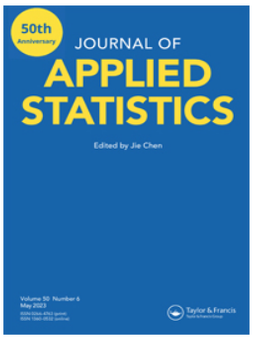

Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Robust scalar-on-function partial quantile regression

Ufuk Beyaztas, Mujgan Tez & Han Lin Shang

To cite this article: Ufuk Beyaztas, Mujgan Tez & Han Lin Shang (2023): Robust scalar-on-function partial quantile regression, Journal of Applied Statistics, DOI: 10.1080/02664763.2023.2202464

To link to this article: https://doi.org/10.1080/02664763.2023.2202464

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2023.2202464

Taylor & Francis
Taylor & Francis Group

Robust scalar-on-function partial quantile regression

Ufuk Beyaztas [a], Mujgan Tez [a] and Han Lin Shang [b]

[a]Department of Statistics, Marmara University, Kadikoy-Istanbul, Turkey; [b]Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, Australia

ABSTRACT

Compared with the conditional mean regression-based scalar-on-function regression model, the scalar-on-function quantile regression is robust to outliers in the response variable. However, it is susceptible to outliers in the functional predictor (called leverage points). This is because the influence function of the regression quantiles is bounded in the response variable but unbounded in the predictor space. The leverage points may alter the eigenstructure of the predictor matrix, leading to poor estimation and prediction results. This study proposes a robust procedure to estimate the model parameters in the scalar-on-function quantile regression method and produce reliable predictions in the presence of both outliers and leverage points. The proposed method is based on a functional partial quantile regression procedure. We propose a weighted partial quantile covariance to obtain functional partial quantile components of the scalar-on-function quantile regression model. After the decomposition, the model parameters are estimated via a weighted loss function, where the robustness is obtained by iteratively reweighting the partial quantile components. The estimation and prediction performance of the proposed method is evaluated by a series of Monte-Carlo experiments and an empirical data example. The results are compared favorably with several existing methods. The method is implemented in an `R` package `robfpqr`.

ARTICLE HISTORY
Received 16 December 2022
Accepted 8 April 2023

KEYWORDS
Functional data; iteratively reweighting; partial quantile covariance; robust estimation

1. Introduction

The advance in technology and data storage have increased the availability of high-dimensional and complex-structured data, whose characteristics vary over a continuum such as time and space, called functional data. Accordingly, the need for reliable tools to analyze this data type has increased in various scientific fields. Consult [15,19,26,27,31,41] for comprehensive theoretical results and applied problems in functional data. Among many others, the scalar-on-function linear regression model, where the response variable is scalar-valued, and the predictor consists of random trajectories, has greatly interested many scientific fields, such as bioscience, chemometrics, meteorology, and population health [see, e.g. 2,3,5,7,17,28,42].

CONTACT Ufuk Beyaztas ufuk.beyaztas@marmara.edu.tr Department of Statistics, Marmara University, Kadikoy-Istanbul 34722, Turkey

The models mentioned above are based on conditional mean regression, i.e. they help explore the average relationship between the response and predictor variables. On the other hand, there is also a strong need to explore the effects of predictors on the response at non-central locations of the response variable's distribution in some cases [see, e.g. 6]. In such cases, the conditional mean regression models are limited. Still, the quantile regression proposed by [30] is helpful to characterize the entire conditional distribution of the response for a given predictor, i.e. it can be used to evaluate the impact of the predictor on the response at different quantile levels.

Quantile regression for the scalar-on-function linear regression model (abbreviated as 'fQR') was first proposed by [8]. By extending the standard quantile regression framework into the functional data, fQR studies the impact of the functional predictor on the scalar response for a given quantile level, which is included in the model via the inner product between the functional predictor and an unknown regression parameter function. Then, several extensions of this model have been proposed [see, e.g. 8–11,18,29,34,44,49–51]. More recently, Li *et al.* [33] studied the statistical inference in functional quantile regression for scalar and functional responses. In addition to characterizing the entire conditional distribution of the scalar response, the fQR is more ambidextrous than the conditional mean-based scalar-on-function linear regression. For example, compared with conditional mean-based regression, the fQR provides a more robust inference when outlying observations and heteroscedasticity are present in the response variable or under non-Gaussian and heavy-tailed distributed error terms.

In fQR, the regression quantiles are obtained as a solution to linear programming algorithms, such as simplex and Frisch-Newton interior-point [see, e.g. 25, 30, for more detailed information]. These regression quantiles are robust to atypical observations in the response variable, called 'outliers' [see, e.g. 1]. However, they are not robust to atypical observations in the predictor space, called 'leverage points'. This is because the influence function of the regression quantiles is bounded in the response variable but unbounded in the predictor space. In the case of outliers, the existing fQR models provide robust inference in characterizing the entire conditional distribution of the scalar response. On the other hand, they may provide poor estimation and prediction results when leverage points are present in the functional predictor since they may alter the predictor matrix's eigenstructure. To our knowledge, a version of fQR that is robust to both leverages and outliers in response has yet to be studied.

In this paper, we propose a robust fQR (called 'RfPQR') to robustly estimate the regression quantiles in the presence of both outliers and leverage points. Our proposed method is based on a functional partial quantile regression (fPQR), which is an extension of the partial quantile regression (PQR) of [16] to the functional data. In the method, we propose a weighted quantile covariance to obtain functional partial quantile components of the fQR. In other words, we consider approximating the infinite-dimensional regression problem fQR with the finite-dimensional quantile regression of the scalar response on the partial quantile components. The regression quantiles are obtained via a weighted loss function, where the effects of leverage points on the predictor matrix are down-weighted with an iterative reweighting algorithm. The regression quantiles obtained by our proposed method are robust to both outliers and leverage points. In addition, our proposed method provides consistent regression quantiles and parameter estimates to the existing fQRs when neither outliers nor leverage points are present in the data.

The remainder of this paper is organized as follows. Section 2 summarizes the fQR model. We introduce the fPQR method in Section 3. The details of the proposed method are given in Section 4. A series of Monte-Carlo experiments are conducted to investigate the proposed method's estimation and predictive performance, and the results are presented in Section 5. The diffusion tensor imaging data analysis results are given in Section 6. Section 7 concludes the paper, along with some possible future directions of the proposed methodology. In a supplement, we introduce the created R package and present a reproducible R code for the Monte Carlo experiments in this study.

2. Scalar-on-function linear quantile regression

We consider a random sample $\{Y_i, \mathcal{X}_i(t) : i = 1, \ldots, n\}$ from the pair $(Y, \mathcal{X})$, where $Y \in \mathbb{R}$ is a scalar response variable and $\mathcal{X} \in \mathcal{L}_2(\mathcal{I})$ is a functional predictor variable belonging to $\mathcal{L}_2$ Hilbert space defined on the bounded and closed interval $t \in \mathcal{I}$. Without loss of generality, we postulate that $\mathcal{I} = [0, 1]$. For simplicity, we further postulate that the scalar response and functional predictor have been centered to mean zero. For a given $\tau \in (0, 1)$, let $Q_{Y_i|\mathcal{X}_i}(\tau)$ denote the τth conditional quantile function of the scalar response $\mathcal{Y}_i$ given the functional predictor $\mathcal{X}_i(t)$. Then, the fQR is defined as follows:

$$Q_{Y_i \,|\, \mathcal{X}_i}(\tau) = \int_0^1 \mathcal{X}_i(t)\beta(t, \tau)\, \mathrm{d}t, \tag{1}$$

where $\beta(t, \tau) \in \mathcal{L}_2[0, 1]$ is the regression parameter function measuring the effect of $\mathcal{X}_i(t)$ on the τth quantile of Y_i.

Model (1) is a straightforward extension of the standard quantile regression model proposed by [30] to functional data, and the main aim in this model is to estimate the regression coefficient function $\beta(t, \tau)$. Similar to the standard quantile regression model, the τth conditional quantile function can be solved by minimizing the check loss function $\rho_\tau(u) = u\{\tau - \mathbb{I}(u < 0)\}$, where $\mathbb{I}$ denotes the indicator function [30]:

$$\underset{\beta(t,\tau)}{\arg\min} \sum_{i=1}^{n} \rho_\tau \left[Y_i - \int_0^1 \mathcal{X}_i(t)\beta(t, \tau)\, \mathrm{d}t \right]. \tag{2}$$

However, the minimization problem in (2) is an ill-posed problem because of the infinite-dimensional nature of the functional predictor $\mathcal{X}_i(t)$. Several methods, such as smoothing splines [8], functional principal component analysis (fPCA) [29,34,50], nonparametric quantile regression estimation [13,14], and PQR [51]. While these methods provide a robust estimate for the regression parameter function $\beta(t, \tau)$ when outlying observations contaminate the scalar response, they may fail to produce robust estimates when outliers are present in the functional predictor $\mathcal{X}(t)$. This is because the influence function of the regression quantiles is bounded in the response variable but unbounded in the predictor space. This paper proposes a robust estimation procedure based on the fPQR.

3. The fPQR

The PQR algorithm of [16] follows the standard partial least squares regression exactly with the redefinition of the expectation and covariance operators. In PQR, Dodge and Whittaker

[16] defined the quantile expectation $\mathrm{E}_\tau(\cdot)$ and quantile covariance $\mathrm{Cov}_\tau(\cdot,\cdot)$ by replacing the least-squares loss function with the check loss function. Let $\{Y_i, \boldsymbol{Z}_i : i = 1,\ldots,n\}$ denote a sample of scalar response Y and P-dimensional matrix of scalar-valued predictors where $\boldsymbol{Z}_i = [Z_{i1},\ldots,Z_{iP}]^\top$ from the pair $(Y, \boldsymbol{Z})$. Then $\mathrm{E}_\tau(\cdot)$ and $\mathrm{Cov}_\tau(\cdot,\cdot)$ are defined as follows:

$$\begin{aligned}\mathrm{E}_\tau(Y) &= \arg\inf_{\alpha} E[\rho_\tau(Y-\alpha)],\\ \mathrm{Cov}_\tau(Y,\boldsymbol{Z})^\top &= \arg\inf_{\alpha,\beta} E[\rho_\tau(Y-\alpha-\beta^\top \mathrm{Var}(\boldsymbol{Z})^{-1}(\boldsymbol{Z}-E(\boldsymbol{Z}))],\end{aligned}$$

where minimizing $E[\rho_\tau(Y-\alpha)]$ with respect to α corresponds to the τth quantile of Y.

Let $\mathcal{C}^\tau_{Y\mathcal{X}}$ and $\mathcal{C}^\tau_{\mathcal{X}Y}$ denote the cross-quantile covariance operator evaluating the contribution of $\mathcal{X}(t)$ to the τth quantile of Y and its adjoint, respectively, as follows:

$$\begin{aligned}&\mathcal{C}^\tau_{Y\mathcal{X}} : \mathcal{L}_2[0,1] \to \mathbb{R}, \quad f \xrightarrow{\mathcal{C}^\tau_{Y\mathcal{X}}} x = \int_0^1 \mathrm{Cov}_\tau\left(Y, \mathcal{X}(t)\right) f(t)\,\mathrm{d}t,\\ &\mathcal{C}^\tau_{\mathcal{X}Y} : \mathbb{R} \to \mathcal{L}_2[0,1], \quad x \xrightarrow{\mathcal{C}^\tau_{\mathcal{X}Y}} f(t) = x\,\mathrm{Cov}_\tau\left(Y, \mathcal{X}(t)\right).\end{aligned}$$

Then, the spectral analysis of self-adjoint, positive, and compact operator $\mathcal{U} = \mathcal{C}^\tau_{\mathcal{X}Y} \circ \mathcal{C}^\tau_{Y\mathcal{X}}$ leads to a countable set of positive eigenvalues $\lambda(\tau)$ associated with orthonormal eigenfunctions $w(t,\tau)$ as a solution of $\mathcal{U}w(t,\tau) = \lambda(\tau)w(t,\tau)$, where $\int_0^1 w(t,\tau)w^\top(t,\tau)\,\mathrm{d}t = 1$ [see, e.g. 40]. Similar to [3,40], the fPQR components of the fQR in (1), can be obtained by optimizing the squared quantile covariance between Y and $\mathcal{X}(t)$ as follows:

$$\arg\max_{\substack{w(t,\tau)\in\mathcal{L}_2[0,1]\\ \|w(t,\tau)\|=1}} \mathrm{Cov}^2_\tau\left(Y, \int_0^1 \mathcal{X}(t)w(t,\tau)\,\mathrm{d}t\right), \tag{3}$$

$$\Leftrightarrow \arg\min_{w(t,\tau)\in\mathcal{L}_2[0,1]} \mathrm{E}^2\left[\rho_\tau\left(Y - \int_0^1 \mathcal{X}(t)w(t,\tau)\,\mathrm{d}t\right)\right], \tag{4}$$

where $w(t,\tau) \in \mathcal{L}_2[0,1]$ is the eigenfunction for quantile level τ associated to the largest eigenvalue of $\mathcal{U}$.

Let us denote by $w^{(1)}(t,\tau)$ the eigenfunction of $\mathcal{U}$ for quantile level τ associated with its largest eigenvalue $\lambda_{\max}(\tau)$, i.e. $\mathcal{U}w^{(1)}(t,\tau) = \lambda_{\max}(\tau)w^{(1)}(t,\tau)$. Then, the first fPQR component of the fQR is obtained as a linear function of $\mathcal{X}(t)$ as follows:

$$T^{(1)}(\tau) = \int_0^1 \mathcal{X}(t)w^{(1)}(t,\tau)\,\mathrm{d}t.$$

We consider an iterative procedure to obtain the subsequent fPQR components, where at each iteration, the fPQR components are obtained by subtracting the information gathered from the previous iteration. In more detail, let $h = 1,2,\ldots$ denote the iteration step and let $Y^{(0)} = Y$ and $\mathcal{X}^{(0)}(t) = \mathcal{X}(t)$. Then, at iteration h, the hth eigenfunction is obtained as

follows:

$$w^{(h)}(t,\tau) = \underset{\substack{w(t,\tau)\in\mathcal{L}_2[0,1]\\ \|w(t,\tau)\|=1}}{\arg\max}\ \mathrm{Cov}_\tau^2\left(Y^{(h-1)}, \int_0^1 \mathcal{X}^{(h-1)}(t)w(t,\tau)\,\mathrm{d}t\right),$$

where

$$Y^{(h)} = Y^{(h-1)} - c^{(h)}T^{(h)}(\tau),$$
$$\mathcal{X}(t)^{(h)} = \mathcal{X}(t)^{(h-1)} - \xi^{(h)}(t)T^{(h)}(\tau),$$

with $c^{(h)} = \mathrm{E}[Y^{(h-1)}T^{(h)}(\tau)]/\mathrm{E}[(T^{(h)}(\tau))^2]$ and $\xi^{(h)}(t) = \mathrm{E}[\mathcal{X}(t)^{(h-1)}T^{(h)}(\tau)]/\mathrm{E}[(T^{(h)}(\tau))^2]$. Accordingly, the hth fPQR component is obtained as follows:

$$T^{(h)}(\tau) = \int_0^1 \mathcal{X}(t)w^{(h)}(t,\tau)\,\mathrm{d}t.$$

Let $\{T^{(1)}(\tau),\ldots,T^{(H)}(\tau)\}$ denote the set of fPQR components obtained after H iterations. Then, the fPQR algorithm is finalized by conducting a quantile regression of Y on $\{T^{(1)}(\tau),\ldots,T^{(H)}(\tau)\}$.

3.1. Estimation of regression coefficient function via basis expansion

In fQR (1), while the functional predictor belongs to infinite-dimensional $\mathcal{L}_2$ Hilbert space, in practice, its elements are observed in a finite set of discrete time points. Therefore, the fPQR components cannot be computed directly from the discrete data. To remedy this problem, as usual in practice, we consider approximating the functional form of discretely observed sample elements of functional predictor via a basis expansion method. With this approach, the eigenanalysis problem $\mathcal{U}w_\tau = \lambda_\tau w_\tau$ reduces to the finite-dimensional eigenanalysis problem. Let us assume that the sample elements of the functional predictor belong to a finite-dimensional space generated by a set of basis expansion functions $\{\phi_1(t),\ldots,\phi_K(t)\}$. Then, $\mathcal{X}(t)$, for $i = 1,\ldots,n$, can be represented by

$$\mathcal{X}_i(t) = \sum_{k=1}^{K} a_{ik}\phi_k(t) = \boldsymbol{A}_i^\top \boldsymbol{\Phi}(t),$$

where $\boldsymbol{A}_i = [a_{i1},\ldots,a_{iK}]$ denotes the basis expansion coefficients.

From $\mathcal{U}w(\tau) = \lambda(\tau)w(\tau)$, the eigenfunction $w(t,\tau)$ and the regression parameter function $\beta(t,\ \tau)$ can be represented in the basis $\boldsymbol{\Phi}(t)$:

$$w(t,\tau) = \sum_{k=1}^{K} w_k(\tau)\phi_k(t) = \boldsymbol{w}^\top(\tau)\boldsymbol{\Phi}(t), \quad \beta(t,\tau) = \sum_{k=1}^{K} \beta_k(\tau)\phi_k(t) = \boldsymbol{\beta}^\top(\tau)\boldsymbol{\Phi}(t),$$

where $\boldsymbol{w}(\tau) = [w_1(\tau),\ldots w_K(\tau)]$ and $\boldsymbol{\beta}(\tau) = [\boldsymbol{\beta}_1(\tau),\ldots,\boldsymbol{\beta}_K(\tau)]$ are the basis expansion coefficients. In addition, the cross-quantile covariance operators can be represented on the

same basis as follows:

$$\mathcal{C}^{\tau}_{Y\mathcal{X}} : \mathcal{L}_2[0,1] \to \mathbb{R}, \quad f \xrightarrow{\mathcal{C}^{\tau}_{Y\mathcal{X}}} x = \Sigma^{\top}_{\boldsymbol{A}Y}(\tau)\boldsymbol{\Phi} f,$$

$$\mathcal{C}^{\tau}_{\mathcal{X}Y} : \mathbb{R} \to \mathcal{L}_2[0,1], \quad x \xrightarrow{\mathcal{C}^{\tau}_{\mathcal{X}Y}} f(t) = x\,\Sigma_{\boldsymbol{A}Y}(\tau),$$

where $\Sigma_{\boldsymbol{A}Y}(\tau) = (\sigma(\tau))_{K\times 1}$ with $\sigma(\tau) = \mathrm{E}[A_k Y]$ for $k = 1, \ldots, K$ is the cross-quantile covariance vector between $\boldsymbol{A}$ and Y and $\boldsymbol{\Phi} = \int_0^1 \boldsymbol{\Phi}(t)\boldsymbol{\Phi}^{\top}(t)\,\mathrm{d}t$ is the matrix of inner products between basis expansion functions.

Based on the basis expansion representations of the functions, the eigenvectors are obtained in the finite dimension of basis expansion coefficients by solving

$$\Sigma_{\boldsymbol{A}Y}(\tau)\Sigma^{\top}_{\boldsymbol{A}Y}(\tau)\boldsymbol{\Phi} w(\tau) = \lambda(\tau)\boldsymbol{w}(\tau). \tag{5}$$

Let $\boldsymbol{w}^{(1)}(\tau)$ denote the first eigenvector obtained by solving (5). In addition, let us consider the decomposition $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{1/2}(\boldsymbol{\Phi}^{1/2})^{\top}$. In what follows, the followings hold

$$\begin{aligned}(\boldsymbol{w}^{(1)}(\tau))^{\top}\boldsymbol{\Phi}\boldsymbol{w}^{(1)}(\tau) &= (\boldsymbol{w}^{(1)}(\tau))^{\top}\boldsymbol{\Phi}^{1/2}(\boldsymbol{\Phi}^{1/2})^{\top}\boldsymbol{w}^{(1)}(\tau),\\ &= (\widetilde{\boldsymbol{w}}^{(1)}(\tau))^{\top}\widetilde{\boldsymbol{w}}^{(1)}(\tau),\end{aligned}$$

where $\widetilde{\boldsymbol{w}}^{(1)}(\tau) = (\boldsymbol{\Phi}^{1/2})^{\top}\boldsymbol{w}^{(1)}(\tau)$ and $\boldsymbol{w}^{(1)}(\tau) = (\Phi^{-1/2})^{\top}\widetilde{\boldsymbol{w}}^{(1)}(\tau)$. Then, for (5), we have

$$(\boldsymbol{\Phi}^{1/2})^{\top}\Sigma_{\boldsymbol{A}Y}(\tau)\Sigma^{\top}_{\boldsymbol{A}Y}(\tau)\boldsymbol{\Phi}^{1/2}\widetilde{\boldsymbol{w}}^{(1)}(\tau) = \lambda_{\max}(\tau)\widetilde{\boldsymbol{w}}^{(1)}(\tau).$$

Accordingly, the first PQR component can be obtained as follows:

$$T^{(1)}(\tau) = \boldsymbol{A}(\boldsymbol{\Phi}^{1/2})^{\top}\widetilde{\boldsymbol{w}}^{(1)}(\tau).$$

The iterative process given in Section 3, along with the basis expansion coefficients, can be used to obtain the subsequent PQR components. The results above show that the fPQR constructed using Y and $\mathcal{X}(t)$ and the standard PQR constructed based on Y and basis expansion coefficients produce the same PQR components at each iteration.

Let us now denote by $\widetilde{\boldsymbol{W}}^{(H)}$ and $\boldsymbol{T}^{(H)}$ the matrices of eigenvectors and PQR components obtained after $h = 1, \ldots, H$ iterations from the PQR of Y on the basis expansion coefficients $\boldsymbol{A}(\boldsymbol{\Phi}^{1/2})^{\top}$. Then the approximate form of fQR in (1) in the finite-dimensional space is given by

$$Q_{Y_i|\boldsymbol{T}}(\tau) = \boldsymbol{T}^{(H)}\boldsymbol{\gamma}(\tau) = \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{\beta}(\tau), \tag{6}$$

where $\boldsymbol{\beta}(\tau) = (\boldsymbol{\Phi}^{-1/2})^{\top}\widetilde{\boldsymbol{W}}^{(H)}\boldsymbol{\gamma}(\tau)$. The estimate of parameter $\boldsymbol{\gamma}(\tau)$ can be obtained by solving

$$\widehat{\boldsymbol{\gamma}}(\tau) = \underset{\boldsymbol{\gamma}(\tau)}{\arg\min} \sum_{i=1}^{n} \rho_{\tau}\left[Y_i - \boldsymbol{T}^{(H)}\boldsymbol{\gamma}(\tau)\right].$$

Finally, the regression parameter function $\beta(t,\tau)$ is approximated by

$$\widehat{\beta}(t,\tau) = \boldsymbol{\Phi}(t)\widehat{\boldsymbol{\beta}}(\tau), \tag{7}$$

where $\widehat{\boldsymbol{\beta}}(\tau) = (\boldsymbol{\Phi}^{-1/2})^{\top}\widetilde{\boldsymbol{W}}^{(H)}\widehat{\boldsymbol{\gamma}}(\tau)$.

4. The proposed RfPQR

From (3) and (4), the eigenfunctions $w(t,\tau)$ defining the fPQR components $T(\tau)$ are obtained using the quantile covariance. In addition, from (5), the eigenvectors used to compute the PQR components in the finite-dimensional space of basis expansion coefficients are obtained via the quantile covariance. With the robustness of the quantile regression, the fPQR and PQR components correctly reflect the underlying structure of the functional predictor, even if outliers contaminate the scalar response. On the other hand, they may not reflect the underlying structure of the predictor in the case of leverage points. This is because the quantile regression is sensitive to leverage points and the effects of such points present in the PQR components. Moreover, in the last steps of the PQR algorithm, the regression parameter $\gamma(\tau)$ in (6) is used to obtain the estimate of regression parameter function $\widehat{\beta}(t,\ \tau)$ via the quantile regression. Again, the quantile regression may produce a biased estimate for $\gamma(\tau)$ when leverage points are present in the predictor. Consequently, in the case of leverage points, the fPQR/PQR may produce biased estimates for the regression parameter function $\beta(t,\ \tau)$, leading to poor predictions of the scalar response. To overcome this problem, we propose a robust alternative to the fPQR.

In the proposed RfPQR method, we consider optimizing the following weighted quantile covariance to obtain the eigenfunctions (and, therefore, the fPQR components):

$$\underset{\substack{w_\psi(t,\tau)\in\mathcal{L}_2[0,1]\\ \|w_\psi(t,\tau)\|=1}}{\arg\max}\ \mathrm{Cov}^2_{\psi\ \tau}\left(Y, \int_0^1 \mathcal{X}(t) w_\psi(t,\tau)\,\mathrm{d}t\right),$$

$$\Leftrightarrow \underset{w_\psi(t,\tau)\in\mathcal{L}_2[0,1]}{\arg\min}\ \mathrm{E}^2\left[\psi\ \rho_\tau\left(Y - \int_0^1 \mathcal{X}(t) w_\psi(t,\tau)\,\mathrm{d}t\right)\right],$$

where $\psi = \psi^y\psi^x$ denotes the vector-valued weights. Here, ψ^y is used to down-weight the effects of outliers while ψ^x is used to down-weight the effects of leverage points (and, therefore, the effects of outliers in the fPQR components). When $\tau = 0.5$, i.e. when we consider the median regression, ψ^y does not have a significant impact on down-weighting the effects of outliers because the median regression is already robust to outliers, but ψ^x has a significant impact on down-weighting the effects of leverage points. If all the weights are taken as 1, then the optimization problem above results in standard fPQR. In addition, if all the weights are taken equally, then the optimization problem results in a weighted fPQR approach, which may not be robust. Similar to [45], we pretend ψ_i, for $i = 1,\ldots,n$, as non-fixed weights, depending on unknown quantities. The underlying idea behind our approach is to compute the first approximations of the fPQR components and parameter estimate via appropriate starting values for ψ. Then, the anterior parameter estimate is used to recompute the weights, from which the second approximations of the fPQR components and parameter estimate are computed by the weighted fPQR. This process, which we call iteratively reweighted fPRQ, is continued until a suitable conditional convergence is achieved.

Because the fQR model belongs to an infinite-dimensional space, similar to fPQR, we consider approximating the RfPQR components as well as the estimate of regression parameter function in the finite-dimensional space of basis expansion coefficients, i.e. we consider the PQR of Y on $\boldsymbol{D} = \boldsymbol{A}(\boldsymbol{\Phi}^{1/2})^\top$. For this model, the weighted eigenvectors,

denoted by $\widetilde{\boldsymbol{w}}_{\psi}(\tau)$, from the proposed method are obtained by

$$\widetilde{\boldsymbol{w}}_{\psi}(\tau) = \mathrm{Cov}_{\psi}\left(Y, \boldsymbol{A}(\boldsymbol{\Phi}^{1/2})^{\top}\right),$$
$$\Leftrightarrow \underset{\widetilde{\boldsymbol{w}}(\tau)}{\arg\min} \sum_{i=1}^{n} \psi_i \rho_{\tau}\left(Y_i - \boldsymbol{D}_i \widetilde{\boldsymbol{w}}(\tau)\right).$$

Then, the RfPQR components are approximated by $\boldsymbol{T}_{\psi}(\tau) = \boldsymbol{D}\widetilde{\boldsymbol{w}}_{\psi}(\tau)$. In the iteratively reweighted fPRQ algorithm, the weights are computed as $\psi_i = f(r_i/\widehat{\sigma}_m)$, where $\widehat{\sigma}_m = \underset{i}{\mathrm{median}}|r_i - \underset{j}{\mathrm{median}}\, r_j|$ and $f(\cdot)$ is the Hampel's loss function

$$f(x) = \begin{cases} x, & 0 \le |x| \le c_1 \\ c_1 \mathrm{sign}(x), & c_1 \le |x| \le c_2 \\ \dfrac{c_1(c_3 - |x|)}{c_3 - c_2}\mathrm{sign}(x), & c_2 \le |x| \le c_3 \\ 0, & c_3 \le |x| \end{cases}$$

where $c_1 = 1.65$, $c_2 = 1.96$ and $c_3 = 3.09$. In the algorithm, ψ_i^y is computed as ψ_i while ψ_i^x is computed as follows:

$$\psi_i^x = f\left(\frac{\|\boldsymbol{T}_{i\psi}(\tau) - \mathrm{med}_{L_1}(\boldsymbol{T}_{\psi}(\tau))\|}{\mathrm{median}_i\|\boldsymbol{T}_{i\psi}(\tau) - \mathrm{med}_{L_1}(\boldsymbol{T}_{\psi}(\tau))\|}\right), \tag{8}$$

where $\mathrm{med}_{L_1}(\boldsymbol{T}_{\psi}(\tau))$ is the L_1 median of the vector of PQR components. The starting values for ψ_i^y is computed as $\psi_i^y = Y_i - \underset{j}{\mathrm{median}}\mathcal{Y}_j$ and ψ_i^x is computed by (8) but replacing $\boldsymbol{T}_{i\psi}(\tau)$ with $\boldsymbol{D}_i$. Based on the definitions above, the detailed algorithm of the proposed RfQR method to estimate the fQR in (1) is given in Algorithm 1.

Let $\widetilde{\boldsymbol{W}}_{\psi}^{(H)}$ and $\boldsymbol{T}_{\psi}^{(H)}$ denote the matrices of the weighted eigenvectors and RfPQR components obtained after $h = 1, \ldots, H$ iterations by the Algorithm 1. Then, the estimated fQR of Y in terms of $\boldsymbol{T}_{\psi}^{(H)}$ is given by

$$\widehat{Q}_{Y_i|\boldsymbol{T}}(\tau) = \boldsymbol{T}_{\psi}^{(H)}\widehat{\boldsymbol{\gamma}}_{\psi}(\tau), \tag{9}$$

where

$$\widehat{\boldsymbol{\gamma}}_{\psi}(\tau) = \underset{\boldsymbol{\gamma}(\tau)}{\arg\min} \sum_{i=1}^{n} \rho_{\tau}\left[Y_i - \boldsymbol{T}_{\psi}^{(H)}\boldsymbol{\gamma}(\tau)\right].$$

Finally, the RfPQR estimation of the regression parameter function $\beta(t,\ \tau)$ by the proposed method is given by

$$\widehat{\beta}_{\psi}(t,\ \tau) = \boldsymbol{\Phi}(t)\widehat{\boldsymbol{\beta}}_{\psi}(\tau), \tag{10}$$

where $\widehat{\boldsymbol{\beta}}_{\psi}(\tau) = (\boldsymbol{\Phi}^{-1/2})^{\top}\widetilde{\boldsymbol{W}}_{\psi}^{(H)}\widehat{\boldsymbol{\gamma}}_{\psi}(\tau)$.

Comparing the fPQR and RfPQR estimates of $\beta(t,\ \tau)$ respectively given by (7) and (10), the fPQR estimate, $\widehat{\beta}(t,\ \tau)$ is robust to only outliers. In contrast, the RfPQR estimate $\widehat{\beta}_{\psi}(t,\ \tau)$ is robust to both outliers and leverage points. This is because while the effects of

Algorithm 1: RfPQR algorithm

(1) Obtain the basis expansion coefficient $\boldsymbol{D} = \boldsymbol{A}(\boldsymbol{\Phi}^{1/2})^{\top}$ from the functional predictor $\mathcal{X}(t)$.
(2) Compute the starting values for ψ_i^y and ψ_i^x for $i = 1, \ldots, n$ and compute $\psi_i = \psi_i^y \psi_i^x$.
(3) Perform a PQR on the weighted Y and $\boldsymbol{D}$ (Y_ψ and $\boldsymbol{D}_\psi$, respectively) to obtain the RfPQR components $\boldsymbol{T}_\psi(\tau)$, where Y_ψ and $\boldsymbol{D}_\psi$ are obtained by multiplying each row of Y and $\boldsymbol{D}$ with $\sqrt{\psi_i}$.
(4) Perform a quantile regression of Y_ψ on the RfPQR components $\boldsymbol{T}_\psi(\tau)$, i.e., Model (6), to obtain a robust estimate of $\gamma(\tau)$, denoted by $\widehat{\gamma}_\psi(\tau)$.
(5) Correct the RfPQR components by dividing each row of $\boldsymbol{T}_\psi(\tau)$ by $\sqrt{\psi_i}$. Compute the residuals, i.e., $e_i = Y_i - \boldsymbol{T}_{i\psi}(\tau)\widehat{\gamma}_\psi(\tau)$ to update ψ_i^y as $\psi_i^y = f(e_i/\widehat{\sigma}_m)$. In addition, update ψ_i^x by applying (8) on the corrected RfPQR components. Then, update the weights as $\psi = \psi_i^y \psi_i^x$.
(6) Repeat steps 3-5 until convergence, where convergence is achieved if the relative difference in norm between two consecutive $\widehat{\gamma}_\psi(\tau)$ is smaller than a threshold value, e.g., 10^{-2}.

leverage points are down-weighted in the RfPQR algorithm, this is not done in the fPQR algorithm. In addition, comparing the approximate models of fQR in (1) obtained with fPQR and RfPQR respectively given by (6) and (9), while the latter one produces robust predictions of Y to both outliers and leverage points, the former produces robust predictions of Y to only outliers.

5. Monte Carlo experiments

We conduct a series of Monte-Carlo experiments under different data generation processes to investigate the estimation and predictive performance of the proposed RfPQR method. The empirical performance of the RfPQR is compared with those of the functional linear model (fLM) based on the principal component analysis, fPCA, and fPQR. Throughout the experiments, the dataset is generated under two scenarios (S1 and S2). In S1, the dataset is generated from a relatively smooth process, where no atypical observations are present. For this scenario, we aim to show whether the proposed method performs similarly to its competitors, i.e. the aim is to show the correctness of the proposed RfPQR method. In S2, on the other hand, the generated dataset is contaminated by atypical observations at 5% and 10% contamination levels. For this scenario, we consider two cases (Cases-I and-II). In Case-I, the generated dataset is contaminated by outliers, i.e. only the values of the scalar response variable are contaminated by atypical observations. In this case, the proposed method is expected to provide similar empirical performance to fPCA and fPQR. In Case-II, the generated dataset is contaminated by outliers and leverage points, and the proposed RfPQR is expected to provide improved empirical performance than its competitors. Throughout the experiments, we set $\tau = 0.5$ to evaluate the robust nature of the proposed method.

In our numerical analyses, 20 B-spline basis expansion functions are used to approximate the components for all the methods, as described in Section 3.1. Also, for all the methods, $K = 4$ components are used to estimate the models.

We use the same data generating process as in [5] to generate the dataset. For the functional predictor, $n = 600$ trajectories at 201 equally spaced points in the interval $[0, 1]$ are generated from the following process:

$$\mathcal{X}(t) = \sum_{j=1}^{5} \kappa_j \nu_j(t),$$

where $\kappa_j \sim \mathcal{N}(0, 4j^{-3/2})$ and $\nu_j(t) = \sin(j\pi t) - \cos(j\pi t)$. The regression parameter function $\beta(t,\ \tau = 0.5)$ is generated from the sinusoidal process $\beta(t,\ \tau = 0.5) \sim 2\sin(2\pi t)$. Then, the observations of the scalar response are generated as follows:

$$Y = \int_0^1 \mathcal{X}(t)\beta(t,\ \tau = 0.5) + \epsilon, \tag{11}$$

where $\epsilon \sim \mathcal{N}(0, 1)$.

To evaluate the empirical performance of the methods, the generated data is first divided into a training sample with size $n_{\text{train}} = 420$ and a test sample with size $n_{\text{test}} = 180$. Then, the models are constructed using the training sample to predict the observations of the scalar response in the test sample. The process mentioned above is repeated 1,000 times, and in each experiment, the following means squared prediction error (MSPE) is computed to compare the predictive performance of the methods:

$$\text{MSPE} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} (Y_i - \widehat{Y}_i)^2,$$

where $\widehat{Y}_i$ denotes the ith predicted response in the test sample. In addition, the following relative integrated squared estimation error (RISEE) is computed in each experiment to compare the estimation performance of the methods [52]:

$$\text{RISEE} = \frac{\|\beta(t, \tau = 0.5) - \widehat{\beta}(t, \tau = 0.5)\|_2^2}{\|\beta(t, \tau = 0.5)\|_2^2},$$

where $\|\cdot\|_2^2$ denotes the $\mathcal{L}_2$ norm.

In Case-I, we consider a contaminated normal distribution for ϵ to generate the outliers. In doing so, before computing the values of the scalar response, the randomly selected $n_{\text{train}} \times [5\%,\ 10\%]$ observations of ϵ are replaced with the observations generated from $\mathcal{N}(15, 1)$. Then, the response values are computed using (11). In Case-II, on the other hand, first, $n_{\text{train}} \times [5\%,\ 10\%]$ trajectories of $\mathcal{X}(t)$ are replaced with $\widetilde{\mathcal{X}}(t) = \mathcal{X}(t) + \upsilon(t)$, where $\upsilon(t)$ are generated from the Ornstein-Uhlenbeck process:

$$\upsilon(t) = \eta[\upsilon_0(t) - \eta]e^{-\theta t} + \sigma \int_0^1 e^{-\theta(t-u)}\, \mathrm{d}W_u,$$

where η, θ, and σ are constant, $\upsilon_0(t)$ is the initial value of $\upsilon(t)$, and W_u is the Wiener process. We consider generating the initial value of $\upsilon(t)$ from a normal distribution with a

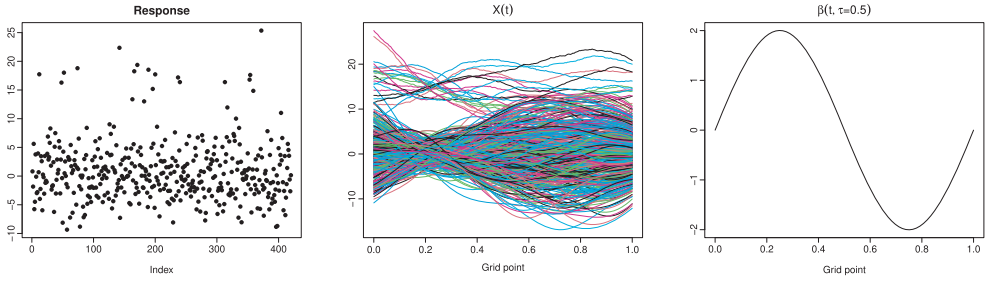

Figure 1. Graphical display of the generated scalar response (left panel), functional predictor (middle panel), and regression parameter function (right panel). The data are generated under S2-Case-II.

mean of 15 and unit variance to generate magnitude outliers for the functional predictor. Then, similar to Case-I, $n_{\text{train}} \times [5\%, 10\%]$ observations of ϵ are generated from $\mathcal{N}(15, 1)$. Finally, the observations of the scalar response are obtained by applying (11). A graphical display of the generated data is presented in Figure 1.

The median values of the computed RISEE values over 1,000 Monte-Carlo experiments for both scenarios and cases are reported in Table 1. From this table, when no outliers are present in the data (i.e. S1), it is evident that all the methods tend to produce similar results. When outliers contaminate only the values of the scalar response variable, i.e. Case-I, the quantile regression models are unaffected by the outliers and produce similar median RISEE values to those they produce under S1. On the other hand, fLM is seriously affected by the outliers and produces median RISEE values about 1.6 and 2 times larger (respectively for the 5% and 10% contamination levels) than those it produces under S1. When both the scalar response and functional predictor variables are contaminated by the outliers and leverage points (Case-II), the fPCA and fPQR methods exhibit the same behavior as the fLM and produce considerably larger median RISEE values compared with those obtained under S1. These results indicate that the presence of leverage points may cause the loss of the robust nature of the fPCA and fPQR methods. On the other hand, our proposed RfPQR method is unaffected by the outliers and produces median RISEE values similar to those obtained under S1. Note that we consider several different regression parameter functions, including sinusoidal and exponential, in the Monte-Carlo experiments (the results are not presented in the paper to save space but can be obtained from the corresponding author upon request). Our results obtained with different regression parameter functions indicate that the proposed method produces similar RISEE values with fLM and fPCA when no outlier is present in the data. It provides improved RISEE values than its competitors

Table 1. Comparison of the parameter estimation performance: Computed median REISEE values, best performances in bold.

		Method			
Scenario/Case	Contamination level	fLM	fPCA	fPQR	RfPQR
S1	0%	**0.101**	0.104	**0.101**	0.108
Case-I	5%	0.161	0.104	**0.103**	0.108
	10%	0.217	0.107	**0.105**	0.108
Case-II	5%	0.933	0.937	1.053	**0.084**
	10%	0.934	0.935	0.955	**0.105**

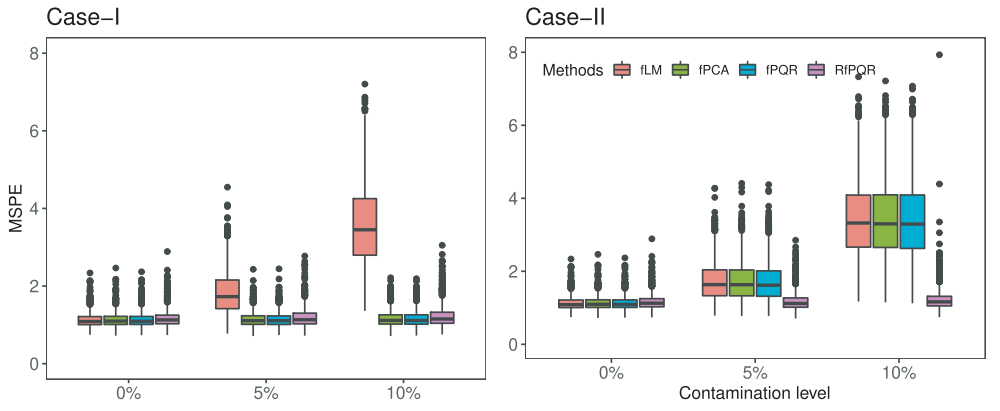

Figure 2. Predictive model performances: Computed MSPE values of the fLM, fPCA, fPQR, and RfPQR under S1 (0% contamination level in each plot), S2-Case-I (left panel), and S2-Case-II (right panel).

when the data are contaminated by outliers and leverage points (i.e. the performance of the proposed method is not sensitive to the choice of regression parameter function).

The MSPE values recorded from the Monte-Carlo experiments are presented in Figure 2. From this figure, when the dataset is generated under S1, i.e. 0% contamination level, all the methods produce similar out-of-sample predictive performance. When only the scalar response is contaminated by the outliers (Case-I), the quantile regression methods produce a similar performance, which is significantly better than fLM. On the other hand, when outliers and leverage points contaminate both the scalar response and functional predictor variables, our proposed method produces considerably improved out-of-sample predictive performance than all the methods. In this case, the results support the findings presented in Table 1 so that the fPCA and fPQR methods are affected by the leverage points and produce similar results with the fLM. A reproducible R code to obtain the MSPE results under Case-II is given in the supplement.

6. Diffusion tensor imaging data

We consider a diffusion tensor imaging (DTI) dataset used in [48], which is available in the R package 'refund' [21] and it has been used in several functional regression model-based studies [see, e.g. 20,22,23,28,36,46]. The dataset was collected from 100 patients with multiple sclerosis (MS) during their visits to Johns Hopkins University and the Kennedy-Krieger Institute. MS is one of the most common causes of non-traumatic disability in young and middle-aged adults [43]. It is a demyelinating autoimmune-mediated disease that occurs with the formation of multiple plaques in the brain and spinal cord resulting from a disorder in the person's immune system. DTI quantifies the magnitude and direction of water diffusion and is one of the most commonly used magnetic resonance imaging techniques to identify MS lesions. Fractional anisotropy (FA) measures the degree of anisotropy of water molecules. FA is a DTI-derived metric and is used as a proxy for demyelination. At each visit of the MS patients in the dataset, patients underwent a DTI scan of their brains. Then, FA for each of the patients was extracted along with two major tracts; (1) the corpus callosum (CCA) and (2) the right corticospinal (RCS). The CCA is the most prominent white

matter tract that connects the two hemispheres of the brain, and damage to this structure has previously been linked to a decline in cognitive performance among MS patients [39,48]. FA-CCA and FA-RCS are viewed as functions of spatial distance along the tract. In addition, the MS patients completed a paced auditory serial addition test (PASAT), a measure of cognitive function. They assess acoustic information processing speed, flexibility, and calculation ability, resulting in a scalar score [24]. PASAT values differ between MS patients and healthy individuals, and evaluation of FA-CCA functions may benefit from linking the FA-CCA tract to PASAT values. Thus, with this dataset, we aim to characterize the conditional distribution of the PASAT scores of the MS patients given their FA profiles along the CCA tract (FA-CCA). From the estimated distribution, lower quantiles indicate that an additional neurological follow-up might be needed.

We consider the DTI dataset obtained at the first visit of MS patients (66 PASAT scores and 66 curves for FA-CCA). The graphical display of observed PASAT scores and functional FA values for the CCA tract are presented in Figure 3. The following model is considered:

$$Q_{Y_i|\mathcal{X}_i}(\tau) = \beta_0(\tau) + \int_{t=1}^{93} \mathcal{X}_i(t)\beta(t,\tau)\,\mathrm{d}t, \quad i = 1,\ldots,66,$$

where $\beta_0(\tau)$ is the intercept varying over τ, Y_i is the ith PASAT score and $\mathcal{X}_i(t)$ is the ith FA-CCA. However, as seen from Figure 3, there are potential outlying curves in both PASAT and FA-CCA values (mostly in FA-CCA). These outliers may prevent accurately predicting the distribution function of PASAT scores conditional on the FA-CCA profiles. We apply the functional boxplot method proposed by [47] available in the R package 'fdaoutlier' to determine the outliers in the FA-CCA. According to the functional boxplot results, four outliers (6th, 13th, 20th, and 55th observations) exist in the functional predictor. This motivates us to apply our proposed RfPQR approach to robustly predict the conditional distribution of PASAT scores.

As stated by [11], assessing the quality of the estimated quantiles is not straightforward because only one pair of the data $(Y_i, \mathcal{X}_i(t))$ is available at a particular predictor level. We follow a similar approach as in [11] to evaluate the empirical performance of the quantile

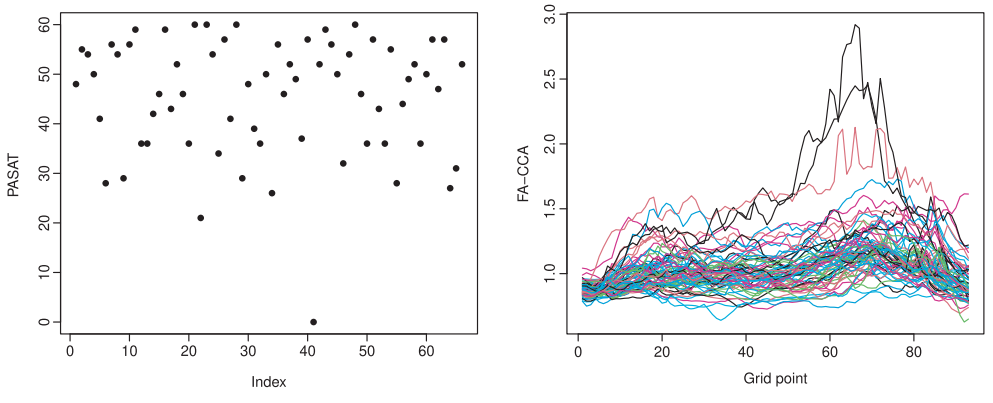

Figure 3. Graphical display of PASAT score (left panel) and FA in the CCA tract (right panel) obtained at the first visit of MS patients. FA-CCA profiles are registered to a dense grid of 93 points.

regression methods in predicting the distribution function of PASAT scores conditional on the FA-CCA profiles. Let $Q_{Y_i|\mathcal{X}_i}(\tau)$ denote the 100τ%th quantile function of PASAT score conditional on the FA-CCA profiles. Then, for the indicator function $\mathbb{I}_i(\tau) = \mathbb{I}[Y_i \leq Q_{Y_i|\mathcal{X}_i}(\tau)]$, it is expected that $\mathrm{E}[\mathbb{I}_i(\tau)|\mathcal{X}_i(t)] = \tau$. For a set of τ levels, $\tau \in \{0.05 \times s : s = 1, \ldots, 19\}$, we obtain the predicted 100τ%th quantile function of PASAT score conditional on the FA-CCA profiles, $\widehat{Q}_{Y_i|\mathcal{X}_i}(\tau)$, for each quantile regression methods (i.e. fPCA, fPQR, and RfPQR). Then, for each of the methods, we compute the following measure:

$$\widehat{\mathbb{I}}(\tau) = \frac{1}{n^*} \sum_{i=1}^{n^*} \mathbb{I}[Y_i^* \leq \widehat{Q}_{Y_i|\mathcal{X}_i}(\tau)],$$

where $n^* = 62$ (i.e. the number of observations after discarding the outlying points) and Y_i^* are the PASAT score values that do not contain the observations corresponding to the outlying functional predictors. $\widehat{\mathbb{I}}(\tau)$ values close to τ can be interpreted as an indication that the quantile regression method produces reasonable estimates.

The scatter plots of $\mathbb{I}(\tau)$ vs $\widehat{\mathbb{I}}(\tau)$ are presented in Figure 4. From this figure, while it seems that all the methods produce $\widehat{\mathbb{I}}(\tau)$ values close to the identity line, a closer look at the plots indicates that the proposed method produces improved $\widehat{\mathbb{I}}(\tau)$ values compared with the fPCA and fPQR.

To compare the methods, we compute the mean absolute error (MAE) between $\mathbb{I}(\tau)$ and $\widehat{\mathbb{I}}(\tau)$, i.e. we compute

$$\mathrm{MAE} = \frac{1}{19} \sum_{s=1}^{19} |\mathbb{I}(\tau_s) - \widehat{\mathbb{I}}(\tau_s)|, \quad \tau_s = 0.05 \times s.$$

According to our results, the MAE values are 0.0198, 0.0239, and 0.0147 for the fPCA, fPQR, and RfPQR methods, respectively. These results indicate that our proposed method produces a more accurate distribution function for the PASAT scores conditional on the FA-CCA profiles.

To assess the individual PASAT scores, we use PASAT score 32 as a reference (we consider this reference value for all age groups based on the results of [35,38]). Based on the reference value, ten individuals (individuals with IDs $= [6, 9, 22, 29, 34, 41, 46, 55, 64, 65]$) with PASAT scores less than 32. In Figure 5, we present the estimated conditional quantiles of PASAT scores obtained by the proposed RfPQR method and actual PASAT scores for four individuals (IDs $= [6, 22, 34, 41]$). These individuals have PASAT scores less than the

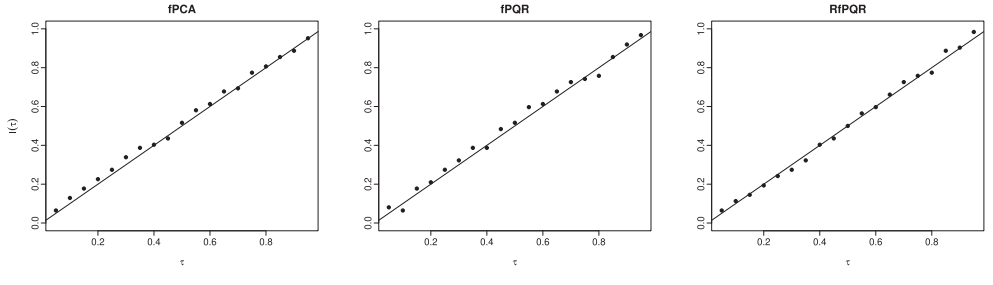

Figure 4. Scatter plots of $\mathbb{I}(\tau)$ vs. $\widehat{\mathbb{I}}(\tau)$ for fPCA (left panel), fPQR (middle panel), and RfPQR (right panel).

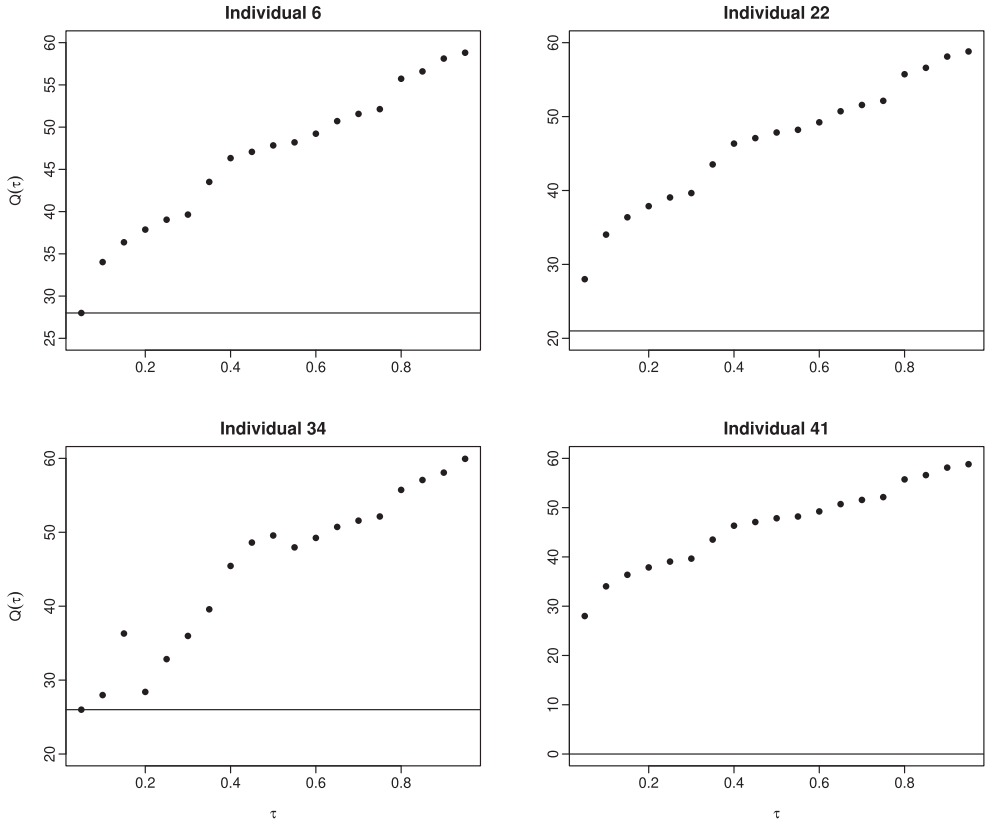

Figure 5. Estimated conditional quantiles of PASAT scores obtained by the proposed RfPQR method. (—– denotes the actual PASAT score).

estimated conditional quantiles of PASAT scores by the proposed method at level $\tau = 0.1$, and they may need an additional neurological follow-up.

7. Conclusion

The fQR has become a general framework to characterize the entire conditional distribution of a scalar response variable for a given functional predictor. Compared with fLM, the fQR has several important advantages, such as robustness to outliers in the response variable and heteroscedasticity. On the other hand, it is susceptible to leverage points, i.e. outlying observations in the predictor space. In the case of leverage points, the existing strategies used to estimate the conditional quantiles may produce biased parameter estimates and poor prediction results for the conditional quantiles of the scalar response. This paper proposes an estimation strategy based on the fPQR approach, robust to both outliers and leverage points. In our method, the fPQR components are obtained via a weighted covariance function, and the effects of outliers and leverage points are down-weighted using an iterative reweighting algorithm. The estimation and predictive performance of the proposed method are investigated via a series of Monte-Carlo experiments and

empirical data analysis. The results produced by the proposed method are favorably compared with those of several existing methods. The results obtained from our numerical analyses indicate that the proposed method produces similar empirical performance to existing quantile regression methods when no outliers are present in the data or when outliers are present in the scalar response. On the other hand, the proposed method produces improved empirical performance over the existing methods when leverage points contaminate the dataset.

The current methodology can be extended further by the following five research directions: (1) In the proposed method, we consider the unpenalized B-spline basis expansion coefficients to approximate the infinite-dimensional fQR in the finite-dimensional space. However, when the unpenalized basis expansion coefficients are used in the model, optimization without imposing restrictions on the function may cause over-fitting. In addition, the regression parameter function approximated by the unpenalized basis expansion coefficients may be strong under-smoothing when the regression parameter function is much smoother than the higher-order fPQR components since the smoothness of the function parameter is induced by choice of basis dimension of the functional predictor [see, e.g. 28]. A penalized version of the current methodology can be proposed similar to those of [2] to control the smoothness degree of the regression parameter function and make a better inference about the fQR model. (2) In the current study, we only consider a single functional predictor. The proposed method can easily be extended to multiple functional predictors cases similar to [5]. (3) In the fQR model, we only consider functional predictors. However, in real-world problems, a mixed data type, including both functional and scalar predictors, may be needed to obtain improved estimates for the conditional quantiles for the scalar response. The proposed method can easily be extended to partially functional linear quantile regression models, which include both functional and scalar predictors, as an alternative to [34,49–51]. (4) The present methodology can also be extended to the function-on-function linear quantile regression models as an alternative to [4,6]. (5) In our method, we consider the PQR algorithm of [16] to obtain the fPQR components. Alternative PQR algorithms, including the ones proposed by [12,32,37], can be implemented into the proposed method.

Acknowledgments

We gratefully acknowledge the comments of two anonymous referees, which led to an improved version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant (grant no: 120F270).

ORCID

Ufuk Beyaztas http://orcid.org/0000-0002-5208-4950
Mujgan Tez http://orcid.org/0000-0002-8633-1980

Han Lin Shang http://orcid.org/0000-0003-1769-6430

References

[1] J. Adrover, R.A. Maronna, and V.J. Yohai, *Robust regression quantiles*, J. Stat. Plan. Inference. 122 (2004), pp. 187–202.

[2] A.M. Aguilera, M. Aguilera-Morillo, and C. Preda, *Penalized versions of functional PLS regression*, Chemometr. Intell. Lab. Syst. 154 (2016), pp. 8–92.

[3] A.M. Aguilera, M. Escabias, C. Preda, and G. Saporta, *Using basis expansions for estimating functional PLS regression applications with chemometric data*, Chemometr. Intell. Lab. Syst. 104 (2010), pp. 289–305.

[4] U. Beyaztas and H.L. Shang, *Function-on-function linear quantile regression*, Math. Modell. Anal. 27 (2022), pp. 322–341.

[5] U. Beyaztas and H.L. Shang, *A robust functional partial least squares for scalar-on-multiple-function regression*, J. Chemom. 36 (2022), p. e3394.

[6] U. Beyaztas, H.L. Shang, and A. Alin, *Function-on-function partial quantile regression*, J. Agricultural, Biol. Environmental Statist. 27 (2022), pp. 149–174.

[7] C. Capezza, A. Lepore, A. Menafoglio, B. Palumbo, and S. Vantini, *Control charts for monitoring ship operating conditions and CO2 emissions based on scalar-on-function regression*, Appl. Stoch. Models. Bus. Ind. 36 (2020), pp. 477–500.

[8] H. Cardot, C. Crambes, and P. Sarda, *Quantile regression when the covariates are functions*, J. Nonparametr. Stat. 17 (2005), pp. 841–856.

[9] H. Cardot, C. Crambes, and P. Sarda, *Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates*, in *Statistical Methods for Biostatistics and Related Fields*, W. Hardle, Y. Mori, and P. Vieu, eds., Springer, Berlin, 2007, pp. 221–243.

[10] M. Chaouch, A.A. Bouchentouf, A. Traore, and A. Rabhi, *Single functional index quantile regression under general dependence structure*, J. Nonparametr. Stat. 32 (2020), pp. 725–755.

[11] K. Chen and H.G. Müller, *Conditional quantile analysis when covariates are functions, with application to growth data*, J. Royal Statist. Soc.: Ser. B 74 (2012), pp. 67–89.

[12] J.E. Choi and D.W. Shin, *Quantile correlation coefficient: a new tail dependence measure*, Statist. Papers 63 (2022), pp. 1075–1104.

[13] C. Crambes, A. Gannoun, and Y. Henchiri, *Weak consistency of the support vector machine quantile regression approach when covariates are functions*, Statist. Probab. Lett. 81 (2011), pp. 1847–1858.

[14] C. Crambes, A. Gannoun, and Y. Henchiri, *Support vector machine quantile regression approach for functional data: simulation and application studies*, J. Multivar. Anal. 121 (2013), pp. 50–68.

[15] A. Cuevas, *A partial overview of the theory of statistics with functional data*, J. Stat. Plan. Inference. 147 (2014), pp. 1–23.

[16] Y. Dodge and J. Whittaker, *Partial quantile regression*, Metrika 70 (2009), pp. 35–57.

[17] J.J. Dziak, D.L. Coffman, M. Reimherr, J. Petrovich, R. Li, S. Shiffman, and M.P. Shiyko, *Scalar-on-function regression for predicting distal outcomes from intensively gathered longitudinal data: interpretability for applied scientists*, Statist. Survey 36 (2019), pp. 150–180.

[18] F. Ferraty, A. Rabhi, and P. Vieu, *Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon*, Sankhya: Indian J. Statist. 67 (2005), pp. 378–398.

[19] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis*, Springer, New York, 2006.

[20] J. Goldsmith, J. Bobb, C.M. Crainiceanu, B. Caffo, and D. Reich, *Penalized functional regression*, J. Comput. Graph. Stat. 20 (2011), pp. 830–851.

[21] J. Goldsmith, F. Scheipl, L. Huang, J. Wrobel, C. Di, J. Gellar, J. Harezlak, M.W. McLean, B. Swihart, L. Xiao, C. Crainiceanu, and P.T. Reiss, *Refund: regression with functional data*, R package version 0.1-22, 2020. Available at https://CRAN.R-project.org/package=refund.

[22] S. Greven, C.M. Crainiceanu, and D. Reich, *Longitudinal functional principal component analysis*, Electron. J. Stat. 4 (2010), pp. 1022–1054.

[23] S. Greven and F. Scheipl, *A general framework for functional regression modelling*, Stat. Modell. 17 (2017), pp. 1–35.
[24] D.M.A. Gronwall, *Paced auditory serial-addition task: a measure of recovery from concussion*, Percept. Mot. Skills. 44 (1977), pp. 367–373.
[25] L. Hao and D.Q. Naiman, *Quantile Regression*, Sage Publications, California, 2007.
[26] L. Horvath and P. Kokoszka, *Inference for Functional Data with Applications*, Springer, New York, 2012.
[27] T. Hsing and R. Eubank, *Theoretical Foundations of Functional Data Analysis, with An Introduction to Linear Operators*, John Wiley & Sons, Chennai, India, 2015.
[28] A.E. Ivanescu, A.M. Staicu, F. Scheipl, and S. Greven, *Penalized function-on-function regression*, Comput. Stat. 30 (2015), pp. 539–568.
[29] K. Kato, *Estimation in functional linear quantile regression*, Ann. Statist. 40 (2012), pp. 3108–3136.
[30] R. Koenker, *Quantile Regression*, Cambridge University Press, New York, 2005.
[31] P. Kokoszka and M. Reimherr, *Introduction to Functional Data Analysis*, CRC Press, Boca Raton, 2017.
[32] G. Li, Y. Li, and C.L. Tsai, *Quantile correlations and quantile autoregressive modeling*, J. Amer. Statist. Assoc.: Theory Methods 110 (2015), pp. 246–261.
[33] M. Li, K. Wang, A. Maity, and A.M. Staicu, *Inference in functional linear quantile regression*, J. Multivar. Anal. 190 (2022), p. 104985.
[34] H. Ma, T. Li, H. Zhu, and Z. Zhu, *Quantile regression for functional partially linear model in ultra-high dimensions*, Comput. Statist. Data Anal. 129 (2019), pp. 135–147.
[35] J.A. Matias-Guiu, A. Cortes-Martinez, P. Montero, V. Pytel, T. Moreno-Ramos, M. Jorquera, M. Yus, J. Arrazola, and J. Matias-Guiu, *Structural mri correlates of pasat performance in multiple sclerosis*, BMC. Neurol. 18 (2018), p. 214.
[36] M.W. McLean, G. Hooker, A.M. Staicu, F. Scheipl, and D. Ruppert, *Functional generalized additive models*, J. Comput. Graph. Stat. 23 (2014), pp. 249–269.
[37] A. Mendez-Civieta, M.C. Aguilera-Morillo, and R.E. Lillo, *Fast partial quantile regression*, Chemometr. Intell. Lab. Syst. 223 (2020), p. 104533.
[38] S. Ozakbas, B.P. Cinar, M.A. Gurkan, O. Ozturk, D. Oz, and B.B. Kursun, *Paced auditory serial addition test: national normative data*, Clin. Neurol. Neurosurg. 140 (2016), pp. 97–99.
[39] A. Ozturk, S.A. Smith, E.M. Gordon-Lipkin, D.M. Harrison, N. Shiee, D.L. Pham, B.S. Caffo, P.A. Calabresi, and D.S. Reich, *Mri of the corpus callosum in multiple sclerosis: association with disability*, Multiple Sclerosis 16 (2010), pp. 166–177.
[40] C. Preda and G. Saporta, *PLS regression on a stochastic process*, Comput. Stat. Data. Anal. 48 (2005), pp. 149–158.
[41] J.O. Ramsay and B.W. Silverman, *Functional Data Analysis*, Springer, New York, 2006.
[42] P.T. Reiss, J. Goldsmith, H.L. Shang, and R.T. Odgen, *Methods for scalar-on-function regression*, Int. Statist. Rev. 85 (2017), pp. 228–249.
[43] M. Rodriguez, A. Siva, J. Ward, K. Stolp-Smith, P. O'Brien, and L. Kurland, *Impairment, disability, and handicap in multiple sclerosis: a population-based study in Olmsted County, Minnesota*, Neurology 44 (1994), pp. 28–33.
[44] P. Sang and J. Cao, *Functional single-index quantile regression models*, Stat. Comput. 30 (2020), pp. 771–781.
[45] S. Serneels, C. Croux, P. Filzmoser, and P.J.V. Espen, *Partial robust M-regression*, Chemometr. Intell. Lab. Syst. 79 (2005), pp. 55–64.
[46] A.M. Staicu, C.M. Crainiceanu, D.S. Reich, and D. Ruppert, *Modeling functional data with spatially heterogeneous shape characteristics*, Biometrics 68 (2012), pp. 331–343.
[47] Y. Sun and M.G. Genton, *Functional boxplots*, J. Comput. Graph. Stat. 20 (2011), pp. 316–334.
[48] B.J. Swihart, J. Goldsmith, and C.M. Crainiceanu, *Restricted likelihood ratio test for functional effects in the functional linear model*, Technometrics 56 (2014), pp. 483–493.

[49] Q. Tang and L. Cheng, *Partial functional linear quantile regression*, Sci. China Math. 57 (2014), pp. 2589–2608.
[50] F. Yao, S. Sue-Chee, and F. Wang, *Regularized partially functional quantile regression*, J. Multivar. Anal. 156 (2017), pp. 39–56.
[51] D. Yu, L. Kong, and I. Mizera, *Partial functional linear quantile regression for neuroimaging data analysis*, Neurocomputing 195 (2016), pp. 74–87.
[52] Z. Zhou, *Fast implementation of partial least squares for function-on-function regression*, J. Multivar. Anal. 185 (2021), p. 104769.