Simultaneous Spline Quantile Regression Under Shape Constraints

Daichi Kitahara*, Ke Leng*, Yuji Tezuka†, and Akira Hirabayashi*
*Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga, Japan
†Department of Diabetes and Endocrinology, Kusatsu General Hospital, Shiga, Japan

***Abstract*—As data analysis methods, hypothesis testing and regression analysis are famous. However, the hypothesis testing can only detect *significant differences* between two groups divided by some characteristic or some empirical threshold, and the regression analysis can only construct one *averaged model* whose information is limited. Quantile regression is a robust and flexible analysis method, and can construct *multilevel models*, e.g., the median and the first and third quartiles. To make the most of the quantile regression, existing papers employed spline regression models as generalizations of polynomial regression models, but the regression of each level was individually executed. In this paper, we propose simultaneous spline quantile regression which considers the similarity between the adjacent quantiles. Further, the proposed method enforces the non-crossing and one shape (non-decreasing/non-increasing/convex/concave) constraints. Experiments demonstrate that the proposed method recovers harmonious quantiles.**

***Index Terms*—Quantile regression, spline function, simultaneous regression, shape constraint, convex optimization.**

I. Introduction

Data analysis [1] is becoming more and more important in the big data era. In the simplest case, we analyze a pair of random variables from bivariate observations. One famous analysis method is hypothesis testing [2]–[4]. In this approach, we first divide the observations into two groups by using a certain empirically determined threshold on one random variable, and then we check whether or not the distributions of the other random variable are *significantly different between the two groups*. However, the hypothesis testing cannot detect small differences between the two groups and any difference within one group.

Another famous method is regression analysis [1], [2], [5]–[7]. In this approach, we create a univariate continuous function which maps one random variable to the other one. Thus, we can analyze the continuous relation between the two random variables. This regression function is often constructed as a low-order polynomial having the *least square errors* due to the simplicity. In robust statistics [4], [8]–[11], a polynomial having the *least absolute errors*, which leads to the regression of the median, is used for suppressing the influence of outliers.

Quantile regression [12]–[15], which is seen as a generalization of the median regression, enables robust and flexible analysis. In this approach, by minimizing certain *asymmetric absolute errors*, we can construct multiple *quantile lines*, i.e., *percentile lines*. Therefore, we can realize continuous and two-dimensional analysis using the multilevel regression results. Although the quantile regression is an effective analysis method, there is a possibility that a simple regression model such as a polynomial cannot approximate the true quantiles enough. To make the most of the quantile regression, a spline regression model is adopted [16]–[22]. Splines are defined as piecewise polynomials and have been widely utilized for construction of smooth functions, including regression analysis, due to their flexibility and optimality [16]. However, in the existing methods [14], [15], [22], the quantile of each level was individually estimated even though all the quantiles are defined from only one conditional probability density function (see Sect. II-B).

This work was supported in part by JSPS Grants-in-Aid Grant Number for Early-Career Scientists JP19K20361 (e-mail: d-kita@fc.ritsumei.ac.jp).

In this paper, we propose a novel spline quantile regression technique. Differently from the methods in [14], [15], [22], we simultaneously estimate multilevel quantiles while considering the smoothness of the conditional probability density function, which makes the first derivatives of the adjacent quantiles similar. Further, we enforce the *non-crossing* condition [23]–[26], which the true quantiles always satisfy, and an optional shape constraint such as *non-decreasing*, *non-increasing*, *convex*, or *concave* property [27]–[30], which linear or quadratic polynomials have. Numerical experiments demonstrate that the proposed quantiles are more harmonious than the previous ones.

II. Preliminaries

Let $\mathbb{R}$ and $\mathbb{N}$ be the sets of all real numbers and nonnegative integers, respectively. For $\rho \in \mathbb{N} \cup \{\infty\}$ and any open interval $(a, b) \subseteq \mathbb{R}$, $C^{\rho}(a, b)$ stands for the set of all ρ-times continuously differentiable real-valued functions on (a, b). For any $d \in \mathbb{N}$, $\mathbb{P}_d$ ($\subsetneq C^{\infty}(\mathbb{R})$) denotes the set of all univariate real polynomials of degree d at most, i.e., $\mathbb{P}_d := \{u : \mathbb{R} \to \mathbb{R} : x \mapsto \sum_{k=0}^{d} c_k x^k \mid c_k \in \mathbb{R}\}$. Boldface small and capital letters express vectors and matrices, respectively.

A. Regression Analysis

Suppose that we have finite observations $(x_i, y_i) \in \mathbb{R}^2$ ($i = 1, 2, \ldots, n$) of a pair of real-valued random variables (X, Y), where the joint probability density function $f_{X,Y}(x, y)$ satisfies $\iint_{\Omega} f_{X,Y}(x, y)\,\mathrm{d}x\mathrm{d}y = 1$ and $f_{X,Y}(x, y) > 0$ for all (x, y) in a certain rectangular domain $\Omega := (x_{\inf}, x_{\sup}) \times (y_{\inf}, y_{\sup}) \subseteq \mathbb{R}^2$. Therefore, the *conditional probability density function of Y given X* is defined by $f_{Y|X}(y|x) := f_{X,Y}(x, y)/f_X(x) := f_{X,Y}(x, y)/\int_{y_{\inf}}^{y_{\sup}} f_{X,Y}(x, y)\,\mathrm{d}y > 0$ for all $(x, y) \in \Omega$.

When we want to analyze a continuous relation between the two random variables X and Y, the least squares regression

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \sum_{i=1}^{n} |y_i - r_{\boldsymbol{\theta}}(x_i)|^2 \tag{1}$$

is often used due to its low computational cost [1], [2], [5]–[7], where the function $r_{\boldsymbol{\theta}}(x)$ is some regression model such as a polynomial and the vector $\boldsymbol{\theta}$ denotes adjustable parameters to be optimized. The optimal solution $r_{\boldsymbol{\theta}^*}(x)$ of (1) expresses one average relation because, as n approaches infinity, $r_{\boldsymbol{\theta}^*}(x)$ converges to the *conditional mean* $\mu_Y(x)$ *of* Y *given* $X = x$:

$$r_{\boldsymbol{\theta}^*}(x) \to \mu_Y(x) := E[Y|X = x] = \int_{y_{\inf}}^{y_{\sup}} y f_{Y|X}(y|x)\,\mathrm{d}y$$

under the assumption that $r_{\boldsymbol{\theta}}(x)$ can exactly express $\mu_Y(x)$ if we choose the appropriate $\boldsymbol{\theta}$ (see, e.g., [1] for proof).

It is well-known that the square error as in (1) is sensitive to outliers and hence the reliability of the optimal solution $r_{\boldsymbol{\theta}^*}(x)$ of (1) significantly decreases for long-tailed data [4], [8]–[11]. In such situations, the least absolute deviations regression

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \sum_{i=1}^{n} |y_i - r_{\boldsymbol{\theta}}(x_i)| \tag{2}$$

is utilized [10], [11]. Since the absolute error as in (2) does not over-evaluate the outliers differently from the square error, this regression is robust for long-tailed data even if n is not large. As n approaches infinity, the optimal solution $r_{\boldsymbol{\theta}^*}(x)$ of (2) converges to the *conditional median* $m_Y(x)$ *of* Y *given* $X = x$:

$$r_{\boldsymbol{\theta}^*}(x) \to m_Y(x) \text{ satisfying } \int_{y_{\inf}}^{m_Y(x)} f_{Y|X}(y|x)\,\mathrm{d}y = 0.5$$

under the assumption that $r_{\boldsymbol{\theta}}(x)$ can exactly express $m_Y(x)$ if we choose the appropriate $\boldsymbol{\theta}$ (see, e.g., [10] for proof).

B. Quantile Regression

By generalizing the fact that the least absolute deviations regression leads to the conditional median as shown in Sect. II-A, we can estimate any *quantile line* as follows. Define the *conditional cumulative distribution function of* Y *given* $X = x$ by

$$F_{Y|x}(y) := \int_{y_{\inf}}^{y} f_{Y|X}(t|x)\,\mathrm{d}t \quad \text{for } y \in (y_{\inf}, y_{\sup}).$$

$F_{Y|x}(y)$ becomes a strictly increasing function due to the positivity of $f_{Y|X}(y|x)$, and its inverse function $F_{Y|x}^{-1}(p)$ is well-defined for $p \in (0, 1)$. Actually, the *conditional quantile function of* Y *given* $X = x$ is equivalent to $F_{Y|x}^{-1}(p)$ [2], [31]:

$$Q_{Y|x}(p) := F_{Y|x}^{-1}(p) \quad \text{for } p \in (0, 1).$$

The value of $q_{p,Y}(x) := Q_{Y|x}(p)$ s.t. $x \in (x_{\inf}, x_{\sup})$ is called the *pth conditional quantile of* Y *given* $X = x$. Note that the pth quantile $q_{p,Y}(x)$ is also called the *$100p$th percentile* or *centile*, e.g., in [32]–[34]. When $p = 0.5$, the quantile $q_{p,Y}(x)$ is equivalent to the conditional median $m_Y(x)$ of Y given $X = x$.

For $p \in (0, 1)$, define an *asymmetric absolute value function*

$$J_p(t) := \begin{cases} pt & \text{if } t \geq 0, \\ -(1-p)t & \text{if } t < 0, \end{cases}$$

and consider the following optimization problem [12]–[15]

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \sum_{i=1}^{n} J_p(y_i - r_{\boldsymbol{\theta}}(x_i)). \tag{3}$$

Then, as n approaches infinity, the optimal solution $r_{\boldsymbol{\theta}^*}(x)$ of (3) converges to the pth conditional quantile $q_{p,Y}(x)$ of Y:

$$r_{\boldsymbol{\theta}^*}(x) \to q_{p,Y}(x) \text{ satisfying } \int_{y_{\inf}}^{q_{p,Y}(x)} f_{Y|X}(y|x)\,\mathrm{d}y = p$$

under the assumption that $r_{\boldsymbol{\theta}}(x)$ can exactly express $q_{p,Y}(x)$ if we choose the appropriate $\boldsymbol{\theta}$ (see, e.g., [14] for proof). This is because the solution $r_{\boldsymbol{\theta}^*}(x)$ of (3) yields lower-side errors $e_i = y_i - r_{\boldsymbol{\theta}}(x_i) < 0$ and upper-side errors $e_i = y_i - r_{\boldsymbol{\theta}}(x_i) > 0$, at a ratio of $p : 1 - p$. Therefore, we can easily estimate any quantile line $q_{p,Y}(x)$ only by changing p in (3).

C. Spline Function

Let $\sqcup_b := \{I_i := (\xi_{i-1}, \xi_i)\}_{i=1}^{b}$ be a set of b subintervals I_i on an open interval $I := (\xi_0, \xi_b)$ ($\subsetneq \mathbb{R}$) s.t. $\xi_i - \xi_{i-1} =: h_i > 0$ ($i = 1, 2, \ldots, b$). For $\sqcup_b$ and $\rho, d \in \mathbb{N}$ s.t. $0 \leq \rho \leq d$, define

$$\mathcal{S}_d^{\rho}(\sqcup_b) := \{s \in C^{\rho}(I) \mid s = u_i \in \mathbb{P}_d \text{ on } I_i \in \sqcup_b\}$$

as the set of all univariate spline functions of degree d at most and smoothness ρ on $\sqcup_b$. In this paper, we express each spline function $s \in \mathcal{S}_d^{\rho}(\sqcup_b)$ in the interval normalization form:

$$s(x) := u_i(x) := \sum_{k=0}^{d} c_k^{\langle i \rangle} \left(\frac{x - \xi_{i-1}}{h_i}\right)^k \quad \text{for } x \in I_i, \tag{4}$$

where $c_k^{\langle i \rangle} \in \mathbb{R}$ ($k = 0, 1, \ldots, d$) are coefficients of $u_i \in \mathbb{P}_d$. Spline functions are widely utilized to construct smooth functions, e.g., for interpolation, computer aided design, and regression analysis [16]–[22], due to the following optimality.

Fact 1 (Spline as the unique solution of a variational problem): There exists the unique solution of the following problem

$$\underset{g \in C^2(\mathbb{R})}{\text{minimize}} \sum_{i=1}^{n} |y_i - g(x_i)|^2 + \lambda \int_{\mathbb{R}} |g''(x)|^2\,\mathrm{d}x, \tag{5}$$

and it is a natural cubic spline, which is a kind of spline function of degree 3 at most and smoothness 2 [16], where $g''(x)$ is the second derivative of $g(x)$. In the optimization problem of (5), the smoothing parameter $\lambda > 0$ controls the trade-off between the data fidelity and the smoothness of the solution.

By using the coefficients $c_k^{\langle i \rangle}$ in (4), we can easily evaluate the characteristics of spline functions as follows.

1) Quadratic Form of the Roughness Penalty Term: In (5), if we restrict the domain of interest from $\mathbb{R} = (-\infty, \infty)$ to $I = (\xi_0, \xi_b)$ ($\supsetneq (x_{\min}, x_{\max}) := (\min\{x_i\}, \max\{x_i\})$) and the function space from $C^2(\mathbb{R})$ to $\mathcal{S}_d^{\rho}(\sqcup_b)$ s.t. $2 \leq \rho \leq d$, then the roughness penalty term used in (5) can be decomposed as

$$\int_I |s''(x)|^2\,\mathrm{d}x = \sum_{i=1}^{b} \int_{I_i} |s''(x)|^2\,\mathrm{d}x. \tag{6}$$

By using the expression in (4), each roughness penalty on I_i is expressed as the following quadratic form

$$\int_{I_i} |s''(x)|^2\,\mathrm{d}x = \sum_{k=2}^{d} \sum_{j=2}^{d} \frac{k(k-1)j(j-1)}{h_i^3(k+j-3)} c_k^{\langle i \rangle} c_j^{\langle i \rangle}. \tag{7}$$

From (6) and (7), the total roughness penalty on I can be expressed as a certain quadratic form by $\int_I |s''(x)|^2 \,\mathrm{d}x = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{c}$, where $\boldsymbol{c} = (c_d^{\langle 1\rangle}, c_{d-1}^{\langle 1\rangle}, \ldots, c_0^{\langle 1\rangle}, c_d^{\langle 2\rangle}, c_{d-1}^{\langle 2\rangle}, \ldots, c_0^{\langle 2\rangle}, \ldots, c_0^{\langle b\rangle})^{\mathrm{T}} \in \mathbb{R}^{b(d+1)}$ is the coefficient vector of a spline $s \in \mathcal{S}_d^\rho(\sqcup_b)$ and $\boldsymbol{Q} \in \mathbb{R}^{b(d+1)\times b(d+1)}$ is a positive semidefinite matrix.

2) Linear Equation for the ρ-Times Differentiability: For a spline $s \in \mathcal{S}_d^\rho(\sqcup_b)$ in (4), to guarantee the ρ-times continuous differentiability over $I = (\xi_0, \xi_b)$, i.e., $s \in C^\rho(I)$, every pair of the coefficients of the adjacent polynomial pieces $u_i \in \mathbb{P}_d$ and $u_{i+1} \in \mathbb{P}_d$ has to satisfy the following linear equation

$$\frac{1}{h_i^j}\sum_{k=j}^{d}\frac{k!}{(k-j)!}c_k^{\langle i\rangle} - \frac{j!}{h_{i+1}^j}c_j^{\langle i+1\rangle} = 0 \quad (j = 0, 1, \ldots, \rho)$$
$$\Leftrightarrow s \in C^\rho(\xi_{i-1}, \xi_{i+1}). \qquad (8)$$

From (8), there exists some matrix $\boldsymbol{H} \in \mathbb{R}^{(b-1)(\rho+1)\times b(d+1)}$ satisfying $\boldsymbol{H}\boldsymbol{c} = \boldsymbol{0} \Leftrightarrow s \in C^\rho(I)$.

3) Linear Inequality for the Nonnegativity: In [35] and [36], the first author estimated probability density functions by spline smoothing. It was very difficult to give a useful necessary and sufficient condition for the nonnegativity of $s \in \mathcal{S}_d^\rho(\sqcup_b)$ over each I_i. Instead, the first author used a sufficient condition in [37], which can be expressed as the following linear inequality

$$\sum_{k=0}^{j}\frac{(d-k)!}{(j-k)!(d-j)!}c_k^{\langle i\rangle} \geq 0 \quad (j = 0, 1, \ldots, d)$$
$$\Rightarrow s(x) \geq 0 \text{ for all } x \in I_i. \qquad (9)$$

From (9), there exists some matrix $\boldsymbol{G} \in \mathbb{R}^{b(d+1)\times b(d+1)}$ satisfying $\boldsymbol{G}\boldsymbol{c} \geq \boldsymbol{0} \Rightarrow s(x) \geq 0$ for all $x \in I$.

D. Quantile Regression via Spline Smoothing

In the problems of (1), (2), and (3), the most commonly used regression model is a polynomial $r_{\boldsymbol{\theta}}(x) = \sum_{k=0}^{d} c_k x^k$ of degree $d = 1$ or $d = 2$ [1], [2], [5]–[7]. In this situation, coefficients of the polynomial are adjustable parameters, i.e., $\boldsymbol{\theta} = (c_d, c_{d-1}, \ldots, c_0)^{\mathrm{T}} \in \mathbb{R}^{d+1}$. However, there is a high probability that such a simple model cannot approximate the true conditional quantile lines $q_{p,Y}(x)$ enough.

To deal with quantiles of various shapes flexibly, we employ a spline regression model $r_{\boldsymbol{\theta}} = s \in \mathcal{S}_d^\rho(\sqcup_b)$ as a generalization of the polynomial regression model [14], [15], [22]. In this situation, the adjustable parameters become the coefficient vector $\boldsymbol{\theta} = \boldsymbol{c} \in \mathbb{R}^{b(d+1)}$ of $s \in \mathcal{S}_d^\rho(\sqcup_b)$ in Sect. II-C-1. Although the spline regression model is really flexible, overfitting will arise when the number n of observations is not enough. Therefore, by assuming that the energy of local change of $q_{p,Y}(x)$ is small and using the roughness penalty as a regularization term as in (5), we solve, instead of (3), the following problem

$$\underset{s \in \mathcal{S}_d^\rho(\sqcup_b)}{\text{minimize}} \sum_{i=1}^{n} J_p(y_i - s(x_i)) + \lambda \int_I |s''(x)|^2 \,\mathrm{d}x, \qquad (10)$$

where $\lambda > 0$. Note that when $\rho = d$ and λ approaches $+0$ in (10), the solution $s^*(x)$ converges to the optimal polynomial regression model in (3). By repeatedly solving the problem of (10) for different $p = p_l$ (s.t. $p_1 < p_2 < \cdots < p_L$), we can estimate and construct the quantiles of L levels.

III. Simultaneous Quantile Regression with Shape-Constrained Splines

In the basic spline quantile regression [14], [15], [22] of (10), the *non-crossing* conditions as $\forall x\ q_{p_{l+1},Y}(x) > q_{p_l,Y}(x)$ ($l = 1, 2, \ldots, L-1$) are ignored. Although these conditions are satisfied automatically when h_i is small or $p_{l+1} - p_l$ is not small, some papers incorporated them into constraints and guaranteed the non-crossing regression results [23]–[26]. Other papers enforced the *non-decreasing*, *non-increasing*, *convex*, or *concave* property on each spline regression model $s_l(x)$ [27]–[30] since the polynomial regression model of degree 1 is non-decreasing or non-increasing, and that of degree 2 is convex or concave.

In the conventional methods [14], [15], [22]–[30], the quantile $q_{p_l,Y}(x)$ of each level was individually estimated by solving a problem similar to (10), even though all the quantiles are defined from only one conditional probability density function $f_{Y|X}(y|x)$ as mentioned in Sect. II-B. Hence, estimates $s_l^*(x)$ of $q_{p_l,Y}(x)$ ($l = 1, 2, \ldots, L$) do not consider the smoothness of $f_{Y|X}(y|x)$ along the y-axis while the smoothness along the x-axis is considered implicitly by the roughness penalty term in (6). As a result, the similarity of the adjacent quantiles are lost (see, e.g., Fig. 2 in this paper and Fig. 2(a) in [24]).

In this paper, we simultaneously estimate the multiple quantiles by considering the smoothness of $f_{Y|X}(y|x)$ along the y-axis, i.e., the similarity of the adjacent quantiles. Thus we solve

$$\underset{(s_l \in \mathcal{S}_d^\rho(\sqcup_b))_{l=1}^L}{\text{minimize}} \sum_{l=1}^{L}\sum_{i=1}^{n} w_i J_{p_l}(y_i - s_l(x_i))$$
$$+ \lambda\sum_{l=1}^{L}\int_I |s_l''(x)|^2 \,\mathrm{d}x + \kappa\sum_{l=1}^{L-1}\int_I |s_{l+1}'(x) - s_l'(x)|^2 \,\mathrm{d}x$$
$$\text{subject to} \quad \forall x\ s_{l+1}(x) \geq s_l(x) \quad (l = 1, 2, \ldots, L-1)$$
$$(\text{and } \ \forall l\,\forall x\ s_l'(x) \geq 0,\ s_l'(x) \leq 0,\ s_l''(x) \geq 0, \text{ or } s_l''(x) \leq 0), \qquad (11)$$

where $\lambda > 0$, $\kappa > 0$, and $w_i > 0$ is often set to a value similar to $f_X(x_i)$. On the first term, $s_l(x_i)$ can be expressed as $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{c}_l$ with the coefficient vector $\boldsymbol{c}_l = (c_{l,d}^{\langle 1\rangle}, \ldots, c_{l,0}^{\langle 1\rangle}, c_{l,d}^{\langle 2\rangle}, \ldots, c_{l,0}^{\langle b\rangle})^{\mathrm{T}} \in \mathbb{R}^{b(d+1)}$ of $s_l \in \mathcal{S}_d^\rho(\sqcup_b)$ and a certain vector $\boldsymbol{a}_i \in \mathbb{R}^{b(d+1)}$. The second term is expressed as $\lambda\sum_{l=1}^{L}\boldsymbol{c}_l^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{c}_l = \lambda\bar{\boldsymbol{c}}^{\mathrm{T}}\boldsymbol{Q}_1\bar{\boldsymbol{c}}$, from (6) and (7), by using the concatenated coefficient vector $\bar{\boldsymbol{c}} := (\boldsymbol{c}_1^{\mathrm{T}}, \boldsymbol{c}_2^{\mathrm{T}}, \ldots, \boldsymbol{c}_L^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{b(d+1)L}$ and some block diagonal matrix $\boldsymbol{Q}_1 := \mathrm{diag}(\boldsymbol{Q}, \boldsymbol{Q}, \ldots, \boldsymbol{Q}) \in \mathbb{R}^{b(d+1)L\times b(d+1)L}$. From (6) and

$$\int_{I_i} |s_{l+1}'(x) - s_l'(x)|^2 \,\mathrm{d}x = \sum_{k=1}^{d}\sum_{j=1}^{d}\frac{kj}{h_i(k+j-1)}$$
$$\cdot\left(c_{l+1,k}^{\langle i\rangle}c_{l+1,j}^{\langle i\rangle} - c_{l+1,k}^{\langle i\rangle}c_{l,j}^{\langle i\rangle} - c_{l,k}^{\langle i\rangle}c_{l+1,j}^{\langle i\rangle} + c_{l,k}^{\langle i\rangle}c_{l,j}^{\langle i\rangle}\right), \qquad (12)$$

the third term is expressed as $\kappa\bar{\boldsymbol{c}}^{\mathrm{T}}\boldsymbol{Q}_2\bar{\boldsymbol{c}}$ with $\bar{\boldsymbol{c}}$ and some positive semidefinite matrix $\boldsymbol{Q}_2 \in \mathbb{R}^{b(d+1)L\times b(d+1)L}$. From (9), a sufficient condition for the non-crossing constraints is given by

$$\sum_{k=0}^{j}\frac{(d-k)!}{(j-k)!(d-j)!}(c_{l+1,k}^{\langle i\rangle} - c_{l,k}^{\langle i\rangle}) \geq 0 \quad (j = 0, 1, \ldots, d)$$
$$\Rightarrow s_{l+1}(x) \geq s_l(x) \text{ for all } x \in I_i, \qquad (13)$$

that for the non-decreasing property is

$$\sum_{k=0}^{j} \frac{(d-k-1)!(k+1)}{(j-k)!(d-j-1)!} c_{l,k+1}^{\langle i \rangle} \geq 0 \quad (j=0,1,\ldots,d-1)$$
$$\Rightarrow s_l'(x) \geq 0 \text{ for all } x \in I_i, \qquad (14)$$

and that for the convex property is

$$\sum_{k=0}^{j} \frac{(d-k-2)!(k+2)(k+1)}{(j-k)!(d-j-2)!} c_{l,k+2}^{\langle i \rangle} \geq 0 \quad (j=0,1,\ldots,d-2)$$
$$\Rightarrow s_l''(x) \geq 0 \text{ for all } x \in I_i. \qquad (15)$$

The above sufficient conditions in (13), (14), and (15) can be expressed as $\boldsymbol{G}_1\bar{\boldsymbol{c}} \geq \boldsymbol{0}$, $\boldsymbol{G}_2\boldsymbol{c}_l \geq \boldsymbol{0}$, and $\boldsymbol{G}_3\boldsymbol{c}_l \geq \boldsymbol{0}$, respectively, where $\boldsymbol{G}_1 \in \mathbb{R}^{b(d+1)(L-1)\times b(d+1)L}$, $\boldsymbol{G}_2 \in \mathbb{R}^{bd\times b(d+1)}$, and $\boldsymbol{G}_3 \in \mathbb{R}^{b(d-1)\times b(d+1)}$ are appropriately defined matrices, and $\boldsymbol{0}$ is the zero vector. Sufficient conditions for the non-increasing and concave properties are given by $\boldsymbol{G}_2\boldsymbol{c}_l \leq \boldsymbol{0}$ and $\boldsymbol{G}_3\boldsymbol{c}_l \leq \boldsymbol{0}$.

From (6), (7), (8), (12), (13), (14) and (15), the problem of (11) is reduced to the following convex optimization problem

$$\underset{\bar{\boldsymbol{c}}\in\mathbb{R}^{b(d+1)L}}{\text{minimize}} \sum_{l=1}^{L}\sum_{i=1}^{n} w_i J_{p_l}(y_i - \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{c}_l) + \lambda\bar{\boldsymbol{c}}^{\mathrm{T}}\boldsymbol{Q}_1\bar{\boldsymbol{c}} + \kappa\bar{\boldsymbol{c}}^{\mathrm{T}}\boldsymbol{Q}_2\bar{\boldsymbol{c}}$$
$$\text{subject to} \qquad \forall l \;\; \boldsymbol{H}\boldsymbol{c}_l = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{G}_1\bar{\boldsymbol{c}} \geq \boldsymbol{0}$$
$$(\text{and} \quad \forall l \;\; \boldsymbol{G}_2\boldsymbol{c}_l \geq \boldsymbol{0},\, \boldsymbol{G}_2\boldsymbol{c}_l \leq \boldsymbol{0},\, \boldsymbol{G}_3\boldsymbol{c}_l \geq \boldsymbol{0}, \text{ or } \boldsymbol{G}_3\boldsymbol{c}_l \leq \boldsymbol{0}).$$

The optimal solution $\bar{\boldsymbol{c}}^*$ can be obtained, e.g., by the alternating direction method of multipliers (ADMM) [38]. Note that the *proximity operator* of the data fidelity term is computed by

$$\mathrm{prox}_{wJ_p(y-\cdot)}(z) = \begin{cases} z + pw & \text{if } y - z \geq pw, \\ z - (1-p)w & \text{if } y - z \leq -(1-p)w, \\ y & \text{otherwise.} \end{cases}$$

IV. Numerical Experiments

To show the effectiveness of the proposed method, we estimate $L = 51$ quantiles of a conditional probability density

$$f_{Y|X}(y|x) := \frac{1}{\sqrt{2\pi}\widehat{\sigma}y} e^{-\frac{(\log y - \widehat{\mu}(x))^2}{2\widehat{\sigma}^2}} \quad \text{for } y \in (0,\infty) \quad (16)$$

from $n = 1000$ observations (x_i, y_i) by the conventional and proposed methods in (10) and in (11), respectively, where $\widehat{\sigma} = 1$, $p_l = 0.01(l-1)+0.25$ $(l = 1,2,\ldots,51)$, $d = 5$, $\rho = 2$, $b = 40$, $I = (-1,1)$, $h_i = \frac{2}{b} = \frac{1}{20}$ $(i = 1,2,\ldots,40)$, and $x_i \in I$ $(i = 1,2,\ldots,1000)$ are generated from a probability density

$$f_X(x) := \frac{0.3}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \frac{0.7}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

with $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (-0.7, 0.4, 0.3, 0.45)$. First, we define

$$\widehat{\mu}(x) := 0.5\sin(\pi x) + 0.5$$

in (16), and estimate the quantiles $q_{p_l,Y}(x)$ $(l = 1,2,\ldots,51)$ by (10) with $\lambda = 0.04$ and by (11) with $w_i = f_X(x_i)$, $\lambda = 0.0075$ and $\kappa = 25$. Figure 1 shows only 11 quantiles $q_{p_l,Y}(x)$ $(p_l \in \{0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75\})$, where blue circles denote the observations (x_i, y_i), the darkest solid line is the median $q_{0.5,Y}(x)$, and the top and bottom lines are the first and third quartiles $q_{0.25,Y}(x)$, $q_{0.75,Y}(x)$. Figures 2 and 3 shows estimates by (10) and those by (11), respectively. From Figs. 1–3, we observe that some estimates by the conventional method are too close to or too far from each other while those by the proposed method become harmonious like the true quantiles by utilizing the similarity of the adjacent quantiles.

Next, for estimation of convex quantiles, we define

$$\widehat{\mu}(x) := \begin{cases} (x+0.15)^2 + 0.25 & \text{if } x \in (-1,-0.15], \\ 0.25 & \text{if } x \in (-0.15, 0.15], \\ 0.5(x-0.15)^2 + 0.25 & \text{if } x \in (0.15, 1), \end{cases}$$

in (16), and estimate $q_{p_l,Y}(x)$ $(l = 1,2,\ldots,51)$ by (11) with $w_i = f_X(x_i)$, $\lambda = 0.0025$ and $\kappa = 15$ under the convex constraint in (15). Figure 4 shows the convex quantiles $q_{p_l,Y}(x)$ $(p_l \in \{0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75\})$, which are black lines, and estimates by the proposed method, which are red lines. From Fig. 4, we can see that the proposed method reconstructs the harmonious convex quantiles well.

V. Conclusion

In this paper, we proposed a novel spline smoothing technique for simultaneous quantile regression. Differently from the other methods, we estimated multiple quantiles simultaneously by considering the smoothness of a conditional probability density function along the y-axis, i.e., the similarity of the adjacent quantiles. We also considered the non-crossing conditions and a shape (non-decreasing, non-increasing, convex, or concave) constraint. Numerical experiments demonstrated that the proposed method reconstructs multilevel harmonious quantiles.

Acknowledgment

We are very grateful to Dr. Atsunori Kashiwagi and Dr. Osamu Sekine in Kusatsu General Hospital for fruitful discussions.

References

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., ser. Springer Series in Statistics. New York, NY: Springer, 2009.

[2] J. Shao, *Mathematical Statistics*, 2nd ed., ser. Springer Texts in Statistics. New York, NY: Springer, 2003.

[3] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed., ser. Springer Texts in Statistics. New York, NY: Springer, 2008.

[4] R. R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*, 4th ed., ser. Statistical Modeling and Decision Science. London, UK: Academic Press, 2016.

[5] B. Abraham and J. Ledolter, *Introduction to Regression Modeling*, ser. Duxbury Advanced Series. Belmont, CA: Duxbury, 2005.

[6] J. A. Rice, *Mathematical Statistics and Data Analysis*, 3rd ed., ser. Duxbury Advanced Series. Belmont, CA: Duxbury, 2006.

[7] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 5th ed., ser. Wiley Series in Probability and Statistics. New York, NY: Wiley, 2013.

[8] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed., ser. Wiley Series in Probability and Statistics. New York, NY: Wiley, 2009.

[9] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust Statistics: Theory and Methods (with R)*, 2nd ed., ser. Wiley Series in Probability and Statistics. New York, NY: Wiley, 2019.

[10] G. Bassett Jr. and R. Koenker, "Asymptotic theory of least absolute error regression," *J. Am. Stat. Assoc.*, vol. 73, no. 363, pp. 618–622, 1978.

[11] P. Bloomfield and W. L. Steiger, *Least Absolute Deviations: Theory, Applications, and Algorithms*, ser. Progress in Probability and Statistics, P. Huber and M. Rosenblatt Eds. Boston, MA: Birkhäuser, 1983, vol. 6.

[12] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

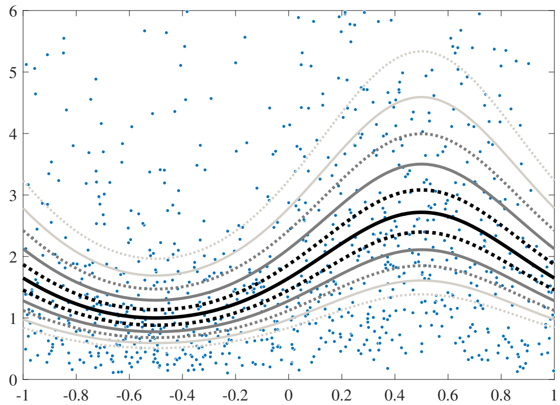

Fig. 1. The $0.25/0.3/0.35/0.4/0.45/0.5/0.55/0.6/0.65/0.7/0.75$th quantiles.

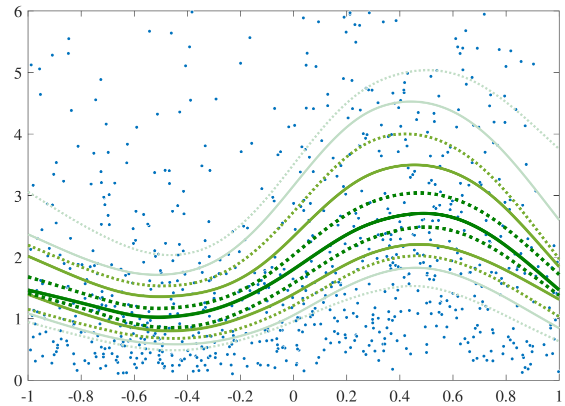

Fig. 2. Regression results by (10) for the quantiles in Fig. 1.

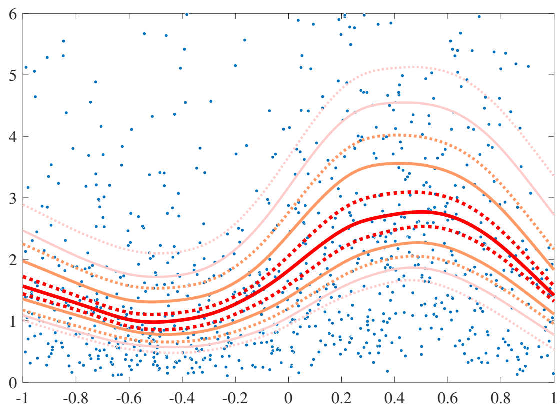

Fig. 3. Regression results by (11) for the quantiles in Fig. 1.

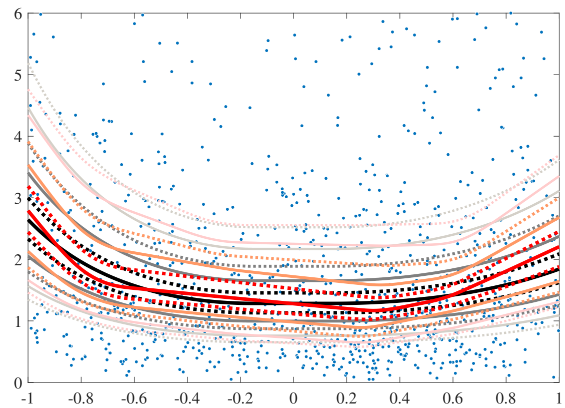

Fig. 4. Regression results by (11) for the convex quantiles.

[13] R. Koenker and K. F. Hallock, "Quantile regression," *J. Econ. Perspect.*, vol. 15, no. 4, pp. 143–156, 2001.

[14] R. Koenker, *Quantile regression*. New York, NY: Cambridge University Press, 2005.

[15] R. Koenker, V. Chernozhukov, X. He, and L. Peng, *Handbook of Quantile Regression*, ser. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. New York, NY: Chapman & Hall, 2017.

[16] C. de Boor, "Best approximation properties of spline functions of odd degree," *J. Math. Mech.*, vol. 12, no. 5, pp. 747–749, 1963.

[17] S. Wold, "Spline functions in data analysis," *Technomet.*, vol. 16, no. 1, pp. 87–111, 1974.

[18] B. W. Silverman, "Some aspects of the spline smoothing approach to non-parametric regression curve fitting," *J. Royal Stat. Soc. B*, vol. 47, no. 1, pp. 1–52, 1985.

[19] G. Wahba, *Spline Models for Observational Data*, ser. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: SIAM, 1990, vol. 59.

[20] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[21] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd ed. New York, NY, USA: Springer, 2005.

[22] R. J. Bosch, Y. Ye, and G. G Woodworth, "A convergent algorithm for quantile regression with smoothing splines," *Comput. Stat. Data Anal.*, vol. 19, no. 6, pp. 613–630, Jun. 1995.

[23] X. He, "Quantile curves without crossing," *Am. Stat.*, vol. 51, no. 2, pp. 186–192, 1997.

[24] I. Takeuchi, Q. V. Le, T. Sears, and A. J. Smola, "Nonparametric quantile regression," *J. Mach. Learn. Res.*, vol. 7, pp. 1231–1264, 2006.

[25] V. M. R. Muggeo, M. Sciandra, A. Tomasello, and S. Calvo, "Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology," *Environ. Ecol. Stat.*, vol. 20, no. 4, pp. 519–531, 2013.

[26] Y. Yuan, N. Chen, and S. Zhou, "Modeling regression quantile process using monotone B-splines," *Technomet.*, vol. 59, no. 3, pp. 338–350, 2017.

[27] X. He and P. Shi, "Monotone B-spline smoothing," *J. Am. Stat. Assoc.*, vol. 93, no. 442, pp. 643–650, 1998.

[28] R. Koenker and P. Ng, "Inequality constrained quantile regression," *Sankhyā: Indian J. Stat.*, vol. 67, no. 2, pp. 418–440, 2005.

[29] K. Bollaerts, P. H. C. Eilers, and M. Aerts, "Quantile regression with monotonicity restrictions using P-splines and the L_1-norm," *Stat. Modelling*, vol. 6, pp. 189–207, 2006.

[30] T. Neocleous and S. Portnoy, "On monotonicity of regression quantile functions," *Stat. Prob. Lett.*, vol. 78, no. 10, pp. 1226–1229, 2008.

[31] G. R. Shorack, *Probability for Statistics*, 2nd ed., ser. Springer Texts in Statistics. New York, NY: Springer, 2017.

[32] R. J. Casady and J. D. Cryer, "Monotone percentile regression," *Ann. Stat.*, vol. 4, no. 3, pp. 532–541, 1976.

[33] D. Griffiths and M. Willcox, "Percentile regression: A parametric approach," *J. Am. Stat. Assoc.*, vol. 73, no. 363, pp. 496–498, 1978.

[34] T. J. Cole, "Fitting smoothed centile curves to reference data," *J. Royal Stat. Soc. A*, vol. 151, no. 3, pp. 385–418, 1988.

[35] D. Kitahara and I. Yamada, "Probability density function estimation by positive quartic C^2-spline functions," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Brisbane, Australia, 2015, pp. 3556–3560.

[36] D. Kitahara and I. Yamada, "Two-dimensional positive spline smoothing and its application to probability density estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Shanghai, China, 2016, pp. 4219–4223.

[37] W. Heß and J. W. Schmidt, "Positive quartic, monotone quintic C^2-spline interpolation in one and two dimensions," *J. Comput. Appl. Math.*, vol. 55, no. 1, pp. 51–67, 1994.

[38] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.