High-Dimensional Spatial Quantile Function-on-Scalar Regression

Zhengwu Zhang, Xiao Wang, Linglong Kong & Hongtu Zhu

Taylor & Francis
Taylor & Francis Group

Check for updates

High-Dimensional Spatial Quantile Function-on-Scalar Regression

Zhengwu Zhang[a], Xiao Wang[b], Linglong Kong[c], and Hongtu Zhu[d]

[a]Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC; [b]Department of Statistics, Purdue University, West Lafayette, IN; [c]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada; [d]Department of Biostatistics, University of North Carolina, Chapel Hill, NC

ABSTRACT

This article develops a novel spatial quantile function-on-scalar regression model, which studies the conditional spatial distribution of a high-dimensional functional response given scalar predictors. With the strength of both quantile regression and copula modeling, we are able to explicitly characterize the conditional distribution of the functional or image response on the whole spatial domain. Our method provides a comprehensive understanding of the effect of scalar covariates on functional responses across different quantile levels and also gives a practical way to generate new images for given covariate values. Theoretically, we establish the minimax rates of convergence for estimating coefficient functions under both fixed and random designs. We further develop an efficient primal-dual algorithm to handle high-dimensional image data. Simulations and real data analysis are conducted to examine the finite-sample performance.

1. Introduction

Functional data analysis (FDA) has been an active area of research in the past decade. Well-known monographs in this area include Ramsay and Silverman (2005, 2007), Bowman (2010), and Ferraty and Vieu (2006). Functional regression, particularly the functional linear regression model (FLM), has been extensively studied. Functional data can be treated as either response variables or covariate predictors (Greven and Scheipl 2017). In FLM, the mean dependence is modeled through a linear model generalizing the standard multiple linear regression model. The literature on FLM is vast. For example, scalar-on-function regression (a continuous response variable regressed on functional covariates) has been studied by Cai and Hall (2006), Crambes, Kneip, and Sarda (2009), Yuan and Cai (2010), Hall and Horowitz (2007), and Wang, Zhu, and Alzheimer's Disease Neuroimaging Initiative (2017). Function-on-function regression (a functional response regressed on functional predictors) has been investigated by Yao, Müller, and Wang (2005), Ramsay and Silverman (2005), Ivanescu et al. (2015), and Sun et al. (2018). Function-on-scalar regression (a functional response regressed on scalar predictors) has been studied by Ramsay and Silverman (2005, chap. 13), Reiss, Huang, and Mennes (2010), and Goldsmith and Kitago (2016). The aim of this article is to develop a spatial quantile function-on-scalar regression model and investigate its related computational and theoretical issues.

Different from the ordinary regression that only models conditional mean dependence, quantile regression as a more comprehensive procedure estimates conditional quantiles of the response variable. Since Koenker and Bassett (1978), quantile regression has been an emerging field of statistical research

and has been widely applied in many disciplines including neuroimaging, finance, and economics. Quantile regression is attractive because it estimates conditional quantiles and allows statistical inference on the response's entire conditional distribution. Most importantly, quantile regression does not require a specified error distribution, providing a flexible framework for modeling complex non-Gaussian data. To date, while many studies have examined different quantile regression models, a limited number of these consider functional data. Cai and Xu (2008) studied quantile regression methods for a set of smooth coefficient time series models. Koenker (2004) and Yi and He (2009) developed quantile regression methods for longitudinal data. Wang, Zhu, and Zhou (2009) studied semiparametric quantile autoregression models in partially linear varying coefficient models using splines. Li, Liu, and Zhu (2007) proposed a kernel quantile regression to estimate conditional quantile functions given training scalar covariates and responses. Kato (2012) studied functional quantile regression with functional covariates and a scalar response.

In this article, motivated by medical imaging analysis, we propose a novel spatial quantile function-on-scalar regression model (denoted by SQR) that studies the conditional spatial distribution of a functional response given scalar predictors. We are particularly interested in the case where the responses are high-dimensional functions or images obtained from n independent subjects. Let the functional response be $\{Y(s) : s \in \mathcal{S}\}$ on a field $\mathcal{S}$ and the scalar predictors be $\mathbf{x} \in \mathbb{R}^p$. Our SQR model includes two major components. The first component is to model the marginal conditional distribution of $Y(s)$ given $\mathbf{x}$ at a fixed location s. This task can

be achieved by assuming that the τ-level conditional quantile of $Y(s)$ is assumed to be a linear function of $\mathbf{x}$ such that $Q_{Y(s)}(\tau|x) = \mathbf{x}^T\boldsymbol{\beta}_\tau(s)$, where $Q_{Y(s)}(\cdot|\mathbf{x})$ is the conditional quantile function of $Y(s)$ given $\mathbf{x}$ at a location s and $\boldsymbol{\beta}_\tau(s) = (\beta_{\tau 1}(s),\ldots,\beta_{\tau p}(s))^T$ contains p unknown coefficient functions. The second component is to model the joint conditional distribution of $Y(s)$ among spatial locations via a copula model. We adopt a parametric copula to specify the joint distribution of $F_{(s,\mathbf{x})}(Y(s))$, where $F_{(s,\mathbf{x})}(\cdot)$ is the conditional cumulative distribution function (CDF) of $Y(s)$ given $\mathbf{x}$ at s and $F_{(s,\mathbf{x})}(Y(s))$ follows a uniform distribution marginally. With these two components, we are able to explicitly characterize the conditional joint distribution of $Y(s)$ on the whole spatial domain $\mathcal{S}$, forming a generative model that given $\mathbf{x}$, we can simulate the image data Y. The proposed method provides a framework to comprehensively understand the effects of scalar covariates (e.g., age, gender, and disease status) on an image response, and a practical way to generate new images given covariates.

Although SQR has been studied in the literature (Su and Yang 2007; Hallin, Lu, and Yu 2009; Kostov 2009; Reich, Fuentes, and Dunson 2011; Reich 2012; Lu, Tang, and Cheng 2014; Yang and He 2015), our method differs from them in several important aspects. First, our SQR is primarily developed for handling large-scale image data $Y(s)$ with dense grids across multiple subjects. In contrast, most existing works on SQR focus on the development of statistical methods for analyzing simple spatial or longitudinal data (Hallin, Lu, and Yu 2009; Reich, Fuentes, and Dunson 2011; Reich 2012; Yang and He 2015). For instance Hallin, Lu, and Yu (2009) considered SQR for a strictly stationary real random field and proposed a local linear estimator of spatial quantile function. Second, we employ a parametric copula model to directly delineate the spatial dependence of image data within each subject. In contrast, most existing copula models were applied to quantile regression for different purposes and data types (Chen, Koenker, and Xiao 2009; Bouyé and Salmon 2013; De Backer, Ghouch, and Van Keilegom 2017; Kraus and Czado 2017; Wang, Feng, and Dong 2019). For instance, in Wang, Feng, and Dong (2019), the copula was used to model the temporal dependence of longitudinal data, while it is assumed a linear quantile regression model with $Q_{Y(t)}(\tau|\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}_0(\tau)$ at a fixed time point t and for given covariates $\mathbf{x}$. However, our model assumes that the components of $\boldsymbol{\beta}_\tau(s)$ are functions of both τ and the spatial location. Third, the unknown coefficient functions $\boldsymbol{\beta}_\tau(s)$ are assumed to reside in a reproducing kernel Hilbert space and estimate them through minimizing a quantile check function (Koenker and Bassett 1978) plus a roughness penalty. Furthermore, we estimate the unknown parameters of the copula model by using generalized least squares (Cressie 1985; Genton 1998). Most of the aforementioned articles involving coefficient function estimation utilize a finite-dimensional approximation, such as Fourier basis or through a roughness penalty, to regularize model complexity. It often leads to reasonable functional estimate, but at a price of complicating theoretical investigation.

Our main contributions are summarized as follows. First, it allows us to establish a representation theorem stating that, although the optimization is defined on an infinite-dimensional function space, its solution actually resides in a data-adaptive finite-dimensional subspace. This result guarantees an exact solution when optimization is carried out on this finite dimensional subspace. Second, we develop and implement an efficient primal-dual algorithm to handle large image data efficiently, as computation is generally very challenging with complex, high-dimensional images in quantile regression. We use extensive numerical studies to demonstrate the computational advantages of our method over other popular quantile regression optimization solvers such as the alternating direction method of multipliers (ADMM) algorithm (Boyd et al. 2011). Third, an important theoretical result of our work establishes the optimal convergence rate of the error in estimating these coefficient functions under both fixed and random designs. This generalizes the results in Cai and Yuan (2012) and Du and Wang (2014) for functional linear regression with a scalar response to the functional response scenario.

The rest of this article is organized as follows. In Section 2, we introduce the mathematical details of the proposed SQR model, including the quantile regression and copula models. Section 3 introduces an efficient primal-dual algorithm to solve the high-dimensional quantile regression model and a generalized least square method to estimate parameters in the copula model. In Section 4, we studied the optimal convergence rate of the error in estimating the coefficient functions under both fixed and random designs. In Section 5, extensive numerical studies are used to illustrate the advantages of the proposed method.

2. Spatial Quantile Regression and Copula Models

The SQR model studies the conditional spatial distribution of an imaging response $\{Y(s) : s \in \mathcal{S}\}$ given scalar predictors $\mathbf{x} \in \mathbb{R}^p$. The conditional quantile function of $Y(s)$ given $\mathbf{x}$ for a fixed location s at any quantile level $\tau \in (0, 1)$ is assumed to be

$$Q_{Y(s)}(\tau|\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}_\tau(s), \tag{1}$$

where $\boldsymbol{\beta}_\tau(s) = (\beta_{\tau 1}(s),\ldots,\beta_{\tau p}(s))^T \in \mathcal{F}^p$ is the vector of the unknown coefficient functions with each component of $\boldsymbol{\beta}_\tau$ residing in the function space $\mathcal{F}$. Assume that $\mathcal{F}$ is a reproduce kernel Hilbert space (RKHS) $\mathcal{H}(K)$ with a reproducing kernel K. Equation (1) completely specifies the conditional marginal distribution of $Y(s)$ given $\mathbf{x}$ and s.

In practice, the image response is often observed at discrete locations $s_1,\ldots,s_m$ in $\mathcal{S}$, and the image values across different points in $\mathcal{S}$ are highly correlated. The conditional marginal distribution of $Y(s)$ is fully captured by (1), however, the spatial dependence among $Y(s)$ at different points in $\mathcal{S}$ is not modeled. To capture the joint distribution of $(Y(s_1),\ldots,Y(s_m))$ given scalar covariates $\mathbf{x}$, we propose using copula models. Let the conditional CDF of $Y(s_j)$ given $\mathbf{x}$ be $F_{(s_j,\mathbf{x})}(y) = \mathbb{P}(Y(s_j) \leq y|\mathbf{x})$, $j = 1,\ldots,m$. The random variable $U_{\mathbf{x}}(s_j) = F_{(s_j,\mathbf{x})}(y)$ has a uniform marginal distribution. The conditional copula of $(Y(s_1),\ldots,Y(s_m))$ given $\mathbf{x}$ is defined as the conditional joint CDF of $(U_{\mathbf{x}}(s_1),\ldots,U_{\mathbf{x}}(s_m))$ given by

$$C_{\boldsymbol{\theta}}(u_1,\ldots,u_m) = \mathbb{P}\Big(U_{\mathbf{x}}(s_1) \leq u_1,\ldots,U_{\mathbf{x}}(s_m) \leq u_m\Big|\mathbf{x}\Big). \tag{2}$$

The copula C parameterized by $\boldsymbol{\theta}$ contains all information on the spatial dependence structure of $Y(\cdot)$. In this article, we adopt a

parametric family of copulas C_θ with $\theta = \theta(\mathbf{x})$ to characterize the joint conditional distribution of $(Y(s_1), \ldots, Y(s_m))$. Therefore, the conditional joint distribution of $(Y(s_1), \ldots, Y(s_m))$ given $\mathbf{x}$ can be written as

$$\mathbb{P}\Big(Y(s_1) \le y_1, \ldots, Y(s_m) \le y_m \Big| \mathbf{x}\Big)$$
$$= C_{\theta(\mathbf{x})}\Big(F_{(s_1,\mathbf{x})}(y_1), \ldots, F_{(s_m,\mathbf{x})}(y_m)\Big).$$

There are many choices of copula models such as the Student's t copula (Demarta and McNeil 2005), the noncentral χ^2-copula (Bárdossy 2006), and the Gaussian copula (Kazianka and Pilz 2010). It is a challenging task to determine the optimal choice of the copula. In our numerical analysis, we have examined 21 different copula models for the real diffusion tensor imaging (DTI) dataset from the ADNI and found that the Student's t copula fits the data well in most of the times. Therefore, we adopt the Student's t copula in this article, and denote it by

$$C_\theta\Big(u_1, \ldots, u_m\Big) = \mathbf{t}_{\varrho,\Sigma_\theta}\Big(t_\varrho^{-1}(u_1), \ldots, t_\varrho^{-1}(u_m)\Big),$$

where $\mathbf{t}_{\varrho,\Sigma_\theta}$ represents the multivariate t-distribution function with parameter $(\Sigma_\theta, \varrho, \mathbf{0})$, and t_ϱ^{-1} is the inverse CDF of t-distribution with ϱ degrees of freedom. Note that $\varrho/(\varrho - 2)\Sigma_\theta$ is the covariance matrix of the t-distribution for $\varrho > 2$ in t-distribution. Furthermore, we use the Matérn family (Guttorp and Gneiting 2006; Matérn 2013) to specify a class of isotropic correlation functions given by

$$\text{corr}\Big(t_\varrho^{-1}(U(s+h)), t_\varrho^{-1}(U(s)) \Big| \mathbf{x}\Big) = M(h)$$

that depends on the spatial separation parameter h, where

$$M(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\alpha\|h\|)^\nu K_\nu(\alpha\|h\|). \tag{3}$$

Here K_ν is the modified Bessel function of the second kind, $\nu > 0$ is a smoothness parameter defining the smoothness of the sample path, and $\alpha > 0$ is a spatial scale parameter. For our numerical studies, the parameters ϱ and ν are selected and kept fixed, and the parameter α is treated as unknown. We assume that α depends on $\mathbf{x}$. With an abuse of notation, we write the scale parameter as $\exp(\alpha^T \mathbf{x})$, where $\alpha \in \mathbb{R}^p$.

Suppose that we observe $(\mathbf{x}_i, Y_i(s_{ij}))$ for subject i and location s_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, m$. In this article, we consider two different designs for the location points $\{s_{ij}\}$. The first one is called a *fixed design*, where the functional response are observed at the same locations across curves or images. That is, $s_{1j} = s_{2j} = \cdots = s_{nj} := s_j$ for $j = 1, \ldots, m$. A second design, called *random design*, occurs when $\{s_{i1}, \ldots, s_{im}\}$ are independently sampled from a distribution $\{\pi(s) : s \in \mathcal{S}\}$. Our goal is to estimate the coefficient functions $\beta_{\tau k}(\cdot)$, $k = 1, \ldots, p$ and the parameters in the copula model. There are two main challenges related to this problem. The first one is to construct statistically efficient estimates of unknown parameters, in particularly the infinite dimensional coefficient functions. The second one is to develop a numerically efficient algorithm to estimate unknown parameters to handle the high dimensionality of functional or image data.

3. Algorithms

3.1. Efficient Primal-Dual Algorithm for Estimating Coefficient Functions

For given observations $(\mathbf{x}_i, Y_i(s_{ij}))$, we estimate the unknown parameters in the spatial quantile regression model (1) by solving the following optimization problem for a given $\tau \in (0, 1)$,

$$\min_{\beta \in \mathcal{F}^p} \; \sum_{i=1}^n \sum_{j=1}^m \rho_\tau(Y_i(s_{ij}) - \mathbf{x}_i^T \beta_\tau(s_{ij})) + \frac{\lambda}{2}\sum_{k=1}^p \mathcal{R}(\beta_{\tau k}), \tag{4}$$

where $\rho_\tau(r) = \tau r I(r > 0) - (1-\tau) r I(r \le 0)$ is the check function (Koenker and Bassett 1978) and $I(\cdot)$ is the indicator function, $\beta_\tau(s) = (\beta_{\tau 1}(s), \ldots, \beta_{\tau p}(s))^T$ are the coefficient functions, $\mathcal{R}$ is a roughness penalty on β_τ, and $\lambda > 0$ is a parameter controlling the smoothness penalty. We let $\mathcal{R}(\beta_{\tau k}) = \|\beta_{\tau k}\|_K^2$, where $\|\cdot\|_K$ is a seminorm in the RKHS $\mathcal{H}(K)$ (Wahba 1990). For simplicity, we also assume that the null space of $\mathcal{H}(K)$ is $\{0\}$. In the following derivations, we illustrate the algorithm in the case of fixed design, so that $Y(s_{ij})$ can be denoted as $Y(s_j)$. An extension to the case of a random design is straightforward.

Throughout the article, Gaussian kernels are used for $K(\cdot, \cdot)$ just for simplicity. Other kernels can be easily incorporated. For example, some common options include the Laplace kernel $K(x, y) = \exp(-\sigma\|x - y\|_1)$, the polynomial kernel $K(x, y) = (\langle x, y\rangle + \sigma^2)^d$, and the inverse-quadratic kernel $K(x, y) = \sigma^2/(\sigma^2 + \|x - y\|_2^2)$. Learning kernels is definitely a nontrivial question. A common way for learning kernels is to combine different kernels to improve them. For example, let $K(x, y) = c_1 K_1(x, y) + c_2 K_2(x, y)$ with $c_1, c_2 \ge 0$, where K_1 and K_2 are potential kernels. The hyperparameters c_1 and c_2 can be set by cross-validation. Moreover, based on both simulations and real data analysis reported below, the use of Gaussian kernels leads to accurate estimation results.

Let $\widehat{\beta}_\tau = (\widehat{\beta}_{\tau 1}, \ldots, \widehat{\beta}_{\tau p})$ be the optimal solution of (4). For notational simplicity, we drop the subscript τ from the coefficient functions β when the context is clear. It is straightforward to establish a representation theorem (Wahba 1990) stating that $\widehat{\beta}$ actually resides in a finite-dimensional subspace of $\mathcal{H}$, which facilitates computation by reducing an infinite-dimensional optimization problem to a finite-dimensional one. Specifically,

$$\widehat{\beta}_k(s) = \mu_k + \boldsymbol{b}_k^T \mathbf{k}_s, \tag{5}$$

for $k = 1, \ldots, p$, where μ_k is a scalar, $\boldsymbol{b}_k = (b_{k1}, \ldots, b_{km})^T \in \mathbb{R}^m$, and $\mathbf{k}_s = (K(s, s_1), \ldots, K(s, s_m))^T \in \mathbb{R}^m$. The spatial smoothness of $\widehat{\beta}_k(s)$ comes from the RKHS assumption and the nature of image data. The use of regularization term $\lambda\mathcal{R}(\beta_k)$ is to regularize the coefficients $\{b_{kj}\}_{j=1}^m$ to avoid overfitting and the use of GACV allows us to implicitly incorporate spatial correlations to select the tuning parameter λ. Specifically, if there is high correlation among observations on the same subject, then GACV would lead to the selection of a large λ.

The challenges of solving (4) under the parameterized form in (5) come from several aspects. First, the number of parameters to be estimated is generally large in medical imaging applications. In particular, each of the p coefficient functions is represented by $m + 1$ unknown parameters (m depends on the

dimensionality of Y). For example, if Y is a two-dimensional image with a resolution of 50×50, we have $m = 2500$, and so we will need to estimate $p(2500 + 1)$ parameters. Second, the nondifferentiability of the check function significantly increases the difficulty of the optimization problem. A straightforward optimization strategy for (4) is to use the popular ADMM algorithm (Boyd et al. 2011) that divides the optimization (4) into a few simpler optimization subproblems that can be solved iteratively. Details of the ADMM algorithm applied for solving (4) are included in Supplement III. However, the convergence of ADMM can be very slow, and when the dimension of Y is large, the computational cost of some subproblems of ADMM can be huge due to the large matrix inversion.

To overcome this computational challenge, we propose a primal-dual algorithm that converges much faster than ADMM and avoids large matrix inversion in each optimization iteration. The primal-dual kind of optimization approaches has been applied in the field of quantile regression. For instance, Koenker and Park (1996) described a primal-dual approach for computation of nonlinear quantile regression estimators based on the interior point method. Portnoy and Koenker (1997) and Portnoy (1997) compared the interior point method with existing simplex-based methods for quantile regression, and showed that the interior point approach was competitive and exhibited a rapidly increasing advantage for large problems. In our problem, to estimate coefficient function $\widehat{\beta}_k$ corresponding to the covariate x_k for $k = 1, \ldots, p$, we rely on the formula (5). However, the traditional interior point algorithm for quantile regression (Koenker and Park 1996; Portnoy 1997; Portnoy and Koenker 1997) does not involve $\{\mathbf{b}_k\}$ that are difficult to solve.

We make a critical observation that the primal variable $\mu = (\mu_1, \ldots, \mu_p)^T$ is also a dual variable for the dual problem, which leads to the following efficient optimization strategy. Similar to Li, Liu, and Zhu (2007), we introduce positive residuals $\{\xi_{ij} \in \mathbb{R}_+\}$ and negative residuals $\{\zeta_{ij} \in \mathbb{R}_+\}$ such that $-\zeta_{ij} \leq Y_i(s_j) - \widehat{Y}_i(s_j) \leq \xi_{ij}$, where $\widehat{Y}_i(s_j) = \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}(s_j)$. We obtain an optimization problem that is equivalent to (4) as follows:

$$\min_{\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\mu}, \boldsymbol{b}} \tau \sum_{i=1}^n \sum_{j=1}^m \xi_{ij} + (1 - \tau) \sum_{i=1}^n \sum_{j=1}^m \zeta_{ij} + \frac{\lambda}{2} \sum_{k=1}^p \boldsymbol{b}_k^T \boldsymbol{\Sigma} \boldsymbol{b}_k$$

$$\text{s.t. } Y_i(s_j) = \mathbf{x}_i^T \boldsymbol{\mu} + \sum_{k=1}^p x_{ik} \mathbf{k}_{s_j}^T \boldsymbol{b}_k + \xi_{ij} - \zeta_{ij},$$

$$i = 1, \ldots, n, \ j = 1, \ldots, m, \ (\boldsymbol{\xi}, \boldsymbol{\zeta}) \in \mathbb{R}_+^{nm},$$

where $\boldsymbol{\Sigma} = (K(s_i, s_j)) \in \mathbb{R}^{m \times m}$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T \in \mathbb{R}^p$, x_{ik} is the kth element of $\mathbf{x}_i$, and ξ_{ij} and ζ_{ij} are elements of $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$, respectively. The primal Lagrangian function is defined as

$$\Lambda = \tau \sum_{i=1}^n \sum_{j=1}^m \xi_{ij} + (1 - \tau) \sum_{i=1}^n \sum_{j=1}^m \zeta_{ij} + \frac{\lambda}{2} \sum_{k=1}^p \boldsymbol{b}_k^T \boldsymbol{\Sigma} \boldsymbol{b}_k$$

$$- \sum_{i=1}^n \sum_{j=1}^m d_{ij} (\mathbf{x}_i^T \boldsymbol{\mu} + \sum_{k=1}^p x_{ik} \mathbf{k}_{s_j}^T \boldsymbol{b}_k + \xi_{ij} - \zeta_{ij} - Y_i(s_j))$$

$$- \sum_{i=1}^n \sum_{j=1}^m \kappa_{ij} \xi_{ij} - \sum_{i=1}^n \sum_{j=1}^m \tilde{\kappa}_{ij} \zeta_{ij},$$

where $d_{ij}, \kappa_{ij} \geq 0$, and $\tilde{\kappa}_{ij} \geq 0$ are Lagrange multipliers. Setting the derivatives of Λ to zero leads to

$$\frac{\partial}{\partial \boldsymbol{\mu}} : \sum_{i=1}^n \sum_{j=1}^m d_{ij} \mathbf{x}_i = \mathbf{0}, \tag{6}$$

$$\frac{\partial}{\partial \boldsymbol{b}_k} : \boldsymbol{b}_k = \frac{1}{\lambda} \sum_{i=1}^n \sum_{j=1}^m d_{ij} x_{ik} \boldsymbol{\Sigma}^{-1} \mathbf{k}_{s_j}, \tag{7}$$

$$\frac{\partial}{\partial \xi_{ij}} : d_{ij} = \tau - \kappa_{ij}, \tag{8}$$

$$\frac{\partial}{\partial \zeta_{ij}} : d_{ij} = -(1 - \tau) + \tilde{\kappa}_{ij}, \tag{9}$$

and the Karush–Kuhn–Tucker conditions are

$$d_{ij}(\mathbf{x}_i^T \boldsymbol{\mu} + \sum_{k=1}^p x_{ik} \mathbf{k}_{s_j}^T \boldsymbol{b}_k + \xi_{ij} - \zeta_{ij} - Y_i(s_j)) = 0, \tag{10}$$

$$\kappa_{ij} \xi_{ij} = 0, \tag{11}$$

$$\tilde{\kappa}_{ij} \zeta_{ij} = 0. \tag{12}$$

Since κ_{ij} and $\tilde{\kappa}_{ij}$ must be nonnegative, following (8) and (9), we have $d_{ij} \in [-(1 - \tau), \tau]$. Furthermore, we have

$$d_{ij} \in (-(1 - \tau), \tau) \Rightarrow \kappa_{ij}, \tilde{\kappa}_{ij} > 0 \Rightarrow \xi_{ij} = \zeta_{ij} = 0 \Rightarrow Y_i(s_j)$$

$$= \mathbf{x}_i^T \boldsymbol{\mu} + \sum_{k=1}^p x_{ik} \mathbf{k}_{s_j}^T \boldsymbol{b}_k.$$

Define Se as the index set

$$Se = \{(i, j) : d_{ij} \in (-(1 - \tau), \tau), i = 1, \ldots, n, \ j = 1, \ldots, m\}.$$

If $(i, j) \in Se$, we have

$$Y_i(s_j) = \mathbf{x}_i^T \boldsymbol{\mu} + \sum_{k=1}^p x_{ik} \mathbf{k}_{s_j}^T \boldsymbol{b}_k \tag{13}$$

$$= \mathbf{x}_i^T \boldsymbol{\mu} + \frac{1}{\lambda} \sum_{i'=1}^n \sum_{j'=1}^m d_{i'j'} \mathbf{x}_{i'}^T \mathbf{x}_i K(s_{j'}, s_j).$$

If the $\{d_{ij}\}$ are known, then we may solve (4) by obtaining $\boldsymbol{\mu}$ and $\boldsymbol{b}_k$ through (13) and (7). However, it is difficult to obtain d_{ij} in the presence of linear constraints (6).

Next, we consider the Lagrange dual problem where the dual function is defined as $\inf_{\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\kappa}, \tilde{\boldsymbol{\kappa}}} \Lambda$. To simplify the notation, we vectorize our data by letting $\mathbf{y} = (Y_i(s_j)) \in \mathbb{R}^{nm}$, $\boldsymbol{d} = (d_{ij}) \in \mathbb{R}^{nm}$, and $\mathbf{X} = [\mathbf{x}_1; \ldots; \mathbf{x}_n] \in \mathbb{R}^{p \times n}$. With these notations, from (6)–(9), the dual problem can be written as

$$\max_{\boldsymbol{d}} \ -\frac{1}{2\lambda} \boldsymbol{d}^T \mathbf{Q} \boldsymbol{d} + \boldsymbol{d}^T \mathbf{y}, \tag{14}$$

$$\text{s.t. } \boldsymbol{d} \in [-(1 - \tau), \tau]^{nm}, \ \widetilde{\mathbf{X}} \boldsymbol{d} = \mathbf{0},$$

where $\mathbf{Q} = \boldsymbol{\Sigma} \otimes \mathbf{X}^T \mathbf{X} \in \mathbb{R}^{nm \times nm}$ and $\widetilde{\mathbf{X}} = (e \otimes \mathbf{X}^T)^T \in \mathbb{R}^{p \times nm}$ with $e = [1, \ldots, 1]^T \in \mathbb{R}^m$ and $\otimes$ indicating the Kronecker product. The optimization problem (14) is a high-dimensional quadratic programming problem with both box and linear constraints. Because of the linear constraints, we may not be able to solve this problem efficiently. However, these linear constraints can be removed using the primal information

because the primal variable $\boldsymbol{\mu}$ is also a dual variable for the dual problem. Hence, if $\boldsymbol{\mu}$ is known, the dual variable $\boldsymbol{d}$ can be obtained by solving

$$
\begin{aligned}
\max_{\boldsymbol{d}} \quad & -\frac{1}{2\lambda}\boldsymbol{d}^T \mathbf{Q}\boldsymbol{d} + \boldsymbol{d}^T(\mathbf{y} - \widetilde{\mathbf{X}}^T\boldsymbol{\mu}), && (15)\\
\text{s.t.} \quad & \boldsymbol{d} \in [-(1-\tau),\tau]^{nm}.
\end{aligned}
$$

The quadratic program (15) with the box constraints can be solved efficiently with the algorithm introduced in De Angelis, Pardalos, and Toraldo (1997). Therefore, our optimization strategy for (4) is to solve the primal and dual problems alternatively until convergence.

Next, we discuss the effective degrees of freedom of our model and smoothing parameter selection. The divergence

$$
\text{div}(\widehat{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\partial \widehat{Y}_i(s_j)}{\partial Y_i(s_j)}
$$

has been used by many authors (Efron 1986; Meyer and Woodroofe 2000; Koenker 2005; Li, Liu, and Zhu 2007) to estimate the effective dimension for a general modeling procedure. This idea arises from the framework of Stein's unbiased risk estimation theory (Stein 1981). Under the setting of nonparametric regression with homoscedastic normal errors, divergence is an unbiased estimate of the sum of the covariance between individual fitted values and the corresponding observed values. Koenker, Ng, and Portnoy (1994) heuristically argued that, under the one-dimensional nonparametric quantile smoothing spline model, the number of interpolated observations is a plausible estimate for the effective dimension of the fitted model. Li, Liu, and Zhu (2007) proved that, under the one-dimensional nonparametric kernel quantile regression model, the divergence is exactly equal to the number of interpolated observations. In this article, we formally prove that the divergence is exactly the same as the number of interpolated $Y_i(s_j)$'s, thus, justifying its use for the selection of λ.

Theorem 3.1. Let $Se = \{(i,j) : Y_i(s_j) = \widehat{Y}_i(s_j)\}$. For any fixed $\lambda > 0$ and any $Y_i(s_j)$, we have

$$
\text{div}(\widehat{Y}) = |Se|.
$$

The proof is presented in the Appendix. The choice of the smoothing parameter λ is a critical but difficult question. Commonly used criteria for quantile regression include the Schwarz information criterion (SIC) (Schwarz 1978; Koenker, Ng, and Portnoy 1994) and the generalized approximate cross validation (GACV) (Yuan 2006). We adopt the GACV criteria in this article to select λ:

$$
\text{GACV}(\lambda) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} \rho_\tau\left(Y_i(s_j) - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}}(s_j)\right)}{nm - df}, \quad (16)
$$

where df is a measure of the effective dimensionality of the fitted model that can be unbiasedly estimated by the divergence as defined in Theorem 3.1.

3.2. Generalized Least Squares for the Copula

The quantile regression model in (1) only gives the marginal distribution of $Y(s_j)$ given $\mathbf{x}$. To obtain the joint distribution of $(Y(s_1),\ldots,Y(s_m))$ given $\mathbf{x}$, we use the copula model. Let $U_\mathbf{x}(s_j) = F_{(s_j,\mathbf{x})}(Y(s_j))$, $j = 1,\ldots,m$. We characterize the joint distribution of $(U_\mathbf{x}(s_1),\ldots,U_\mathbf{x}(s_m))$ using the Student's t copula model

$$
\begin{aligned}
C_{\boldsymbol{\theta}}\left(u_1,\ldots,u_m \middle| \mathbf{x}\right) &= \mathbb{P}\left(U_\mathbf{x}(s_1) < u_1,\ldots,U_\mathbf{x}(s_m) < u_m \middle| \mathbf{x}\right)\\
&= \boldsymbol{t}_{\varrho,\boldsymbol{\Sigma}_{\boldsymbol{\theta}(\mathbf{x})}}\left(t_\varrho^{-1}(u_1),\ldots,t_\varrho^{-1}(u_m)\right).
\end{aligned}
$$

The covariance function follows the Matérn model (Guttorp and Gneiting 2006; Matérn 2013), which specifies a class of isotropic correlation functions

$$
\text{corr}(t_\varrho^{-1}(U(s_j)), t_\varrho^{-1}(U(s_k))|\mathbf{x}) = M(s_j - s_k), \quad s_j \neq s_k.
$$

For the Matérn model, we can easily verify that

$$
\text{var}\left(t_\varrho^{-1}(U(s_j)) - t_\varrho^{-1}(U(s_k))\right) = 2\varrho/(\varrho-2)(1 - M(h)),
$$

where $h = s_j - s_k$. We denote $N(h) = \{(s_j,s_k) : s_j - s_k = h; j,k = 1,\ldots,m\}$ to be the set of all pairs of locations having lag difference h. Denote $\gamma(h) = 1 - M(h)$. Matheron (1963) defined the earliest unbiased nonparametric variogram estimator of $\gamma(h)$ for a fixed lag h as

$$
2\widehat{\gamma}(h) = \frac{\varrho - 2}{\varrho|N(h)|}\sum_{(s_j,s_k)\in N(h)}\left|t_\varrho^{-1}(U(s_j)) - t_\varrho^{-1}(U(s_k))\right|^2.
$$

The method of generalized least squares determines an estimator $\widehat{\boldsymbol{\theta}}$ by minimizing

$$
G(\boldsymbol{\theta}) = (2\widehat{\gamma} - 2\gamma(\boldsymbol{\theta}))^T \Omega^{-1}(2\widehat{\gamma} - 2\gamma(\boldsymbol{\theta})).
$$

Cressie (1985) studied the method of generalized least squares for variogram fitting in the case where the variance-covariance matrix Ω is diagonal, leading to the method of weighted least squares (WLS). Genton (1998) discussed a more general framework to approximate the variance-covariance matrix and significantly improved the fit. We first discuss how to construct pseudo copula observations. Pseudo copula observations $\widetilde{U}_{\mathbf{x}_i}(s_j)$ can be constructed as $\widehat{F}_{(s_j,\mathbf{x})}(Y_i(s_j))$, for $i = 1,\ldots,n$, $j = 1,\ldots,m$, where $\widehat{F}_j$ is the estimated conditional CDF of $Y_i(s_j)$ given $\mathbf{x}_i$. The information contained in the dual problem is the key to obtain these pseudo copula observations. Let $\tilde{d}_{ij} = d_{ij} + (1-\tau)$ such that $\tilde{d}_{ij}(\tau;\mathbf{x}_i,s_j) : [0,1] \to [0,1]$. This $\tilde{d}_{ij}$ plays a crucial role in connecting the statistical theory of quantile regression to the classical theory of rank tests (Gutenbrunner and Jurecková 1992; Gutenbrunner et al. 1993). In particular, let $\widehat{Y}_i(s_j) = \mathbf{x}_i^T\widehat{\boldsymbol{\beta}}_\tau(s_j)$ be the fitted value. We have

$$
\tilde{d}(\tau;\mathbf{x}_i,s_j) = \begin{cases} 1 & \text{if } Y_i(s_j) > \widehat{Y}_i(s_j),\\ (0,1) & \text{if } Y_i(s_j) = \widehat{Y}_i(s_j),\\ 0 & \text{if } Y_i(s_j) < \widehat{Y}_i(s_j). \end{cases} \quad (17)
$$

The integral $\int_0^1 \tilde{d}(\tau;\mathbf{x}_i,s_j)d\tau$ provides a natural estimate of conditional quantile level of the observed response $Y_i(s_j)$ given $\mathbf{x}_i$, that is

$$
\widetilde{U}_{\mathbf{x}_i}(s_j) = \int_0^1 \tilde{d}(\tau;\mathbf{x}_i,s_j)d\tau, \quad i = 1,\ldots,n \text{ and } j = 1,\ldots,m. \quad (18)
$$

We denote different lags h by h_k for $k = 1,\ldots,K$, and let $g_{\mathbf{x}}(\boldsymbol{\theta}) = (2\widehat{\gamma}_{\mathbf{x}}(h_1) - 2\gamma_{\mathbf{x}}(h_1,\boldsymbol{\theta}),\ldots,2\widehat{\gamma}_{\mathbf{x}}(h_K) - 2\gamma_{\mathbf{x}}(h_K,\boldsymbol{\theta}))^T$. We use Genton's method (Genton 1998) to estimate Ω in the generalized least squares. Let $V_{\mathbf{x}}(\boldsymbol{\theta})$ be the inverse of the dispersion matrix of the sample variogram $\mathbb{E}(g_{\mathbf{x}}(\boldsymbol{\theta})g_{\mathbf{x}}(\boldsymbol{\theta})^T|\mathbf{x})$. We can obtain $\boldsymbol{\theta}$ via:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} g_{\mathbf{x}_i}^T(\boldsymbol{\theta}) V_{\mathbf{x}_i}(\boldsymbol{\theta}) g_{\mathbf{x}_i}(\boldsymbol{\theta}). \tag{19}$$

The asymptotic properties of this estimator have been established in Cressie (1985, 2015) and Lahiri, Lee, and Cressie (2002). We omit the details here.

4. Optimal Rate of Convergence

In this section, we establish the minimax rate of convergence of estimating the coefficient functions. We consider the two different designs introduced previously. For the fixed design, the spatial functional response is observed at the same locations across curves, that is, $m_1 = m_2 = \cdots = m_n := m$ and $s_{1j} = s_{2j} = \cdots = s_{nj} := s_j$ for $j = 1,\ldots,m$. Assume that, as $m \to \infty$, the empirical distribution of s_j's converges to a fixed distribution $\pi(s)$. For the random design, the s_{ij} are independently sampled from a distribution. With an abuse of notation, we also denote it by $\pi(s)$. For any two p-dimensional vector functions $f_1, f_2 \in \mathcal{F}^p$, we define the $\mathbb{L}_2$-distance as

$$\left\| f_1 - f_2 \right\|_{s,2}^2 = \int_{\mathcal{S}} \sum_{k=1}^{p} (f_{1k}(s) - f_{2k}(s))^2 \pi(s) ds.$$

For any fixed quantile level $\tau \in (0,1)$, let $\beta_{\tau k}$ be the true coefficient function. We measure the accuracy of the estimation of $\widehat{\beta}_{\tau k}$ by

$$\mathcal{E}_{\tau}(\widehat{\beta}_{\tau k}, \beta_{\tau k}) = \left\| \widehat{\beta}_{\tau k} - \beta_{\tau k} \right\|_{s,2}^2.$$

The rate of convergence of $\mathcal{E}_{\tau}(\widehat{\beta}_{\tau k}, \beta_{\tau k})$ as the sample size n and the location sampling frequency m increase reflects the difficulty of the estimation problem.

4.1. Minimax Lower Bound

The following result establishes the minimax lower bound for estimating $\boldsymbol{\beta}_{\tau}$ over $\mathcal{F}^p$ under both fixed and random designs. The minimax lower bound is given in the following theorem.

Theorem 4.1. Fix $\tau \in (0,1)$, and suppose the eigenvalues $\{\rho_k : k \geq 1\}$ of the reproducing kernel K satisfies $\rho_k \asymp k^{-2r}$ for some constant $0 < r < \infty$. Then

(a) for the fixed design,

$$\lim_{a_{\tau} \to 0} \lim_{n,m \to \infty} \inf_{\widetilde{\boldsymbol{\beta}}_{\tau} \in \mathcal{F}^p} \sup_{\boldsymbol{\beta}_{\tau} \in \mathcal{F}^p} \mathbb{P}\Big(\mathcal{E}_{\tau}(\widetilde{\boldsymbol{\beta}}_{\tau}, \boldsymbol{\beta}_{\tau}) \geq a_{\tau}(n^{-1} + m^{-2r}) \Big)$$
$$= 1; \tag{20}$$

(b) and for the random design,

$$\lim_{a_{\tau} \to 0} \lim_{n,m \to \infty} \inf_{\widetilde{\boldsymbol{\beta}}_{\tau} \in \mathcal{F}^p} \sup_{\boldsymbol{\beta}_{\tau} \in \mathcal{F}^p} \mathbb{P}\Big(\mathcal{E}_{\tau}(\widetilde{\boldsymbol{\beta}}_{\tau}, \boldsymbol{\beta}_{\tau}) \geq a_{\tau}((nm)^{-\frac{2r}{2r+1}} + n^{-1}) \Big)$$
$$= 1. \tag{21}$$

The above infimums are taken over all possible estimators $\widetilde{\boldsymbol{\beta}}_{\tau}$ based on the training data.

Remark 4.1. The lower bounds established in Theorem 4.1 depend on the decay rate of the eigenvalues of the reproducing kernel K. The lower bounds are different between the fixed design and the random design. For both designs, when the number of locations m is large, it has no effect on the rate of convergence and the optimal rate is of order n^{-1}. On the other hand, a phrase transition phenomenon happens when m is of order $n^{1/2r}$. If m is below this order, the optimal rate for the fixed design is of order m^{-2r} and for the random design is of order $(mn)^{-2r/(2r+1)}$. We may conclude that the random design leads to a better result in terms of the rate of convergence. Similar phenomenon has been studied when estimating the mean of functional data (Cai and Yuan 2012).

Remark 4.2. The constants a_{τ} in (20) and (21) depend on the quantile level τ. If we assume $\tau \in \mathcal{T} \subset (0,1)$ with $\mathcal{T}$ being a compact subset of $(0,1)$, it is possible to extend Theorem 4.1 to hold uniformly for all τ, that is, we may choose a constant a in both (20) and (21) that does not depend on τ. This result can be established by a slight modification of the proof of Theorem 4.1.

4.2. Minimax Upper Bound

In this section, we consider the upper bound for the minimax risk and construct specific rate optimal estimators under both designs. The upper bound shows that the rates given in Theorem 4.1 are sharp. Specifically, we adopt the roughness regularization method to estimate the coefficient function vector $\boldsymbol{\beta}_{\tau} \in \mathcal{F}^p$ by minimizing

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \rho_{\tau}\left(Y_i(s_{ij}) - \mathbf{x}_i^T \boldsymbol{\beta}_{\tau}(s_{ij}) \right) + \lambda \sum_{k=1}^{p} \|\beta_{\tau k}\|_K^2, \tag{22}$$

where $\lambda > 0$ is a tuning parameter balancing fidelity to the data and smoothness of the estimate. Let $\widehat{\boldsymbol{\beta}}_{\tau}$ be the estimate from (22).

We now introduce the following main assumptions, which put some constraints on the conditional CDF of $Y(s)$ given $\mathbf{x}$ and s and the design matrix from the covariates:

A1. Let the CDF of $Y(s)$ given $\mathbf{x}$ and s be $F_{(s,\mathbf{x})}(y) = \mathbb{P}(Y(s) \leq y|\mathbf{x})$. Assume that there exist constants $c_0 > 0$ and $c_1 > 0$ such that for any u satisfying $|u| \leq c_0$, $\left| F_{(s,\mathbf{x})}(u) - F_{(s,\mathbf{x})}(0) \right| \geq c_1 |u|^2$.

A2. Assume $\mathbf{x}$ belongs to a compact subset of $\mathbb{R}^p$ and that the eigenvalues of $\mathbb{E}(\mathbf{xx}^T)$ are bounded below and above by some positive constants c_2 and $1/c_2$, respectively.

Theorem 4.2. Fix $\tau \in (0,1)$. Suppose the eigenvalues $\{\rho_k : k \geq 1\}$ of the reproducing kernel K satisfies $\rho_k \asymp k^{-2r}$ for some constant $0 < r < \infty$. If A1 and A2 hold, then

(a) for the fixed design,

$$\lim_{A_{\tau} \to \infty} \lim_{n,m \to \infty} \sup_{\boldsymbol{\beta}_{\tau} \in \mathcal{F}^p} \mathbb{P}\Big(\mathcal{E}_n(\widehat{\boldsymbol{\beta}}_{\tau}, \boldsymbol{\beta}_{\tau}) \geq A_{\tau}(n^{-1} + m^{-2r}) \Big) = 0 \tag{23}$$

for any $\lambda = O(nm(n^{-1} + m^{-2r}))$;

(b) and for the random design,

$$\lim_{A_\tau \to \infty} \lim_{n,m \to \infty} \sup_{\boldsymbol{\beta}_\tau \in \mathcal{F}^p} \mathbb{P}\left(\mathcal{E}_n(\widehat{\boldsymbol{\beta}}_\tau, \boldsymbol{\beta}_\tau) \geq A_\tau((nm)^{-\frac{2r}{2r+1}} + n^{-1})\right) = 0 \tag{24}$$

for any $\lambda = O(nm(n^{-1} + m^{-2r/(2r+1)}))$.

Remark 4.3. Combining Theorems 4.1 and 4.2 demonstrates that $\widehat{\boldsymbol{\beta}}_\lambda$ is rate-optimal. Similar to Theorem 4.1, the rate of convergence shows both similarities and significant differences between fixed and random designs. In particular, for the random design, the optimal rate is $(mn)^{-2r/(2r+1)}$ when $m = O(n^{1/2r})$ and it is interesting to note that the dependency among $Y_i(s_{ij})$ does not affect the rate of convergence.

It is possible to extend Theorem 4.2 to hold uniformly for all τ in some domain of interest. In the following, we assume that $\tau \in \mathcal{T}$, with $\mathcal{T}$ being a compact subset of $(0, 1)$.

Corollary 4.1. Assume $\tau \in \mathcal{T}$, and assume the same conditions of Theorem 4.2.

(a) For the fixed design, we have

$$\sup_{\tau \in \mathcal{T}} \left\| \widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau \right\|_{s,2}^2 = O_p\left(n^{-1} + m^{-2r}\right) \tag{25}$$

for any $\lambda = O(nm(n^{-1} + m^{-2r}))$;

(b) For the random design, we have

$$\sup_{\tau \in \mathcal{T}} \left\| \widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau \right\|_{s,2}^2 = O_p\left((nm)^{-\frac{2r}{2r+1}} + n^{-1}\right) \tag{26}$$

for any $\lambda = O(nm(n^{-1} + m^{-2r/(2r+1)}))$.

5. Numerical Analysis

In the following section, we evaluate the finite sample performance of our method by using both simulated and real datasets, and compare with other spatial quantile regression methods.

5.1. Algorithm Complexity

We first study the computational complexity of the proposed primal-dual algorithm for SQR. As a comparison, we developed another optimization algorithm based on the popular ADMM (Boyd et al. 2011). Please refer to Supplement III for more details on the ADMM algorithm. The iterative ADMM algorithm divides the optimization problem in (4) into multiple subproblems, each of which has an explicit solution. In the high-dimensional scenario, however, explicitly solving each step is not computationally easy since it requires the multiplication and inversion of large matrices. Another disadvantage of using ADMM is its convergence rate, which can be poor when high accuracy is desired (Boyd et al. 2011). In contrast, the proposed primal-dual algorithm converges very quick and does not require matrix inversion. We implemented and compared both algorithms in MATLAB on a MacBook Pro with a 2.5 GHz Intel Core i7 CPU and 16 GB of RAM. In addition, an Bayesian spatial quantile regression (BSQR) method proposed in Reich, Fuentes, and Dunson (2011) was implemented in R and compared with our method. In our implementation, 5000 MCMC samples were

drawn and the final results were summarized based on a burn-in of first 1000 samples.

Simulation datasets were generated from the following model:

$$Y_i(s_j) = x_{i1}\beta_{\tau 1}(s_j) + x_{i2}\beta_{\tau 2}(s_j) + x_{i3}\beta_{\tau 3}(s_j) + \eta_i(s_j, \tau),$$

for $i = 1, \ldots, n$; $j = 1, \ldots, m$, where $x_{i1} = 1$, $x_{i2} \sim$ Bernoulli(0.5), $x_{i3} \sim$ Uniform(0, 1), and $\{s_j\}$ are evenly sampled on $[0, 1]$. We also set $\eta_i(s_j) = v_i(s_j) + \epsilon_i(s_j)$, where $\epsilon_i(s_j) \sim N(0, 0.1)$, and $(v_i(s_1), \ldots, v_i(s_m))$ follows a multivariate normal distribution with zero mean and a covariance matrix taking the form $\text{cov}(v_i(s_j), v_i(s_l)) = a * \exp(-((s_j - s_l)/h)^2)$ with $a = 0.6$ and $h = 0.8$. We first construct $\eta_i(s_1), \ldots, \eta_i(s_m)$ for each i, and then let $\eta_i(s_j, \tau) = \eta_i(s_j) - F^{-1}(\tau)$ with F being the marginal density function of $\eta_i(s_j)$ to make the τth quantile of $\eta_i(s_j, \tau)$ zero for identifiability. We set $\beta_{\tau 1}(s) = 5s^2$, $\beta_{\tau 2}(s) = 2(1 - s)^4$, and $\beta_{\tau 3} = 2 + 20 \sin(6s) + 2s^3$ for $s \in [0, 1]$. A Gaussian kernel with $\sigma = 0.2$ is used for $\mathcal{H}(K)$. In all simulations, we fix the smoothing parameter $\lambda = 1$.

To compare the computational complexity, we generated different datasets with a combination of $n = \{50, 100, 200, 500\}$ and $m = \{50, 100, 1000\}$. For each simulated dataset, we ran each algorithm for 100 times and computed the averages and standard errors of the elapsed CPU times. Table 1 summarizes the results at $\tau = 0.1, 0.5$, and 0.8. Note that we did not directly compare with the BSQR in Table 1, which as a Bayesian method estimates $\boldsymbol{\beta}_\tau$ at L quantile levels simultaneously. Using the same computer, it took BSQR about 1860 sec to get 5000 MCMC samples when $m = 50$ and $n = 50$. The computational time significantly increases as m increases (i.e., >3000 sec when $m = 100$).

To compare the estimation precision, we use the root mean integrated squared error (RMISE), which is defined as

$$\text{RMISE}_\tau = \left(m^{-1} \sum_{j=1}^m (\widehat{\beta}_{\tau k}(s_j) - \beta_{\tau k}(s_j))^2\right)^{1/2} \quad \text{for } k = 1, 2, 3.$$

Table 2 reports RMISE, averaged over 100 runs, and its standard deviation (SD). Table 1.1 in Supplement II reports RMISEs in a more noisy scenario.

The results in Tables 1 and 2 clearly indicate that the primal-dual algorithm outperforms the ADMM algorithm and BSQR in both computational efficiency and accuracy. Also the proposed algorithm is robust to high noise based on the results in Table 1.1 in Supplement II. When the sample size n and dimension of data m are large, it takes an extremely long time for ADMM to obtain a solution, whereas the primal-dual algorithm is much more efficient. Such computational efficiency is particularly important for our high-dimensional neuroimaging applications. Moreover, given a fixed m and an increasing number of the observations n, the estimated $\boldsymbol{\beta}$ gets better, following our theoretical results. Experiments with the random design also produce similar results. The computational bottleneck of our current implementation is the quadratic programming with box constraints, which can be improved by using C++ or parallel computing.

**Table 1.** Elapsed CPU times of ADMM and primal-dual algorithms.

	Primal-dual				ADMM		
(n, m)	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.8$		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.8$
(50,50)	$0.042_{(0.009)}$	$0.045_{(0.008)}$	$0.059_{(0.065)}$		$18.5_{(0.221)}$	$18.1_{(0.595)}$	$19.10_{(0.468)}$
(50,100)	$0.130_{(0.008)}$	$0.152_{(0.007)}$	$0.131_{(0.012)}$		$41.1_{(0.409)}$	$41.41_{(0.433)}$	$41.97_{(0.646)}$
(100,50)	$0.133_{(0.007)}$	$0.148_{(0.007)}$	$0.126_{(0.005)}$		$39.01_{(0.920)}$	$38.98_{(0.909)}$	$38.98_{(0.908)}$
(100,100)	$0.293_{(0.037)}$	$0.400_{(0.089)}$	$0.358_{(0.068)}$		$89.57_{(0.273)}$	$89.61_{(0.292)}$	$89.61_{(0.292)}$
(200,50)	$0.295_{(0.036)}$	$0.450_{(0.088)}$	$0.330_{(0.043)}$		$82.49_{(0.356)}$	$82.51_{(0.340)}$	$82.45_{(0.339)}$
(200,100)	$0.703_{(0.046)}$	$0.849_{(0.095)}$	$0.737_{(0.096)}$		$178.80_{(323)}$	$179.45_{(0.400)}$	$179.20_{(0.399)}$
(500,50)	$1.10_{(0.076)}$	$1.24_{(0.137)}$	$1.07_{(0.058)}$		$201.62_{(0.211)}$	$209.07_{(0.349)}$	$200.07_{(0.349)}$
(500,100)	$2.71_{(0.230)}$	$3.11_{(0.544)}$	$2.92_{(0.562)}$		$440.60_{(0.490)}$	$444.49_{(0.552)}$	$443.40_{(0.553)}$
(500,1000)	$120_{(1.36)}$	$118_{(0.798)}$	$119_{(0.570)}$		*	*	*

NOTE: Mean and standard deviation (in bracket) are displayed in each entry.
  * Missing value due to significant computing time. The unit is second.

Table 2. Comparison of the RMISE of the estimators obtained from different methods.

	Primal-dual ($\tau = 0.5$)			ADMM ($\tau = 0.5$)			BSQR ($\tau = 0.5$)		
(n, m)	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
(50,50)	$0.065_{(0.034)}$	$0.092_{(0.041)}$	$0.272_{(0.069)}$	$0.282_{(0.071)}$	$0.210_{(0.067)}$	$0.560_{(0.112)}$	$0.482_{(0.097)}$	$0.340_{(0.081)}$	$0.584_{(0.191)}$
(50,100)	$0.042_{(0.027)}$	$0.054_{(0.030)}$	$0.150_{(0.056)}$	$0.078_{(0.034)}$	$0.148_{(0.046)}$	$0.212_{(0.069)}$	$0.206_{(0.141)}$	$0.342_{(0.163)}$	$0.542_{(0.123)}$
(100,50)	$0.053_{(0.035)}$	$0.063_{(0.034)}$	$0.189_{(0.056)}$	$0.105_{(0.035)}$	$0.205_{(0.065)}$	$0.311_{(0.064)}$	$0.235_{(0.055)}$	$0.200_{(0.144)}$	$0.337_{(0.123)}$
(100,100)	$0.028_{(0.019)}$	$0.025_{(0.017)}$	$0.087_{(0.032)}$	$0.033_{(0.019)}$	$0.136_{(0.045)}$	$0.151_{(0.065)}$	$0.231_{(0.024)}$	$0.157_{(0.072)}$	$0.354_{(0.143)}$
(200,50)	$0.039_{(0.021)}$	$0.029_{(0.018)}$	$0.110_{(0.031)}$	$0.067_{(0.033)}$	$0.164_{(0.047)}$	$0.227_{(0.053)}$	$0.231_{(0.097)}$	$0.128_{(0.056)}$	$0.310_{(0.091)}$
(200,100)	$0.029_{(0.013)}$	$0.018_{(0.010)}$	$0.079_{(0.020)}$	$0.025_{(0.020)}$	$0.122_{(0.053)}$	$0.082_{(0.042)}$	$0.137_{(0.025)}$	$0.069_{(0.015)}$	$0.251_{(0.025)}$
(500,50)	$0.036_{(0.012)}$	$0.019_{(0.010)}$	$0.090_{(0.019)}$	$0.028_{(0.019)}$	$0.166_{(0.057)}$	$0.083_{(0.043)}$	$0.077_{(0.039)}$	$0.083_{(0.009)}$	$0.157_{(0.041)}$
(500,100)	$0.029_{(0.007)}$	$0.018_{(0.007)}$	$0.078_{(0.011)}$	$0.020_{(0.018)}$	$0.097_{(0.029)}$	$0.062_{(0.035)}$	$0.073_{(0.017)}$	$0.098_{(0.007)}$	$0.143_{(0.073)}$

NOTE: Mean and standard deviation (in bracket) are displayed in each entry.

(a) Simulated β^μ's   (b) Simulated β^σ's   (c) Simulated Y under the copula model
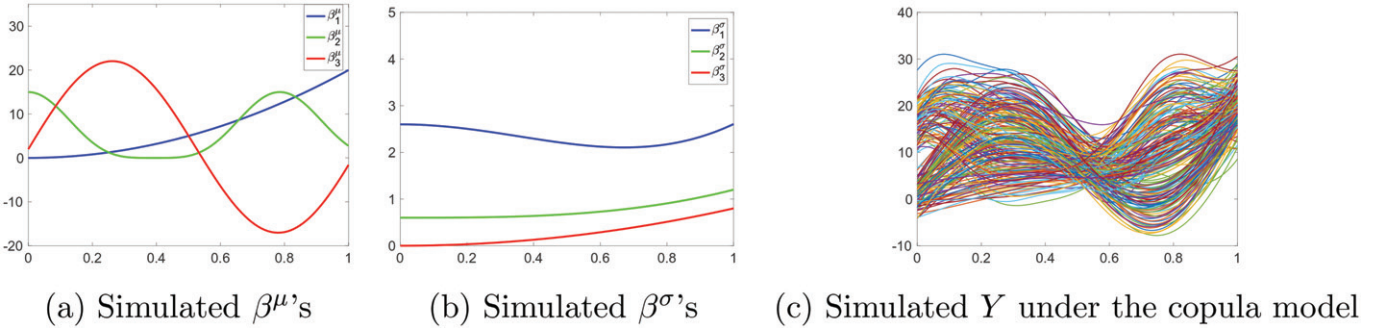
Figure 1. β^μ, β^δ and simulated Y in the copula model.

5.2. Simulation Study on Copula Model

In this section, we evaluate the finite sample performance of the proposed copula model in capturing the joint distribution of $\{Y(s_j)\}$ given $\mathbf{x}$. Similar to Section 5.1, we simulate $\mathbf{x} = [x_1, x_2, x_3]^T \in \mathbb{R}^3$ according to $x_1 = 1, x_2 \sim$ Bernoulli(0.5), and $x_3 \sim$ Uniform(0, 1). To adapt the copula model, we simulate the response image Y using a different procedure. We let $Y(s)$ conditional on $\mathbf{x}$ be a normal distribution with mean $\mu_{\mathbf{x}}(s)$ and variance $\sigma_{\mathbf{x}}^2(s)$ and the quantile of $Y(s)$ at different s follow a Gaussian process. More specifically, the τth quantile of $Y(s_j)$ is given as

$$F_{(s_j, \mathbf{x})}^{-1}(\tau) = \mu_{\mathbf{x}}(s_j) + \sigma_{\mathbf{x}}(s_j)\Phi^{-1}(\tau).$$

Since in our mode we assume linear relations, both $\mu_{\mathbf{x}}(s_j)$ and $\sigma_{\mathbf{x}}(s_j)$ should be linear functions of $\mathbf{x}$ in the simulation. We let $\mu_{\mathbf{x}}(s_j) = x_1\beta_1^\mu(s_j) + x_2\beta_2^\mu(s_j) + x_3\beta_3^\mu(s_j)$ and $\sigma_{\mathbf{x}}(s_j) = x_1\beta_1^\sigma(s_j) + x_2\beta_2^\sigma(s_j) + x_3\beta_3^\sigma(s_j)$. That is, we write the τth

quantile of $Y(s_j)$ as $F_{(s_j, \mathbf{x})}^{-1}(\tau) = \mathbf{x}^T \boldsymbol{\beta}_\tau(s_j)$, where $\boldsymbol{\beta}_\tau(s) = (\beta_1^\mu + \beta_1^\sigma \Phi^{-1}(\tau), \ldots, \beta_3^\mu + \beta_3^\sigma \Phi^{-1}(\tau))^T$. Figure 1 plots the simulated $\boldsymbol{\beta}$'s. To simulate $Y(\cdot)$ with spatial dependence, we use the introduced copula model. We let $\Phi^{-1}(U(s))$ with $U(s) = F_{(s, \mathbf{x})}(Y(s))$ be a Gaussian process generated from the Matérn family

$$(\Phi^{-1}(U(s_1)), \ldots, \Phi^{-1}(U(s_m)))|\mathbf{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{(\mathbf{x}, \boldsymbol{\alpha})}),$$

with $\boldsymbol{\Sigma}_{(\mathbf{x}, \boldsymbol{\alpha})}(s, s) = \sigma_{(\mathbf{x}, \boldsymbol{\alpha})}^2(s) = (x_1\beta_1^\sigma(s) + x_2\beta_2^\sigma(s) + x_3\beta_3^\sigma(s))^2$, $\boldsymbol{\Sigma}_{(\mathbf{x}, \boldsymbol{\alpha})}(s_i, s_j) = \sigma_{(\mathbf{x}, \boldsymbol{\alpha})}(s_i)\sigma_{(\mathbf{x}, \boldsymbol{\alpha})}(s_j) M(|s_i - s_j|)$, and $M(h) = 2^{1-\nu} \Gamma(\nu)(\exp(\boldsymbol{\alpha}^T\mathbf{x})\|h\|)^\nu K_\nu(\exp(\boldsymbol{\alpha}^T\mathbf{x})\|h\|)$. We set $\nu = 5/2$, $\boldsymbol{\alpha} = (0.8, 0.8, 0.8)^T$. We now can simulate $(Y(s_1), \ldots, Y(s_m))$ from this model. Figure 1(c) shows 200 simulated Y's.

We generated 200 training observations from the above model, along with 500 observations for validation and 500 for testing. We used a fixed Gaussian kernel with $\sigma = 0.2$ for $\mathcal{H}(K)$. First, we evaluate the λ selection criterion. Specifically,
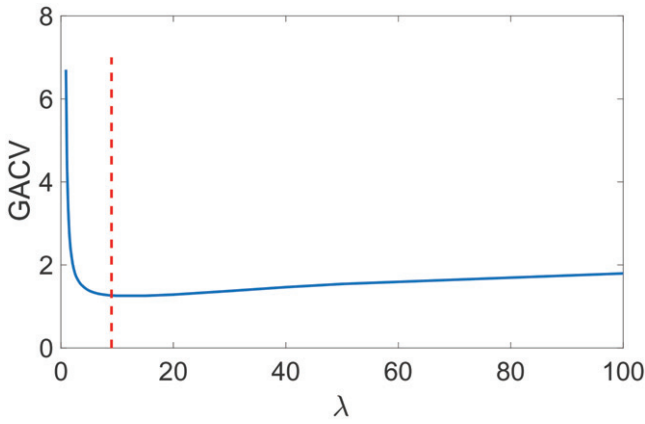
Figure 2. GACV values versus λ.

we compare GACV with a gold standard method that minimizes the sum of RMISEs (sRMISE) defined as

$$\text{sRMISE}(\tau) = \sum_{k=1}^{p} \left(m^{-1} \sum_{j=1}^{m} (\widehat{\beta}_{\tau k}(s_j) - \beta_{\tau k}(s_j))^2 \right)^{1/2}.$$

Figure 2 shows the selected λ versus the corresponding GACV value. We then compare parameter estimation accuracy under GACV and the gold standard. Table 3 shows the $\mathbb{L}_2$ distance between $\beta_{\tau i}$ and $\widehat{\beta}_{\tau i}$ across different τ values. It is not surprising that the gold standard is better than GACV since we used the ground truth information in the gold standard method. However, we also can observe that GACV has very similar performances as the gold standard method, especially at $\tau = 0.5$, where we have more data available for parameter estimation in this simulation. The smoothness level goes up from β_1 to β_3 in our simulation. We observe that GACV tends to select a simpler model than does the gold standard—that is the smoothest β_1 is best estimated with GACV.

Next, we consider the copula model. Based on a randomly selected observation pair $(Y_i, \mathbf{x}_i = (1, 1, 0.2926)^T)$, we obtained the following results in Figure 3: (a) the ground truth marginal distribution given the $\mathbf{x}_i$ (different quantiles are plotted with different colors), (b) the estimated quantiles based on the λ selected by the gold standard method, and (c) the estimated quantiles based on the GACV. The red curve in (a)–(c) shows Y_i. In Figure 3(d), we show the quantile functions $U(s)$ of the observation Y_i in different scenarios (distributions) of (a),(b) and (c).

To quantitatively evaluate the estimated quantile functions $U(s)$ in different λ selection scenarios, we calculate its $\mathbb{L}^2$ distance to the ground truth quantile function. In average, we got $\|U(s) - \widetilde{U(s)}\| = 0.081(\pm 0.0351)$ and $0.101(\pm 0.0398)$ for GACV and gold standard methods, respectively. This result indicates that in terms of recovering the quantile function $U(s)$, GACV is not worse than the gold standard. GACV is used in all of the following experiments.

We then compare different couple models on fitting the bivariate empirical observation pair $(\hat{F}_{(s_i,\mathbf{x})}(y_i), \hat{F}_{(s_j,\mathbf{x})}(y_j))$ with different Δs ($\Delta s = \|s_j - s_i\|$). Using the R package "VineCopula" (Nikoloulopoulos, Joe, and Li 2012), we compared 21 different copula models, including the Gaussian copula and Student's t copula. The results are shown in Figure 2.2 in Supplement II. Although the pair $(F_{(s_i,\mathbf{x})}(y_i), F_{(s_j,\mathbf{x})}(y_j))$ is simulated from Gaussian, t-distribution fits the empirical data better in most of the cases, justifying our choice of t-copula. Note that we obtained similar conclusion with the real data (refer to Figure 2.1 in Supplement II).

We also evaluate the generalized least square algorithm for estimating parameters in the t-copula model, whose degree of free parameter ϱ was estimated based on $(\hat{F}_{(s_i,\mathbf{x})}(y_i), \hat{F}_{(s_j,\mathbf{x})}(y_j))$ using the R package "VineCopula." We fixed $\nu = 5/2$ in all experiments. Figure 4(a) shows the 20 ground truth functions of $\Phi^{-1}(U(s))$, and (b) shows the corresponding estimated $\widehat{\Phi}^{-1}(U(s))$. Moreover, some flat regions in $\widehat{\Phi}^{-1}(U(s))$

Table 3. Comparison of different methods on selecting λ, evaluated through $\|\beta_i - \widehat{\beta}_i\|_0$.

($n = 200$,	GACV			Gold standard		
$m = 100$)	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.8$
β_1	$0.139_{(0.050)}$	$0.045_{(0.035)}$	$0.061_{(0.039)}$	$0.095_{(0.061)}$	$0.054_{(0.042)}$	$0.054_{(0.051)}$
β_2	$0.233_{(0.058)}$	$0.143_{(0.043)}$	$0.199_{(0.047)}$	$0.101_{(0.059)}$	$0.073_{(0.052)}$	$0.041_{(0.041)}$
β_3	$0.191_{(0.075)}$	$0.242_{(0.069)}$	$0.302_{(0.065)}$	$0.159_{(0.093)}$	$0.155_{(0.092)}$	$0.162_{(0.091)}$

NOTE: The mean and standard deviation (in bracket) are displayed in each entry.

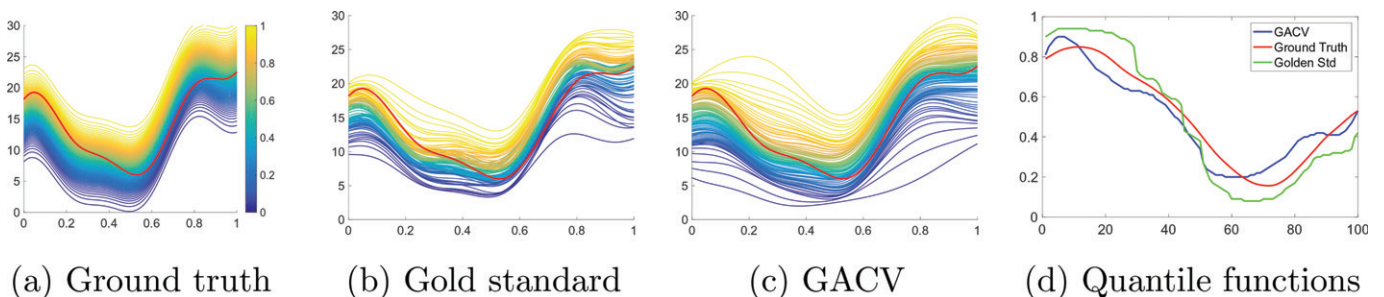(a) Ground truth   (b) Gold standard   (c) GACV   (d) Quantile functions

Figure 3. Comparison of marginal density estimation under different λ selection criteria. Given an observation $(Y_i, \mathbf{x}_i = (1, 1, 0.2926)^T)$ (Y_i is plotted in the red curve in panels (a)–(c)), panels (a)–(c) show the marginal distribution of $Y|\mathbf{x}_i$ with the ground truth parameters, parameters estimated based the gold standard method, and parameters based on GACV, respectively. Panel (d) shows the quantile function $U(s)$ of the $Y_i(s)$ in the distributions in panels (a)–(c).

(a) $\Phi^{-1}(U(s))$ (b) $\widehat{\Phi}^{-1}(U(s))$ (c) truth cov   (d) estimated cov
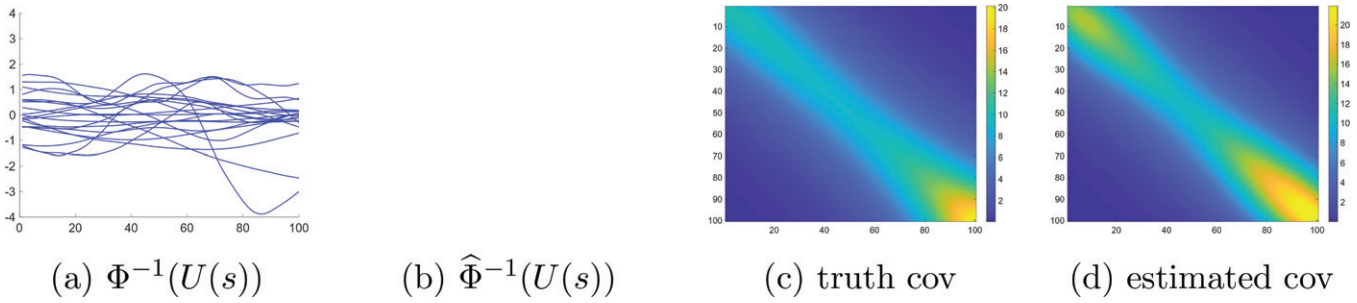
Figure 4. Parameters estimation in the copula model. Panels (a) and (b) show some ground truth quantile functions $\Phi^{-1}(U(s))$ and the corresponding estimated $\widehat{\Phi}^{-1}(U(s))$, respectively. The blue lines mark different quantile ranges for calculating difference between $\Phi^{-1}(U(s))$ and $\widehat{\Phi}^{-1}(U(s))$. Panels (c) and (d) show the ground truth covariance and the estimated covariance, respectively.

(a) Samples using ground truth parameters   (b) Samples using estimated parameters
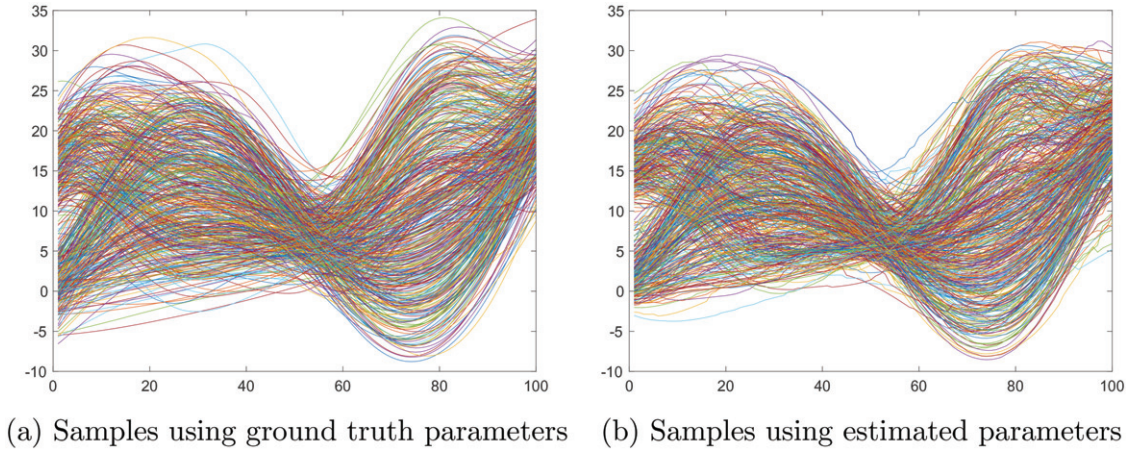
Figure 5. Examples of simulating $\{Y_i\}$ with given $\{x_i\}$ in the simulated test dataset.

are caused by the estimation precision because we only estimated quantiles from 1 to 99. The average point-wise distances between $\Phi^{-1}(U(s))$ and $\widehat{\Phi}^{-1}(U(s))$ shown in Figure 4(b) show that the estimates are better in the middle than in the tail. We then calculated $t_\varrho^{-1}(U(s))$, and estimate $\boldsymbol{\Sigma}_{\boldsymbol{\theta}(\mathbf{x})}$ in our copula model. For $\mathbf{x} = (1, 1, 0.8507)^T$, Figures 4(c) and (d) show the ground truth covariance function and the estimated covariance function $\varrho/(\varrho-2)\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}(\mathbf{x})}$, respectively. The estimation algorithm can work very well even when the domain is irregular (see additional simulation results in the supplementary materials).

Next, we can obtain the joint distribution of $Y|\mathbf{x}$. Using $\{\mathbf{x}_i\}$ ($i = 1, \ldots, 500$) in the test dataset, we simulated Y_i. In Figure 5, (a) shows the randomly sampled $\{Y_i\}$ using the ground truth parameters, and (b) shows the simulated Y_i using estimated parameters based on GACV which we can see follow similar patterns of $\{Y_i\}$ sampled using the ground truth parameters.

5.3. Real Data Analysis

5.3.1. ADNI DTI Data

We apply the proposed method to analyze the diffusion tensor imaging (DTI) data in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. In 2003, the ADNI was started by National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, and some private pharmaceutical companies and nonprofit organizations. This multisite study assesses clinical, imaging,

genetic and biospecimen biomarkers through the process of normal aging to early mild cognitive impairment, to late mild cognitive impairment, to dementia or Alzheimer's disease (AD). Participants were recruited across North America to participant in the project and a variety of imaging and clinical assessments were conducted for each participant. Results were shared by ADNI through the Laboratory of Neuro Imaging's Image Data Archive (*https://ida.loni.usc.edu/*).

In our study, 203 subjects' diffusion weighted MRI and demographic data were downloaded and processed. DTI data for each subject were extracted using two steps. First, estimate a diffusion tensor at each voxel using a weighted least square estimation (Koay et al. 2006; Zhu et al. 2007). Second, register DTI images from multiple subjects using the FSL TBSS pipeline (Smith et al. 2006) to create a mean image and a mean skeleton. To be more specific, after estimating the diffusion tensor, fractional anisotropy (FA), a scalar measure of the degree of anisotropy, was calculated for each voxel. Next, FA maps of all subjects were fed into the TBSS tool in the FSL software and were aligned non-linearly. The mean FA image was then calculated and thinned to obtain a mean FA skeleton representing the centers of all white matter tracts common to the group. Subsequently, each aligned FA data were projected onto this skeleton. We focus on the midsagittal corpus callosum skeleton and the associated FA curves from all subjects. The corpus callosum is the largest fiber bundle in the human brain and is responsible for much of the communication between the two cerebral hemispheres. Figure 6(a) shows the FA curves from all 203 subjects.
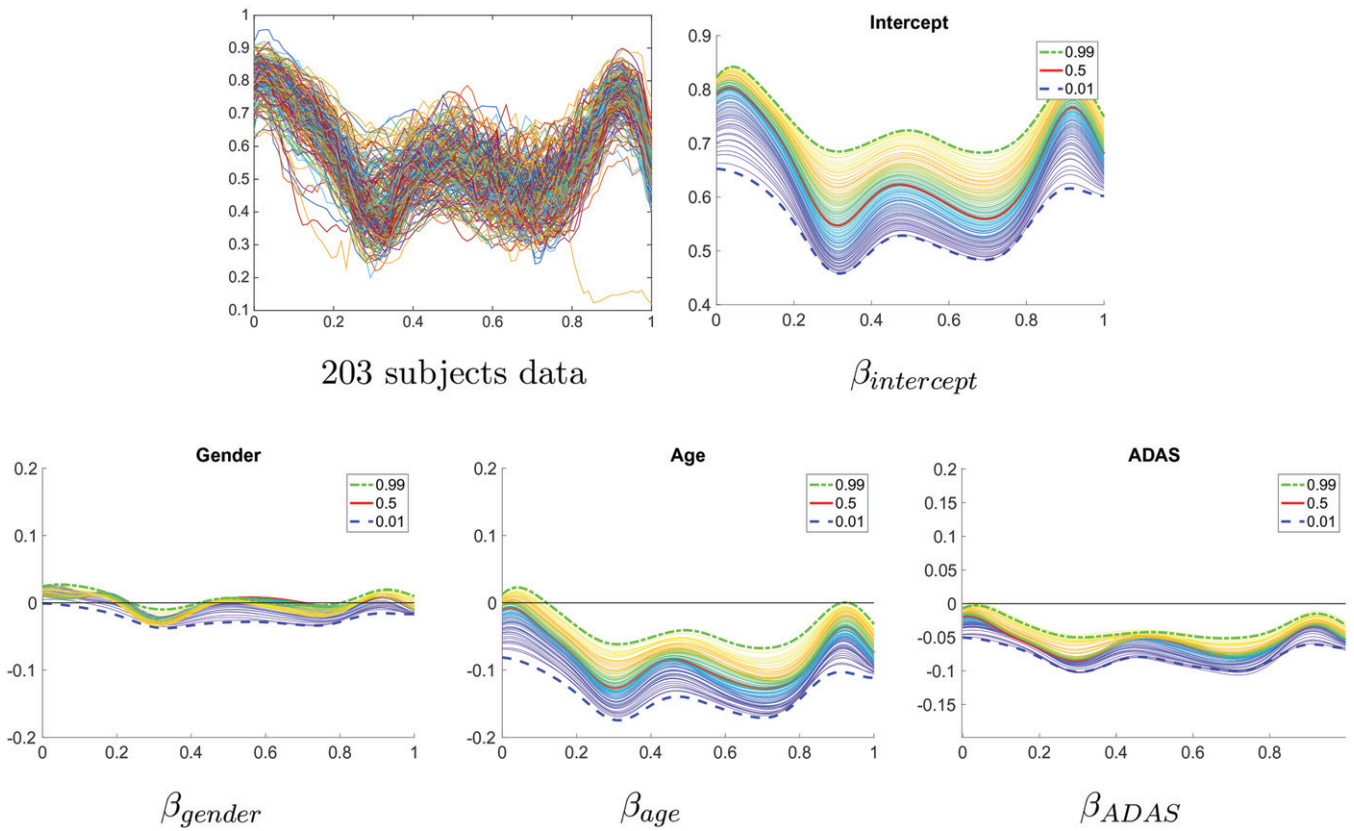
Figure 6. Our SQR results on corpus callosum FA curves from ADNI data. The top left panel shows the original FA curves of 203 subjects in ADNI. The remaining panels show our estimated $\boldsymbol{\beta}_\tau$ at different τ (τ increases from blue to yellow).
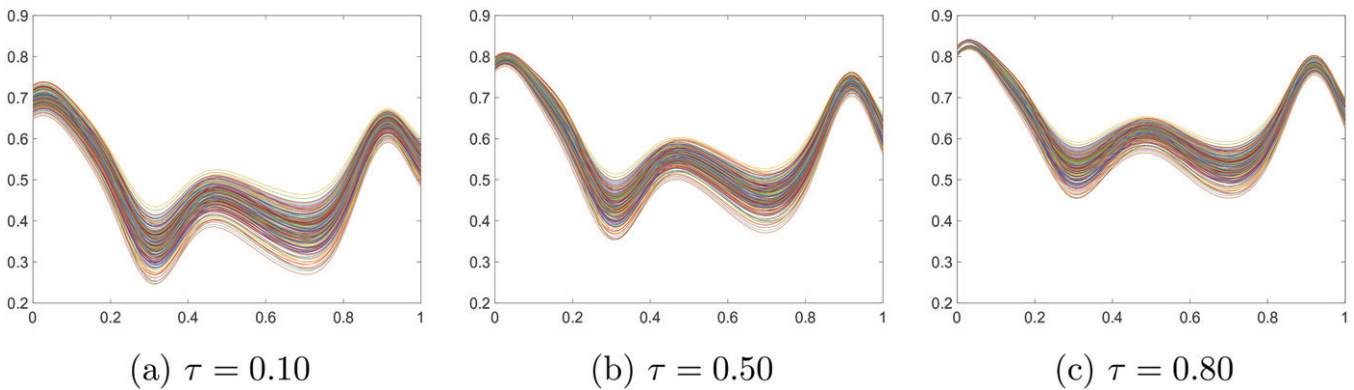

Figure 7. Predicted FA curves $\mathbf{x}_i^T \boldsymbol{\beta}$ at different τ given $\mathbf{x}_i$ in ADNI DTI dataset.

We are interested in exploring the effects of gender, age, and Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS) (Rosen, Mohs, and Davis 1984) score on the diffusion properties along corpus callosum tract at different quantile levels. Specifically, we have $\mathbf{x}_i = (1, \text{gender}, \text{age}, \text{ADAS})^T$, and Y_i representing an FA curve. For gender, males are set as 0 and females are set as 1. To make comparisons with Reich, Fuentes, and Dunson (2011), we standardized both age and ADAS to be in $(0, 1)$. The fitted coefficient functions $\hat{\boldsymbol{\beta}}_\tau$ are shown in Figure 6 at different quantile levels ($\tau = 0.01, 0.02, \ldots, 0.99$). Figure 7 shows the predicted FA curves $\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\tau$ across $\tau = 0.1, 0.5$, and 0.8. It is well-known that aging deteriorates brain structure (Wyss-Coray 2016). The ADAS is widely used to detect cognitive deficits in people suffering from AD. The range of

ADAS in our dataset is $(0, 51)$ with higher scores indicating greater degrees of cognitive deficit. From Figure 6, it follows that both aging and ADAS have a negative effect on the diffusivity on the midsagittal corpus callosum skeleton.

Across different quantiles, we observe a nice layout of $\beta_{\text{intercept}}$—the median is in the middle between quantile 1 and quantile 99. We also observe some interesting structures in the coefficient functions. From β_{gender}, we see that there is not much difference between males and females at quantiles ranging from 10 to 99. However, if we consider lower quantiles (e.g., from 1 to 10), the FA for females along corpus callosum skeleton has smaller values compared with those for males. Biologically, this indicates in the population that the lower FA values along the midsagittal corpus callcosum in females are smaller than
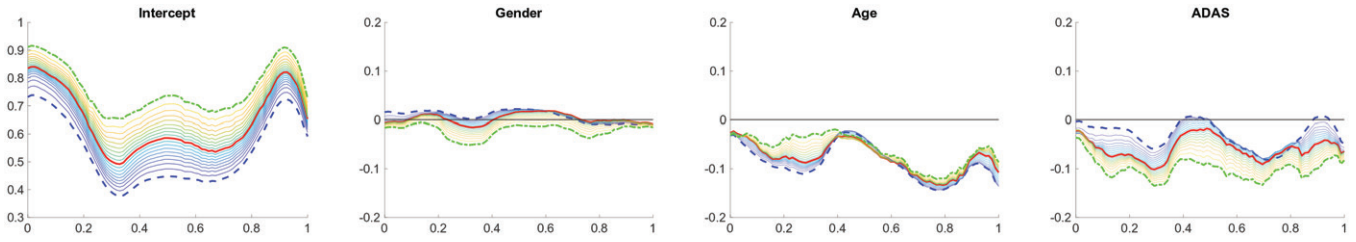
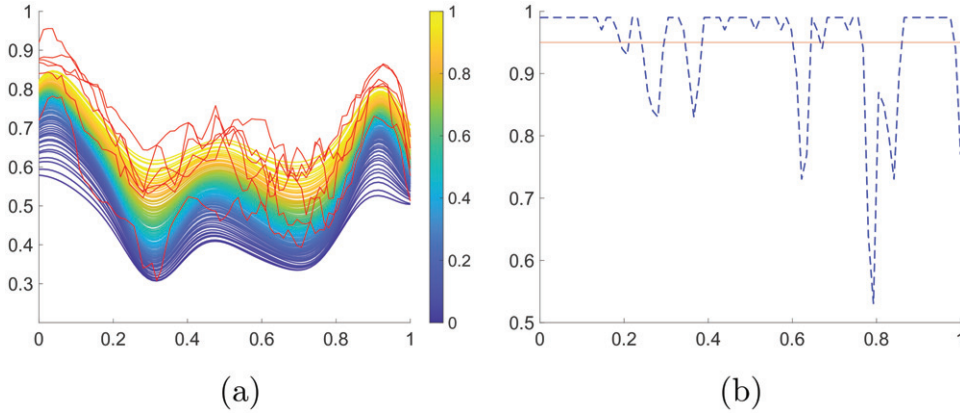**Figure 8.** BSQR results on on corpus callosum FA curves from ADNI data.

Figure 9. Illustration of marginal distribution of $Y|\mathbf{x}$ and statistical inference using p-values. (a) Marginal distribution of $Y(s_j)$ given $\mathbf{x} = (1, \text{male}, \text{age} = 60, \text{ADAS} = 51)$. Red curves show Y_i of subjects having much smaller ADAS scores (<10). (b) The p-value curves of one subject in (a) with ADAS = 8. The solid horizontal line represents $p = 0.95$.

the lower FA values in males, but their means might not have much difference (Inano et al. 2011). We utilized the package FADTTS (Zhu et al. 2011) to perform a mean regression and the results are presented in Supplement II, where we can see that the coefficient function for gender fluctuates around 0. From β_{age}, we see that for people with worse FA values (at lower quantiles), the same amount of aging can contribute to more FA reduction (indicating worse white matter deterioration (Kochunov et al. 2007)) than those at high quantiles. The mean regression results from FADTTS cannot give this information. For β_{ADAS}, the deterioration of Alzheimer's Disease (measured by ADAS) is more similar across different quantiles, acting differently compared with age.

As another comparison, we ran the BSQR model (Reich, Fuentes, and Dunson 2011) on the DTI data. Figure 8 shows the results at 20 different quantile levels. These results are based on 5000 MCMC runs after burn-in of the first 1000 samples. In a laptop with a 2.3 GHz Intel i9 CPU and 32 GB memory, it took about 1400 sec to get $\boldsymbol{\beta}_\tau$ at 20 quantile levels with the BSQR, and 13 sec to get $\boldsymbol{\beta}_\tau$ at 99 quantile levels with our method. BSQR contains a two-stage approach to approximate $\boldsymbol{\beta}_\tau$. First, independent quantile regressions at different locations are done to obtain estimates of the quantile process and their asymptotic covariance. Next, a Gaussian process model is fitted based on the initial estimates of $\boldsymbol{\beta}_\tau(s)$ to introduce spatial dependence structure. As a consequence of this two-stage solution, we see from Figure 8 that the estimated $\boldsymbol{\beta}_\tau$ can be rough and sub-optimal.

Next we show how to use SQR to make statistical inference. First, we selected one male patient with age 60 and ADAS

51 and calculated FA curves at different quantiles along the midsagittal corpus callosum skeleton with the fitted model. Figure 9(a) shows the marginal distribution $Y(s_j)$ given $\mathbf{x} = (1, \text{male}, \text{age} = 60, \text{ADAS} = 51)$. In addition, we plot a subset of subjects having much smaller ADAS scores (<10) in red in Figure 9(a). From this plot, we see that, in general, people with smaller ADAS scores have higher FA values than the one with ADAS = 51. FA represents white matter integrity. A smaller FA along the midsagittal corpus callosum indicates deteriorations of the corpus callosum fiber bundle, and thus weakened bilateral communications. Our finding is consistent with the literature (Biegon et al. 1994; Ardekani et al. 2014). Second, we carried out statistical inference using p-values. Figure 9(b) shows the p-value curve of the subject with ADAS = 8 under the estimated marginal distribution of the subject with ADAS = 51. In most locations, the FA value differs significantly from the two subjects ($1 - p < 0.05$). Third, Figure 10 shows more results obtained from the copula model, where panels (a) and (b) show $\varrho/(\varrho - 2)\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}(\mathbf{x}_1)}$ and $\varrho/(\varrho - 2)\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}(\mathbf{x}_2)}$ for $\mathbf{x}_1 = (\text{male}, \text{age} = 61, \text{ADAS} = 5)$ and $\mathbf{x}_2 = (\text{male}, \text{age} = 60, \text{ADAS} = 51)$, respectively, panel (c) shows their difference, and panel (d) shows random samples simulated from the distributions of $Y|\mathbf{x}_1$ (blue) and $Y|\mathbf{x}_2$ (red). The covariance matrices $\varrho/(\varrho - 2)\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}(\mathbf{x})}$ and their difference in panels (a)–(c) reveal that subjects with large ADAS values have longer range correlation between $t_\varrho^{-1}(u_i)$ and $t_\varrho^{-1}(u_j)$ at certain locations, indicating that they have more smoothed FA curves compared with those with small ADAS values. Moreover, the results in panel (d) reveals that healthy subjects in general have higher FA values than AD subjects with the same age and gender.
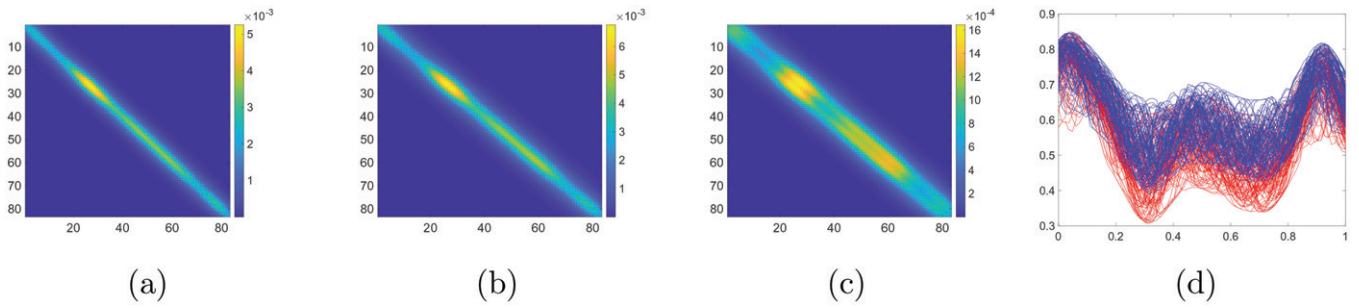
Figure 10. Results of the copula model in ANDI DTI dataset. (a) $\rho/(\rho-2)\widehat{\Sigma}_{\theta(x_1)}$ with $x_1 = $ (male, age $= 61$, ADAS $= 5$). (b) $\rho/(\rho-2)\widehat{\Sigma}_{\theta(x_2)}$ with $x_2 = $ (male, age $= 60$, ADAS $= 51$). (c) $\rho/(\rho-2)(\widehat{\Sigma}_{\theta(x_1)} - \widehat{\Sigma}_{\theta(x_2)})$. (d) Random samples from the distributions of $Y|x_1$ (blue) and $Y|x_2$ (red).
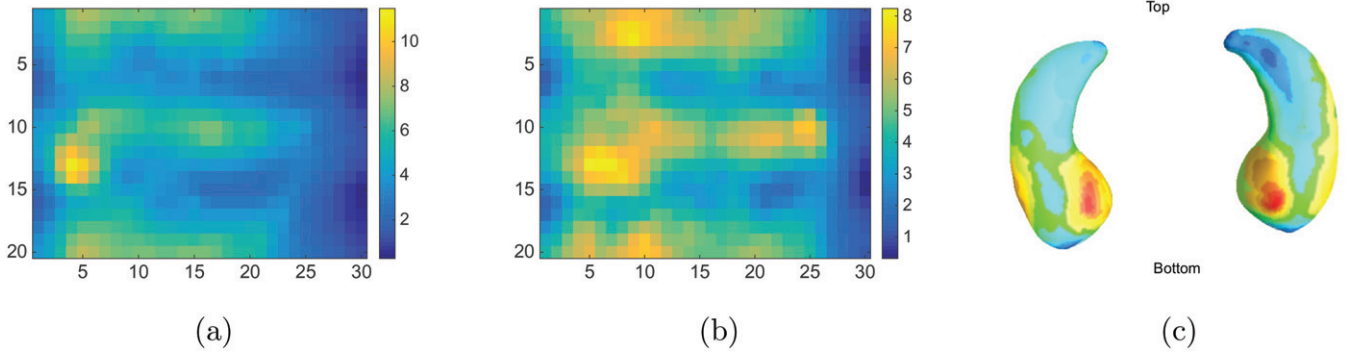

Figure 11. Observed conformally mapped left (a) and right (b) hippocampal images, and the original statistics on the surface (c).

5.3.2. Hippocampus Surface Data

In another example, we analyze the hippocampal substructures extracted from MRI scans in the ADNI study. The hippocampus locates in the medial temporal lobe underneath the cortical surface. It belongs to the limbic system and plays important roles in the consolidation of information from short-term memory to long-term memory and spatial navigation (Colom et al. 2013; Luders et al. 2013). The neurodegenerative activity of AD is evident in the hippocampus region.

In our study, given the MRI scans, the hippocampal substructures were segmented with FSL FIRST (Patenaude et al. 2011) and the hippocampal surfaces were automatically reconstructed with the marching cube method (Lorensen and Cline 1987). We used a surface fluid registration based hippocampal subregional analysis package (Wang et al. 2011; Shi et al. 2013, 2014) that uses isothermal coordinates and fluid registration to generate the correspondence between hippocampal surfaces and the statistics computed on the surface. This method introduces two cuts on a hippocampal surface to convert it into a genus zero surface with two open boundaries. The locations of the two cuts were at the front and back of the hippocampal surface. By using conformal parameterization, it converts a three-dimensional surface registration problem into a two-dimensional image registration problem. The flow induced in the parameter domain establishes high-order correspondences between three-dimensional surfaces. After the registration, various surface statistics were computed on the registered surface, such as multivariate tensor-based morphometry statistics (Wang et al. 2011) that retain the full tensor information of the deformation Jacobian matrix, and the radial distance (Pizer et al. 1999). This software package has been applied in various studies

(Wang et al. 2011, 2013; Shi et al. 2013, 2014). The radial distance feature is used in this article. An example of one subject's left and right hippocampus images are shown in Figure 11, where the left side of (a) or (b) corresponds to the bottom of the hippocampus in (c), and the right side corresponds to the top.

We applied our SQR to the hippocampus dataset. We have 403 subjects, 223 of which are healthy controls (107 females and 116 males), and 180 of which have AD (87 females and 93 males). Scalar covariates include the subject's gender, age, and behavior score (1–36, where lower scores correspond to healthy controls and higher scores corresponds to ADs). The response variable is the two-dimensional hippocampus image. Preliminary analysis indicates that gender does not have a significant effect on hippocampus and therefore, in our analysis, we only include two covariates—age and behavior score, both of which are normalized. In the following, we present results for the right hippocampal surfaces. Figure 12 shows the coefficient images at $\tau = 0.5$, with more results for $\tau = 0.25$ and 0.75 shown in Figure 2.1 in Supplement II. Our results indicate that both aging and AD will degenerate the hippocampus, especially the bottom part, and the AD has more adverse effect than the aging (Scher et al. 2007; Frisoni et al. 2008).

In another experiment, we compared marginal distributions given different covariates. We first got the marginal distribution $Y(s)|x = (1.47, 2.48)$ (corresponding to age $= 85$, behavior $= 30.3$) and then calculated the p-value maps of some randomly selected observations Y_i. The results are shown in Figure 13. We can see that the hippocampus in younger adults with good behavior scores have significantly bigger hippocampus (especially the part close to the bottom) than subjects with age 85 and behavior 30.3.
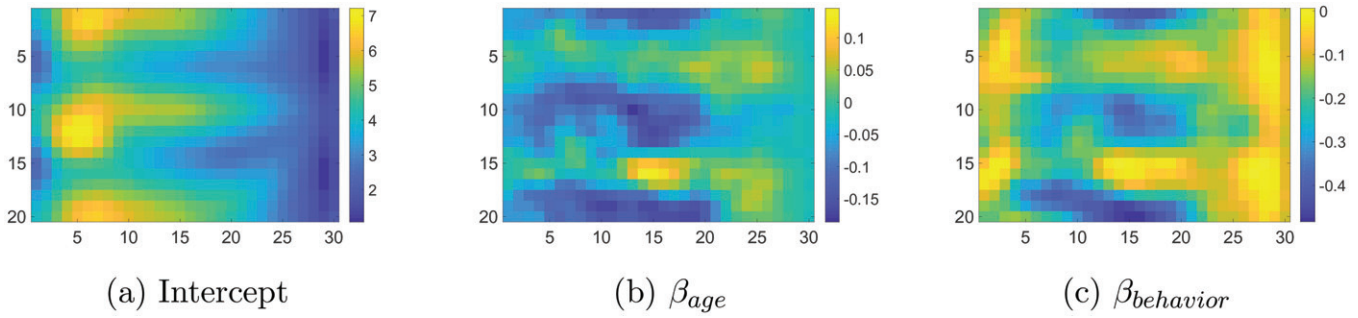
(a) Intercept (b) β_{age} (c) $\beta_{behavior}$

Figure 12. Coefficient images β at $\tau = 0.5$.

(62,3.3) (70,3) (66,2.7)
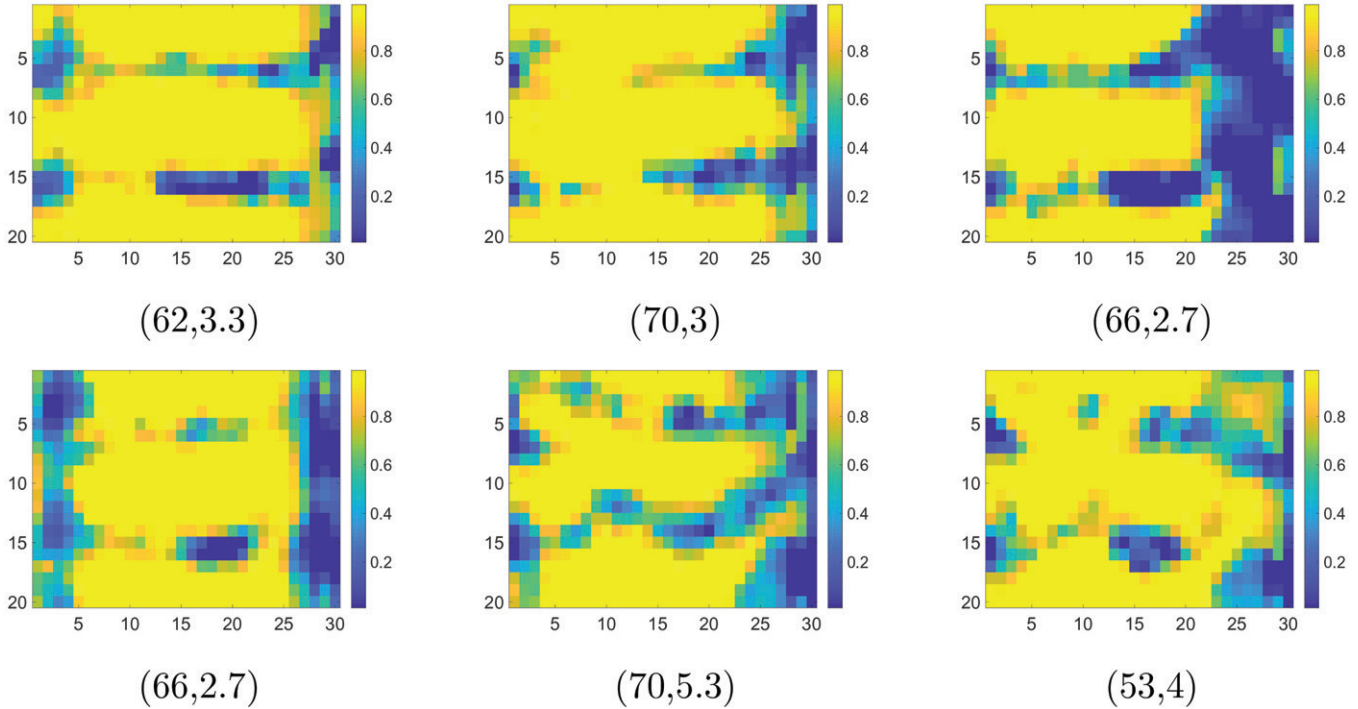
(66,2.7) (70,5.3) (53,4)

Figure 13. p-value maps for some observed Y_i under the marginal distribution of $Y(s)|\mathbf{x}$ with $\mathbf{x} = (1.47, 2.48)$, corresponding to an age of 85 and a behavior score of 30.3. Numbers in the parentheses correspond to age and behavior score.

6. Discussions

We have developed a class of SQR models for function-on-scalar regression, which explicitly characterizes the conditional spatial distribution of a high-dimensional functional/image response given scalar predictors. We have borrowed the strength from both quantile regression and copula modeling, and have developed an efficient primal-dual algorithm to estimate unknown parameters. Simulations and real data analysis are used to show that SQR is efficient to obtain a comprehensive understanding of the effect of scalar covariates on the functional response at different quantile levels. We also established the optimal rate of convergence on the estimation of the coefficient functions.

Many important issues need to be addressed in future research. First, although we focus on the analysis of one- and two-dimensional images, it is straightforward to apply our methods to k-dimensional images with $k > 2$. Under this circumstance, the coefficient functions $\beta_\tau(\cdot)$ are k-dimensional unknown functions, and the spatial location sampling frequency m will be much larger than that of two-dimensional images. In this case, the computational efficiency is the key

challenge. Second, instead of analyzing image data in a raw space directly, we may achieve better results by working on a transformed space, such as frequency. In this case, some further development may be necessary. Third, it is interesting to consider several alternatives to the quantile used in this article. For instance, we may consider a new multivariate concept of quantile based on a directional version of Koenker and Bassett's traditional regression quantiles for multivariate location and multiple-output regression problems (Koenker et al. 2017). In their empirical version, those quantiles can be computed efficiently via linear programming techniques. The contours generated by those quantiles are shown to coincide with the classical halfspace depth contours associated with the name of Tukey. This depth approach is an interesting alternative and we will leave it for future research.

Supplementary Materials

The supplementary materials contain proofs of main theorems and lemmas, additional numerical experiments, and the details of the ADMM algorithm.

Funding

References

Ardekani, B. A., Bachman, A. H., Figarsky, K., and Sidtis, J. J. (2014), "Corpus Callosum Shape Changes in Early Alzheimer's Disease: An MRI Study Using the OASIS Brain Database," *Brain Structure and Function*, 219, 343–352. [12]

Bárdossy, A. (2006), "Copula-Based Geostatistical Models for Groundwater Quality Parameters," *Water Resources Research*, 42, W11416. [3]

Biegon, A., Eberling, J., Richardson, B., Roos, M., Wong, S., Reed, B. R., and Jagust, W. (1994), "Human Corpus Callosum in Aging and Alzheimer's Disease: A Magnetic Resonance Imaging Study," *Neurobiology of Aging*, 15, 393–397. [12]

Bouyé, E., and Salmon, M. (2013), "Dynamic Copula Quantile Regressions and Tail Area Dynamic Dependence in Forex Markets," in *Copulae and Multivariate Probability Distributions in Finance*, eds. A. Dias, M. Salmon, and C. Adcock, London: Routledge, pp. 125–154. [2]

Bowman, A. (2010), "Functional Data Analysis With R and MATLAB," *Journal of Statistical Software*, 34, 1–2. [1]

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011), "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends* in Machine Learning, 3, 1–122. [2,4,7]

Cai, T. T., and Hall, P. (2006), "Prediction in Functional Linear Regression," *The Annals of Statistics*, 34, 2159–2179. [1]

Cai, T. T., and Yuan, M. (2012), "Minimax and Adaptive Prediction for Functional Linear Regression," *Journal of the American Statistical Association*, 107, 1201–1216. [2,6]

Cai, Z., and Xu, X. (2008), "Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models," *Journal of the American Statistical Association*, 103, 1595–1608. [1]

Chen, X., Koenker, R., and Xiao, Z. (2009), "Copula Based Nonlinear Quantile Autoregression," *The Econometrics Journal*, 12, S50–S67. [2]

Colom, R., Stein, J. L., Rajagopalan, P., Martínez, K., Hermel, D., Wang, Y., Álvarez-Linera, J., Burgaleta, M., Quiroga, M. Á., Shih, P. C., and Thompson, P. M. (2013), "Hippocampal Structure and Human Cognition: Key Role of Spatial Processing and Evidence Supporting the Efficiency Hypothesis in Females," *Intelligence*, 41, 129–140. [13]

Crambes, C., Kneip, A., and Sarda, P. (2009), "Smoothing Splines Estimators for Functional Linear Regression," *The Annals of Statistics*, 37, 35–72. [1]

Cressie, N. (1985), "Fitting Variogram Models by Weighted Least Squares," *Mathematical Geology*, 17, 563–586. [2,5,6]

—— (2015), *Statistics for Spatial Data* (Revised Edition), New York: Wiley. [6]

De Angelis, P. L., Pardalos, P. M., and Toraldo, G. (1997), "Quadratic Programming With Box Constraints," in *Developments in Global Optimization*, eds. I. M. Bomze, T. Csendes, R. Horst, and P. M. Pardalos, Boston, MA: Springer, pp. 73–93. [5]

De Backer, M., Ghouch, A. E., and Van Keilegom, I. (2017), "Semiparametric Copula Quantile Regression for Complete or Censored Data," *Electronic Journal of Statistics*, 11, 1660–1698. [2]

Demarta, S., and McNeil, A. J. (2005), "The *t* Copula and Related Copulas," *International Statistical Review*, 73, 111–129. [3]

Du, P., and Wang, X. (2014), "Penalized Likelihood Functional Regression," *Statistica Sinica*, 24, 1017–1041. [1]

Efron, B. (1986), "How Biased Is the Apparent Error Rate of a Prediction Rule?," *Journal of the American Statistical Association*, 81, 461–470. [5]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations*, New York: Springer. [1]

Frisoni, G. B., Ganzola, R., Canu, E., Rüb, U., Pizzini, F. B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C., and Thompson, P. M. (2008), "Mapping Local Hippocampal Changes in Alzheimer's Disease and Normal Ageing With MRI at 3 Tesla," *Brain*, 131, 3266–3276. [13]

Genton, M. G. (1998), "Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure," *Mathematical Geology*, 30, 323–345. [2,5,6]

Goldsmith, J., and Kitago, T. (2016), "Assessing Systematic Effects of Stroke on Motor Control by Using Hierarchical Function-on-Scalar Regression," *Journal of the Royal Statistical Society*, Series C, 65, 215–236. [1]

Greven, S., and Scheipl, F. (2017), "A General Framework for Functional Regression Modelling," *Statistical Modelling*, 17, 1–35. [1]

Gutenbrunner, C., and Jurecková, J. (1992), "Regression Rank Scores and Regression Quantiles," *The Annals of Statistics*, 20, 305–330. [5]

Gutenbrunner, C., Jurečková, J., Koenker, R., and Portnoy, S. (1993), "Tests of Linear Hypotheses Based on Regression Rank Scores," *Journal of Nonparametric Statistics*, 2, 307–331. [5]

Guttorp, P., and Gneiting, T. (2006), "Studies in the History of Probability and Statistics XLIX on the Matérn Correlation Family," *Biometrika*, 93, 989–995. [3,5]

Hall, P., and Horowitz, J. L. (2007), "Methodology and Convergence Rates for Functional Linear Regression," *The Annals of Statistics*, 35, 70–91. [1]

Hallin, M., Lu, Z. L., and Yu, K. (2009), "Local Linear Spatial Quantile Regression," *Bernoulli*, 15, 659–686. [2]

Inano, S., Takao, H., Hayashi, N., Abe, O., and Ohtomo, K. (2011), "Effects of Age and Gender on White Matter Integrity," *American Journal of Neuroradiology*, 32, 2103–2109. [12]

Ivanescu, A. E., Staicu, A.-M., Scheipl, F., and Greven, S. (2015), "Penalized Function-on-Function Regression," *Computational Statistics*, 30, 539–568. [1]

Kato, K. (2012), "Estimation in Functional Linear Quantile Regression," *The Annals of Statistics*, 40, 3108–3136. [1]

Kazianka, H., and Pilz, J. (2010), "Copula-Based Geostatistical Modeling of Continuous and Discrete Data Including Covariates," *Stochastic Environmental Research and Risk Assessment*, 24, 661–673. [3]

Koay, C. G., Chang, L.-C., Carew, J. D., Pierpaoli, C., and Basser, P. J. (2006), "A Unifying Theoretical and Algorithmic Framework for Least Squares Methods of Estimation in Diffusion Tensor Imaging," *Journal of Magnetic Resonance*, 182, 115–125. [10]

Kochunov, P., Thompson, P. M., Lancaster, J. L., Bartzokis, G., Smith, S., Coyle, T., Royall, D. R., Laird, A., and Fox, P. T. (2007), "Relationship Between White Matter Fractional Anisotropy and Other Indices of Cerebral Health in Normal Aging: Tract-Based Spatial Statistics Study of Aging," *Neuroimage*, 35, 478–487. [12]

Koenker, R. (2004), "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89. [1]

—— (2005), *Quantile Regression*, New York: Cambridge University Press. [5]

Koenker, R., and Bassett, G., Jr. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50. [1,2,3]

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017), *Handbook of Quantile Regression*, New York: Chapman and Hall/CRC. [14]

Koenker, R., Ng, P., and Portnoy, S. (1994), "Quantile Smoothing Splines," *Biometrika*, 81, 673–680. [5]

Koenker, R., and Park, B. J. (1996), "An Interior Point Algorithm for Nonlinear Quantile Regression," *Journal of Econometrics*, 71, 265–283. [4]

Kostov, P. (2009), "A Spatial Quantile Regression Hedonic Model of Agricultural Land Prices," *Spatial Economic Analysis*, 4, 53–72. [2]

Kraus, D., and Czado, C. (2017), "D-Vine Copula Based Quantile Regression," *Computational Statistics & Data Analysis*, 110, 1–18. [2]

Lahiri, S. N., Lee, Y., and Cressie, N. (2002), "On Asymptotic Distribution and Asymptotic Efficiency of Least Squares Estimators of Spatial Variogram Parameters," *Journal of Statistical Planning and Inference*, 103, 65–85. [6]

Li, Y., Liu, Y., and Zhu, J. (2007), "Quantile Regression in Reproducing Kernel Hilbert Spaces," *Journal of the American Statistical Association*, 102, 255–268. [1,4,5]

Lorensen, W. E., and Cline, H. E. (1987), "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," in *ACM Siggraph Computer Graphics* (Vol. 21), pp. 163–169. [13]

Lu, Z., Tang, Q., and Cheng, L. (2014, 02), "Estimating Spatial Quantile Regression With Functional Coefficients: A Robust Semiparametric Framework," *Bernoulli*, 20, 164–189. [2]

Luders, E., Thompson, P. M., Kurth, F., Hong, J.-Y., Phillips, O. R., Wang, Y., Gutman, B. A., Chou, Y.-Y., Narr, K. L., and Toga, A. W. (2013), "Global and Regional Alterations of Hippocampal Anatomy in Long-Term Meditation Practitioners," *Human Brain Mapping*, 34, 3369–3375. [13]

Matérn, B. (2013), *Spatial Variation* (Vol. 36), New York: Springer-Verlag. [3,5]

Matheron, G. (1963), "Principles of Geostatistics," *Economic Geology*, 58, 1246–1266. [5]

Meyer, M., and Woodroofe, M. (2000), "On the Degrees of Freedom in Shape-Restricted Regression," *The Annals of Statistics*, 28, 1083–1104. [5]

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2012), "Vine Copulas With Asymmetric Tail Dependence and Applications to Financial Return Data," *Computational Statistics & Data Analysis*, 56, 3659–3673. [9]

Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011), "A Bayesian Model of Shape and Appearance for Subcortical Brain Segmentation," *Neuroimage*, 56, 907–922. [13]

Pizer, S. M., Fritsch, D. S., Yushkevich, P. A., Johnson, V. E., and Chaney, E. L. (1999), "Segmentation, Registration, and Measurement of Shape Variation via Image Object Shape," *IEEE Transactions on Medical Imaging*, 18, 851–865. [13]

Portnoy, S. (1997), "On Computation of Regression Quantiles: Making the Laplacian Tortoise Faster," in *Lecture Notes-Monograph Series*, pp. 187–200. [4]

Portnoy, S., and Koenker, R. (1997), "The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error Versus Absolute-Error Estimators," *Statistical Science*, 12, 279–300. [4]

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer-Verlag. [1]

——— (2007), *Applied Functional Data Analysis: Methods and Case Studies*, New York: Springer-Verlag. [1]

Reich, B. J. (2012), "Spatiotemporal Quantile Regression for Detecting Distributional Changes in Environmental Processes," *Journal of the Royal Statistical Society*, Series C, 61, 535–553. [2]

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011), "Bayesian Spatial Quantile Regression," *Journal of the American Statistical Association*, 106, 6–20. [2,7,11,12]

Reiss, P. T., Huang, L., and Mennes, M. (2010), "Fast Function-on-Scalar Regression With Penalized Basis Expansions," *The International Journal of Biostatistics*, 6, 1–27. [1]

Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984), "A New Rating Scale for Alzheimer's Disease," *The American Journal of Psychiatry*, 141, 1356–1364. [11]

Scher, A., Xu, Y., Korf, E., White, L., Scheltens, P., Toga, A., Thompson, P., Hartley, S., Witter, M., Valentino, D., and Launer, L. (2007), "Hippocampal Shape Analysis in Alzheimer's Disease: A Population-Based Study," *Neuroimage*, 36, 8–18. [13]

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [5]

Shi, J., Lepore, N., Gutman, B. A., Thompson, P. M., Baxter, L. C., Caselli, R. J., and Wang, Y. (2014), "Genetic Influence of Apolipoprotein E4 Genotype on Hippocampal Morphometry: An N = 725 Surface-Based Alzheimer's Disease Neuroimaging Initiative Study," *Human Brain Mapping*, 35, 3903–3918. [13]

Shi, J., Thompson, P. M., Gutman, B., and Wang, Y. (2013), "Surface Fluid Registration of Conformal Representation: Application to Detect Disease Burden and Genetic Influence on Hippocampus," *NeuroImage*, 78, 111–134. [13]

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., and Behrens, T. E. (2006), "Tract-Based Spatial Statistics: Voxelwise Analysis of Multi-Subject Diffusion Data," *NeuroImage*, 31, 1487–1505. [10]

Stein, C. M. (1981), "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, 9, 1135–1151. [5]

Su, L., and Yang, Z. (2007), "Instrumental Variable Quantile Estimation of Spatial Autoregressive Models," Development Economics Working Papers 22476, East Asian Bureau of Economic Research. [2]

Sun, X., Du, P., Wang, X., and Ma, P. (2018), "Optimal Penalized Function-on-Function Regression Under a Reproducing Kernel Hilbert Space Framework," *Journal of the American Statistical Association*, 113, 1601–1611. [1]

Wahba, G. (1990), *Spline Models for Observational Data* (Vol. 59), Philadelphia, PA: SIAM. [3]

Wang, H. J., Feng, X., and Dong, C. (2019), "Copula-Based Quantile Regression for Longitudinal Data," *Statistica Sinica*, 29, 245–264. [2]

Wang, H. J., Zhu, Z., and Zhou, J. (2009), "Quantile Regression in Partially Linear Varying Coefficient Models," *The Annals of Statistics*, 37, 3841–3866. [1]

Wang, X., Zhu, H., and Alzheimer's Disease Neuroimaging Initiative (2017), "Generalized Scalar-on-Image Regression Models via Total Variation," *Journal of the American Statistical Association*, 112, 1156–1168. [1]

Wang, Y., Song, Y., Rajagopalan, P., An, T., Liu, K., Chou, Y.-Y., Gutman, B., Toga, A. W., and Thompson, P. M. (2011), "Surface-Based TBM Boosts Power to Detect Disease Effects on the Brain: An N = 804 ADNI Study," *Neuroimage*, 56, 1993–2010. [13]

Wang, Y., Yuan, L., Shi, J., Greve, A., Ye, J., Toga, A. W., Reiss, A. L., and Thompson, P. M. (2013), "Applying Tensor-Based Morphometry to Parametric Surfaces Can Improve MRI-Based Disease Diagnosis," *Neuroimage*, 74, 209–230. [13]

Wyss-Coray, T. (2016), "Ageing, Neurodegeneration and Brain Rejuvenation," *Nature*, 539, 180–186. [11]

Yang, Y., and He, X. (2015), "Quantile Regression for Spatially Correlated Data: An Empirical Likelihood Approach," *Statistica Sinica*, 25, 261–274. [2]

Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903. [1]

Yi, G. Y., and He, W. (2009), "Median Regression Models for Longitudinal Data With Dropouts," *Biometrics*, 65, 618–625. [1]

Yuan, M. (2006), "GACV for Quantile Smoothing Splines," *Computational Statistics & Data Analysis*, 50, 813–829. [5]

Yuan, M., and Cai, T. T. (2010), "A Reproducing Kernel Hilbert Space Approach to Functional Linear Regression," *The Annals of Statistics*, 38, 3412–3444. [1]

Zhu, H., Kong, L., Li, R., Styner, M., Gerig, G., Lin, W., and Gilmore, J. H. (2011), "FADTTS: Functional Analysis of Diffusion Tensor Tract Statistics," *NeuroImage*, 56, 1412–1425. [12]

Zhu, H., Zhang, H., Ibrahim, J. G., and Peterson, B. S. (2007), "Statistical Analysis of Diffusion Tensors in Diffusion-Weighted Magnetic Resonance Imaging Data," *Journal of the American Statistical Association*, 102, 1085–1102. [10]