	0.9937	1	1
0.01	0	0.7634	0.9937	1	1

Finally, we examine the type I and II error rates of the global test statistic T_n. In particular, we fixed $\beta_1(s)$ and set $\beta_2(s) = c(1-s)^2$, where c is a scalar specified below. We want to test the following hypotheses

$$\mathbb{H}_0 : \beta_2(s) = 0 \text{ for all } s \in [0,1] \text{ vs. } \mathbb{H}_1 : \beta_2(s) \neq 0 \text{ for some } s \in [0,1], \tag{19}$$

where we set $c = 0$ to assess the type I error rates for T_n, and set $c = 0.1, 0.2, 0.3$, and 0.4 to examine the power of T_n. For each simulation, the significance levels were set at $\alpha = 0.05$ and 0.01, 500 bootstrap replications were generated to construct the empirical distribution of T_n under $\mathbb{H}_0$, and 200 replications were used to estimate the rejection rates, which are summarized in Table 2(b). It shows that the type I error rates for T_n based on the bootstrap method are accurate, and the power increases rapidly while increasing the value of c.

5 Real data analysis

In this real data analysis, we consider the pre-operative FLAIR brain images of 101 patients from TCIA-GBM collection. The tumor segmentation annotation is also extracted via GLISTRboost, a machine learning method based on a hybrid generative-discriminative model (Bakas et al. 2017*b*), after registration of the whole brain. Illustration of the segmentation annotation is shown in Figure 1. Similar to Yang et al. (2020), for each patient, we only focused on the image slice with largest tumor area, extracted the pixel intensities within the tumor area, where the number of pixels within the tumor ranged from 503 to 5053. The goal of this data analysis is to derive functional representations of extracted pixel intensities and investigate their relationship with some covariates of interests, including gender, age, survival length, and the area proportions of tumor sub-regions, i.e., necrosis-tumor ratio and edema-tumor ratio. The summary statistics of these covariates are shown in Table 3.

Table 3: Summary statistics of clinical and demographic variables from 101 patients.

	Female (38 patients)				Male (63 patients)				Overall			
	min	max	median	mean	min	max	median	mean	min	max	median	mean
age	18.0	84.0	55.0	54.7	23.0	81.0	60.0	59.6	18.0	84.0	59.0	57.7
survival length	71.0	1757.0	306.5	490.1	5.0	2768.0	430.0	595.0	5.0	2768.0	405.0	555.5
necrosis-tumor ratio	0.00	0.58	0.12	0.18	0.00	0.74	0.10	0.15	0.00	0.74	0.11	0.16
edema-tumor ratio	0.16	0.91	0.52	0.52	0.15	0.98	0.52	0.54	0.15	0.98	0.52	0.53

We first derived the boundary-corrected density estimators from these pixel intensities, and applied the LQD transformation on the estimated density functions. Then we fitted our model to the transformed curves based on different values of τ. Figure 3 shows the estimation (solid curves) and 95% simultaneous confidence bands (dashed curves) of the coefficient functions and index function at $\tau = 0.5$. In particular, nonlinear patterns can be found in the estimated link

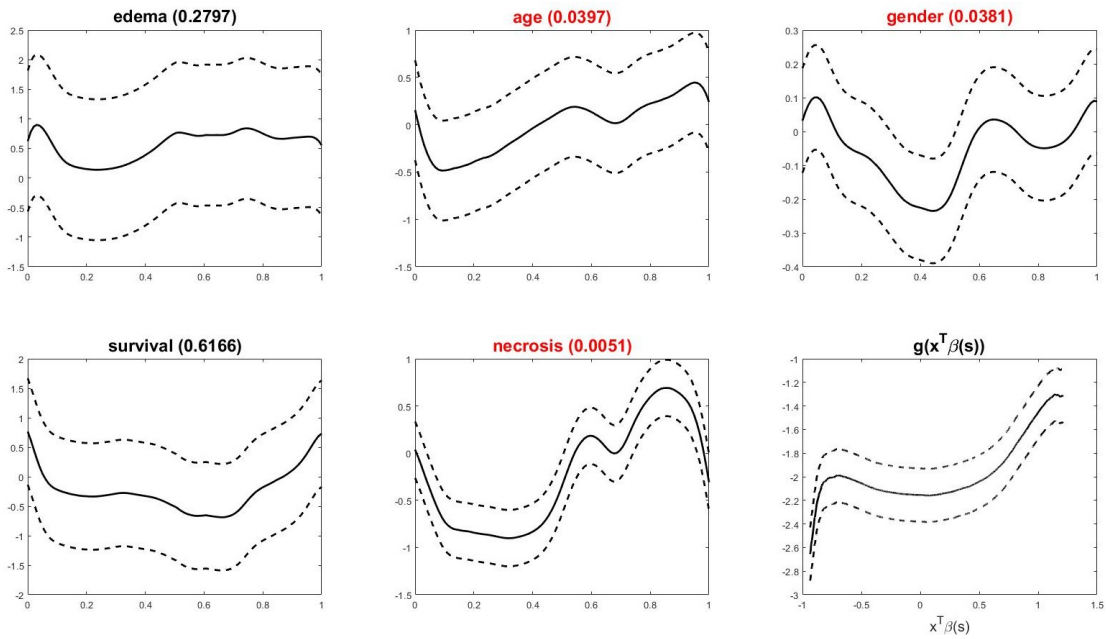

Figure 3: The estimated coefficients functions and index functions (solid curves) with corresponding 95% simultaneous confidence bands (dashed curves) at $\tau = 0.5$.

function. Next, we tested the hypothesis:

$$\mathbb{H}_0 : \beta_k(s) = 0 \text{ for all } s \text{ vs. } \mathbb{H}_1 : \beta_k(s) \neq 0 \text{ for at least one } s,$$

where $k = 1, 2, 3, 4$, and 5, corresponding to the effects of age, gender, survival length, edema-tumor ratio, and necrosis-tumor ratio, respectively. The corresponding p-values are summarized in Fig.3 with those p-values less than the significant level 5% highlighted in red. Given the significant level 0.05, age, gender and necrosis-tumor ratio are all found to be significant on the LQD transformed curves. Consistent findings were also reported in some existing literature. For example, some intensity based radiomic features, e.g., the minimum of intensity, were found to be significantly correlated to the age division (Li et al. 2017); and females were found to have higher pixel intensities than males, especially in the upper tails of the distribution (Yang et al. 2020). The significance of necrosis-tumor ratio is mainly caused by the FLAIR image acquisition techniques, i.e., pixel intensities with necrosis are higher than other tumor sub-regions, e.g.,

edema, in FLAIR brain images (Henker et al. 2017). As the necrosis-tumor ratio is a helpful, reliable sign of poor prognosis in supratentorial glioblastoma (Pierallini et al. 1996), our derived LQD transformation based functional representations have the potential to be important biomarkers in defining prognosis and distinguishing different glioblastoma subtypes.

Furthermore, the prediction performance of our method in terms of the quantile functions was assessed at different levels of τ. Another two methods, i.e., the quantile function on scalar regression model (Yang et al. 2020) and the single-index varying coefficient model (Luo et al. 2016), were also considered for comparison. For our method and the method in Luo et al. (2016), the inverse transformation (Petersen & Müller 2016) was applied on the predicted LQD transformed curves to get the corresponding quantile functions. The two-fold cross-validation strategy is adopted to calculate the prediction errors. Specifically, the original dataset was first randomly partitioned into two equal sized sub-datasets, i.e., $\mathcal{D}_1$ and $\mathcal{D}_2$. Next, we trained the model on $\mathcal{D}_1$ and predicted the quantile functions on $\mathcal{D}_2$, followed by training on $\mathcal{D}_2$ and predicting on $\mathcal{D}_1$. Then, the the difference between predicted quantile functions and the empirical ones in each of the two sub-datasets are averaged as the final prediction errors. Here we consider both the integrated squared error (ISE) and the integrated absolute erros (IAE), which are shown in Table 4. It can be found that, when $\tau = 0.7$ and 0.8, our method achieves the best performance among the three competing methods in terms of both ISE and IAE, which indicates the advantages of our method in introducing (i) the nonlinear link function compared to the method in Yang et al. (2020) and (ii) the quantile regression methodology compared to the method in Luo et al. (2016).

6 Discussion

One of the key techniques in our method is to map the density functions to functions in an unconstrained linear Hilbert space by the LQD transformation. However, as discussed in Chen et al.

Table 4: Prediction performance comparison in terms of the quantile functions.

Method	our method					Yang et al. (2020)	Luo et al. (2016)
	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$		
ISE	0.1425	0.1306	0.1124	0.1066	0.1572	0.1368	0.1246
IAE	0.1021	0.0892	0.0747	0.0757	0.1171	0.1012	0.0833

(2021), the LQD transformation does not take into account the geometry of the space of probability distributions and therefore the corresponding transformation map is not isometric and leads to deformations that change distances between pairs of densities. Some metrics have been considered to investigate the geometry of the space of density functions, including the Wasserstein metric (Panaretos & Zemel 2020) and the Fisher-Rao metric (Srivastava & Klassen 2016). The Wasserstein metric provides an optimal transport distance that measures the cost of transporting one density to another while the Fisher-Rao metric can be viewed as the geodesic distance between the square-roots of the densities (Petersen et al. 2021). Recently, these two metrics have been widely adopted in establishing regression models with random distributions served as responses and/or predictors (Chen et al. 2021, Ghosal et al. 2021, Mohammed et al. 2021, Bhattacharjee & Müller 2021). However, few of these methods focuses on the statistical inference procedures such as hypothesis testing and constructing simultaneous confidence band (or region) on the coefficient parameters. Therefore, in our future research work, it will be critically important to extend our density-on-scalar single-index quantile regression model to some novel ones that can successfully incorporate the geometric properties of densities and establish the corresponding estimation and inference procedures simultaneously.

SUPPLEMENTARY MATERIAL

Supplementary_Material: Supplementary material includes bandwidth selection, assumptions

of theorems, proofs of the theoretical results, and additional simulation results. (.pdf file)

References

Bai, X., Wang, X., Liu, X., Liu, Q., Song, J., Sebe, N. & Kim, B. (2021), 'Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments', *Pattern Recognition* **120**, 108102.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K. & Davatzikos, C. (2017*a*), Segmentation labels for the pre-operative scans of the tcga-gbm collection, The Cancer Imaging Archive. https://wiki.cancerimagingarchive.net/x/KoZyAQ.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K. & Davatzikos, C. (2017*b*), 'Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features', *Scientific data* **4**(1), 1–13.

Bhattacharjee, S. & Müller, H.-G. (2021), 'Single index fréchet regression', *arXiv: 2108.05437* .

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. et al. (2018), 'The uk biobank resource with deep phenotyping and genomic data', *Nature* **562**(7726), 203–209.

Chen, Y., Lin, Z. & Müller, H.-G. (2021), 'Wasserstein regression', *Journal of the American Statistical Association* pp. 1–14.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M. et al. (2013), 'The cancer imaging archive (tcia): maintaining and operating a public information repository', *Journal of digital imaging* **26**(6), 1045–1057.

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

Ghosal, A., Meiring, W. & Petersen, A. (2021), 'Fr\'echet single index models for object response regression', *arXiv:2108.06058* .

Han, K., Müller, H.-G. & Park, B. U. (2019), 'Additive functional regression for densities as responses', *Journal of the American Statistical Association* .

Henker, C., Kriesen, T., Glass, Ä., Schneider, B. & Piek, J. (2017), 'Volumetric quantification of glioblastoma: experiences with different measurement techniques and impact on survival', *Journal of neuro-oncology* **135**(2), 391–402.

Ichimura, H. (1993), 'Semiparametric least squares (sls) and weighted sls estimation of single-index models', *Journal of econometrics* **58**(1-2), 71–120.

Koenker, R. & Bassett, G. J. R. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., Dekker, A. et al. (2012), 'Radiomics: extracting more information from medical images using advanced feature analysis', *European journal of cancer* **48**(4), 441–446.

Li, Z., Wang, Y., Yu, J., Guo, Y. & Zhang, Q. (2017), 'Age groups related glioblastoma study based on radiomics approach', *Computer Assisted Surgery* **22**(sup1), 18–25.

Liu, Y., Li, M. & Morris, J. S. (2020), 'Function-on-scalar quantile regression with application to mass spectrometry proteomics data', *The Annals of Applied Statistics* **14**(2), 521–541.

Luo, X., Zhu, L. & Zhu, H. (2016), 'Single-index varying coefficient model for functional responses', *Biometrics* **72**(4), 1275–1284.

Ma, Y. & Zhu, L. (2014), 'On estimation efficiency of the central mean subspace', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(5), 885–901.

Mechelli, A., Price, C. J., Friston, K. J. & Ashburner, J. (2005), 'Voxel-based morphometry of the human brain: methods and applications', *Current Medical Imaging* **1**(2), 105–113.

Mohammed, S., Bharath, K., Kurtek, S., Rao, A. & Baladandayuthapani, V. (2021), 'Radiohead: Radiogenomic analysis incorporating tumor heterogeneity in imaging through densities', *The Annals of Applied Statistics* **15**(4), 1808–1830.

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

Müller, U. U., Schick, A. & Wefelmeyer, W. (2007), 'Estimating the error distribution function in semiparametric regression', *Statistics & Decisions* **25**, 1–18.

Panaretos, V. M. & Zemel, Y. (2020), *An invitation to statistics in Wasserstein space*, Springer Nature.

Petersen, A., Chen, C.-J. & Müller, H.-G. (2019), 'Quantifying and visualizing intraregional connectivity in resting-state functional magnetic resonance imaging with correlation densities', *Brain connectivity* **9**(1), 37–47.

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

Petersen, A., Zhang, C. & Kokoszka, P. (2021), 'Modeling probability density functions as data objects', *Econometrics and Statistics* .

Pierallini, A., Bonamini, M., Osti, M., Pantano, P., Palmeggiani, F., Santoro, A., Maurizi Enrici,

R. & Bozzao, L. (1996), 'Supratentorial glioblastoma: neuroradiological findings and survival after surgery and radiotherapy', *Neuroradiology* **38**(1), S26–S30.

Roy, S., Bhattacharyya, D., Bandyopadhyay, S. K. & Kim, T.-H. (2018), 'Heterogeneity of human brain tumor with lesion identification, localization, and analysis from mri', *Informatics in Medicine Unlocked* **13**, 139–150.

Srivastava, A. & Klassen, E. P. (2016), *Functional and shape data analysis*, Springer.

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

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

Yao, F. & Lee, T. C. M. (2006), 'Penalized spline models for functional principal component analysis', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(1), 3–25.

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

You, J., Chen, G. & Zhou, Y. (2006), 'Block empirical likelihood for longitudinal partially linear regression models', *Canadian Journal of Statistics* **34**(1), 79–96.

Zhu, H. T., Li, R. & Kong, L. (2012), 'Multivariate varying coefficient model for functional responses', *The Annals of Statistics* **40**(5), 2634–2666.