Computational Statistics
https://doi.org/10.1007/s00180-020-00998-w

ORIGINAL PAPER

Bayesian joint-quantile regression

Yingying Hu[1] · Huixia Judy Wang[2] · Xuming He[3] · Jianhua Guo[4]

Received: 30 May 2019 / Accepted: 3 June 2020

Abstract
Estimation of low or high conditional quantiles is called for in many applications, but commonly encountered data sparsity at the tails of distributions makes this a challenging task. We develop a Bayesian joint-quantile regression method to borrow information across tail quantiles through a linear approximation of quantile coefficients. Motivated by a working likelihood linked to the asymmetric Laplace distributions, we propose a new Bayesian estimator for high quantiles by using a delayed rejection and adaptive Metropolis and Gibbs algorithm. We demonstrate through numerical studies that the proposed estimator is generally more stable and efficient than conventional methods for estimating tail quantiles, especially at small and modest sample sizes.

Keywords Adaptive Metropolis · Asymmetric Laplace distribution · Delayed rejection · High quantile · Quantile regression

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00180-020-00998-w) contains supplementary material, which is available to authorized users.

✉ Yingying Hu
hyysunlinxiao@126.com

Huixia Judy Wang
judywang@email.gwu.ed

Xuming He
xmhe@umich.edu

Jianhua Guo
jhguo@nenu.edu.cn

1 College of Mathematics and System Science, Xinjiang University, Urumqi 830046, Xinjiang Uygur Autonomous Region, People's Republic of China

2 Department of Statistics, The George Washington University, Washington, DC 20052, USA

3 Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

4 School of Mathematics and Statistics and KLAS, Northeast Normal University, Changchun 130024, Jilin Province, People's Republic of China

Published online: 15 June 2020

1 Introduction

The modeling and prediction of extreme and rare events, e.g., heavy rainfall and big financial loss, is an important and difficult task in many fields. For such events, the interest is often on the characteristics of tail quantiles rather than the averages or the median. Quantile regression, formally proposed in Koenker and Bassett (1978), provides a useful tool for studying the conditional distribution of the response. In finite samples, however, the conventional quantile regression estimator is often unstable at tails due to data sparsity. Improving the estimation of quantiles at tail regions would require additional model assumptions or information, leading to two classes of methods in the literature.

The first class is based on extreme value theory by making additional assumptions on the tail properties of the conditional distribution of the response. For instance, in the reference (Wang et al. 2012), the authors assumed that the conditional distribution is heavy-tailed and lies in the maximum domain of attraction of an extreme value distribution. Based on this assumption, estimates of intermediate quantiles from the conventional quantile regression can be extrapolated to the far tail to estimate extremely high quantiles. Some other developments in this direction include the references (Chernozhukov and Du 2008; Gardes and Girard 2011; Wang and Li 2013), to name a few. All the methods in this class are based on extreme value theory, which would require relatively large samples for the approximation to work well.

The other class of work is based on joint-quantile analysis by assuming and sharing some common information across quantiles. For instance, for linear quantile regression, one may assume that the slope coefficients are constant across quantiles (in a given region). Under this assumption, the composite quantile estimator can be obtained by minimizing the combined quantile loss function at multiple quantiles to estimate the common slopes. Some developments along this line include the references (Jiang et al. 2013a, 2018; Koenker 1984; Wang and Wang 2016; Wu et al. 2020; Zhao and Xiao 2014; Zou and Yuan 2008) for linear models, and (Guo et al. 2012; Jiang et al. 2013b; Kai et al. 2010; Tang et al. 2018) for semiparametric regression models, among others.

We develop a new method for estimating tail conditional quantiles under the Bayesian framework. The Bayesian approach provides a convenient way to incorporate prior information, if available. The idea is to improve the tail quantile estimation by borrowing information across quantiles through a linear approximation of quantile coefficients in the tail quantile region. The proposed linear expansion covers the common-slope assumption considered in the literature as a special case, but offers more flexibility and thus is useful in broader applications. In addition, under the Bayesian framework, common features across quantiles can be captured and incorporated in a more natural way.

Bayesian methods have been developed for various quantile regression models. Some of these focused on the analysis at a single quantile level; see for instance, (Kozumi and Kobayashi 2011; Li et al. 2010; Yu and Moyeed 2001; Yu and Stander 2007). There have been a number of Bayesian studies for multiple-quantile analysis. Specifically, Alhamzawi (2016), Huang and Chen (2015) and Tian et al. (2017) considered Bayesian composite quantile regression by assuming that quantile slopes are

constant across quantile levels, Reich et al. (2011), Reich and Smith (2013), Rodrigues and Fan (2017), Rodrigues et al. (2019) and Yang and Tokdar (2017) developed methods for modeling quantile process to overcome challenges such as quantile crossing and spatial dependence.

The proposed method is based on a working composite-quantile likelihood, motivated by the connection between quantile regression at a single quantile level and the asymmetric Laplace distribution proposed in Yu and Moyeed (2001). The posterior distributions of the quantile regression parameters do not have closed expressions, and we propose posterior sampling by adopting the delayed rejection and adaptive Metropolis (DRAM) algorithm, which is the combination of the delayed rejection (Green and Mira 2001; Mira 2002) and adaptive Metropolis (Haario et al. 1999, 2001) methods. Compared to the standard Metropolis–Hastings algorithm, the DRAM algorithm provides better protection against both over- and under-calibrated proposals, and thus relaxes the burden of identifying suitable proposal distributions for each parameter involved. We note that the problem of multiple quantile estimation in a pseudo-Bayesian framework has been studied in Sriram et al. (2016), where the authors proved posterior consistency under correct specifications of the quantile functions. In contrast, the present paper uses local linear approximations to the quantile coefficients with which we aim to borrow information across quantiles in a small region for better bias-variance trade-off in quantile estimation.

The rest of the paper is organized as follows. In Sect. 2, we present the proposed working likelihood and posterior sampling algorithm. In Sect. 3, we conduct simulation studies to compare the performance of the proposed estimator with existing methods. We further demonstrate the practical value of the proposed method through the analysis of a Chicago precipitation data in Sect. 4. Some final conclusions are given in Sect. 5.

2 The proposed method

2.1 Basic setup and the target posterior distribution

Consider the linear quantile regression model

$$Q_Y(\tau|\mathbf{X}=\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}(\tau),\ \ \tau \in [\omega, 1), \tag{1}$$

where Y is the response variable, $\mathbf{X}$ is the p-dimensional covariate vector with the first element one to include the intercept term, $Q_Y(\tau|\mathbf{X})$ is the τ-th quantile of Y given $\mathbf{X}$, $\boldsymbol{\beta}(\tau) = (\beta_1(\tau), \ldots, \beta_p(\tau))^T$ is the unknown quantile coefficient vector, and $\omega \in (0, 1)$ is a constant close to one.

Based on a random sample $\mathcal{D} = \{(y_i, \mathbf{x}_i), i = 1, \ldots, n\}$ of $(Y, \mathbf{X})$, the coefficient $\boldsymbol{\beta}(\tau)$ can be estimated by

$$\widehat{\boldsymbol{\beta}}(\tau) = \arg\min_{\boldsymbol{\beta}\in R^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T\boldsymbol{\beta}), \tag{2}$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function. At a fixed quantile level $0 < \tau < 1$, the conventional estimator $\widehat{\boldsymbol{\beta}}(\tau)$ is known to be consistent and asymptotically normal under some regularity conditions; see Koenker (2005). However, in finite samples, the estimator $\widehat{\boldsymbol{\beta}}(\tau)$ is often unstable for τ close to one or zero due to the lack of sufficient data in tail regions. To reduce the instability at high quantiles, we propose to conduct a joint-quantile analysis through the linear expansion of $\boldsymbol{\beta}(\tau)$ in a small region of τ.

Let $\boldsymbol{\tau} = \{\tau_1, \cdots, \tau_L\}$, $\omega \le \tau_1 < \cdots < \tau_L < 1$ be a grid of quantile levels in the upper tail, and $\tau^\star \ge \tau_L$ be the high quantile level of interest. Suppose that $\beta_j(\tau)$, $j = 1, \cdots, p$, are smooth functions of τ. Then by the Taylor expansion at a given quantile level τ_0, it follows that

$$\beta_j(\tau) = \beta_j(\tau_0) + \beta_j'(\tau_0)(\tau - \tau_0) + o(\tau - \tau_0) \tag{3}$$

for τ near τ_0. In practice, we let $\tau_0 = \sum_{l=1}^{L} \tau_l / L$. Therefore, we have

$$\beta_j(\tau) \approx \theta_{1j} + \theta_{2j}(\tau - \tau_0),$$

where $\theta_{1j} = \beta_j(\tau_0)$ and $\theta_{2j} = \beta_j'(\tau_0)$. Under this approximation, the linear quantile regression model (1) can be rewritten as

$$Q_Y(\tau|\mathbf{x}) = \mathbf{x}^T(\boldsymbol{\vartheta}_1 + \boldsymbol{\vartheta}_2\delta), \tag{4}$$

where $\boldsymbol{\vartheta}_1 = (\theta_{11}, \cdots, \theta_{1p})^T$, $\boldsymbol{\vartheta}_2 = (\theta_{21}, \cdots, \theta_{2p})^T$ and $\delta = \tau - \tau_0$. By applying the linear approximation of $\boldsymbol{\beta}(\tau_l)$ at τ_l around τ_0 for $l = 1, \ldots, L$, we can estimate the common parameters $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ by minimizing the combined quantile objective function, that is, by

$$(\widetilde{\boldsymbol{\theta}}_1, \widetilde{\boldsymbol{\theta}}_2) = \arg\min \sum_{i=1}^{n} \sum_{l=1}^{L} \rho_{\tau_l}\{y_i - \mathbf{x}_i^T(\boldsymbol{\vartheta}_1 + \boldsymbol{\vartheta}_2\delta_l)\}, \tag{5}$$

where $\delta_l = \tau_l - \tau_0$, $l = 1, \ldots, L$. Hereafter, we will refer to the estimator $(\widetilde{\boldsymbol{\theta}}_1, \widetilde{\boldsymbol{\theta}}_2)$ as the frequentist linear composite quantile estimator.

At a single quantile level, that is, when $L = 1$, the quantile loss function in (5) is proportional to the negative log density of the asymmetric Laplace distribution (ALD) (Yu and Moyeed 2001). This motivated researchers to consider the AL working likelihood for Bayesian quantile regression at a single quantile level $\tau \in (0, 1)$:

$$L(\boldsymbol{\beta}(\tau); \mathcal{D}) = \prod_{i=1}^{n} \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\frac{\rho_\tau(y_i - \mathbf{x}_i^T\boldsymbol{\beta}(\tau))}{\sigma}\right\}, \tag{6}$$

where $\sigma > 0$ is a scale parameter. More discussions can be found in Geraci and Bottai (2007), Yang et al. (2016), Yu and Stander (2007), and other works reviewed in Wang

and Yang (2018). For the joint-quantile analysis, a natural and analogous choice is the following working likelihood,

$$\prod_{i=1}^{n} C(\mathbf{x}_i, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \sigma, \boldsymbol{\tau}) \frac{1}{\sigma} \exp\left\{ - \frac{\sum_{l=1}^{L} \rho_{\tau_l}(y_i - \mathbf{x}_i^T(\boldsymbol{\vartheta}_1 + \boldsymbol{\vartheta}_2 \delta_l))}{\sigma} \right\}, \tag{7}$$

where

$$C(\mathbf{x}, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \sigma, \boldsymbol{\tau}) \times \int \frac{1}{\sigma} \exp\left\{ - \frac{\sum_{l=1}^{L} \rho_{\tau_l}(y - \mathbf{x}^T(\boldsymbol{\vartheta}_1 + \boldsymbol{\vartheta}_2 \delta_l))}{\sigma} \right\} dy = 1.$$

Note that the normalizing constant $C(\mathbf{x}, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \sigma, \boldsymbol{\tau})$ depends on $\mathbf{x}$ and the parameters in a rather intractable way. It was argued in Sriram et al. (2016) that a pseudo-likelihood without this normalizing factor would ensure posterior consistency for a broad range of priors. The pseudo-likelihood takes the form

$$L(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2; \mathcal{D}) \propto \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left\{ - \frac{\sum_{l=1}^{L} \rho_{\tau_l}(y_i - \mathbf{x}_i^T(\boldsymbol{\vartheta}_1 + \boldsymbol{\vartheta}_2 \delta_l))}{\sigma} \right\}. \tag{8}$$

Then we can use independent flat priors for $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$, i.e. $\pi(\boldsymbol{\vartheta}_1) = \prod_{j=1}^{p} \pi(\theta_{1j}) \propto 1$, $\pi(\boldsymbol{\vartheta}_2) = \prod_{j=1}^{p} \pi(\theta_{2j}) \propto 1$, and for the scale parameter σ, $\pi(\sigma) \propto I(\sigma > 0)/\sigma$. The posterior for $(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \sigma)$ can be derived as

$$\pi(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \sigma | \mathcal{D}) \propto \frac{1}{\sigma^{n+1}} \exp\left\{ - \frac{1}{\sigma} \sum_{i=1}^{n} \sum_{l=1}^{L} \rho_{\tau_l}(y_i - \mathbf{x}_i^T(\boldsymbol{\vartheta}_1 + \boldsymbol{\vartheta}_2 \delta_l)) \right\}. \tag{9}$$

We can arrive at the same posterior from the working likelihood (7) but with a prior density that is proportional to $1/C(\mathbf{x}_i, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \sigma, \boldsymbol{\tau})$. Whatever the motivation, the posterior mode from (9) is equivalent to the frequentist linear composite quantile estimator $(\widetilde{\boldsymbol{\theta}}_1, \widetilde{\boldsymbol{\theta}}_2)$ in (5). Following the similar proof as in Sriram et al. (2016) and Zou and Yuan (2008), we can show that $(\widetilde{\boldsymbol{\theta}}_1, \widetilde{\boldsymbol{\theta}}_2)$ is consistent to $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

2.2 Computation

The posterior of $(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \sigma)$ is analytically intractable. It is natural to use the MCMC method for posterior sampling. For quantile regression at a single quantile level, the authors of Yu and Moyeed (2001) proposed to use a random-walk Metropolis algorithm with a Gaussian proposal density centered at the current state of the chain to update each parameter. The authors of Kozumi and Kobayashi (2011) proposed a three-variable Gibbs sampling algorithm that can simplify the posterior sampling by using the Normal-mixture representation of ALD, and as shown in Khare and Hobert (2012), the Markov chain underlying this three-variable Gibbs sampling algorithm

converges at a geometric rate. However, this mixture representation is not applicable to the joint-quantile working likelihood.

In this paper, we suggest posterior sampling by adopting the delayed rejection and adaptive Metropolis (DRAM) and Gibbs algorithms. The DRAM algorithm, proposed in Haario et al. (2006), is the combination of delayed rejection (DR) and adaptive Metropolis (AM). The AM can enhance the efficiency of DR if no good proposal distributions are available, and the DR provides a systematic remedy when AM has a slow start. To be more specific, upon rejection in a Metropolis–Hastings (M–H) algorithm, the DR proposes a second stage move instead of staying at the current position. The process of DR can be iterated for a fixed or random number of stages. The basic idea of AM is to create a Gaussian proposal distribution with a covariance matrix depending on the history of the chain. The DRAM algorithm used in this paper is a direct way of combining AM adaptation with a J-stage DR algorithm. Our numerical results show that the DRAM algorithm is computationally more efficient and leads to chains with less autocorrelation than the M–H algorithm.

The detailed procedure of posterior sampling with DRAM and Gibbs algorithms is as follows. Let M be the length of the chain and m_0 be an initial non-adaptation period. Also let $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, \boldsymbol{\vartheta}_2^T)^T$. Suppose that $(\boldsymbol{\vartheta}^0, \sigma^0)$ is the initial value sampled from the uniform distribution $U(0, 1)$. At the step m ($m = 1, \ldots, M$), sample $\boldsymbol{\vartheta}^{(m)}$ and $\sigma^{(m)}$ based on the following procedure.

(i) Sample $\boldsymbol{\vartheta}^{(m)}$:

(a) At the first stage of DR, we take the Gaussian proposal to be centered at the current position $\boldsymbol{\vartheta}^{(m-1)}$ and set the covariance C_m^1 to be

$$C_m^1 = \begin{cases} C_0, & m \le m_0; \\ s_{\tilde{p}} Cov(\boldsymbol{\vartheta}^0, \ldots, \boldsymbol{\vartheta}^{(m-1)}) + s_{\tilde{p}} \varepsilon I_{\tilde{p}}, & m > m_0, \end{cases}$$

where C_0 is the initial covariance and in this paper we set $C_0 = I_{\tilde{p}}$, the $\tilde{p}$-dimensional identity matrix with $\tilde{p}$ being the dimension of $\boldsymbol{\vartheta}$, $s_{\tilde{p}}$ is a scaling parameter depending on $\tilde{p}$, and $\varepsilon \ge 0$ is a chosen constant to ensure that C_m will not become singular. We take $s_{\tilde{p}} = 2.38^2/\tilde{p}$ according to Haario et al. (2006) and Gelman et al. (1996) because this value optimizes the mixing properties in the case of Gaussian targets and Gaussian proposals. Throughout our numerical studies, we simply let $\varepsilon = 0$ since this choice did not lead to any singular C_m in our setup.

(b) For the jth ($j = 2, \ldots, J$) stage, the covariance C_m^j of the proposal is computed simply as a scaled version of the proposal of the first stage, that is $C_m^j = \gamma_j C_m^1$ for some constants γ_j. We simply set $J = 2$ and $\gamma_2 = 0.01$ in this paper.

(ii) Sample $\sigma^{(m)}$:

$\sigma^{(m)}$ is sampled from IG $\left(n, \sum_{i=1}^n \sum_{l=1}^L \rho_{\tau_l}\{y_i - \mathbf{x}_i^T(\boldsymbol{\vartheta}_1^{(m)} + \boldsymbol{\vartheta}_2^{(m)}\delta_l)\}\right)$, where IG$(a, b)$ is the Inverse Gamma distribution with shape parameter a and scale parameter b.

2.3 Estimation of the tail conditional quantile

Based on the Markov chains, we can calculate the posterior means of $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$, denoted as $\widehat{\boldsymbol{\vartheta}}_1$ and $\widehat{\boldsymbol{\vartheta}}_2$. The quantile coefficient vector at the quantile level $\tau^\star$ can then be estimated through extrapolation by

$$\widehat{\boldsymbol{\beta}}(\tau^\star) = \widehat{\boldsymbol{\vartheta}}_1 + \widehat{\boldsymbol{\vartheta}}_2 \left(\tau^\star - \sum_{l=1}^{L} \tau_l / L \right).$$

Then the conditional quantile at the quantile level $\tau^\star$ can be estimated by

$$\widehat{Q}_Y(\tau^\star|\mathbf{x}) = \mathbf{x}^T \widehat{\boldsymbol{\beta}}(\tau^\star).$$

When the interest is on a single high quantile level $\tau^\star$, our empirical study shows that letting $\tau_L = \tau^\star$, that is, including $\tau^\star$ in the quantile grid, often leads to more stable estimation than extrapolating from $\tau_L < \tau^\star$.

3 Simulation study

We conduct a simulation study to assess the finite sample performance of the proposed Bayesian joint-quantile regression estimator, referred to as BJQR. For comparison, we also include (1) the conventional linear quantile regression estimator (QR), computed by using the R function "rq" in the *quantreg* package; (2) the constant composite quantile regression estimator (CCQR), which assumes that the quantile slope coefficients are constant and estimates the constant slope by combining information from quantile levels $\tau_1 < \cdots < \tau_L$; (3) the extreme-value-based method in Wang et al. (2012), which extrapolates from intermediate quantiles to estimate conditional quantiles at the extreme tail, referred to as EXQR.

We consider the following three examples. In all the cases, x_1, x_2, ε (or u) are mutually independent.

– Example 1:

$$y_i = 1 + 3x_{1i} + 3x_{2i} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $x_{1i}, x_{2i} \sim N(0, 1)$ and $\varepsilon_i \sim t_3$.

– Example 2:

$$y_i = \beta_0(u_i) + \beta_1(u_i)x_{1i} + \beta_2(u_i)x_{2i}, \quad i = 1, \ldots, n,$$

where $x_{1i}, x_{2i} \sim U(0, 4)$, $u_i \sim U(0, 1)$. In addition, the coefficients $\beta_j(u)$ ($j = 0, 1, 2$) are defined in the following way:

1. if $u < 0.93$, $\beta_0(u) = 1 + F^{-1}(u) - F^{-1}(0.93)$, $\beta_1(u) = 2 - F^{-1}(0.93) + F^{-1}(u)$, $\beta_2(u) = 3 - F^{-1}(0.93) + F^{-1}(u)$, where F is the CDF of t(3) distribution;

2. if $u \geq 0.93$, $\beta_0(u) = 1 + u - 0.93$, $\beta_1(u) = 2 + 4(u - 0.93)u$, $\beta_2(u) = 3 + 5(u - 0.93)$.

– Example 3:

$$y_i = 1 + x_{1i} + 3x_{2i} + (2 + 1.5x_{1i})\varepsilon_i, \quad i = 1, \ldots, n,$$

where $x_{1i} \sim U(-1, 1)$ and $x_{2i}, \varepsilon_i \sim N(0, 1)$.

For the three examples, the true conditional quantile function is $Q_Y(\tau|x_1, x_2) = \beta_0(\tau)+x_1\beta_1(\tau)+x_2\beta_2(\tau)$. Example 1 contains independent and identically distributed (i.i.d.) errors so the quantile slope effects $\beta_1(\tau) = \beta_2(\tau) = 3$ are constant across τ. Both examples 2 and 3 have heteroscedastic errors. In Example 2, the three quantile coefficients $\beta_j(\tau)$, $j = 0, 1, 2$, are all linear in $\tau \in [0.93, 1)$. In Example 3, $\beta_2(\tau) = 3$ is constant and $\beta_1(\tau) = 1 + 1.5\Phi^{-1}(\tau)$ is a nonlinear function of τ.

We consider four sample sizes $n \in \{50, 200, 500, 2000\}$. The simulation is repeated 500 times for each scenario. We focus on two high quantile levels $\tau^\star = 0.99$ and 0.995. For both BJQR and CCQR, we let the L quantile levels in the grid be equally spaced in the interval in $[\tau_1, \tau_L = \tau^\star]$, and we choose $\tau_1 = \tau_L - d$ with $d \in \{0.01, 0.02, 0.05\}$ to assess the sensitivity of the methods against the length of the quantile interval. Results for both BJQR and CCQR are based on $L = 5$.

For the BJQR method, we obtain MCMC chains using the DRAM sampling procedure with an initial non-adaptation period $m_0 = 100$ and a 2-stage DR algorithm. We let the initial covariance C_0 be the identity matrix, $s_{\tilde{p}} = 2.38^2/\tilde{p}$ with $\tilde{p} = 6$, and $\gamma = 0.01$. In all cases, the initial values of the parameters are randomly drawn from $U(0, 1)$, the length of the chain is set as $M = 40{,}000$ with a burn-in period of 10,000.

Tables 1, 2 and 3 summarize the bias and mean squared error of the quantile coefficient estimation, and the mean integrated squared error (MISE) of the conditional quantile estimation at $\tau^\star = 0.995$ in three examples, respectively. Results for $\tau^\star = 0.99$ are similar and thus omitted. The MISE is defined to be the average of ISE across simulations, where

$$\text{ISE} = \frac{1}{n}\sum_{i=1}^{n}\{\widehat{Q}_Y(\tau^\star|\mathbf{x}_i) - Q_Y(\tau^\star|\mathbf{x}_i)\}^2.$$

For both CCQR and BJQR, the tuning parameter d measures the range of quantile levels, over which a constant and linear expansion are used to approximate the quantile coefficients, respectively. The parameter d plays a similar role as the bandwidth parameter in kernel smoothing for balancing between bias and variance: a larger d tends to give smaller variance but larger bias. To achieve a better balance between bias and variance, we suggest a smaller d for larger samples, and our numerical studies showed that the rule of thumb $d = \max(0.7n^{-0.7}, 0.01)$ worked well; the corresponding choices of d are marked bold in Tables 1, 2 and 3.

EXQR method is not included in Example 2, since in this case the conditional distributions of Y given X have bounded ranges so the extreme value theory based method EXQR does not really apply. Results show that the EXQR method from Wang

Table 1 The Bias, MSE and MISE of three methods at $\tau^\star = 0.995$ and $L = 5$ in Example 1

n	d	Method	100×Bias			MSE			MISE
			$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	
50		QR	− 233.11	3.38	− 9.48	11.55	2.08	1.78	14.33 (2.25)
		EXQR	516.10	− 0.72	− 2.28	276.66	0.32	0.28	282.46 (163.49)
	0.01	CCQR	− 233.11	3.38	− 9.48	11.55	2.08	1.78	14.33 (2.25)
		BJQR	− 183.88	1.98	− 8.49	10.94	1.18	1.14	12.46 (2.64)
	0.02	CCQR	− 233.10	3.43	− 9.49	11.55	2.08	1.78	14.33 (2.25)
		BJQR	− 171.89	1.80	− 8.50	11.37	1.20	1.18	12.85 (2.99)
	0.05	CCQR	− 208.37	4.14	− 7.20	12.84	0.84	0.73	14.31 (3.17)
		BJQR	− 156.47	1.04	− 8.96	11.18	1.50	1.37	13.05 (3.12)
200		QR	18.94	− 17.62	13.24	8.54	2.93	2.44	13.58 (2.51)
		EXQR	27.09	− 3.11	0.09	2.69	0.07	0.07	2.89 (0.30)
	0.01	CCQR	73.83	− 12.34	− 4.77	14.10	1.17	0.93	16.14 (3.71)
		BJQR	− 16.19	− 13.89	2.80	4.64	1.69	1.41	7.64 (1.14)
	0.02	CCQR	109.15	− 5.94	− 3.72	15.94	0.56	0.50	16.98 (3.58)
		BJQR	− 50.47	− 12.13	− 2.08	3.57	1.32	1.06	5.87 (0.95)
	0.05	CCQR	145.30	− 2.95	− 0.78	17.68	0.16	0.15	18.02 (3.56)
		BJQR	− 126.79	− 7.18	− 2.56	2.94	0.68	0.61	4.20 (0.21)

Table 1 continued

n	d	Method	100×Bias			MSE			MISE
			$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	
500		QR	19.25	2.94	− 7.54	2.38	1.19	1.17	4.74 (0.38)
		EXQR	10.10	− 0.05	0.26	1.00	0.03	0.02	1.07 (0.07)
	0.01	CCQR	− 23.79	− 0.01	− 3.89	1.56	0.37	0.36	2.27 (0.13)
		BJQR	− 20.45	2.68	− 6.35	1.33	0.70	0.80	2.80 (0.15)
	0.02	CCQR	− 15.11	− 0.84	− 1.60	1.63	0.18	0.18	1.99 (0.16)
		BJQR	− 58.51	2.71	− 5.85	1.14	0.52	0.55	2.20 (0.08)
	0.05	CCQR	− 4.25	− 1.33	0.10	1.64	0.06	0.05	1.75 (0.17)
		BJQR	− 137.17	− 0.29	− 2.54	2.23	0.28	0.28	2.78 (0.06)
2000		QR	2.35	4.42	0.36	0.56	0.37	0.36	1.29 (0.08)
		EXQR	3.74	0.22	0.28	0.40	0.01	0.01	0.41 (0.03)
	0.01	CCQR	101.67	2.47	0.01	4.07	0.12	0.11	4.30 (0.46)
		BJQR	− 28.84	3.53	− 0.39	0.45	0.28	0.26	0.99 (0.04)
	0.02	CCQR	114.27	1.99	0.27	5.09	0.06	0.05	5.19 (0.58)
		BJQR	− 63.87	3.40	− 0.16	0.63	0.21	0.18	1.02 (0.03)
	0.05	CCQR	57.88	0.88	− 0.10	2.24	0.02	0.01	2.27 (0.28)
		BJQR	− 141.91	2.11	− 0.06	2.11	0.10	0.08	2.29 (0.04)

Values in parentheses are standard errors of MISE. The bold value of d corresponds to the rule of thumb $d = \max(0.7n^{-0.7}, 0.01)$

Table 2 The Bias, MSE and MISE of three methods at $\tau^\star = 0.995$ and $L = 5$ in Example 2

n	d	Method	100 × Bias			100 × MSE			100 × MISE
			$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	
50		QR	42.49	− 40.97	− 36.84	1036.46	171.96	159.92	625.55 (43.90)
	0.01	CCQR	42.49	− 40.97	− 36.84	1036.46	171.96	159.92	625.55(43.90)
		BJQR	218.98	− 49.18	− 41.99	1167.01	136.54	122.92	443.51 (26.91)
	0.02	CCQR	43.27	− 41.18	− 36.95	1044.11	172.46	160.13	626.86 (43.91)
		BJQR	246.66	− 49.62	− 42.86	1307.52	137.00	123.26	471.02 (25.37)
	0.05	CCQR	71.69	− 44.25	− 39.85	1089.24	164.60	147.72	571.77 (40.46)
		BJQR	278.37	− 47.89	− 40.43	1502.35	140.70	124.78	551.01 (23.76)
200		QR	9.99	− 6.08	− 7.35	42.53	5.65	5.87	20.87 (2.33)
	0.01	CCQR	18.28	− 6.98	− 8.71	43.80	5.35	6.08	20.13 (2.05)
		BJQR	48.84	− 8.90	− 10.49	57.42	5.34	5.74	18.70 (1.47)
	0.02	CCQR	32.51	− 9.25	− 10.73	64.61	5.91	6.88	21.42 (2.27)
		BJQR	56.47	− 9.16	− 10.88	67.21	5.52	6.04	21.12 (1.46)
	0.05	CCQR	84.19	− 16.58	− 20.12	225.11	13.16	16.10	53.32 (5.61)
		BJQR	70.75	− 7.97	− 10.86	94.62	6.40	7.27	33.33(1.73)

Table 2 continued

n	d	Method	100 × Bias			100 × MSE			100 × MISE
			$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	
500		QR	4.07	− 2.38	− 2.50	4.12	0.67	0.64	2.64 (0.21)
	0.01	CCQR	9.48	− 4.05	− 4.66	5.10	0.81	0.95	3.65 (0.25)
		BJQR	18.95	− 3.70	− 4.51	7.17	0.71	0.80	2.60 (0.16)
	0.02	CCQR	20.99	− 6.27	− 7.49	10.37	1.21	1.45	4.73 (0.28)
		BJQR	22.78	− 4.07	− 4.96	9.02	0.78	0.89	3.09 (0.16)
	0.05	CCQR	57.14	− 12.55	− 15.59	49.73	3.56	4.27	11.87 (0.60)
		BJQR	32.87	− 4.61	− 5.35	15.45	0.94	0.99	5.16 (0.17)
2000		QR	0.19	− 0.34	− 0.41	0.25	0.06	0.07	0.29 (0.02)
	0.01	CCQR	8.35	− 2.40	− 3.04	1.23	0.15	0.21	0.66 (0.03)
		BJQR	3.68	− 0.89	− 1.10	0.45	0.07	0.09	0.30 (0.02)
	0.02	CCQR	19.13	− 4.62	− 6.05	4.77	0.35	0.56	1.39 (0.06)
		BJQR	5.16	− 1.08	− 1.33	0.64	0.09	0.11	0.38 (0.02)
	0.05	CCQR	54.27	− 11.15	− 14.27	32.94	1.55	2.42	5.71 (0.15)
		BJQR	9.15	− 1.35	− 1.92	1.38	0.13	0.17	0.68 (0.03)

Values in parentheses are standard errors of MISE. The bold value of d corresponds to the rule of thumb $d = \max(0.7n^{-0.7}, 0.01)$

Table 3 The Bias, MSE and MISE of three methods at $\tau^\star = 0.995$ and $L = 5$ in Example 3

n	d	Method	100 × Bias			MSE			MISE
			$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	
50		QR	− 130.38	− 134.83	7.57	2.39	4.18	0.60	4.27 (0.15)
		EXQR	601.59	− 260.20	1.10	89.35	7.95	0.33	95.86 (22.86)
	0.01	CCQR	− 130.38	− 134.83	7.57	2.39	4.18	0.60	4.27 (0.15)
		BJQR	− 76.65	− 145.33	5.70	1.42	3.50	0.32	2.82 (0.11)
	0.02	CCQR	− 130.38	− 134.80	7.51	2.39	4.17	0.59	4.27 (0.15)
		BJQR	− 66.11	− 145.36	5.63	1.33	3.50	0.32	2.73 (0.10)
	0.05	CCQR	− 117.65	− 141.99	7.42	2.26	3.56	0.42	3.79 (0.13)
		BJQR	− 50.79	− 138.40	5.66	1.21	3.43	0.34	2.61 (0.10)
200		QR	− 22.87	− 57.87	2.35	0.53	1.74	0.35	1.44 (0.07)
		EXQR	116.50	− 254.67	0.15	2.03	6.74	0.06	4.49 (0.12)
	0.01	CCQR	− 4.60	− 68.33	0.92	0.78	1.24	0.19	1.37 (0.09)
		BJQR	− 20.04	− 61.08	0.23	0.44	1.29	0.20	1.06 (0.05)
	0.02	CCQR	13.30	− 85.94	0.63	0.94	1.22	0.13	1.47 (0.10)
		BJQR	− 26.81	− 62.74	− 0.02	0.40	1.15	0.17	0.94 (0.04)
	0.05	CCQR	43.23	− 129.91	1.10	1.32	1.96	0.06	2.03 (0.12)
		BJQR	− 46.84	− 71.14	0.58	0.46	1.04	0.13	0.93 (0.04)

Table 3 continued

n	d	Method	100 × Bias			MSE			MISE
			$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	$\beta_0(\tau^\star)$	$\beta_1(\tau^\star)$	$\beta_2(\tau^\star)$	
500		QR	− 5.62	− 18.98	− 0.72	0.22	0.57	0.12	0.53 (0.03)
		EXQR	87.77	− 249.29	− 0.87	1.02	6.31	0.03	3.21 (0.05)
	0.01	CCQR	− 20.62	− 51.56	− 0.64	0.21	0.51	0.07	0.44 (0.02)
		BJQR	− 12.17	− 27.15	− 1.10	0.18	0.45	0.08	0.41 (0.02)
	0.02	CCQR	− 15.82	− 73.94	− 0.35	0.22	0.71	0.05	0.49 (0.02)
		BJQR	− 21.60	− 33.26	− 0.94	0.19	0.43	0.07	0.41 (0.02)
	0.05	CCQR	− 0.32	− 121.38	− 0.21	0.25	1.56	0.03	0.79 (0.02)
		BJQR	− 44.75	− 46.84	− 0.48	0.30	0.44	0.05	0.50 (0.02)
2000		QR	1.13	− 1.70	1.54	0.05	0.13	0.03	0.13 (0.01)
		EXQR	61.69	− 247.54	0.51	0.44	6.15	0.01	2.51 (0.02)
	0.01	CCQR	7.67	− 43.69	0.84	0.19	0.26	0.01	0.28 (0.03)
		BJQR	− 7.19	− 9.11	1.13	0.05	0.11	0.02	0.11 (0.01)
	0.02	CCQR	12.38	− 69.58	0.53	0.25	0.53	0.01	0.44 (0.05)
		BJQR	− 17.24	− 16.05	0.86	0.07	0.11	0.02	0.12 (0.01)
	0.05	CCQR	19.81	− 118.56	0.91	0.24	1.43	0.01	0.72 (0.04)
		BJQR	− 43.72	− 34.90	0.51	0.22	0.18	0.01	0.29 (0.01)

Values in parentheses are standard errors of MISE. The bold value of d corresponds to the rule of thumb $d = \max(0.7n^{-0.7}, 0.01)$

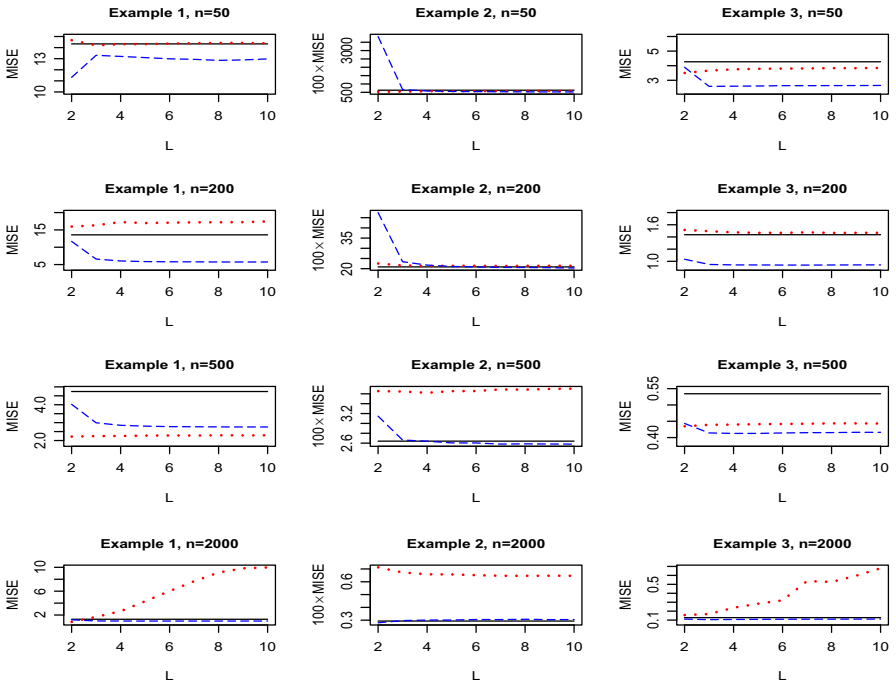

Fig. 1 The mean integrated squared errors of the QR (solid), CCQR (dotted) and BJQR (dashed) estimators at $\tau^\star = 0.995$ across $L \in \{2, 3, \ldots, 10\}$ based on the rule of thumb $d = \max(0.7n^{-0.7}, 0.01)$ in Examples 1–3 for $n = 50$, 200, 500 and 2000

et al. (2012) performs better in Example 1 (with homoscedastic heavy-tailed errors) for larger sample sizes with $n \geq 200$. However, EXQR gives worse performance for small samples with $n = 50$ and in Example 3 with heteroscedastic errors. Such results are not surprising since the EXQR is based on the extreme value theory, whose approximation would require a large sample size. The EXQR method has advantages for estimating extreme quantiles of heavy-tailed distributions with large samples. In contrast, the proposed BJQR method has advantages for estimating tail quantiles with modest samples.

We also assess the sensitivity of the proposed method against the choice of L. Figure 1 presents the MISE of QR, CCQR and BJQR at $\tau^\star = 0.995$ across $L \in [2, 10]$ based on the rule of thumb $d = \max(0.7n^{-0.7}, 0.01)$ in Examples 1–3 with $n \in \{50, 200, 500, 2000\}$. Results show that the BJQR method performs quite stably and better than QR for $L \in [4, 10]$, except in Example 2 with $n = 50$. In contrast, the CCQR method is more sensitive to the choice of L, and it could perform much worse than QR if L is not chosen well, e.g., for $n = 2000$.

For multiple-quantile analysis, quantile crossing may happen in numerical studies, that is, the lower quantiles may be estimated to be larger than the upper quantiles. We compare different methods for checking how often the estimated conditional quantiles cross each other. Table 4 summarizes the percentage of times the estimated quantiles $\{\hat{Q}_Y(\tau_l|\mathbf{x}_i), l = 1, \ldots, L\}$ cross at two or more quantile levels over $\{\mathbf{x}_i, i = 1, \ldots, n\}$

Table 4 The average percentage of times the estimated quantiles $\{\hat{Q}_{\tau_l}(\mathbf{x}_i), l = 1, \ldots, L\}$ cross at two or more quantile levels over $\{\mathbf{x}_i, i = 1, \ldots, n\}$ from different methods in Examples 1–3, where $\tau_L = 0.995$

Example	d	Method	n = 50	n = 200	n = 500	$n = 2000$
1	0.01	QR	25.94	68.66	45.02	13.58
		CCQR	81.40	40.60	5.20	5.60
		BJQR	0.39	2.01	1.92	0.35
	0.02	QR	65.14	52.24	25.75	5.69
		CCQR	57.80	5.79	0.00	1.80
		BJQR	0.09	1.44	0.79	0.09
	0.05	QR	71.90	26.12	11.16	1.16
		CCQR	28.00	0.00	0.00	0.00
		BJQR	0.03	0.53	0.19	0.01
2	0.01	QR	43.45	72.88	56.52	15.56
		CCQR	80.80	42.00	3.05	0.00
		BJQR	0.09	0.08	0.04	0.13
	0.02	QR	60.23	60.84	33.62	6.91
		CCQR	64.00	3.80	0.00	0.00
		BJQR	0.00	0.13	0.04	0.06
	0.05	QR	70.57	35.66	12.93	2.31
		CCQR	29.60	0.00	0.00	0.00
		BJQR	0.00	0.12	0.05	0.02
3	0.01	QR	13.78	69.43	49.71	15.48
		CCQR	81.17	44.78	4.15	1.80
		BJQR	0.16	1.16	1.48	0.51
	0.02	QR	53.80	57.80	28.77	7.75
		CCQR	66.00	5.40	0.00	0.00
		BJQR	0.00	0.84	0.56	0.13
	0.05	QR	67.18	28.91	12.32	2.52
		CCQR	26.40	0.00	0.00	0.00
		BJQR	0.00	0.30	0.15	0.01

from different methods with $L = 5$ and $\tau_L = 0.995$. We do not include EXQR since its estimation is based on extrapolation and is ensured to be increasing in τ. The results show that quantile crossing happens rarely and much less frequently from BJQR than the QR and CCQR methods. This is likely due to the approximation of the quantile coefficient $\beta_j(\tau)$ as a local linear function of τ used in BJQR. Note that to ensure the monotonicity of $Q_Y(\tau|\mathbf{x})$ in τ under the model (4), the sufficient and necessary condition is $\mathbf{x}^T \boldsymbol{\vartheta}_2 > 0$. Therefore, under the Bayesian framework, we can improve the proposed algorithm to avoid quantile crossing by posing priors with constraints on $\boldsymbol{\vartheta}_2$, and one possible prior is the truncated normal as in Feng et al. (2015).

In summary, compared to CCQR, the BJQR method is more flexible and robust against model misspecifications. When the constant slope assumption required by CCQR is correctly specified in Example 1, BJQR gives competitive performance to

CCQR. However, when the assumptions required by both methods are misspecified (Example 3), the performance of CCQR could be even worse than QR, while BJQR still gives better estimation at tails than QR. In terms of computing time, the QR is computationally more efficient since the estimation is obtained at each single quantile separately by minimizing the quantile objective function. In this simulation study, it took about 0.08, 0.202 and 21.7 s for QR, CCQR and BJQR to obtain the quantile estimations for a sample with $n = 2000$ observations, and with $L = 5$ and $M = 40{,}000$. Overall, BJQR provides more stable and efficient estimation than the conventional QR method at tail quantiles for small and modest sample sizes. For larger samples, the gain of reduced variance in the BJQR estimator is offset by the relatively larger bias, especially for larger d, which is caused by the Taylor approximation. Considering both numerical and computational efficiency, we would recommend the proposed BJQR for small to modest samples $n \leq 500$ at quantile levels around 0.99 or higher.

Although we focus on the upper quantiles in this paper, we also conducted simulation at $\tau^\star = 0.005$ and the main observations are similar; the results can be found in the online Supplementary Material.

4 Application to the Chicago precipitation data

In this section, we apply the proposed method to perform statistical downscaling of daily precipitation data in the Chicago/Aurora airport station in Illinois. One aim of the statistical downscaling is to establish a relationship between local-scale historical measurement and large-scale general circulation model predictions, and to extend the relationship to future projections. We focus on the modeling and prediction of extreme heavy precipitations, which is one type of extreme events which happen rarely but have huge societal impacts. For the analysis of extreme events, the estimation of tail quantiles would be more interesting than the central summaries. On the other hand, the problem is challenging due to the sparseness of data in the tail areas, and this motivates us to apply the proposed method to provide more stable tail quantile estimation through joint-quantile analysis.

We take the observed daily precipitation (in inches), at the Aurora Municipal Airport from September 1957 to August 2002 as the response variable Y. The predictors are the simulated daily maximum temperature (X_1), daily humidity (X_2) and daily precipitation (X_3), all generated from the ERA-40 reanalysis model in Uppala and Kallberg (2005). We focus on the wet days only, and this results in total 5132 observations. Our main objective is to predict the high conditional quantiles of local precipitation based on the simulated reanalysis variables generated from the global climate model. We focus on the high quantile $\tau^\star \in \{0.99, 0.995, 0.999, 0.9999\}$. Before analyzing the data, we standardize each predictor by subtracting the mean and dividing by the standard deviation.

For demonstration, we randomly select $m = 200$ observations as training data, and leave the rest as testing data. For each method, we use the training data to estimate the quantile coefficients, which are then used to predict the $\tau^\star$th conditional quantile of the testing data. We then calculate the prediction error (PE) as

Table 5 The average predictor error of three methods for predicting the $\tau^\star$th conditional quantile of the daily precipitation in the Aurora station across 500 cross validations

Method	$\tau^\star$			
	0.99	0.995	0.999	0.9999
QR	126.75	90.25	60.58	53.90
	(1.08)	(0.91)	(1.05)	(1.12)
CCQR	118.72	84.43	51.66	45.03
	(0.79)	(0.86)	(0.88)	(1.02)
BLCQR	118.77	83.27	48.78	40.06
	(0.79)	(0.70)	(0.83)	(0.90)

The values in parentheses are the standard errors

$$\text{PE} = \sum_{i \in \text{testing data}} \rho_\tau\{y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}(\tau^\star)\}.$$

The cross validation is repeated 500 times to report the average prediction error (APE). As suggested by the simulation, we let $\tau_L = \tau^\star$, $L = 5$ and $d = 0.02$ for the proposed BJQR method. For the BJQR method, we let the length of chains be $M = 40{,}000$ and burn-in period be 3000. Table 5 summarizes the average prediction errors of the QR, CCQR and BJQR methods at $\tau^\star \in \{0.99, 0.995, 0.999, 0.9999\}$. Results suggest that both CCQR and BJQR give more accurate prediction than QR at the four high quantiles considered, but BJQR is consistently better than CCQR. This confirms with what we observed in the simulation study that the proposed BJQR method is more flexible than CCQR, and it leads to more accurate tail quantile estimation than QR for small and modest samples.

For illustration, we compare the results from QR, CCQR and BJQR for estimating the conditional quantiles $Q_Y(\tau^\star|\mathbf{X} = \mathbf{x}_u)$ across $\tau^\star \in \{0.99, 0.995, 0.999, 0.9999\}$ based on one training data set, where $\mathbf{x}_u = (x_{u1}, x_{u2}, x_{u3})^T$, and x_{uj} is taken to the uth sample quantile of the jth predictor X_j, $j = 1, 2, 3$. Figure 2 plots the estimated conditional quantiles $\hat{Q}_Y(\tau^\star|\mathbf{X} = \mathbf{x}_u)$ from three methods at $u \in \{0.1, 0.4, 0.6, 0.9\}$. Results show that the QR estimates become a constant when $\tau > 0.995$ and the CCQR estimates even decrease with τ for $\tau > 0.995$ in most of the cases. In contrast, the BJQR estimates increase in τ, confirming the observation from the simulation study that the BJQR avoids the quantile crossing issue better than the QR and CCQR methods.

5 Conclusion and discussion

We develop a new Bayesian method for estimating tail conditional quantiles through joint-quantile estimation and a linear approximation of quantile coefficients in a small quantile interval. Motivated by the connection of quantile regression at a single quantile level with the asymmetric Laplace distribution, we propose a working likelihood for joint-quantile estimation and a MCMC sampling procedure that combines the delayed rejection and adaptive Metropolis and Gibbs algorithms. Numerical results

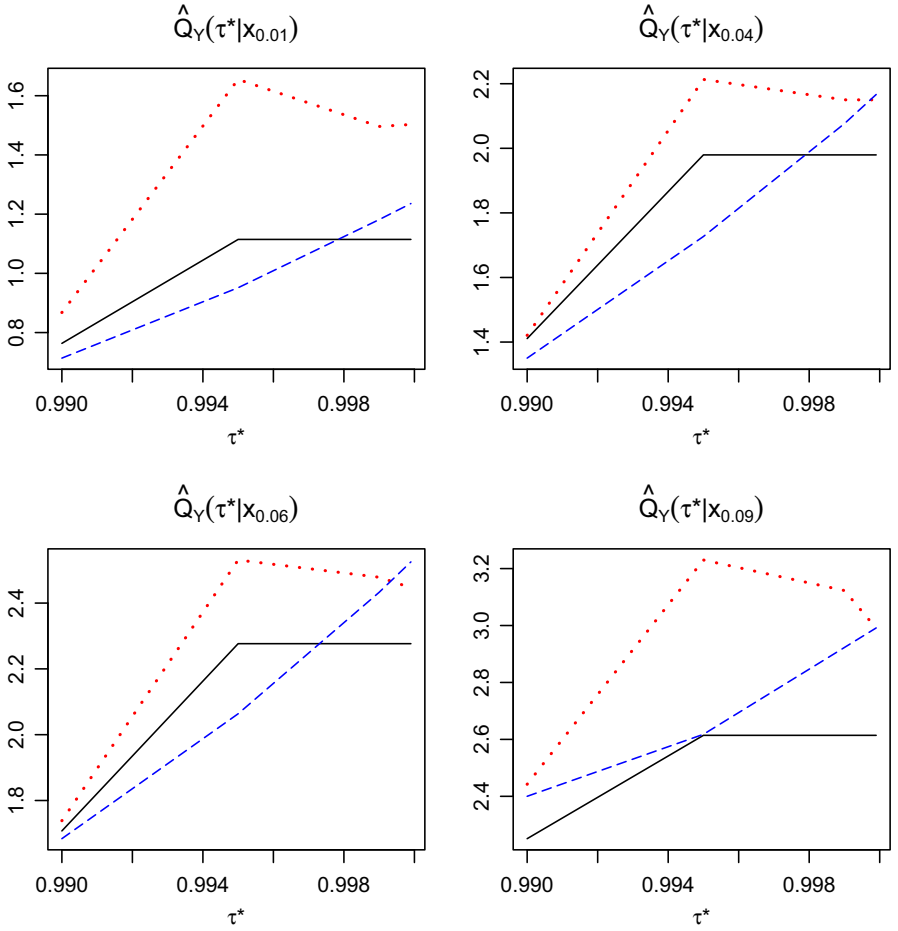

Fig. 2 The estimated conditional quantiles $\hat{Q}_Y(\tau|\mathbf{x}_u)$ from the QR (solid), CCQR (dotted) and BJQR (dashed) methods for the Chicago precipitation data

show that the proposed method leads to more accurate estimation at tail quantiles than the conventional quantile regression method especially for small and modest samples. In addition, the proposed method is more flexible and robust than the existing composite quantile regression estimator that relies on a more stringent assumption of common quantile slopes across quantiles.

The proposed method is based on a working likelihood that gives equal weight to different quantiles. Under the frequentist framework and for the special case of common-quantile-slope models, Wang and Wang (2016) and Zhao and Xiao (2014) studied the optimal choice of weight $w(\tau)$ for composite quantile estimation at central and extreme quantiles, respectively. The optimal weight depends on unknown quantities and often is difficult to estimate. In addition, at tail quantiles, the optimal weight may be negative depending on the distribution and this will lead to computational challenges. It is an interesting future topic to study how to further improve the effi-

ciency by adopting appropriate and convenient quantile-specific weight in the working likelihood for Bayesian joint-quantile analysis.

One advantage of the Bayesian framework is that it can provide inference based on the posterior draws. For the proposed joint-quantile regression, samples from the posterior distribution may provide uncertainty estimates for the parameters. When an ALD working likelihood is used for Bayesian quantile regression at a single quantile level, Yang et al. (2016) showed that the posterior inference is asymptotically valid with a simple posterior variance adjustment. In the Bayesian joint-quantile problem, such an adjustment is not as straightforward but can be done based on the work of Chernozhukov and Hong (2003). The variance estimates for a quantile regression estimator in the frequentist approach would require us to estimate quantities related to the conditional densities or to use the bootstrap methods. Formal inference based on the working likelihood for Bayesian joint-quantile analysis is a challenging problem and requires further studies.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 11631003, 11690012, 11726629 and 11701491), and Grant DMS-1712760 and the IR/D program from the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

Alhamzawi R (2016) Bayesian analysis of composite quantile regression. Stat Biosci 8:358–373

Chernozhukov V, Hong H (2003) An MCMC approach to classical estimation. J Econometrics 115(2):293–346

Chernozhukov V, Du S (2008) Extremal quantiles and value-at-risk. In: Durlauf SN, Blume LE (eds) The new Palgrave dictionary of economics. Palgrave Macmillan, Basingstoke

Feng Y, Chen Y, He X (2015) Bayesian quantile regression with approximate likelihood. Bernoulli 21(2):832–850

Gardes L, Girard S (2011) Functional kernel estimators of conditional extreme quantiles. In: Ferraty F (ed) Recent advances in functional data analysis and related topics. Springer, Berlin, pp 135–140

Gelman A, Roberts GO, Gilks WR (1996) Efficient Metropolis jumping rules. In: Bernardo JM, Berger JO, Dawid AP and Smith AMF (eds) Bayesian statistics 5. Oxford Univ. Press, New York, pp 599–607

Geraci M, Bottai M (2007) Quantile regression for longitudinal data using the asymmetric laplace distribution. Biostatistics 8(1):140–154

Green P, Mira A (2001) Delayed rejection in reversible jump Metropolis–Hastings. Biometrika 88(4):1035–1053

Guo J, Tian M, Zhu K (2012) New efficient and robust estimation in varying-coefficient models with heteroscedasticity. Stat Sin 22(3):1075–1101

Haario H, Saksman E, Tamminen J (1999) Adaptive proposal distribution for random walk metropolis algorithm. Comput Stat 14(3):375–395

Haario H, Saksman E, Tamminen J (2001) An adaptive metropolis algorithm. Bernoulli 7(2):223–242

Haario H, Laine M, Mira A, Saksman E (2006) DRAM: efficient adaptive MCMC. Stat Comput 16(4):339–354

Huang HW, Chen ZX (2015) Bayesian composite quantile regression. J Stat Comput Simul 85(18):3744–3754

Jiang L, Wang H, Bondell HD (2013a) Interquantile shrinkage in regression models. J Comput Graph Stat 22(4):970–986

Jiang R, Zhou Z, Qian W, Chen Y (2013b) Two-step composite quantile regression for single-index models. Comput Stat Data Anal 64:180–191

Jiang R, Hu XP, Yu KM, Qian WM (2018) Composite quantile regression for massive datasets. Statistics 52(5):980–1004
Kai B, Li R, Zou H (2010) Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. J R Stat Soc B 72(1):49–69
Khare K, Hobert JP (2012) Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression. J Multivar Anal 112:108–116
Koenker R (1984) A note on L-estimates for linear models. Stat Probab Lett 2(6):323–325
Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge
Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46(1):33–50
Kozumi H, Kobayashi G (2011) Gibbs sampling methods for Bayesian quantile regression. J Stat Comput Simul 81(11):1565–1578
Li Q, Xi R, Lin N (2010) Bayesian regularized quantile regression. Bayesian Anal 5:533–556
Mira A (2002) On Metropolis–Hastings algorithms with delayed rejection. Metron 59(3–4):231–241
Reich BJ, Smith LB (2013) Bayesian quantile regression for censored data. Biometrics 69:651–660
Reich BJ, Fuentes M, Dunson DB (2011) Bayesian spatial quantile regression. J Am Stat Assoc 106(493):6–20
Rodrigues T, Fan Y (2017) Regression adjustment for noncrossing Bayesian quantile regression. J Comput Graph Stat 26(2):275–284
Rodrigues T, Dortet-Bernadet J-L, Fan Y (2019) Pyramid quantile regression. J Comput Graph Stat 28(3):732–746
Sriram K, Ramamoorthi RV, Ghosh P (2012) Simultaneous Bayesian estimation of multiple quantiles with an extension to hierarchical models. Social Science Electronic Publishing, New York
Sriram K, Ramamoorthi RV, Ghosh P (2016) On Bayesian quantile regression using a pseudo-joint asymmetric Laplace likelihood. Sankhyā Indian J Stat 78–A:87–104
Tang Y, Wang H, Liang H (2018) Composite estimation for single-index models with responses subject to detection limits. Scand J Stat 45(3):444–464
Tian YZ, Lian H, Tian MZ (2017) Bayesian composite quantile regression for linear mixed-effects models. Commun Stat Theory Methods 46(15):7717–7731
Uppala SM, Kallberg PW et al (2005) The era-40 re-analysis. Q J R Meteorol Soc 131:2961–3012
Wang H, Li D (2013) Estimation of extreme conditional quantiles through power trnsformation. J Am Stat Assoc 108:1062–1074
Wang K, Wang H (2016) Optimally combined estimation for tail quantile regression. Stat Sin 26(1):295–311
Wang H, Yang Y (2018) Bayesian quantile regression. In: Koenker R, Chernozhukov V, He X, Peng L (eds) Handbook of quantile regression, Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Boca Raton, pp 41–54
Wang H, Li D, He X (2012) Estimation of high conditional quantiles for heavy-tailed distributions. J Am Stat Assoc 107:1453–1464
Wu YK, Tian MZ, Tang ML (2020) General composite quantile regression: theory and methods. Commun Stat Theory Methods 49(9):2217–2236
Yang Y, Tokdar S (2017) Joint estimation of quantile planes over arbitrary predictor spaces. J Am Stat Assoc 112(519):1107–1120
Yang Y, Wang H, He X (2016) Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. Int Stat Rev 84(3):327–344
Yu K, Moyeed RA (2001) Bayesian quantile regression. Stat Probab Lett 54(4):437–447
Yu K, Stander J (2007) Bayesian analysis of a Tobit quantile regression model. J Econom 137(1):260–276
Zhao Z, Xiao Z (2014) Efficient regressions via optimally combining quantile information. Econom Theory 30(6):1272–1314
Zou H, Yuan M (2008) Composite quantile regression and the oracle model selection theory. Ann Stat 36(3):1108–1126