Nonparametric quantile regression models via majorization minimization-algorithm

YUNLU JIANG

In this paper, we apply the Majorization Minimization (MM)-algorithm to deal with the computational problem of the smoothing nonparametric quantile regression. We show that the proposed MM-algorithm possesses the descent property, and the estimator obtained by the proposed algorithm is smooth. Simulation studies demonstrate that the estimator based on our proposed method is more robust and efficient than the estimator based on the mean smoothing regression and the estimator proposed by Nychka et al. (1995) [14] with GCV scores. Finally, we apply the proposed methodology to analyze the dataset about bone density (BMD) in adolescents.

1. INTRODUCTION

Consider the following general nonparametric model

$$Y = m(X) + \epsilon, \tag{1.1}$$

where Y is scalar response variates, X is 1-dimensional covariate variables and lies in compact intervals, $m(\cdot)$ is an unknown smooth function, ϵ is random error. It is assumed that the error ϵ with zero mean and variance $\sigma^2 = E[\epsilon^2]$, and ϵ is independent of X.

It is one of the most important problems to estimate the smooth regression function $m(x)$. There are many methods introduced and extensively studied in the literature. For example, local polynomial mean regression introduced by Fan and Gijbels [5] is very popular method. However, it is sensitive to outliers in the dataset. Koenker and Bassett Jr. [12] first proposed quantile regression. One advantage of quantile regression is that the quantile regression estimates are robust against outliers in the response measurements. Fan et al. [6] and Yu and Jones [20] proposed local polynomial quantile regression, and Kai et al. (2010) [10] introduced local composite quantile regression. Nevertheless, the computational aspects of nonparametric quantile regression is very challenging. Moveover, the estimated quantile function is not smoothed in many applications, e.g., see Section 7.1 of Koenker [11].

In order to obtain smoothing curves, Wahba (1990) [19] introduced the mean smoothing regression, and pointed out that the solution to the mean smoothing regression was a natural cubic smoothing spline with knots at the observed design points. However, the method was not robust since it depended on the least squares method. Koenker et al. (1994) [13] proposed quantile smoothing splines and Shen (1998) [18] introduced the method of penalization. Some authors studied their asymptotic properties, e.g., Cox (1983) [3] studied asymptotic properties for M-type smoothing splines, and Portnoy (1997) [17] studied local asymptotics for quantile smoothing splines. Furthermore, many algorithms are proposed and studied for quantile smoothing splines. For example, Koenker et al. (1994) [13] used parametric linear programming methods to deal with the L_1 roughness penalty, and Pin (1996) [16] proposed the modified parametric linear programming algorithm based on an algorithm introduced by Bartels and Conn (1980) [2]. Although parametric linear programming provides efficient computation for moderate sample sizes, a large n requires substantial computational resources. Nychka et al. (1995) [14] proposed the smoothing quantile spline with the L_2 roughness penalty, and introduced a pseudo-data algorithm by iteratively solving a weighted smoothing spline problem. Meanwhile, they proposed the approximate Cross-Validation (ACV) as an approximation to the robust cross-validation (RCV) to choose the smoothing parameter. Subsequently, Oh et al. (2004) [15] chose the smoothing parameter by minimizing the generalized cross-validation (GCV), and Yuan (2006) [21] proposed the generalized approximate cross-validation (GACV) to use as a tuning criterion. In this paper, we first apply the MM-algorithm proposed by Hunter and Lange [8, 9] to deal with the computational problem of quantile smoothing splines with the L_2 roughness penalty, and then use GCV scores to choose the smoothing parameter. The proposed MM-algorithm possesses some advantages of MM-algorithm, e.g., the descent property. Moreover, simulation studies show that the proposed method is more robust and efficient than the estimator based on the mean smoothing regression and the estimator proposed by Nychka et al. [14] with GCV scores.

The rest of the paper is organized as follows. In Section 2, we first introduce the smoothing nonparametric quantile regression, and then systematically study the MM-algorithm and its properties in the smoothing nonparametric quantile regression. In Section 3, we propose the data-driven pro-

cedure to select the smooth parameter. Simulation studies and real data application are given in Section 4. We conclude this paper with a brief discussion in Section 5. All proofs are given in Section 6.

2. SMOOTHING NONPARAMETRIC QUANTILE REGRESSION

2.1 Smoothing nonparametric quantile regression

Suppose that $\{(X_i, Y_i), i = 1, \ldots, n\}$ is a random sample generated from the model (1.1). Without loss of generality, we assume the support of X lies in $[a, b]$. Our target is to estimate the τ-conditional quantile function of response Y, given $X = x$. Quantile regression estimates $\hat{m}_\tau(\cdot)$ by minimizing the quantile loss function

$$(2.2) \qquad \mathscr{L}\{m_\tau(\cdot)\} = \sum_{i=1}^n \rho_\tau\{Y_i - m_\tau(X_i)\},$$

where $\rho_\tau(r) = \tau r - rI(r < 0)$ is the check loss function for $\tau \in (0, 1)$. Because Eq. (2.2) involves nonparametric functions, we employ nonparametric natural cubic spline regression techniques to conduct optimization. Suppose $x_1, \ldots, x_n$ are given on $[a, b]$ satisfying $a = x_0 < x_1 < \cdots < x_{n+1} = b$. A function g defined on $[a, b]$ is a cubic spline if the following two conditions are satisfied:
(a) g is a cubic polynomial on each of the intervals $(x_i, x_{i+1}), i = 0, \ldots, n$;
(b) the polynomial pieces fit together at the points x_i in such a way that g itself and its first and second derivatives are continuous at each x_i.

Furthermore, if the second and third derivatives of a function g are zero at a and b, then, a cubic spline on an interval $[a, b]$ is a natural cubic spline. Many merits about a natural cubic spline regression are given in [7].

To consider goodness of fit and smoothness of a curve estimate, we introduce the following smoothing nonparametric quantile regression estimator $\hat{m}_\tau(\cdot)$ which minimizes

$$(2.3) \qquad \mathscr{M}\{m_\tau(\cdot)\} = \mathscr{L}\{m_\tau(\cdot)\} + \alpha \int_a^b \{m_\tau''(x)\}^2 dx,$$

where α is the smoothing parameter controlling the trade-off between fidelity to the data and roughness of the fit and $m_\tau''(x)$ is the second derivative of $m_\tau(x)$. From Eq. (2.3), we can see that when α is too small, $\hat{m}_\tau(\cdot)$ tends to interpolate the τ-th quantiles at the distinct design points. As a consequence, this will lead to undersmoothing. Meanwhile, $\hat{m}_\tau(\cdot)$ is the linear regression quantile fit to the observations since there is too much penalty placed on the estimate for α sufficiently large. We will use the data-driven procedure to choose the smoothing parameter α later. Asymptotic results of $\hat{m}_\tau(\cdot)$ have been given by [18].

2.2 MM-algorithm and its properties

Since the objective function (2.3) are nondifferentiable, we will apply MM-algorithm to optimize it. The key idea is to first construct a surrogate function to majorize (2.3), and then update the estimate of the whole quantile function in the minimization step.

According to surrogate functions proposed by Hunter and Lange [8], for given $m_\tau^{(k)}(\cdot)$, where $m_\tau^{(k)}(\cdot)$ is a given estimated value at iteration k, we construct surrogate functions
(2.4)

$$Q\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} = Q^*\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} + \alpha \int_a^b \{m_\tau''(x)\}^2 dx,$$

where

$$\begin{aligned} &Q^*\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} \\ &= \sum_{i=1}^n \frac{1}{4}\left\{\frac{(Y_i - m_\tau(X_i))^2}{|Y_i - m_\tau^{(k)}(X_i)|} + (4\tau - 2)(Y_i - m_\tau(X_i))\right. \\ &\qquad \left. + |Y_i - m_\tau^{(k)}(X_i)|\right\}. \end{aligned}$$

In the following, we will minimize (2.4) about $m_\tau(\cdot)$. Let the order statistics of a random sample $X_1, \ldots, X_n$ be $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ and $h_i = X_{(i+1)} - X_{(i)}$ for $i = 1, \ldots, n - 1$. According to Theorem 2.1 in Green and Silverman [4], the roughness of $m_\tau(\cdot)$ can be written as a quadratic form

$$\int_a^b \{m_\tau''(x)\}^2 dx = \boldsymbol{\mu}^T \mathbf{K} \boldsymbol{\mu},$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$, $\mu_i = m(x_i)$ is the fit, and $\mathbf{K}$ is an $n \times n$ matrix given by $\mathbf{K} = \mathbf{QR}^{-1}\mathbf{Q}^T$, $\mathbf{Q}$ be the $n \times (n-2)$ matrix of second differences, with entries

$$q_{j-1,j} = h_{j-1}^{-1}, q_{jj} = -h_{j-1}^{-1} - h_j^{-1}, q_{j+1,j} = h_j^{-1},$$
$$\text{for } j = 2, \ldots, n - 1,$$
$$q_{ij} = 0 \quad \text{for } |i - j| \geq 2.$$

$\mathbf{R}$ is a symmetric tridiagonal matrix of order $(n - 2)$ with elements r_{ij}

$$r_{ii} = \frac{1}{3}(h_{i-1} + h_i) \text{ for } i = 2, \ldots, n - 1,$$
$$r_{i,i+1} = r_{i+1,i} = \frac{1}{6}h_i \text{ for } i = 2, \ldots, n - 2.$$

Let

$$\begin{aligned} \hat{\mathbf{Y}} &= \big(Y_1 + (2\tau - 1)|Y_1 - m_\tau^{(k)}(X_1)|, \ldots, Y_n \\ &\quad + (2\tau - 1)|Y_n - m_\tau^{(k)}(X_n)|\big)^T, \\ \mathbf{W} &= diag\{1/|Y_1 - m_\tau^{(k)}(X_1)|, \ldots, 1/|Y_n - m_\tau^{(k)}(X_n)|\}. \end{aligned}$$

Then, the closed form solution of minimizing Eq. (2.4) is given by

$$m_\tau^{(k+1)}(\cdot) = (\mathbf{W} + \alpha\mathbf{K})^{-1}\mathbf{W}\hat{\mathbf{Y}}. \tag{2.5}$$

In the following, we will show that the surrogate functions $Q\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\}$ possess the the descent property [8, 9] of MM-algorithm. The descent property makes the MM algorithm a very remarkable numerical stability.

Theorem 2.1. *The surrogate functions* $Q\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\}$ *satisfies the following inequalities*

$$(1): Q\{m_\tau^{(k)}(\cdot), m_\tau^{(k)}(\cdot)\} = \mathscr{M}\{m_\tau^{(k)}(\cdot)\};$$
$$(2): Q\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} \geq \mathscr{M}\{m_\tau(\cdot)\}.$$

2.3 A perturbed version of MM algorithm

A major disadvantage of minimizing Eq. (2.4) is that, when $Y_i - m_\tau^{(k)}(X_i) = 0$ for some i, then the denominator $|Y_i - m_\tau^{(k)}(X_i)|$ makes $Q\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\}$ undefined. Therefore, this motivates us to consider a perturbed version of MM-algorithm, that is to say, $|Y_i - m_\tau^{(k)}(X_i)|$ is replaced by $|Y_i - m_\tau^{(k)}(X_i)| + \epsilon$ in Eq. (2.4) for some sufficiently small $\epsilon > 0$. In the following, we define a perturbed version of $\mathscr{M}\{m_\tau(\cdot)\}$. Denote $r_{i\tau} = Y_i - m_\tau(X_i)$,

$$\mathscr{M}_\epsilon\{m_\tau(\cdot)\} = \mathscr{L}\{m_\tau(\cdot)\} - \sum_{i=1}^n \frac{\epsilon}{2}\log\left(\epsilon + |r_{i\tau}|\right) \tag{2.6}$$
$$+ \alpha\int_a^b \{m_\tau''(x)\}^2 dx.$$

Next, we study the perturbed version of the MM-algorithm. For given $m_\tau^{(k)}(\cdot)$, $\mathscr{M}_\epsilon\{m_\tau(\cdot)\}$ is majorized by the following surrogate function

$$Q_\epsilon\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} \tag{2.7}$$
$$= Q_\epsilon^*\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} + \alpha\int_a^b \{m_\tau''(x)\}^2 dx,$$

where

$$Q_\epsilon^*\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\}$$
$$= \sum_{i=1}^n \frac{1}{4}\left\{\frac{(Y_i - m_\tau(X_i))^2}{|Y_i - m_\tau^{(k)}(X_i)| + \epsilon} + (4\tau - 2)\left(Y_i - m_\tau(X_i)\right) + c\right\},$$

where c is a constant chosen so that $Q_\epsilon\{m_\tau^{(k)}(\cdot), m_\tau^{(k)}(\cdot)\} = \mathscr{M}_\epsilon\{m_\tau^{(k)}(\cdot)\}$. According to Proposition A.2. in Hunter and Lange [8], $Q_\epsilon\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\}$ also possesses the descent property of MM-algorithm. It is worth it to point out that when we calculate $m_\tau^{(k+1)}(\cdot)$ according to Eq. (2.5), $|Y_i - m_\tau^{(k)}(X_i)|, i = 1, \ldots, n$ of $\hat{\mathbf{Y}}$ and $\mathbf{W}$ should be replaced with $|Y_i - m_\tau^{(k)}(X_i)| + \epsilon, i = 1, \ldots, n$ as well. Next, we

will give some convergence results of the perturbed version of the MM-algorithm. Let $\kappa = 1 + \sup\{|Y_i - m_\tau(X_i)|, i = 1, \ldots, n\}$.

Theorem 2.2. *1. If $\kappa\epsilon < 1$, then*

$$\left|\mathscr{M}\{m_\tau(\cdot)\} - \mathscr{M}_\epsilon\{m_\tau(\cdot)\}\right| \leq -\frac{\epsilon n}{2}\log(\epsilon);$$

2. If $\hat{m}_\tau(\cdot)$ and $\hat{m}_{\tau\epsilon}(\cdot)$ minimize $\mathscr{M}\{m_\tau(\cdot)\}$ and $\mathscr{M}_\epsilon\{m_\tau(\cdot)\}$, respectively, then

$$\mathscr{M}\{\hat{m}_{\tau\epsilon}(\cdot)\} - \mathscr{M}\{\hat{m}_\tau(\cdot)\} \leq -\epsilon n\log(\epsilon);$$

3. If $\hat{m}_{\tau\epsilon}(\cdot)$ minimizes $\mathscr{M}_\epsilon\{m_\tau(\cdot)\}$, then any limit point of $\{\hat{m}_{\tau\epsilon}(\cdot)\}$ as $\epsilon \to 0$ minimizes $\mathscr{M}\{m_\tau(\cdot)\}$.

2.4 Practical implementation issues

The proposed MM-algorithm involves the problem of initial value. Therefore, we first obtain initial value $m_\tau^{(0)}(\cdot)$ via normal approximation. We apply kernel regression to obtain estimate of mean function, i.e., $\hat{m}(x) = \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)}$, and further calculate variance estimate $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \{Y_i - \hat{m}(X_i)\}^2$ by pretending a constant variance. By normal approximation, we set the initial value $m_\tau^{(0)}(x) = \hat{m}(x) + \hat{\sigma}_n\Phi^{-1}(\tau)$, where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function (CDF) of a standard normal distribution. We declare convergence for the proposed MM algorithm whenever

$$\left|Q_\epsilon^*\{m_\tau^{(k+1)}(\cdot), m_\tau^{(k)}(\cdot)\} - Q_\epsilon^*\{m_\tau^{(k)}(\cdot), m_\tau^{(k-1)}(\cdot)\}\right| < \delta \tag{2.8}$$

for a small positive value δ. In this paper, we take $\delta = \epsilon = 10^{-6}$.

3. SELECTION OF THE SMOOTHING PARAMETER

From (2.3), we see that the smoothing parameter α controls the degree of goodness of fit and smoothness of a curve estimate. Therefore, to implement the methods described in previous sections, it is desirable to have a data-driven procedure for selecting the smoothing parameter α. In this paper, we select α via minimizing an approximate generalized cross-validation (GCV) score. Let

$$A(\alpha) = (I + \alpha K)^{-1} = \left(I + \alpha QR^{-1}Q^T\right)^{-1}.$$

Then, GCV score is defined by

$$GCV(\alpha) = \frac{1}{n}\frac{\sum_{i=1}^n (Y_i - \hat{m}_\tau(X_i))^2}{(1 - n^{-1}tr(A(\alpha)))^2}.$$

Therefore,

$$\hat{\alpha} = arg\min_\alpha GCV(\alpha).$$

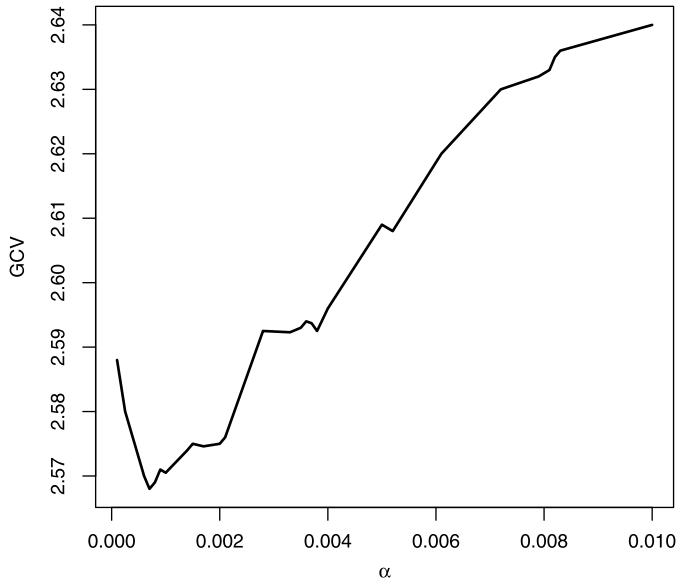The minimization can be carried out by searching over a grid of points for α.

Figure 1. GCV against α.

Figure 2. The fitted curves based on LS-method, WSS-method, and MQR-method.

4. SIMULATION AND APPLICATION

Example 1. We first illustrate how to select the smoothing parameter α. We draw random samples of size $n = 200$ from the model

(4.9) $$Y_i = \sin(2\pi X_i) + \epsilon_i,$$

where X_i are independently sampled from $U(0,1)$. The error term ϵ_i are independent and identically distributed random variables from a Student's t distribution with 2 degrees of freedom. Since the error term is symmetric, we select $\tau = 0.5$. We plot GCV scores against the penalized parameter α as depicted in Figure 1. From Figure 1, we can see that the smoothing parameter α equals 0.0007, selected by minimizing the GCV scores.

Example 2. In this subsection, we conduct simulation studies to evaluate the finite-sample performance of our proposed estimator. We draw random samples of size $n = 50, 100, 200$ from the model (4.9), where X_i are independently sampled from $U(0,1)$. The error term ϵ_i are independent and identically distributed random variables from

1. a standard normal distribution, Φ;
2. a Student's t-distribution with 2 degrees of freedom, t_2;
3. a 5% contaminated normal distribution $CN_{0.05} = 0.95\Phi(x) + 0.05\Phi(x/5)$;
4. a laplace distribution with location parameter $\mu = 0$, and scale parameter $b = 4$, Laplace;
5. a standard Cauchy distribution, Cauchy.

Since the error terms for regression models are symmetric, both conditional mean and median functions represent the true regression function in the usual sense. In the simulation, we sel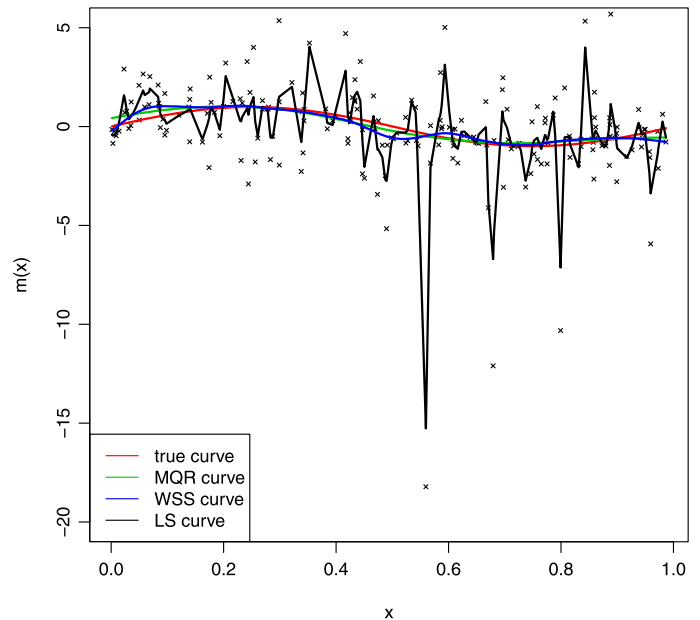ect $\tau = 0.5$ and conduct 1,000 simulations, and calculate the mean-squared errors (MSE) of our proposed estimator (MQR), the mean smoothing regression (LS), and the method (WSS) proposed by Nychka et al. (1995) [14], where the smoothing parameter of these methods is chosen by GCV, and the mean-squared error is defined by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\big(\hat{m}_\tau(x_i) - m(x_i)\big)^2.$$

We compute the estimator based on the method proposed by Nychka et al. [14] using the *qsreg* function of the R package *fields*, and the mean smoothing regression using the *smooth.spline* function of the R package *stats*. The means and standard errors (in parentheses) of MSE are reported in Table 1. From Table 1, we can clearly see that MQR is more robust and efficient than the estimator based on the mean smoothing regression and the estimator proposed by Nychka et al. [14] with GCV scores when the error terms are drawn from any of the thicker-tailed distributions.

Example 3. In this example, we use simulation studies to illustrate the degree of goodness of fit of our proposed method. We draw random samples of size $n = 200$ from the model (4.9), where X_i are independently sampled from $U(0,1)$. The error term ϵ_i are independent and identically distributed random variables from a Student's t-distribution with 2 degrees of freedom. the fitted curves based on LS-method, WSS-method, and MQR-method are drawn in Figure 2. From Figure 2, we see that the fitted curve based on MQR-method is closer to the true curve than those based LS-method and WSS-method when the error terms follow the heavy-tailed distributions.

Table 1. Means and standard errors of MSE for MQR, WSS and LS for the following distributions

n	Method	Φ	t_2	$CN_{0.05}$	Laplace	Cauchy
	MQR	0.1263(0.0776)	0.2174(0.1658)	0.1073(0.0725)	1.8667(1.5759)	0.3256(0.4225)
50	WSS	0.2158(0.1080)	0.4047(0.2691)	0.1870(0.0998)	4.1661(2.6554)	13.865(281.02)
	LS	0.2267(0.3028)	6.4517(43.5986)	0.2377(0.3074)	8.7313(12.456)	7395.7(114992)
	MQR	0.0714(0.0444)	0.1215(0.0736)	0.0569(0.0389)	1.0363(0.7606)	0.1534(0.1057)
100	WSS	0.1122(0.0538)	0.1846(0.0992)	0.0914(0.0498)	1.8500(1.0655)	0.5163(0.5346)
	LS	0.1178(0.2189)	5.1448(22.3873)	0.1150(0.2149)	5.1014(10.136)	21602(414970)
	MQR	0.0405(0.0247)	0.0662(0.0397)	0.0286(0.0195)	0.5364(0.2824)	0.0773(0.0500)
200	WSS	0.0568(0.0287)	0.0917(0.0466)	0.0420(0.0232)	0.8363(0.4447)	0.2156(0.1687)
	LS	0.0607(0.1610)	0.1307(0.2930)	0.0748(0.1957)	2.1906(6.4256)	133190(2424003)

Regression spline with least square criterion   **The estimated conditional quantile curves**

(a)   (b)
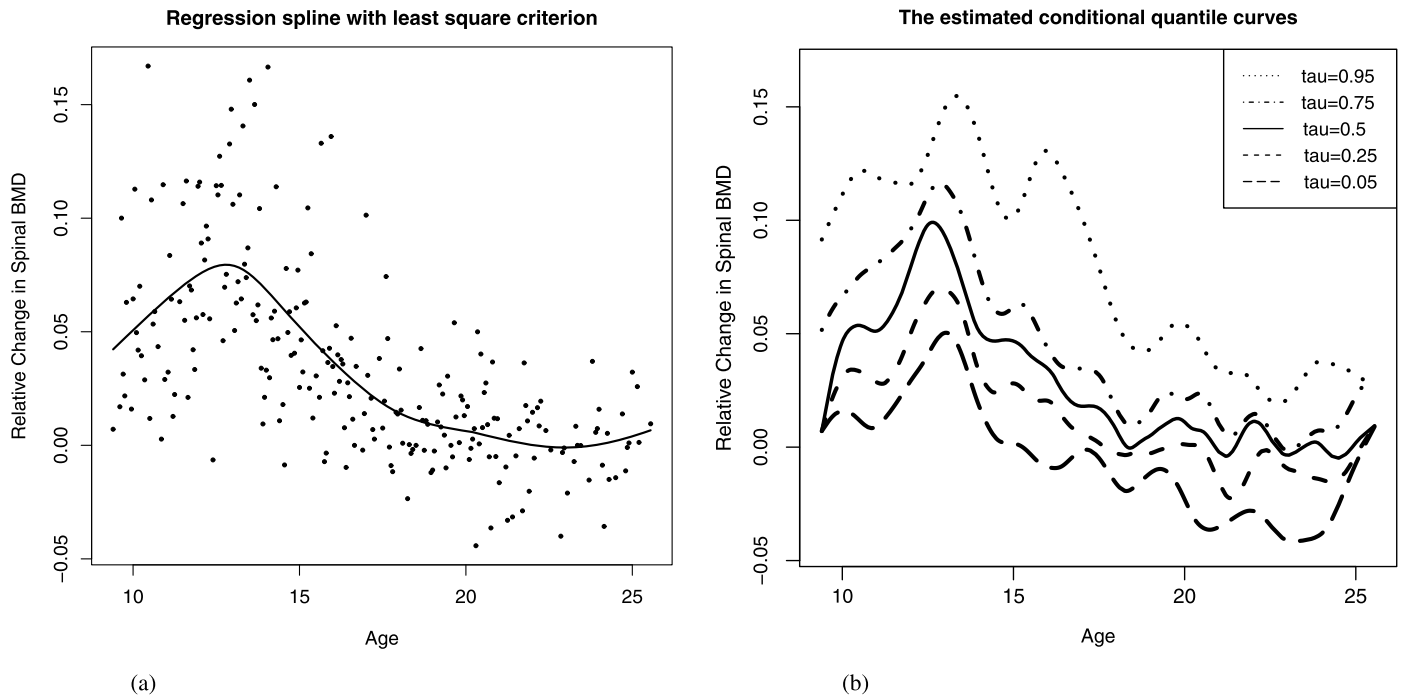
Figure 3. (a) Conditional mean analysis; (b) Conditional quantile analysis.

Example 4. As an illustration, we apply the proposed methodology to analyze the dataset about bone density (BMD) in adolescents. The dataset was originally analyzed in [1] and was also reported in [7]. The data can be downloaded from http://statweb.stanford.edu/~tibs/ElemStatLearn/. The dataset contains the following four variables: idnum, age, gender, and spnbmd (relative spinal bone mineral density measurement). There are 485 observations in the dataset. We will study the relationship between age and spnbmd. Since there are the same age in age variable, we take mean of spnbmd for the same age. We obtain 239 observations, which are showed in Figure 3. In the following, we apply the LS-based method and MQR-method to deal with the dataset. The results are shown in Figure 3. Figure 3(a) shows a conditional mean analysis, and Figure 3(b) describes a conditional quantile analysis with $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$. For MQR-method,

we apply the proposed methodology to select the smoothing parameters with the corresponding quantile, and obtain $\alpha = 0.9, 0.9, 0.9, 0.9, 0.87$, respectively. From Figure 3, we can see that the relative spinal bone mineral density obtains maximum between 12 and 13 years. The conditional mean analysis clearly gives the entire distribution of BMD changes against age, but it provides only the central tendency of the conditional distribution. As controls, the conditional quantile analysis provides us more detailed information of these changes. For example, we can see that the variance of the BMD changes against the age is heteroscedastic, and that the conditional distribution is slightly positively skewed.

5. DISCUSSION

In this paper, we applied MM-algorithm to solve the computational problem of the smoothing nonparametric quantile regression model. Simulation studies and the real data

analysis illustrated that the proposed methodology was robust and efficient. Although the proposed algorithm only dealt with the nonparametric model in this paper, More work remained to be done on the application of the proposed method to other models, like the generalized additive models and semiparametric models.

ACKNOWLEDGEMENTS

APPENDIX

Proof of Theorem 2.1. By Eq. (2.3) and Eq. (2.4), we have

$$Q\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} - \mathscr{M}\{m_\tau(\cdot)\} = Q^*\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} - \mathscr{L}\{m_\tau(\cdot)\}.$$

According to Hunter and Lange [8], we have

$$(A.1) \qquad Q^*\{m_\tau^{(k)}(\cdot), m_\tau^{(k)}(\cdot)\} = \mathscr{L}\{m_\tau^{(k)}(\cdot)\},$$

$$(A.2) \qquad Q^*\{m_\tau(\cdot), m_\tau^{(k)}(\cdot)\} \geq \mathscr{L}\{m_\tau(\cdot)\}.$$

Therefore, by Eq. (A.1) and Eq. (A.2), we complete the proof of Theorem 2.1. $\square$

Proof of Theorem 2.2. If $\epsilon + |Y_i - m_\tau(X_i)| \leq 1$, then $|\log(\epsilon + |Y_i - m_\tau(X_i)|)| \leq -\log(\epsilon)$. Otherwise, $|\log(\epsilon + |Y_i - m_\tau(X_i)|)| < \log(\kappa) < -\log(\epsilon)$ since $\kappa\epsilon < 1$. Therefore,

$$
\begin{aligned}
\left|\mathscr{M}\{m_\tau(\cdot)\} - \mathscr{M}_\epsilon\{m_\tau(\cdot)\}\right| &= \frac{\epsilon}{2}\Big|\sum_{i=1}^n \log\big(\epsilon + |Y_i - m_\tau(X_i)|\big)\Big| \\
&\leq \frac{\epsilon}{2}\sum_{i=1}^n \big|\log\big(\epsilon + |Y_i - m_\tau(X_i)|\big)\big| \\
&\leq -\frac{\epsilon n}{2}\log(\epsilon).
\end{aligned}
$$

$$
\begin{aligned}
&\mathscr{M}\{\hat{m}_{\tau\epsilon}(\cdot)\} - \mathscr{M}\{\hat{m}_\tau(\cdot)\} \\
&= \mathscr{M}\{\hat{m}_{\tau\epsilon}(\cdot)\} - \mathscr{M}_\epsilon\{\hat{m}_{\tau\epsilon}(\cdot)\} + \mathscr{M}_\epsilon\{\hat{m}_{\tau\epsilon}(\cdot)\} - \mathscr{M}\{\hat{m}_\tau(\cdot)\} \\
&\leq \mathscr{M}\{\hat{m}_{\tau\epsilon}(\cdot)\} - \mathscr{M}_\epsilon\{\hat{m}_{\tau\epsilon}(\cdot)\} + \mathscr{M}_\epsilon\{\hat{m}_\tau(\cdot)\} - \mathscr{M}\{\hat{m}_\tau(\cdot)\} \\
&\leq \big|\mathscr{M}\{\hat{m}_{\tau\epsilon}(\cdot)\} - \mathscr{M}_\epsilon\{\hat{m}_{\tau\epsilon}(\cdot)\}\big| + \big|\mathscr{M}_\epsilon\{\hat{m}_\tau(\cdot)\} \\
&\quad - \mathscr{M}\{\hat{m}_\tau(\cdot)\}\big| \leq -\epsilon n \log(\epsilon).
\end{aligned}
$$

Let $\lim_{k\to\infty} \hat{m}_{\tau\epsilon_k}(\cdot) = m_\tau^*(\cdot)$ when $\epsilon_k \to 0$ as $k \to \infty$. In addition, we have $\lim_{k\to\infty} \mathscr{M}_{\epsilon_k}\{m_\tau(\cdot)\} = \mathscr{M}\{m_\tau(\cdot)\}$. Since $\mathscr{M}_{\epsilon_k}\{\hat{m}_{\tau\epsilon_k}(\cdot)\} \leq \mathscr{M}_{\epsilon_k}\{m_\tau(\cdot)\}$, we take limits about this inequality, then, $\mathscr{M}\{m_\tau^*(\cdot)\} \leq \mathscr{M}\{m_\tau(\cdot)\}$. This completes the proof of Theorem 2.2. $\square$

Received 26 September 2012

REFERENCES

[1] BACHRACH, L., HASTIE, T., WANG, M., NARASIMHAN, B. and MARCUS, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, black, and Caucasian youth: A longitudinal study. *Clin. Endocrin. Metab.* **84** 4702–4712.

[2] BARTELS, R. and CONN, A. (1980). Linearly constrained discrete L_1 problems. *ACM Transactions on Mathematical Software* **6** 594–608. MR0599980

[3] COX, D. (1983). Asymptotics for M-type smoothing splines. *Ann. Stat.* **11** 530–551. MR0696065

[4] GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London. MR1270012

[5] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC. MR1383587

[6] FAN, J., HU, T. and TRUONG, Y. (1994). Robust non-parametric function estimation. *Scand. J. Stat.* **21** 433–446. MR1310087

[7] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. MR1851606

[8] HUNTER, D. and LANGE, K. (2000). Quantile regression via an MM algorithm. *J. Comput. Graph. Stat.* **9** 60–77. MR1819866

[9] HUNTER, D. and LANGE, K. (2004). A tutorial on MM algorithms. *Am. Stat.* **58** 30–37. MR2055509

[10] KAI, B., LI, R. and ZOU, H. (2010). Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *J. Roy. Statist. Soc. Ser. B* **72** 49–69. MR2751243

[11] KOENKER, R. (2005). *Quantile Regression*. Cambridge Univ. Pr. MR2268657

[12] KOENKER, R. and BASSETT JR., G. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[13] KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. MR1326417

[14] NYCHKA, D., GRAY, G., HAALAND, P., MARTIN, D. and O'CONNELL, M. (1995). A nonparametric regression approach to syringe grading for quality improvement. *J. Amer. Statist. Assoc.* **90** 1171–1178.

[15] OH, H., NYCHKA, D., BROWN, T. and CHARBONNEAU, P. (2004). Period analysis of variable stars by robust smoothing. *J. Roy. Statist. Soc. Ser. C* **53** 15–30. MR2037881

[16] PIN, T. (1996). An algorithm for quantile smoothing splines. *Comput. Stat. Data. An.* **22** 99–118. MR1410382

[17] PORTNOY, S. (1997). Local asymptotics for quantile smoothing splines. *Ann. Stat.* **25** 414–434. MR1429932

[18] SHEN, X. (1998). On the method of penalization. *Stat. Sinica.* **8** 337–358. MR1624410

[19] WAHBA, G. (1990). *Spline Models for Observational Data*. Society for Industrial Mathematics. MR1045442

[20] YU, K. and JONES, M. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628

[21] YUAN, M. (2006). GACV for quantile smoothing splines. *Comput. Stat. Data. An.* **50** 813–829. MR2207010

Yunlu Jiang
Department of Statistics
College of Economics
Jinan University
Guangzhou, 510632
People's Republic of China
E-mail address: jiangyunlu2011@gmail.com