Statistica Sinica Preprint No: SS-2018-0499	
Title	FUNCTIONAL ADDITIVE QUANTILE REGRESSION
Manuscript ID	SS-2018-0499
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0499
Complete List of Authors	YINGYING ZHANG HENG LIAN GUODONG LI and ZHONGYI ZHU
Corresponding Author	HENG LIAN
E-mail	henglian@cityu.edu.hk
Notice: Accepted version subject to English editing.	

FUNCTIONAL ADDITIVE QUANTILE REGRESSION

YINGYING ZHANG[1], HENG LIAN[2], GUODONG LI[3] ZHONGYI ZHU[4]

East China Normal University[1], *City University of Hong Kong*[2],

University of Hong Kong[3], *Fudan University*[4]

Abstract: We investigate functional additive quantile regression that models the conditional quantile of a scalar response by nonparametric effects of a functional predictor. We model the nonparametric effects of the principal component scores as additive components which are approximated by B-splines. We also select the relevant components using a nonconvex SCAD penalty. We establish that, when the relevant components are known, the convergence rate of the estimator using the estimated principal component scores is the same as that using the true scores. We also show that the estimator based on relevant components is a local solution of the SCAD penalized quantile regression problem. The practical performance of the proposed method is illustrated via simulation studies and an empirical application to the corn yield data.

Key words and phrases: Additive quantile regression; functional data; principal component analysis; splines.

1. Introduction

The functional quantile regression gives us an overall picture of the predictive distribution of the scalar response given the function-valued covariates rather than just focusing on the mean response. The functional linear quantile regression model is adopted, for instance, in Cardot et al. (2005), Chen and Müller (2012), Kato (2012), in which the τth conditional quantile of a scalar response y is constructed from a functional predictor X through a linear operation as

$$Q_y(\tau|X) = \alpha(\tau) + \int_{\mathcal{T}} X(t)\beta(t,\tau)dt, \tag{1.1}$$

where $X(t)$ is a predictor process which is a square integrable random function defined on a compact interval $\mathcal{T}$, and $\beta(t,\tau)$ is the square integrable coefficient function for a given τ. Yao et al. (2017), Ma et al. (2019) further extend the model to accomodate high-dimensional scalar predictors. To deal with functional data which are infinite-dimensional objects, the most widely used approach is to project functional data onto a space spanned by a finite number of basis functions. The basis can be either fixed in advance (e.g. B-splines, Fourier basis, c.f Crambes et al. (2013), Crambes et al. (2014)) or data-driven. One convenient choice for the latter is to use the eigenbasis of the covariance operator of $X(t)$, which often provides a parsimonious and efficient representation. More

specifically, the covariance kernel is defined as $G(s,t) = \text{Cov}(X(s), X(t))$. By the well-known functional principal component analysis (FPCA) (Yao et al. (2005), Hall et al. (2006), Li and Hsing (2010)), we have the spectral expansion $G(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$ as well as the Karhunen-Loeve expansion $X(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are ordered eigenvalues, $\{\phi_k\}_{k=1}^{\infty}$ are eigenfunctions making up an orthonormal basis of $L_2(\mathcal{T})$ and $\xi_k = \int_{\mathcal{T}} X(t)\phi_k(t)dt$ are called the principal component scores for $X(t)$. Using the expansion $\beta(t,\tau) = \sum_{k=1}^{\infty} b_k(\tau)\phi_k(t)$ where $b_k(\tau) = \int_{\mathcal{T}} \beta(t,\tau)\phi_k(t)dt$, the model (1.1) is transformed into a quantile regression model with an infinite number of "regressors":

$$Q_y(\tau|X) = \alpha(\tau) + \sum_{k=1}^{\infty} b_k(\tau)\xi_k, \tag{1.2}$$

and regularization is necessary. In Kato (2012), the regularization is achieved by truncating the eigensequence to the first K leading terms, where K is chosen such that it retains most of the variation in predictor $X(t)$.

There is an obvious limitation for model (1.2) in that the linear relationship can be restrictive for general applications. To make it more flexible, we propose the functional additive quantile regression where the linear components are

replaced by a sum of non-linear functional components, i.e.

$$Q_y(\tau|X) = \alpha(\tau) + \sum_{k=1}^{\infty} f_{k,\tau}(\xi_k),$$

where $\{f_{k,\tau}(\cdot)\}$ are unknown smooth functions. To make the estimation feasible, we assume all useful information is contained in the first s components. That is, we assume $f_{k,\tau} \equiv 0,\ k > s$ for some sufficiently large s. Furthermore, to avoid possible scaling issues, we instead use the standardized version $\zeta_{ik} = \Phi(\lambda_k^{-1/2}\xi_{ik})$, where $\Phi(\cdot)$ is a continuously differentiable map from $\mathbb{R}$ to $[0,1]$. Then the model becomes

$$Q_y(\tau|X) = \alpha(\tau) + \sum_{k=1}^{s} f_{k,\tau}(\zeta_k). \tag{1.3}$$

To ensure identification of the $f_{k,\tau}$'s, we assume that $Ef_{k,\tau}(\zeta_k) = 0,\ k = 1, 2, \ldots, s$. We assume the significant components are contained in the first s components, but not all of the s components are significant.

In this paper, we approximate the component functions using B-splines method which is computationally convenient. To automatically select significant components, we impose a SCAD penalty (Sherwood and Wang (2016)) on the l_1 norm of each coefficient group. With l_1 penalty, the minimization can be solved by linear programming, while the computation is more involved if l_2 penalty is

used. This choice is also made for theoretical convenience. If we penalize the l_2 norm, the sufficient condition for the local minimizer of a convex difference problem in Wang et al. (2012) is too restrictive and hard to satisfy, since the subgradients of $\|\boldsymbol{\theta}\|_1$ only involves the signs of $\boldsymbol{\theta}_k$ while the gradients of $\|\boldsymbol{\theta}\|_2$ are not restricted to $\{-1, 0, 1\}$. Empirically, we find that when we penalize $\|\boldsymbol{\theta}\|_1$ inside a SCAD penalty, the individual components of $\boldsymbol{\theta}$ are not sparse as long as $\boldsymbol{\theta} \neq 0$. Other penalties such as group LASSO or adaptive group LASSO could also be used instead of the SCAD penalty. Our choice is mainly for convenience while many penalties exist in the literature with similar performances.

There is a rich literature on functional additive models when the conditional mean of the scalar response is of interest. Müller and Yao (2008) considered the model $E(y|X) = \alpha + \sum_{k=1}^{\infty} f_k(\xi_k)$. They estimated $\{f_k\}$ by local polynomial smoothing and regularized the model by truncation as we do here. Zhu et al. (2014) estimated and selected the additive components in the framework of reproducing kernel Hilbert space (RKHS) and adopted the COSSO penalty. Furthermore, Wong et al. (2018) extended the model to partial linear functional additive regression with multivariate functional predictors. There are fewer works for conditonal quantile modeling with functional predictors. Kato (2012) pioneered this field by investigating functional linear quantile regression. His analysis is based on fully observed $X(t)$. In contrast, we allow observation er-

rors in the predictor and apply B-spline approximation which is computationally convenient for implementation compared to the kernel method. The selection of the relevant components is then obtained by using a SCAD penalty on the l_1 norm of each coefficient group.

Another related line of research is on additive or sparse additive quantile regression models, such as Horowitz and Lee (2005), Koenker (2011), Kato (2011), Lian (2012), Lv et al. (2018). Our work is different from the above in that the scores, which serve as pseudo-predictors, are estimated by functional principal component analysis. Thus theoretically we need to deal with the error caused by the estimated predictors. This requires new bounds throughout the proof and new conditions constraining the number of components, the number of knots in splines, and the tuning parameter in the penalty.

The rest of the paper is organized as follows. In Section 2, we present the proposed approach and the computational algorithm. In Section 3, we investigate the asymptotic properties of the proposed estimator. We illustrate the method with simulation studies in Section 4 and apply it to a real dataset in section 5. Concluding remarks are provided in Section 6. The proofs are contained in the appendix.

2. Proposed Methodology

Let $\{y_i, X_i(t)\}_{i=1}^n$ be independent and identically distributed (i.i.d.) realizations of the pair $\{y, X(t)\}$. The trajectories $\{X_i(t) : t \in \mathcal{T}\}$ are observed intermittently on possibly irregular grids $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{iN_i})^T$. We also assume that the predictor trajectories are subject to i.i.d. measurement errors, i.e. $X_{ij} = X_i(t_{ij}) + \varepsilon_{ij}$ with $E\varepsilon_{ij} = 0$ and $var(\varepsilon_{ij}) = \sigma^2$, $j = 1, \ldots, N_i$. The sequence of FPC scores of $X_i(t)$ is denoted by $\boldsymbol{\xi}_{i,\infty} = (\xi_{i1}, \xi_{i2}, \ldots)^T$. Denote the s truncated FPC scores as $\boldsymbol{\xi}_i = (\xi_{i1}, , \ldots, \xi_{is})^T$. Similarly, write $\boldsymbol{\zeta}_{i,\infty} = (\zeta_{i1}, \zeta_{i2}, \ldots)^T$ and $\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{is})^T$. The transformed FPC scores $\{\boldsymbol{\zeta}_i\}$ cannot be observed and need to be estimated from discrete observations of $X_i(t)$.

When $X_i(t)$ are fully observed, the mean and covariance of $X(t)$ can be estimated by sample counterpart $\hat{\mu}(t)$ and $\hat{G}(s,t)$ respectively. Spectral decomposition on the estimated covariance function, $\hat{G}(s,t) = \sum_{k=1}^{n-1} \hat{\lambda}_k \hat{\phi}_k(s) \hat{\phi}_k(t)$, provides the estimated eigenvalues and eigenfunctions. The FPC scores are estimated by projecting $X_i(t)$ onto the eigenfunctions, $\hat{\xi}_{ik} = \int_{\mathcal{T}} X_i(t) \hat{\phi}_k(t) dt$, and we set $\hat{\zeta}_{ik} = \Phi(\hat{\lambda}_k^{-1/2} \hat{\xi}_{ik})$ for transformed FPC scores. When only discrete observations are available, we focus on the case where dense measurements are made such that each $X_i(t)$ can be effectively recovered by smoothing. The eigenvalues, eigenfunctions and FPC scores are estimated by replacing $X_i(t)$ with estimated $\hat{X}_i(t)$. For detailed algorithm, the readers can refer to section 3.1 in

Wong et al. (2018). The estimated transformed FPC scores are denoted by $\hat{\boldsymbol{\zeta}}_i$, which serve as the predictors in the following.

The additive components $f_{k,\tau}(\cdot)$, $k = 1, \ldots, s$ are approximated by B-spline basis functions. Let $(b_1(t), \ldots, b_{K_n+l+1}(t))^T$ be the vector of normalized B-spline basis functions of order l with K_n quasi-uniform internal knots on $[0, 1]$. We refer the readers to Schumaker (2007) for details of the B-spline construction. For ease of notation and simplicity of proofs, we use the same number of basis functions for all nonlinear components. To accommodate the identifying restriction $Ef_{k,\tau}(\zeta_k) = 0$, $1 \leq k \leq s$, we use the centered B-spline basis functions $B_m(\zeta_k) = b_{m+1}(\zeta_k) - \frac{1}{n}\sum_{i=1}^n b_{m+1}(\hat{\zeta}_{ik})$ for $m = 1, \ldots, K_n + l$ as Huang et al. (2010a) has done and denote $\boldsymbol{w}(\zeta_k) = (B_1(\zeta_k), \ldots, B_{K_n+l}(\zeta_k))^T$. We can approximate each $f_{k,\tau}(t)$ by $f_{k,\tau}(t) \approx \boldsymbol{w}(t)'\boldsymbol{\theta}_k$. Let $\boldsymbol{\theta} = (\theta_0, \boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_s^T)^T$ be the spline coefficient for the estimation of component functions and define $\boldsymbol{W}(\boldsymbol{\zeta}_i) = (K_n^{-1/2}, \boldsymbol{w}(\zeta_{i1})^T, \ldots, \boldsymbol{w}(\zeta_{is})^T)^T$ where $K_n^{-1/2}$ is used to make the scale of the intercept comparable with the scale of B-spline basis functions. Now we can apply the SCAD penalty on the l_1 norms of $\boldsymbol{\theta}_k$, $1 \leq k \leq s$ to select the significant components. We estimate $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ that minimizes

$$S_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{W}(\hat{\boldsymbol{\zeta}}_i)^T\boldsymbol{\theta}) + \sum_{k=1}^s p_\lambda(\|\boldsymbol{\theta}_k\|_1), \tag{2.1}$$

where $\rho_\tau(u) = |u|+(2\tau-1)u$ is the quantile loss function, $\|\boldsymbol{\theta}_k\|_1 = \sum_{j=1}^{K_n+l} |\theta_{kj}|$ and $p_\lambda(t)$ is the SCAD penalty function defined as

$$p_\lambda(t) = \lambda t I(0 \le t < \lambda) + \frac{a\lambda t - (t^2+\lambda^2)/2}{a-1} I(\lambda \le t \le a\lambda) + \frac{(a+1)\lambda^2}{2} I(t > a\lambda)$$

for some $a > 2$. Then we can estimate the parameters in model (1.3) by $\hat{\alpha}(\tau) = K_n^{-1/2}\hat{\theta}_0$ and $\hat{f}_{k,\tau}(t) = \boldsymbol{w}(t)^T\hat{\boldsymbol{\theta}}_k$.

Owing to the l_1 norm in the penalization, the above penalized problem can be solved by local linear approximation (LLA) proposed in Zou and Li (2008). More specifically, for each step t, we update the estimator by

$$\hat{\boldsymbol{\theta}}^t = \arg\min_{\boldsymbol{\theta}} n^{-1} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{W}(\hat{\boldsymbol{\zeta}}_i)^T\boldsymbol{\theta}) + \sum_{k=1}^{s} p'_\lambda(\|\hat{\boldsymbol{\theta}}_k^{t-1}\|_1)\|\boldsymbol{\theta}_k\|_1.$$

This problem can be transformed to unpenalized weighted quantile regression problem based on the observation $|\theta_{kj}| = \rho_\tau(\theta_{kj}) + \rho_\tau(-\theta_{kj})$ and an augmented data set. Similar algorithm was used in Sherwood and Wang (2016).

In practice, the selection of the tuning parameter λ is important. We choose λ to minimize the Bayesian Information Criterion (BIC, section 3.1 in Lee et al.

(2014)) defined as

$$\text{BIC}(\lambda) = \log \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{W}(\hat{\boldsymbol{\zeta}}_i)^T \hat{\boldsymbol{\theta}}_\lambda) + J_\lambda \frac{\log n}{2n} C_n, \tag{2.2}$$

where $\hat{\boldsymbol{\theta}}_\lambda$ is the SCAD penalized estimator given λ , $J_\lambda = 1 + (K_n + l)|\mathcal{S}_\lambda|$, $|\mathcal{S}_\lambda|$ is the number of selected components and C_n is defined in the asymptotic theory.

3. Asymptotic Properties

In this section, we first study the asymptotic properties of the oracle estimator when the important components are known a priori. Then we show that the oracle estimator is a local minimizer of $S_n(\boldsymbol{\theta})$.

3.1 Oracle estimator

Assume there are q nonzero components in $\{f_{k,\tau}(\cdot), 1 \le k \le s\}$. In particular, we denote by $\mathcal{S}^* = \{k_1, k_2, \ldots, k_q\} \subseteq \{1, 2, \ldots, s\}$ the index set of the important components with $|\mathcal{S}^*| = q$, where $|\cdot|$ denotes the cardinality of a set. Denote the corresponding transformed scores as $\boldsymbol{\zeta}_{i,\mathcal{S}^*} = (\zeta_{i,k_1}, \ldots, \zeta_{i,k_q})^T$. Then $\boldsymbol{\theta}$ can be divided as $\boldsymbol{\theta}_{\mathcal{S}^*} = (\theta_0, \boldsymbol{\theta}_{k_1}^T, \ldots, \boldsymbol{\theta}_{k_q}^T)^T$ and $\boldsymbol{\theta}_{\mathcal{S}^{*c}}$, which is the complement vector of $\boldsymbol{\theta}_{\mathcal{S}^*}$ in $\boldsymbol{\theta}$. Similarly define the B-spline basis $\boldsymbol{W}(\boldsymbol{\zeta}_{i,\mathcal{S}^*}) =$

$(K_n^{-1/2}, \boldsymbol{w}(\zeta_{ik_1})^T, \ldots, \boldsymbol{w}(\zeta_{ik_q})^T)^T$. Then we can obtain the oracle estimator $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^*_{\mathcal{S}^{*c}} = \mathbf{0}$ and

$$\boldsymbol{\theta}^*_{\mathcal{S}^*} = \arg\min_{\boldsymbol{\theta}_{\mathcal{S}^*}} n^{-1} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{W}(\hat{\boldsymbol{\zeta}}_{i,\mathcal{S}^*})^T \boldsymbol{\theta}_{\mathcal{S}^*}). \tag{3.1}$$

The oracle estimation for each component functions are $\alpha^*(\tau) = K_n^{-1/2}\theta_0^*$, $f^*_{k_j,\tau}(t) = \boldsymbol{w}(t)^T\boldsymbol{\theta}^*_{k_j}$ for $j = 1, \ldots, q$ and $f^*_{k,\tau}(t) = 0$ for $k \notin \mathcal{S}^*$. The following technical conditions are imposed for analyzing the asymptotic behavior of $\boldsymbol{\theta}^*$ and $f^*_{k,\tau}(t)$.

(C1) Condition on the functional predictor: $E(\|X(t)\|^4) < \infty$ and there exists a constant $C_\xi > 0$ such that $E(\xi_k^2\xi_{k'}^2) \leq C_\xi\lambda_k\lambda_{k'}$ and $E(\xi_k^2 - \lambda_k)^2 < C_\xi\lambda_k^2$ for all k and $k' \neq k$. In addition, $C_\lambda^{-1}k^{-\beta} \leq \lambda_k \leq C_\lambda k^{-\beta}$, $\lambda_k - \lambda_{k+1} \geq C_\lambda^{-1}k^{-1-\beta}$ for some constant C_λ and $\beta > 1$.

This condition was previously used in Wong et al. (2018). It imposes a weak moment condition on the functional predictor and is satisfied if $X(t)$ is a Gaussian process. In addition, it assumes the eigenvalues decay at a polynomial rate. Under this condition, we have the following lemma is Proposition 1 in Wong et al. (2018).

Lemma 3.1. *Suppose the transformation function $\Phi(\cdot)$ has a bounded derivative. Under condition (C1) and $\min_i N_i > C_1 n^{1/4}$ for some positive constant C_1,*

there exists a constant C *such that* $E(\hat{\zeta}_{ik} - \zeta_{ik})^2 \leq Ck^2/n$ *uniformly for* $k \leq G_n$ *where* $G_n = C_2 n^{1/(2+2\beta)}$ *for some constant* $C_2 > 0$.

The dense condition $\min_i N_i > C_1 n^{1/4}$ is justified in Hall et al. (2006) to ensure that the smoothed function estimators $\hat{X}(t)$ are as good as the true functions $X(t)$, in the sense that the resulting estimators of both λ_k and ϕ_k are first-order equivalent to the estimators that arise on applying conventional principal component analysis to the true curves $X(t)$.

(C2) Condition on the random error: The random error $\epsilon_i = y_i - \alpha(\tau) - \sum_{k=1}^{s} f_{k,\tau}(\zeta_{ik})$ has the conditional distribution function F_i and conditional density function f_i given $X_i(t)$. f_i are uniformly bounded away from zero and infinity in a neighborhood of zero. Its first derivative f_i' has a uniform upper bound in a neighborhood of zero for $i = 1, \ldots, n$.

(C3) Condition on the component functions: $Ef_{k_j,\tau}(\zeta_{ik}) = 0$ and $f_{k_j,\tau} \in \mathcal{H}_r$ for all $1 \leq j \leq q$, where q is fixed and $\mathcal{H}_r$ is the collection of functions f on $[0,1]$ such that the vth order derivative satisfies the Hölder condition of order m with $r = m + v > 3/2$, v is a positive integer and $m \in (0,1]$, i.e., $|f^{(v)}(t) - f^{(v)}(t')| \leq C_h|t - t'|^m$ for all $t, t' \in [0,1]$ for some positive constant C_h. In addition, $f_{k,\tau}(t) = 0$ for all $k \notin \mathcal{S}^*$.

Condition 2 is more relaxed than what is usually imposed on random error for mean regression, which often requires independence with predictors or

homoscedasticity. Conditions 3 and 4 are typical for application of B-splines. From Schumaker (2007), under Condition 3, there exists $\boldsymbol{\theta}^0_{\mathcal{S}^*}$ such that

$$\sup_{\boldsymbol{t}\in[0,1]^q} |\boldsymbol{W}(\boldsymbol{t})^T\boldsymbol{\theta}^0_{\mathcal{S}^*} - \alpha(\tau) - \sum_{j=1}^{q} f_{k_j,\tau}(t_j)| = O(K_n^{-r}),$$

where $\boldsymbol{W}(\boldsymbol{t}) = (K_n^{-1/2}, \boldsymbol{w}(t_1)^T, \ldots, \boldsymbol{w}(t_q)^T)^T$ and $\boldsymbol{t} = (t_1, \ldots, t_q)^T$. The following theorem summarizes the asymptotic properties of the oracle estimator.

Theorem 3.1. *Assume Conditions in Lemma 3.1 and (C2)-(C3) hold. The number of oracle predictors q is fixed and $q \le s \le G_n = C_2 n^{1/(2+2\beta)}$. If $K_n^3 s^2 \ll n$ and $\max\{K_n, s^2, K_n^{-2r}n\} \gg K_n^2\{\frac{s}{\sqrt{n}} + K_n^{-r}\}\log n$, we have*

$$\|\boldsymbol{\theta}^*_{\mathcal{S}^*} - \boldsymbol{\theta}^0_{\mathcal{S}^*}\|_2 = O_p(\frac{K_n}{\sqrt{n}} + \sqrt{\frac{K_n}{n}}s + K_n^{-r+1/2}), \quad (3.2)$$

$$n^{-1}\sum_{i=1}^{n}(g^*(\hat{\boldsymbol{\zeta}}_{i,\mathcal{S}^*}) - g(\boldsymbol{\zeta}_{i,\mathcal{S}^*}))^2 = O_p(\frac{K_n}{n} + \frac{s^2}{n} + K_n^{-2r}), \quad (3.3)$$

where $g^*(\hat{\boldsymbol{\zeta}}_{i,\mathcal{S}^*}) = \alpha^*(\tau) + \sum_{j=1}^q f^*_{k_j,\tau}(\hat{\zeta}_{i,k_j})$ and $g(\boldsymbol{\zeta}_{i,\mathcal{S}^*}) = \alpha(\tau) + \sum_{j=1}^q f_{k_j,\tau}(\zeta_{i,k_j})$.

Remark 1. Except the two familiar terms as in nonparametric regression in the convergence rate, we have a third term s^2/n caused by the error-contaminated predictors $\hat{\boldsymbol{\zeta}}_{i,\mathcal{S}^*}$. The technical condition $K_n^3 s^2 \ll n$ is to guarantee that the B-spline design matrix with estimated scores behaves as well as the one with true scores. And $\max\{K_n, s^2, K_n^{-2r}n\} \gg K_n^2\{\frac{s}{\sqrt{n}} + K_n^{-r}\}\log n$ comes from the

application of Bernstein's inequality in the proof.

Remark 2. Obviously the optimal choice of K_n is $K_n = O(n^{1/(2r+1)})$, which satisfies the order condition $\max\{K_n, s^2, K_n^{-2r}n\} \gg K_n^2\{\frac{s}{\sqrt{n}} + K_n^{-r}\}\log n$ for any s. If s is fixed, the convergence rate for g^* is the traditional nonparametric convergence rate $O_p(\frac{K_n}{n} + K_n^{-2r})$. This implies that $\hat{\boldsymbol{\zeta}}_{i,\mathcal{S}^*}$ converges fast enough and do not influence the global convergence rate. If $s = O(n^{1/(2+2\beta)})$ under the framework laid out in Lemma 3.1, the condition $K_n^3 s^2 \ll n$ becomes $\frac{3}{2r+1} + \frac{1}{1+\beta} < 1$. In this case, there is a trade-off between r and β. With smaller β, the number of scores which can be consistently estimated is larger. Meanwhile, we require stronger smoothness assumption on the functions to be estimated, i.e. $f_{k_j,\tau}$ satisfy the Hölder condition with larger r.

3.2 SCAD penalized estimator

To investigate the asymptotic properties of the SCAD penalized estimator, we need an additional condition on how quickly the nonzero signal can decay.

(C4) Condition on the signal strength: The minimal signal $\min_{k\in\mathcal{S}^*} \|\boldsymbol{\theta}_k^0\|_2 \geq C\sqrt{\frac{K_n}{n}}(K_n^{1/2} + s)n^{\alpha}$ for some positive constants C and α.

Due to the nonsmoothness and nonconvexity of the penalized objective function, in our proof we use a sufficient condition for the local minimizer of a convex difference problem as in Wang et al. (2012) and Sherwood and Wang

(2016). Specifically, the penalized objective function can be represented as the difference of two convex functions. By verifying that the oracle estimator meets the sufficient condition, we obtain the main theorem of the paper. Let $\mathcal{E}(\lambda)$ be the set of local minima of $S_n(\boldsymbol{\theta})$. The oracle estimator belongs to the set $\mathcal{E}(\lambda)$ with probability approaching one.

Theorem 3.2. *Assume Conditions in Theorem 3.1 and (C4) are satisfied. The optimal rate $K_n = O(n^{1/(2r+1)})$ is used. Let $\boldsymbol{\theta}^*$ be the oracle estimator defined in (3.1). If $\max\{\frac{1}{\sqrt{n}}, \frac{s}{\sqrt{K_n n}}\} \ll \lambda \ll \sqrt{\frac{K_n}{n}}(K_n^{1/2} + s)n^{\alpha}$, then $P(\boldsymbol{\theta}^* \in \mathcal{E}(\lambda)) \to 1$ as $n \to \infty$.*

3.3 Consistency of BIC in model selection

In this section, we investigate the consistency of BIC in model selction. The generic notation $\mathcal{S} \subseteq \{1, \cdots, s\}$ denotes an arbitrary candidate model. We define the BIC for model $\mathcal{S}$ as

$$\text{BIC}(\mathcal{S}) = \log \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{W}(\hat{\boldsymbol{\zeta}}_{i,\mathcal{S}})^T \hat{\boldsymbol{\theta}}_{\mathcal{S}}) + J_{\mathcal{S}} \frac{\log n}{2n} C_n, \tag{3.4}$$

where $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ is the estimator under model $\mathcal{S}$, $J_{\mathcal{S}} = (1+(K_n+l)|\mathcal{S}|)$ and C_n diverges to infinity. The diverging order of C_n will be specified later. The main challenge lies in that the number of candidate models increases exponentially with s. We

use additional conditions to establish the model selection consistency of BIC for our model.

(C5) Condition on the spline design covariates:

(i) $\max_{1\le k\le s} \|W(\zeta_{i,k})\| = O_p(1)$,

(ii) $\frac{b_2}{K_n} \le \inf_{\mathcal{S}} E\lambda_{\min}(W(\zeta_{i,\mathcal{S}})W(\zeta_{i,\mathcal{S}})^T) \le \sup_{\mathcal{S}} E\lambda_{\max}(W(\zeta_{i,\mathcal{S}})W(\zeta_{i,\mathcal{S}})^T) \le \frac{b_2^*}{K_n}$ for some positive constant b_2 and b_2^*.

The condition (C5) is not strong and is satisfied in most nonparametric estimation based on B-spline basis approximation (see Lee et al. (2014)). It's well known that $\|W(\zeta_{i,k})\| = O_p(1)$ for every $1 \le k \le s$. Also we have $\frac{b_2}{K_n} \le E\lambda_{\min}(W(\zeta_{i,\mathcal{S}})W(\zeta_{i,\mathcal{S}})^T) \le \sup_{\mathcal{S}} E\lambda_{\max}(W(\zeta_{i,\mathcal{S}})W(\zeta_{i,\mathcal{S}})^T) \le \frac{b_2^*}{K_n}$ for every $\mathcal{S}$ (see proof of Lemma S1.1(2)). The condition (C5) specifies a uniform version. With these conditions, we obtain the model selection consistency of BIC when s is diverging under the framework laid out in Lemma 3.1.

Theorem 3.3. *Assume Conditions in Theorem 3.2 and (C5) are satisfied. The number of candidate components* $s = O(n^{1/2(1+\beta)})$ *and* $\min_{k\in\mathcal{S}^*} \|\boldsymbol{\theta}_k^0\| \ge C\sqrt{K_n}$. *For any sequence* $C_n \to \infty$ *satisfying* $K_n \frac{\log n}{2n} C_n \to 0$ *and* $\frac{s}{K_n C_n} \to 0$, *we have*

$$P(\inf_{\mathcal{S}\neq\mathcal{S}^*} \mathit{BIC}(\mathcal{S}) > \mathit{BIC}(\mathcal{S}^*)) \to 1.$$

Let S_λ denote the model selected by penalized estimation with λ. From the

definitions, BIC(λ) $\geq$ BIC(S_λ) because in the latter unpenalized estimator is used. In addition, Theorem 3.1 implies that, with high probability, the oracle estimator can be produced by some λ^* on the solution path; thus BIC(λ^*) = BIC(S_{λ^*}). Therefore, by Theorem 3.3, for any λ not inducing the oracle model, we have BIC(λ) $\geq$ BIC(S_λ) $>$ BIC(S_{λ^*}) = BIC(λ^*). This suggests that BIC is consistent for tuning parameter selection.

4. Simulation Studies

We conduct simulation studies to illustrate the empirical performance of the proposed method. We generate predictor trajectories $X_i(t)$ over a grid with 100 equally spaced points over $0 \leq t \leq 10$ from

$$X_i(t) = t + sin(t) + \sum_{k=1}^{10} \xi_{ik}\phi_k(t),$$

where $\xi_{ik} \sim N(0, \lambda_k)$, $\lambda_k = 30k^{-2}$, $corr(\xi_{ik}, \xi_{ik'}) = 0$ for $k \neq k'$ and $\phi_k(t) = (1/\sqrt{10})sin(\pi kt/10 + \pi/4)$. This is the Karhunen-Loève expansion and scores are independent. The measurement errors are generated independently from $N(0, 0.2^2)$. For the regression function, we set $f_1(\zeta_1) = 3\zeta_1 - 3/2$, $f_2(\zeta_2) = sin(2\pi(\zeta_2 - 1/2))$, $f_3(\zeta_3) = 8(\zeta_3 - 1/3)^2$, $f_4(\zeta_4) = 8(\zeta_4 - 1/3)^2 - 8/9$. Then we generate responses using five models:

Model 1: $y_i = 1.4 + f_1(\zeta_{i1}) + f_2(\zeta_{i2}) + f_4(\zeta_{i4}) + \epsilon_i$, $\epsilon_i \sim N(0,1)$;

Model 2: $y_i = 1.4 + f_1(\zeta_{i1}) + f_2(\zeta_{i2}) + f_4(\zeta_{i4}) + \epsilon_i$, $\epsilon_i \sim t(5)$.

Model 3: $y_i = 1.4 + f_1(\zeta_{i1}) + f_2(\zeta_{i2}) + f_4(\zeta_{i4}) + f_3(\zeta_{i3})\epsilon_i$, $\epsilon_i \sim N(0,1)$;

Model 4: $y_i = 1.4 + f_1(\zeta_{i1}) + f_2(\zeta_{i2}) + f_4(\zeta_{i10}) + \epsilon_i$, where ϵ_i is $N(0.5, 0.2)$ when $u < 0.3$ and $N(-0.5, 1)$ when $u > 0.7$ with $u \sim$ Uniform$(0,1)$.

Model 5: $y_i = 1.4 + f_1(\zeta_{i1}) + f_2(\zeta_{i2}) + f_{10}(\zeta_{i10}) + \epsilon_i$, where $\epsilon_i \sim N(0,1)$ and $f_{10}(x) = 8(x - 1/3)^2 - 8/9$.

Model 1 is homoscedastic and the τth conditional quantile is $Q_{y_i}(\tau|X_i) = 1.4 + \Phi^{-1}(\tau) + f_1(\zeta_{i1}) + f_2(\zeta_{i2}) + f_4(\zeta_{i4})$. For identification, the functions we fit are actually the demeaned ones $f_1(\zeta_1) - E(f_1(\zeta_1))$, $f_2(\zeta_2) - E(f_2(\zeta_2))$ and $f_4(\zeta_4) - E(f_4(\zeta_4))$. With abuse of notation, we still denote these demeaned functions as $f_1(\zeta_1)$, $f_2(\zeta_2)$ and $f_4(\zeta_4)$ respectively. Model 2 mimics Model 1 with a heavy-tailed error. Model 3 is heteroscedastic and the τth conditional quantile is $Q_{y_i}(\tau|X_i) = 1.4 + f_3(\zeta_{i3})\Phi^{-1}(\tau) + f_1(\zeta_{i1}) + f_2(\zeta_{i2}) + f_4(\zeta_{i4})$. ζ_3 plays a role in the conditional distribution of y given $X(t)$, but does not directly influence the center (mean or median) of the conditional distribution. Model 4 mimics Model 1 with a bimodal distribution. And Model 5 uses the tenth score to replace the fourth score in Model 1. The estimation of the tenth score is more challenging than that of the fourth score.

We consider sample size $n = 200$ and $n = 400$ with three different quantiles $\tau = 0.1,\ 0.5$ and 0.9. We choose s to recover at least 99.9% of the total variation in $X(t)$, use cubic splines with two knots and select tuning parameters by BIC with $C_n = 1.5$ which is demonstrated to perform well when s is small compared with n in Lee et al. (2014). One could select the number of knots using cross-validation or some information criterion but it would increase the computational burden with no appreciable numerical advantages in our experience. In the literature using a pre-fixed number of basis functions is not uncommon, and this is the case, for example, in Huang et al. (2010b); Fan et al. (2011). This choice of the number of knots is small enough to avoid overfitting in typical problems with sample size not too small, and big enough to flexibly approximate many smooth functions accurately.

Based on 500 repetitions, Tables 1-5 report component-selection results for Model 1-5 respectively. We show the selection percentages of the first 6 component functions by our functional additive modeling and quantile linear regression modeling (also with a SCAD penalty for variable selection) for Models 1-5 (for Model 5 we show the result for the first 5 and the tenth component functions). Column "correct set" corresponds to the percentages of exact selection, while column "super set" gives percentages of fittings that include all nonzero functions. For linear regression modeling, we also use the SCAD penalty to select

Table 1: Model 1: percentages of times of selection for the first 6 component and mean of selected model size (last column).

Model 1			1	2	3	4	5	6	correct set	super set	model size
$n = 200$	$\tau = 0.1$	QFAM	1.00	0.84	0.03	0.85	0.07	0.05	0.52	0.73	2.98
		QFLM	1.00	0.84	0.20	0.90	0.27	0.14	0.21	0.76	4.00
	$\tau = 0.5$	QFAM	1.00	0.80	0.01	0.99	0.01	0.00	0.78	0.80	2.82
		QFLM	1.00	0.86	0.12	1.00	0.20	0.05	0.52	0.86	3.33
	$\tau = 0.9$	QFAM	1.00	0.85	0.06	1.00	0.08	0.02	0.67	0.85	3.07
		QFLM	1.00	0.79	0.19	1.00	0.22	0.10	0.27	0.79	3.85
$n = 400$	$\tau = 0.1$	QFAM	1.00	0.99	0.01	1.00	0.05	0.01	0.88	0.99	3.10
		QFLM	1.00	0.98	0.18	0.99	0.34	0.09	0.36	0.98	4.00
	$\tau = 0.5$	QFAM	1.00	0.99	0.00	1.00	0.01	0.00	0.99	0.99	3.00
		QFLM	1.00	0.99	0.14	1.00	0.28	0.04	0.56	0.99	3.55
	$\tau = 0.9$	QFAM	1.00	0.99	0.04	1.00	0.06	0.00	0.88	0.99	3.12
		QFLM	1.00	0.97	0.16	1.00	0.28	0.07	0.38	0.97	3.91

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

Table 2: Model 2: percentages of times of selection for the first 6 component and mean of selected model size (last column).

Model 2			1	2	3	4	5	6	correct set	super set	model size
$n = 200$	$\tau = 0.1$	QFAM	1.00	0.60	0.01	0.64	0.04	0.02	0.34	0.46	2.43
		QFLM	1.00	0.72	0.17	0.83	0.24	0.13	0.17	0.61	3.69
	$\tau = 0.5$	QFAM	1.00	0.64	0.00	0.96	0.01	0.00	0.63	0.64	2.61
		QFLM	1.00	0.83	0.11	0.99	0.17	0.03	0.51	0.82	3.26
	$\tau = 0.9$	QFAM	1.00	0.70	0.04	0.97	0.04	0.02	0.57	0.69	2.84
		QFLM	1.00	0.71	0.16	0.99	0.22	0.10	0.29	0.70	3.68
$n = 400$	$\tau = 0.1$	QFAM	1.00	0.90	0.02	0.96	0.03	0.01	0.78	0.88	2.98
		QFLM	1.00	0.88	0.14	0.97	0.26	0.09	0.32	0.86	3.80
	$\tau = 0.5$	QFAM	1.00	0.98	0.00	1.00	0.01	0.00	0.97	0.98	2.99
		QFLM	1.00	0.97	0.09	1.00	0.23	0.02	0.61	0.97	3.41
	$\tau = 0.9$	QFAM	1.00	0.89	0.03	1.00	0.03	0.01	0.81	0.89	2.97
		QFLM	1.00	0.87	0.12	1.00	0.17	0.09	0.42	0.87	3.64

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

important components and use ordinary BIC to select tuning parameter. One can see that the functional additive model performs slightly better than the functional linear model and the linear model tends to select a little larger model sizes.

As to estimation accuracy of f_k, we calculate the averaged integrated squared errors (AISEs) of the estimated component functions. The integrated squared er-

Table 3: Model 3: percentages of times of selection for the first 6 component and mean of selected model size (last column).

Model 3			1	2	3	4	5	6	correct set	super set	model size
$n = 200$	$\tau = 0.1$	QFAM	1.00	0.70	0.86	0.91	0.06	0.02	0.52	0.63	3.65
		QFLM	1.00	0.58	0.89	0.84	0.24	0.10	0.16	0.43	4.16
	$\tau = 0.5$	QFAM	1.00	0.96	0.02	0.99	0.01	0.00	0.94	0.96	2.98
		QFLM	1.00	0.90	0.12	1.00	0.17	0.02	0.60	0.90	3.27
	$\tau = 0.9$	QFAM	1.00	0.93	0.95	1.00	0.05	0.01	0.80	0.90	4.00
		QFLM	1.00	0.80	0.92	0.99	0.17	0.08	0.37	0.74	4.46
$n = 400$	$\tau = 0.1$	QFAM	1.00	0.98	1.00	1.00	0.08	0.00	0.88	0.98	4.09
		QFLM	1.00	0.83	0.98	0.98	0.30	0.08	0.38	0.80	4.54
	$\tau = 0.5$	QFAM	1.00	1.00	0.01	1.00	0.01	0.00	0.98	1.00	3.02
		QFLM	1.00	0.99	0.15	1.00	0.29	0.02	0.55	0.99	3.53
	$\tau = 0.9$	QFAM	1.00	1.00	1.00	1.00	0.05	0.00	0.94	1.00	4.05
		QFLM	1.00	0.96	0.99	1.00	0.17	0.07	0.54	0.96	4.55

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

Table 4: Model 4: percentages of times of selection for the first 6 component and mean of selected model size (last column).

Model 4			1	2	3	4	5	6	correct set	super set	model size
$n = 200$	$\tau = 0.1$	QFAM	1.00	0.78	0.04	0.82	0.09	0.04	0.45	0.68	2.95
		QFLM	1.00	0.80	0.21	0.90	0.26	0.14	0.20	0.72	3.98
	$\tau = 0.5$	QFAM	1.00	0.82	0.00	0.98	0.02	0.00	0.79	0.81	2.83
		QFLM	1.00	0.91	0.16	1.00	0.20	0.05	0.53	0.90	3.43
	$\tau = 0.9$	QFAM	1.00	0.96	0.10	1.00	0.06	0.00	0.80	0.95	3.13
		QFLM	1.00	0.90	0.19	1.00	0.22	0.07	0.37	0.89	3.82
$n = 400$	$\tau = 0.1$	QFAM	1.00	0.97	0.03	0.99	0.04	0.01	0.85	0.97	3.10
		QFLM	1.00	0.94	0.18	1.00	0.29	0.12	0.32	0.94	4.13
	$\tau = 0.5$	QFAM	1.00	1.00	0.00	1.00	0.01	0.00	0.98	1.00	3.01
		QFLM	1.00	0.99	0.16	1.00	0.30	0.02	0.54	0.99	3.55
	$\tau = 0.9$	QFAM	1.00	1.00	0.08	1.00	0.09	0.00	0.84	1.00	3.17
		QFLM	1.00	0.98	0.13	1.00	0.25	0.09	0.43	0.98	3.84

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

Table 5: Model 5: percentages of times of selection for the first 6 component and mean of selected model size (last column).

Model 5			1	2	3	4	5	10	correct set	super set	model size
$n = 200$	$\tau = 0.1$	QFAM	1.00	0.74	0.02	0.01	0.03	0.61	0.39	0.51	2.52
		QFLM	1.00	0.85	0.12	0.15	0.15	0.82	0.20	0.70	3.92
	$\tau = 0.5$	QFAM	1.00	0.82	0.00	0.00	0.00	0.95	0.78	0.80	2.80
		QFLM	1.00	0.88	0.06	0.04	0.04	1.00	0.45	0.88	3.52
	$\tau = 0.9$	QFAM	1.00	0.80	0.01	0.00	0.02	0.99	0.60	0.79	3.04
		QFLM	1.00	0.83	0.14	0.09	0.11	1.00	0.24	0.83	4.05
$n = 400$	$\tau = 0.1$	QFAM	1.00	0.98	0.02	0.01	0.01	0.97	0.82	0.95	3.10
		QFLM	1.00	0.98	0.15	0.09	0.11	0.98	0.33	0.97	4.08
	$\tau = 0.5$	QFAM	1.00	1.00	0.00	0.00	0.00	1.00	0.97	1.00	3.02
		QFLM	1.00	1.00	0.06	0.03	0.02	1.00	0.53	1.00	3.58
	$\tau = 0.9$	QFAM	1.00	0.99	0.01	0.00	0.00	1.00	0.88	0.99	3.11
		QFLM	1.00	0.96	0.11	0.09	0.09	1.00	0.36	0.96	4.01

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

rors are defined as

$$\text{ISE}(\hat{f}_k) = \int_0^1 \{\hat{f}_k(t) - f_k(t)\}^2 dt.$$

Tables 6-10 show the AISEs of the 6 component functions for Models 1-5. The AISEs for $\hat{g}$ is the summation of AISEs for the first 6 component functions. The last column shows the prediction errors, which is computed by $\frac{1}{n}\sum_{i=1}^{n}\{\hat{Q}_{y_i}(\tau|X_i) - Q_{y_i}(\tau|X_i)\}^2$ on 200 newly generated testing points. To compute the prediction, we first estimate the transformed FPC scores of $X_i(t)$ in the test set using estimates of mean function, eigenvalues and eigenfuncions from the training data, and then plug these scores into the estimated component functions. Notice that functional additive model performs significantly better than functional linear model with smaller estimation errors as well as smaller prediction errors.

5. Empirical Application

Weather has a significant impact on crop yield and many studies developed statistical models to relate weather and crop yield (Cadson et al. (1996), Prasad et al. (2006), Lobell and Burke (2010)). In this section, we apply our proposed method to a corn yield dataset from Wong et al. (2018). We have yield-related variables for corn from 105 counties in Kansas from 1999 to 2011.

Table 6: Model 1: averaged integrated squared errors (AISEs) and standard deviations (in parentheses).

Model 1			$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{g}$	prediction error
$n = 200$	$\tau = 0.1$	QFAM	0.11	0.18	0.01	0.27	0.01	0.01	0.58	0.88
			(0.07)	(0.16)	(0.04)	(0.34)	(0.05)	(0.04)	(0.40)	(0.42)
		QFLM	2.80	0.79	0.09	1.65	0.13	0.04	5.50	1.02
			(1.02)	(0.36)	(0.26)	(0.76)	(0.26)	(0.13)	(1.47)	(0.28)
	$\tau = 0.5$	QFAM	0.07	0.16	0.00	0.13	0.00	0.00	0.36	0.58
			(0.04)	(0.17)	(0.03)	(0.30)	(0.04)	(0.01)	(0.38)	(0.22)
		QFLM	2.67	0.76	0.05	1.76	0.09	0.02	5.35	0.74
			(0.79)	(0.27)	(0.17)	(0.53)	(0.20)	(0.08)	(0.99)	(0.11)
	$\tau = 0.9$	QFAM	0.11	0.18	0.02	0.17	0.02	0.00	0.51	0.77
			(0.07)	(0.15)	(0.07)	(0.32)	(0.08)	(0.03)	(0.41)	(0.27)
		QFLM	2.62	0.77	0.08	1.93	0.09	0.03	5.52	0.96
			(0.98)	(0.33)	(0.20)	(0.67)	(0.20)	(0.12)	(1.25)	(0.22)
$n = 400$	$\tau = 0.1$	QFAM	0.05	0.08	0.00	0.08	0.01	0.00	0.22	0.46
			(0.03)	(0.06)	(0.02)	(0.14)	(0.03)	(0.01)	(0.17)	(0.16)
		QFLM	2.70	0.75	0.05	1.69	0.10	0.02	5.30	0.85
			(0.67)	(0.25)	(0.13)	(0.57)	(0.18)	(0.06)	(0.99)	(0.16)
	$\tau = 0.5$	QFAM	0.03	0.06	0.00	0.05	0.00	0.00	0.14	0.36
			(0.02)	(0.05)	(0.00)	(0.13)	(0.01)	(0.00)	(0.14)	(0.11)
		QFLM	2.68	0.73	0.03	1.78	0.07	0.01	5.30	0.68
			(0.57)	(0.20)	(0.10)	(0.37)	(0.14)	(0.04)	(0.69)	(0.08)
	$\tau = 0.9$	QFAM	0.05	0.08	0.01	0.06	0.01	0.00	0.21	0.45
			(0.03)	(0.06)	(0.03)	(0.13)	(0.05)	(0.01)	(0.16)	(0.13)
		QFLM	2.62	0.72	0.03	1.87	0.07	0.01	5.32	0.85
			(0.68)	(0.24)	(0.09)	(0.40)	(0.15)	(0.06)	(0.84)	(0.16)

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

Table 7: Model 2: averaged integrated squared errors (AISEs) and standard deviations (in parentheses).

Model 2			$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{g}$	prediction error
$n = 200$	$\tau = 0.1$	QFAM	0.16	0.31	0.00	0.44	0.01	0.00	0.92	1.26
			(0.11)	(0.21)	(0.03)	(0.39)	(0.05)	(0.04)	(0.49)	(0.57)
		QFLM	2.65	0.80	0.10	1.68	0.14	0.05	5.43	1.11
			(1.10)	(0.43)	(0.27)	(0.86)	(0.35)	(0.17)	(1.60)	(0.34)
	$\tau = 0.5$	QFAM	0.08	0.23	0.00	0.15	0.00	0.00	0.46	0.66
			(0.05)	(0.23)	(0.00)	(0.31)	(0.02)	(0.00)	(0.42)	(0.27)
		QFLM	2.72	0.79	0.05	1.80	0.08	0.01	5.44	0.76
			(0.84)	(0.31)	(0.16)	(0.60)	(0.19)	(0.07)	(1.10)	(0.12)
	$\tau = 0.9$	QFAM	0.18	0.28	0.02	0.23	0.01	0.00	0.72	1.07
			(0.13)	(0.22)	(0.09)	(0.35)	(0.06)	(0.03)	(0.47)	(0.45)
		QFLM	2.76	0.85	0.09	1.94	0.12	0.04	5.80	1.06
			(1.20)	(0.45)	(0.25)	(0.71)	(0.30)	(0.15)	(1.49)	(0.29)
$n = 400$	$\tau = 0.1$	QFAM	0.08	0.14	0.00	0.12	0.00	0.00	0.35	0.61
			(0.05)	(0.13)	(0.02)	(0.16)	(0.03)	(0.01)	(0.23)	(0.27)
		QFLM	2.69	0.76	0.05	1.68	0.10	0.02	5.30	0.90
			(0.79)	(0.32)	(0.14)	(0.64)	(0.22)	(0.09)	(1.13)	(0.22)
	$\tau = 0.5$	QFAM	0.04	0.07	0.00	0.04	0.00	0.00	0.15	0.37
			(0.02)	(0.07)	(0.01)	(0.08)	(0.02)	(0.00)	(0.11)	(0.12)
		QFLM	2.70	0.74	0.03	1.79	0.07	0.00	5.32	0.68
			(0.57)	(0.22)	(0.09)	(0.35)	(0.14)	(0.02)	(0.72)	(0.08)
	$\tau = 0.9$	QFAM	0.08	0.15	0.01	0.09	0.00	0.00	0.33	0.60
			(0.05)	(0.14)	(0.04)	(0.08)	(0.03)	(0.02)	(0.19)	(0.22)
		QFLM	2.66	0.76	0.04	1.90	0.05	0.02	5.44	0.90
			(0.83)	(0.30)	(0.12)	(0.44)	(0.14)	(0.09)	(0.93)	(0.18)

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

Table 8: Model 3: averaged integrated squared errors (AISEs) and standard deviations (in parentheses).

Model 3			$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{g}$	prediction error
$n = 200$	$\tau = 0.1$	QFAM	0.06	0.21	0.35	0.18	0.01	0.00	0.81	1.52
			(0.05)	(0.20)	(0.33)	(0.25)	(0.04)	(0.01)	(0.59)	(0.74)
		QFLM	2.45	0.60	1.64	1.44	0.10	0.03	6.26	2.10
			(0.98)	(0.21)	(0.66)	(0.90)	(0.22)	(0.10)	(1.60)	(0.45)
	$\tau = 0.5$	QFAM	0.02	0.07	0.00	0.09	0.00	0.00	0.18	0.46
			(0.02)	(0.09)	(0.02)	(0.21)	(0.01)	(0.00)	(0.25)	(0.19)
		QFLM	2.45	0.70	0.05	1.28	0.06	0.00	4.54	0.73
			(0.65)	(0.23)	(0.16)	(0.49)	(0.14)	(0.03)	(0.83)	(0.11)
	$\tau = 0.9$	QFAM	0.06	0.11	0.25	0.12	0.01	0.00	0.55	1.20
			(0.05)	(0.12)	(0.24)	(0.24)	(0.03)	(0.01)	(0.48)	(0.57)
		QFLM	2.51	0.90	1.90	1.14	0.07	0.03	6.56	2.03
			(0.98)	(0.49)	(0.83)	(0.53)	(0.18)	(0.13)	(1.39)	(0.37)
$n = 400$	$\tau = 0.1$	QFAM	0.02	0.07	0.15	0.07	0.00	0.00	0.32	0.85
			(0.02)	(0.07)	(0.10)	(0.06)	(0.02)	(0.00)	(0.15)	(0.31)
		QFLM	2.43	0.55	1.56	1.34	0.08	0.01	5.97	1.89
			(0.62)	(0.12)	(0.52)	(0.59)	(0.15)	(0.06)	(1.06)	(0.31)
	$\tau = 0.5$	QFAM	0.01	0.03	0.00	0.04	0.00	0.00	0.09	0.32
			(0.01)	(0.02)	(0.02)	(0.06)	(0.01)	(0.00)	(0.07)	(0.11)
		QFLM	2.43	0.68	0.04	1.28	0.06	0.00	4.49	0.68
			(0.46)	(0.17)	(0.10)	(0.29)	(0.10)	(0.02)	(0.56)	(0.08)
	$\tau = 0.9$	QFAM	0.03	0.04	0.13	0.05	0.00	0.00	0.25	0.69
			(0.02)	(0.03)	(0.09)	(0.09)	(0.01)	(0.00)	(0.13)	(0.24)
		QFLM	2.49	0.84	1.81	1.11	0.04	0.02	6.30	1.84
			(0.68)	(0.32)	(0.57)	(0.34)	(0.11)	(0.07)	(0.93)	(0.26)

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

Table 9: Model 4: averaged integrated squared errors (AISEs) and standard deviations (in parentheses).

Model 4			$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{g}$	prediction error
$n = 200$	$\tau = 0.1$	QFAM	0.13	0.21	0.01	0.28	0.02	0.01	0.65	0.96
			(0.08)	(0.17)	(0.04)	(0.32)	(0.08)	(0.04)	(0.43)	(0.42)
		QFLM	2.75	0.80	0.10	1.66	0.14	0.04	5.49	1.01
			(1.04)	(0.38)	(0.24)	(0.77)	(0.30)	(0.14)	(1.44)	(0.27)
	$\tau = 0.5$	QFAM	0.06	0.15	0.00	0.11	0.00	0.00	0.33	0.60
			(0.04)	(0.17)	(0.01)	(0.22)	(0.04)	(0.00)	(0.30)	(0.24)
		QFLM	2.73	0.78	0.06	1.75	0.09	0.01	5.43	0.77
			(0.78)	(0.31)	(0.16)	(0.53)	(0.21)	(0.06)	(1.02)	(0.13)
	$\tau = 0.9$	QFAM	0.07	0.11	0.02	0.10	0.01	0.00	0.31	0.63
			(0.05)	(0.10)	(0.06)	(0.23)	(0.06)	(0.01)	(0.29)	(0.24)
		QFLM	2.72	0.77	0.06	1.88	0.08	0.02	5.52	1.11
			(0.89)	(0.33)	(0.14)	(0.52)	(0.19)	(0.07)	(1.07)	(0.27)
$n = 400$	$\tau = 0.1$	QFAM	0.06	0.10	0.00	0.09	0.01	0.00	0.26	0.51
			(0.04)	(0.09)	(0.03)	(0.14)	(0.03)	(0.01)	(0.18)	(0.18)
		QFLM	2.68	0.75	0.05	1.71	0.09	0.02	5.31	0.85
			(0.75)	(0.29)	(0.14)	(0.59)	(0.18)	(0.08)	(1.07)	(0.17)
	$\tau = 0.5$	QFAM	0.03	0.05	0.00	0.05	0.00	0.00	0.14	0.37
			(0.02)	(0.04)	(0.01)	(0.13)	(0.01)	(0.00)	(0.14)	(0.12)
		QFLM	2.66	0.76	0.04	1.78	0.08	0.00	5.31	0.71
			(0.55)	(0.22)	(0.09)	(0.40)	(0.13)	(0.02)	(0.73)	(0.10)
	$\tau = 0.9$	QFAM	0.03	0.06	0.01	0.05	0.01	0.00	0.15	0.43
			(0.02)	(0.04)	(0.03)	(0.16)	(0.03)	(0.00)	(0.18)	(0.16)
		QFLM	2.63	0.74	0.02	1.91	0.06	0.01	5.37	1.03
			(0.60)	(0.22)	(0.08)	(0.36)	(0.12)	(0.05)	(0.73)	(0.21)

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

Table 10: Model 5: averaged integrated squared errors (AISEs) and standard deviations (in parentheses).

Model 5			$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_{10}$	$\hat{g}$	prediction error
$n = 200$	$\tau = 0.1$	QFAM	0.11	0.21	0.00	0.00	0.00	0.45	0.80	1.22
			(0.07)	(0.18)	(0.02)	(0.02)	(0.02)	(0.38)	(0.47)	(0.46)
		QFLM	2.62	0.77	0.04	0.04	0.05	1.33	5.11	1.16
			(0.94)	(0.35)	(0.12)	(0.13)	(0.14)	(0.61)	(1.31)	(0.32)
	$\tau = 0.5$	QFAM	0.07	0.15	0.00	0.00	0.00	0.18	0.41	0.79
			(0.05)	(0.17)	(0.01)	(0.00)	(0.00)	(0.31)	(0.41)	(0.27)
		QFLM	2.56	0.77	0.02	0.01	0.01	1.59	5.17	0.84
			(0.76)	(0.29)	(0.09)	(0.07)	(0.06)	(0.51)	(0.99)	(0.12)
	$\tau = 0.9$	QFAM	0.12	0.21	0.00	0.00	0.00	0.21	0.61	1.05
			(0.08)	(0.17)	(0.02)	(0.01)	(0.04)	(0.34)	(0.45)	(0.36)
		QFLM	2.58	0.79	0.06	0.03	0.04	1.88	5.69	1.08
			(0.98)	(0.36)	(0.17)	(0.11)	(0.13)	(0.68)	(1.26)	(0.24)
$n = 400$	$\tau = 0.1$	QFAM	0.05	0.08	0.00	0.00	0.00	0.14	0.29	0.66
			(0.03)	(0.07)	(0.01)	(0.01)	(0.01)	(0.17)	(0.19)	(0.21)
		QFLM	2.59	0.75	0.03	0.02	0.02	1.30	4.87	0.96
			(0.66)	(0.25)	(0.09)	(0.06)	(0.06)	(0.40)	(0.84)	(0.19)
	$\tau = 0.5$	QFAM	0.03	0.05	0.00	0.00	0.00	0.06	0.15	0.51
			(0.02)	(0.04)	(0.00)	(0.00)	(0.00)	(0.09)	(0.11)	(0.12)
		QFLM	2.57	0.75	0.02	0.01	0.00	1.62	5.08	0.75
			(0.56)	(0.20)	(0.07)	(0.04)	(0.03)	(0.32)	(0.70)	(0.08)
	$\tau = 0.9$	QFAM	0.05	0.08	0.00	0.00	0.00	0.08	0.24	0.64
			(0.04)	(0.06)	(0.01)	(0.01)	(0.01)	(0.12)	(0.16)	(0.17)
		QFLM	2.59	0.74	0.02	0.02	0.02	1.90	5.44	0.96
			(0.70)	(0.26)	(0.07)	(0.07)	(0.06)	(0.47)	(0.88)	(0.17)

QFAM: quantile functional additive model; QFLM: quantile functional linear model.

The dataset contains the average corn yield per acre for a specific year and county which is the scalar outcome of interest. The functional predictor is $X(t) = (X_1(t) + X_2(t))/2$ with $X_1(t)$ the daily maximum temperature trajectories and $X_2(t)$ the daily minimum temperature trajectories with the time domain $\mathcal{T} = [0, 365]$, which are gathered from 1123 weather stations and aggregated at the county level. After deleting missing data, we use the remaining 857 observations in our analysis.

For data exploration, we include 45 principal components in the regression model which account for 99.9% of the variation in the daily temperature trajectories. Then we fit the proposed functional additive quantile regression and linear quantile regression for $\tau = 0.1, 0.5, 0.9$ separately. The linear model is also equipped with a SCAD penalty for variable selection. We choose the tuning parameter using both BIC and 5-fold cross-validation (CV). Results in Table 11 show that tuning parameters chosen by CV produce better prediction. Thus in the following we focus on results with tuning parameter chosen by CV. In Figures 1-3, we show the corresponding additive component functions $\hat{f}_k(\zeta)$ selected for $\tau = 0.1, 0.5, 0.9$ respectively. The dashed curves are the pointwise confidence bands $\hat{f}_k(\zeta) \pm 2 \times se\{\hat{f}_k(\zeta)\}$. The standard errors are estimated by $se\{\hat{f}_k(\zeta)\} = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{f}_k^{(b)}(\zeta) - \bar{f}_k^{(b)}(\zeta))^2$, where $\bar{f}_k^{(b)}(\zeta) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}_k^{(b)}(\zeta)$ and $\hat{f}_k^{(b)}(\zeta)$ is the estimation from pair bootstrap sample $\{y_i^{(b)}, X_i^{(b)}\}$ for $b =$

Table 11: Averaged model size across 500 bootstrap and averaged prediction error (standard deviation in parentheses) for 5-fold cross-validation across 100 random partitions for corn yield data.

	$\tau = 0.1$		$\tau = 0.5$		$\tau = 0.9$	
	QFAM	QFLM	QFAM	QFLM	QFAM	QFLM
MS	7.42	39.89	14.95	40.72	11.69	42.91
PE-BIC	5.77	5.42	14.70	13.13	5.77	5.34
	(0.60)	(0.30)	(0.90)	(0.64)	(0.56)	(0.33)
PE-CV	5.77	5.37	13.94	12.89	5.77	5.25
	(0.60)	(0.34)	(0.71)	(0.63)	(0.56)	(0.32)

MS: model size; PE-BIC: prediction error with BIC tuning; PE-CV: prediction error with CV tuning; QFAM: quantile functional additive model; QFLM: quantile functional linear model.

$1, \ldots, B$. We report the average model size over different bootstrap samples in Table 11. We note that additive regression modeling produces notably smaller models than linear regression without sacrificing much prediction accuracy. Only ζ_1, ζ_2 are shared among the three conditional quantile regressions. Figures 1-3 show that temperature trajectory influences corn yield in different ways for different quantile levels, because different components are selected.

6. Discussion

Many authors focused on functional linear quantile regression to study the relationship between functional predictor and the conditional quantile of response. In this paper, we model the effects of functional covariate nonparametrically to increase model flexibility. We consider nonconvex penalized estimation of functional additive quantile regression model. Under mild conditions, we derive the oracle convergence rate when true relevant components are used.

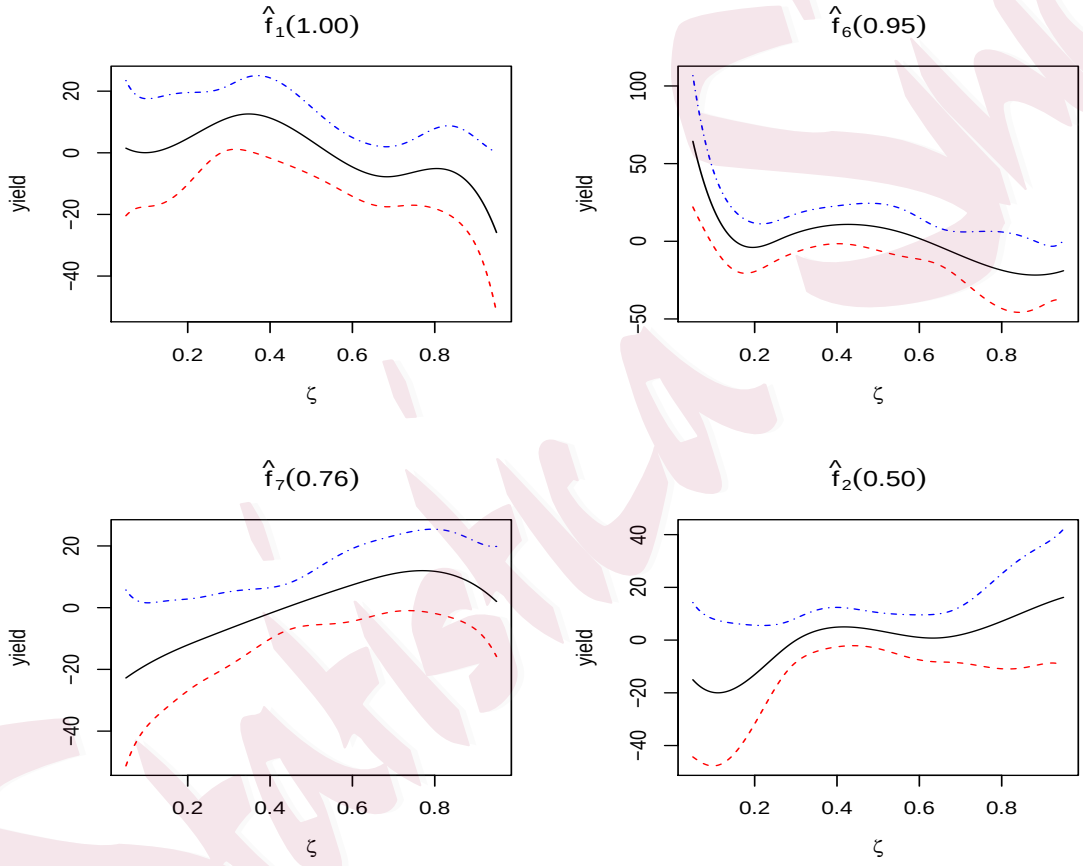

Figure 1: Additive component functions $\hat{f}_k(\zeta)$ selected by FAQR with $\tau = 0.1$, sorted by selection frequencies (in parentheses) across 500 bootstrap.

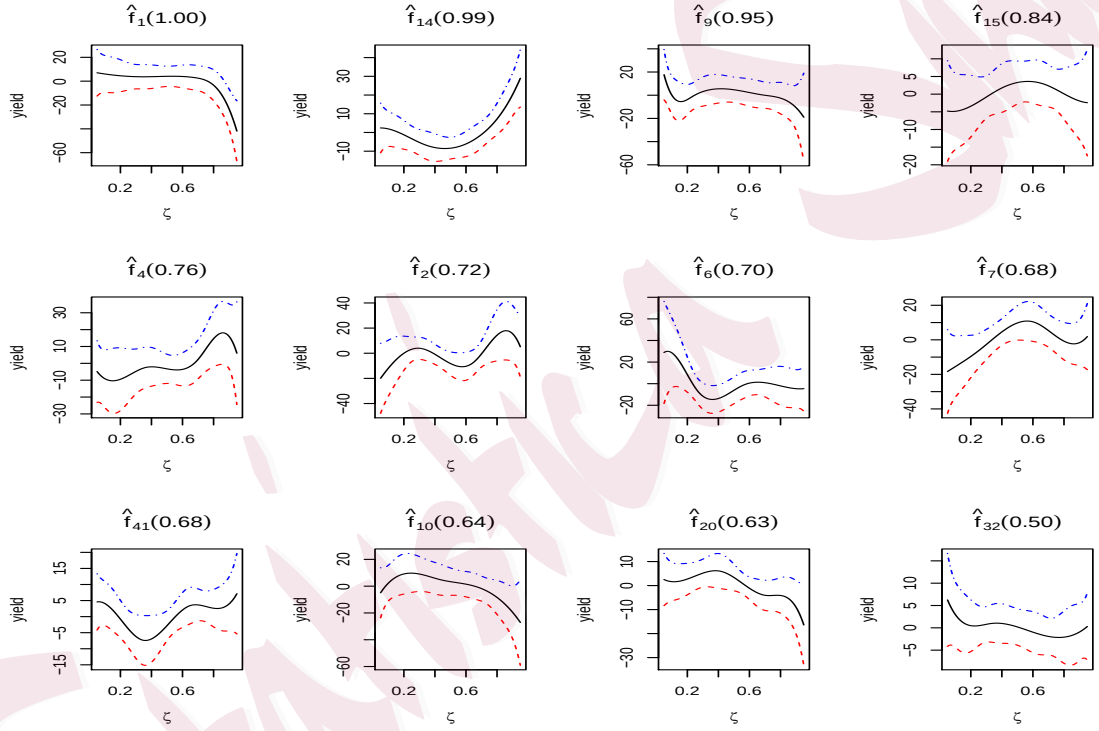

Figure 2: Additive component functions $\hat{f}_k(\zeta)$ selected by FAQR with $\tau = 0.5$, sorted by selection frequencies (in parentheses) across 500 bootstrap.

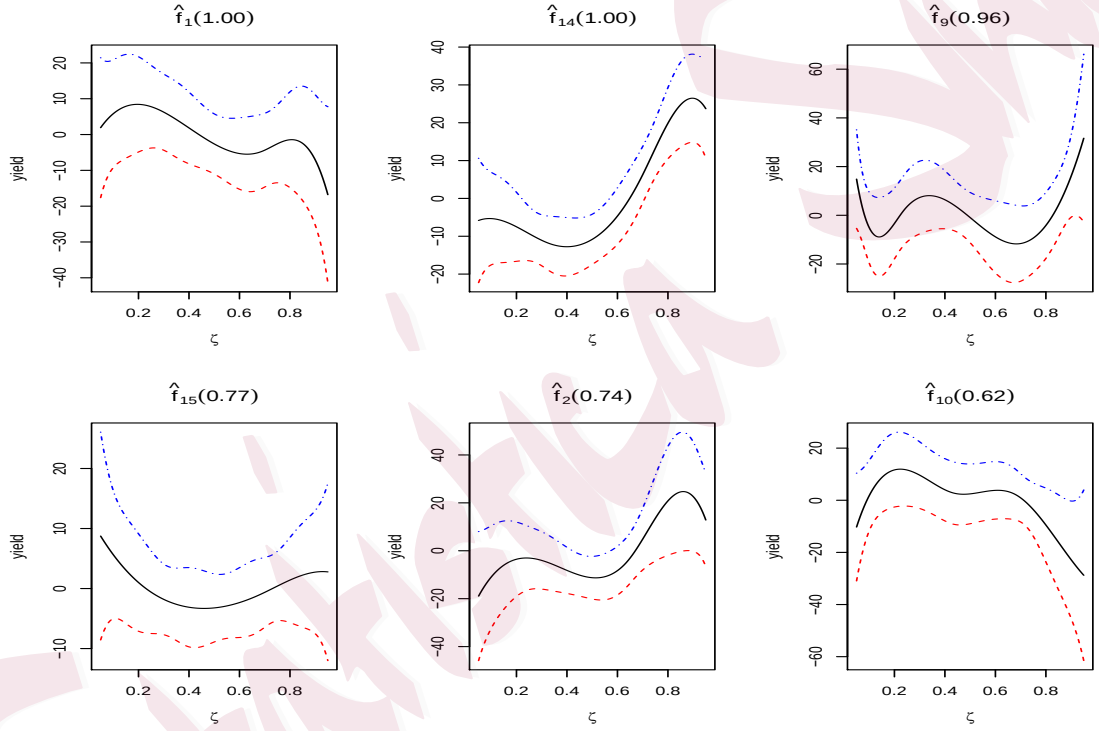

Figure 3: Additive component functions $\hat{f}_k(\zeta)$ selected by FAQR with $\tau = 0.9$, sorted by selection frequencies (in parentheses) across 500 bootstrap.

A problem of important practical interest is to extend our model to take into account scalar covariates and functional covariates simultaneously as Wong et al. (2018) has investigated for mean regression. Another relevant problem is to estimate the conditional quantile function in the reproducing kernel Hilbert space (RKHS) framework and compare the performance with our proposed method. Finally, because we fit each quantile level separately, the phenomenon of quantile crossing may occur. Methods proposed in the literature, such as Koenker and Ng (2005); Bondell et al. (2010); Chernozhukov et al. (2010); Qu and Yoon (2015), may be used to address the problem. These deserve to be investigated in the future.

References

Bondell, H. D., Reich, B. J. and Wang, H. (2010) Noncrossing quantile regression curve estimation. *Biometrika*, **97**, 825–838.

Cadson, R., Todey, D. P. and Taylor, S. E. (1996) Midwestern corn yield and weather in relation to extremes of the southern oscillation. *Journal of production agriculture*, **9**, 347–352.

Cardot, H., Crambes, C. and Sarda, P. (2005) Quantile regression when the covariates are functions. *Journal of Nonparametric Statistics*, **17**, 841–856.

Chen, K. and Müller, H.-G. (2012) Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74**, 67–89.

Chernozhukov, V., Fernandez-Val, I. and Galichon, A. (2010) Quantile and probability curves without crossing. *Econometrica*, **78**, 1093–1125.

Crambes, C., Gannoun, A. and Henchiri, Y. (2013) Support vector machine quantile regression approach for functional data: Simulation and application studies. *Journal of Multivariate Analysis*, **121**, 50–68.

— (2014) Modelling functional additive quantile regression using support vector machines approach. *Journal of Nonparametric Statistics*, **26**, 639–668.

Fan, J., Feng, Y. and Song, R. (2011) Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association*, **106**, 544–557.

Hall, P., Müller, H.-G. and Wang, J.-L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, **34**, 1493–1517.

Horowitz, J. L. and Lee, S. (2005) Nonparametric estimation of an additive quan-

tile regression model. *Journal of the American Statistical Association*, **100**, 1238–1249.

Huang, J., Horowitz, J. L. and Wei, F. (2010a) Variable selection in nonparametric additive models. *The Annals of Statistics*, **38**, 2282–2313.

— (2010b) Variable selection in nonparametric additive models. *Annals of Statistics*, **38**, 2282–2313.

Kato, K. (2011) Group lasso for high dimensional sparse quantile regression models. *arXiv preprint arXiv:1103.1458*.

— (2012) Estimation in functional linear quantile regression. *The Annals of Statistics*, **40**, 3108–3136.

Koenker, R. (2011) Additive models for quantile regression: Model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics*, **25**, 239–262.

Koenker, R. and Ng, P. (2005) Inequality constrained quantile regression. *Sankhya: The Indian Journal of Statistics*, **67**, 418–440.

Lee, E. R., Noh, H. and Park, B. U. (2014) Model selection via Bayesian information criterion for quantile regression models. *Journal of the American Statistical Association*, **109**, 216–229.

Li, Y. and Hsing, T. (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, **38**, 3321–3351.

Lian, H. (2012) Semiparametric estimation of additive quantile regression models by two-fold penalty. *Journal of Business & Economic Statistics*, **30**, 337–350.

Lobell, D. B. and Burke, M. B. (2010) On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, **150**, 1443–1452.

Lv, S., Lin, H., Lian, H. and Huang, J. (2018) Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space. *The Annals of Statistics*, **46**, 781–813.

Ma, H., Li, T., Zhu, H. and Zhu, Z. (2019) Quantile regression for functional partially linear model in ultra-high dimensions. *Computational Statistics & Data Analysis*, **129**, 135–147.

Müller, H.-G. and Yao, F. (2008) Functional additive models. *Journal of the American Statistical Association*, **103**, 1534–1544.

Prasad, A. K., Chai, L., Singh, R. P. and Kafatos, M. (2006) Crop yield estima-

tion model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, **8**, 26–33.

Qu, Z. and Yoon, J. (2015) Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, **185**, 1–19.

Schumaker, L. (2007) *Spline functions: basic theory*. Cambridge University Press.

Sherwood, B. and Wang, L. (2016) Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics*, **44**, 288–317.

Wang, L., Wu, Y. and Li, R. (2012) Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, **107**, 214–222.

Wong, R. K. W., Li, Y. and Zhu, Z. (2018) Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association*, just–accepted.

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

Yao, F., Sue-Chee, S. and Wang, F. (2017) Regularized partially functional quantile regression. *Journal of Multivariate Analysis*, **156**, 39–56.

Zhu, H., Yao, F. and Zhang, H. H. (2014) Structured functional additive regression in reproducing kernel Hilbert spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 581–603.

Zou, H. and Li, R. (2008) One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics*, **36**, 1509–1533.

Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai 200433, China

E-mail: yyzhang@fem.ecnu.edu.cn

Department of Mathematics, City University of Hong Kong, Hong Kong, China

E-mail: henglian@cityu.edu.hk

Department of Statistics & Actuarial Science, University of Hong Kong, Hong Kong, China

E-mail: gdli@hku.hk

Department of Statistics, School of Management, Fudan University, Shanghai 200433, China

E-mail: zhuzy@fudan.edu.cn