Estimation of non-crossing quantile regression curves

Yuzhi Cai[1,*] Swansea University

Tao Jiang[2], University of Cambridge

Summary

Quantile regression methods have been widely used in many research areas in recent years. However, conventional estimation methods for quantile regression models do not guarantee that the estimated quantile curves will be non-crossing. While there are various methods in the literature to deal with this problem, many of these methods force the model parameters to lie within a subset of the parameter space in order for the required monotonicity is satisfied. Note that different methods may use different subspaces of the space of model parameters.

This paper established a relationship between the monotonicity of the estimated conditional quantiles and the comonotonicity of the model parameters. We developed a novel quasi-Bayesian method for parameter estimation which can be used to deal with both time series and independent statistical data. Simulation studies and appli-

*Author to whom correspondence should be addressed.

[1]School of Management, Swansea University, Swansea SA2 8PP, United Kingdom.

Email: y.cai@swansea.ac.uk

[2]Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, United Kingdom

Acknowledgements. We thank the AE and the referees for their very constructive comments and suggestions which have significantly enhanced the quality and the presentation of this paper.

cation to real financial returns show that the proposed method has the potential to be very useful in practice.

Key words: Asymmetric Laplace distribution, comonotonicity, quasi-Bayesian method.

1 Introduction

Quantile regression methods have become increasingly popular recently. Koenker (2005) gives an excellent introduction to quantile regression models. The quantile regression approach offers a mechanism for estimating conditional quantiles, and hence the conditional distribution of the response variable. Compared with statistical models for conditional means, quantile regression models are capable of providing a more flexible statistical analysis of stochastic relationships between random variables.

Various quantile regression models and parameter estimation methods have been proposed in the literature since Koenker & Bassett's (1978) work. For example, Koenker & D'Orey (1987, 1994) developed an optimisation method for estimating parameters, which has been used widely through the package quantreg in statistical software R. Koenker & Xiao (2004, 2006) studied statistical inferences in quantile autoregression models. Geraci & Bottai (2007) developed a new methodology on longitudinal quantile regression. Yu & Moyeed (2001) developed a Bayesian approach to estimating the parameters of a quantile regression model. Cai & Stander (2008) proposed a quantile self-exciting threshold autoregressive time series model and developed a Bayesian method for estimating model parameters. Thompson et al. (2010) proposed a Bayesian non-parametric quantile regression method using splines and Cai et al. (2012) proposed a Bayesian estimation and forecasting method for quantile autoregressive time series models. The Bayesian approaches mentioned above are based on the asymmetric Laplace distribution (ALD). There are also

other Bayesian approaches that have been developed to make the likelihood more nonparametric. For example, Kottas & Krnjajic (2009) proposed a method for Bayesian nonparametric modelling in quantile regression, Lancaster & Jun (2010) studied the application of Bayesian exponentially tilted empirical likelihood to inference about quantile regressions, and Reich et al. (2010) proposed a flexible Bayesian quantile regression method for independent and clustered data.

A well-known problem associated with the estimation methods mentioned above is that the estimated quantiles may not be monotone for all possible covariate values, which leads to an invalid conditional distribution of the response variable. Some methods have been developed to deal with this problem: for example, Koenker (1984) considered parallel quantile planes for linear quantile regressions. He (1997) proposed a sequential algorithm for estimating the quantile curves in order to guarantee the monotonicity of the estimated quantiles. Wu & Liu (2009) proposed a method based on fitting a sequence of ordered quantiles curves. Dette & Volgushev (2008) proposed a non-parametric approach to non-crossing quantile curve estimation. Bondell et al. (2010) proposed a direct correction to the quantile regression optimisation method in order to ensure the monotonicity of the estimated conditional quantiles. Most recently, Schnabel & Eilers (2013) proposed a new methodology to avoid crossing quantile curves.

Koenker (2005) discussed the potential usefulness of the comonotonicity of a group of random variables in the context of quantile regression. One of the main contributions of this paper is to establish a relationship between the monotonicity of the estimated conditional quantiles and the comonotonicity of the model parameters (see Theorem 1 and Theorem 2). Another main contribution of this paper is the development of a quasi-Bayesian method for estimating a sequence of non-crossing conditional quantile curves simultaneously. This method can be applied to both independent and dependent data.

In Section 2 we develop the main theoretical results of the paper and propose a quasi-

Bayesian MCMC method for non-crossing quantile estimation. In Section 3 we present the results of our simulation studies. An empirical application to a real financial returns dataset can be found in Section 4. Finally, in Section 5 some comments and concluding remarks are given.

2 The main results and the quasi-Bayesian approach

2.1 Model and comonotonicity

Let $\mathbf{y}^\top = (y_1, \ldots, y_n)$ be independent samples of a response variable y, $\mathbf{x}_i^\top = (x_{1i}, \ldots, x_{pi})$ $(i = 1, \ldots, n)$ the values of covariates $\mathbf{x}^\top = (x_1, \ldots, x_p)$, and $\mathbf{z}^\top = (1, \mathbf{x}^\top)$. Consider a sequence of quantile regression models

$$q_{y_i|\mathbf{x}_i}^{\tau_k} = \beta_{0\tau_k} + \beta_{1\tau_k} x_{1i} + \ldots + \beta_{p\tau_k} x_{pi} = \mathbf{z}_i^\top \boldsymbol{\beta}_{\tau_k}, \quad k = 1, \ldots, K, \tag{1}$$

where $\boldsymbol{\beta}_{\tau_k}^\top = (\beta_{0\tau_k}, \ldots, \beta_{p\tau_k})$ is the parameter vector of the kth model, and $0 < \tau_1 < \tau_2 < \cdots < \tau_K < 1$ are quantile levels of interest.

We want to estimate the conditional quantiles defined by (1) such that, for any value of $\mathbf{x}$, the following inequality holds:

$$q_{y|\mathbf{x}}^{\tau_1} \leq q_{y|\mathbf{x}}^{\tau_2} \leq \cdots \leq q_{y|\mathbf{x}}^{\tau_K}. \tag{2}$$

Before presenting our method, we remind readers the concept of comonotonicity.

Two random variables U and V are said to be comonotonic if there exists a third random variable W and increasing functions h_1 and h_2 such that $U = h_1(W)$ and $V = h_2(W)$. See Koenker (2005) for details.

Note that in a general quantile regression model $q^{\tau}_{Y|X} = \mathbf{z}^{\top}\boldsymbol{\beta}_{\tau}$ the components of the coefficient vector $\boldsymbol{\beta}_{\tau}$ can be viewed as functions of τ. Therefore, if we treat τ as a uniformly distributed random variable on $[0, 1]$and let $W = \tau$ and $U_j = h_j(\tau) = \beta_{j\tau}$ for $j = 0, \ldots, p$ and if all $h_j(\cdot)$ are increasing functions of τ, then $\beta_{j\tau}$ $(j = 0, \ldots, p)$ are comonotonic. That is, we have

$$\beta_{j\tau_1} \leq \beta_{j\tau_2} \leq \cdots \leq \beta_{j\tau_K}, \quad j = 0, 1, \ldots, p. \tag{3}$$

We now establish the relationship between (2) and (3) with details given below.

2.2 Main theoretical results

To estimate a sequence of model (1) satisfying condition (2), Bondell et al. (2010) proposed estimating the model parameter $\boldsymbol{\beta}^{\top} = (\boldsymbol{\beta}^{\top}_{\tau_1}, \ldots, \boldsymbol{\beta}^{\top}_{\tau_K})$ by solving the optimisation problem

$$\min_{\boldsymbol{\beta}} \sum_{k=1}^{K} \sum_{i=1}^{n} w(\tau_k)\rho_{\tau_k}(u_{ik})$$
$$\text{subject to } \mathbf{z}^{\top}\boldsymbol{\beta}_{\tau_\ell} < \mathbf{z}^{\top}\boldsymbol{\beta}_{\tau_{\ell+1}} \text{for all } \mathbf{x} \in D \text{ and } \ell = 1, \ldots, K-1,$$

for some weight function $w(\tau_k)$ such that $w(\tau_k) > 0$ $(k = 1, \ldots, K)$, where $u_{ik} = y_i - \sum_{j=0}^{p} \beta_{j\tau_k} x_{ji}$, $x_{0i} = 1$ $(i = 1, \ldots, n)$, $D = [0, 1]^p$ and $\rho_{\tau}(u) = u(\tau - I[u < 0])$, where $I[\cdot]$ is an indicator function. To guarantee monotonicity, the covariate needs to take values in D and the model needs to be reparameterised.

Following the approach used by Bondell et al. (2010), in the rest of the paper, we take $w(\tau_k) = 1$ for all k, implying that all $\rho_{\tau_k}(u_{ik})$ are of equal importance. The main differences between our approach and that of Bondell et al. (2010) are that the covariate space and the model parameter space are different, and that our method is based on a Bayesian approach while theirs is an optimisation based method. The capabilities of the two methods are also different which will become clearer later in the paper.

Before presenting the method that we have developed, let us first consider the following theorem.

Theorem 1 *Suppose that $\beta_{j\tau} \leq \beta_{j\tau'}$, for $j = 0, \ldots, p$ and that at least one of the inequalities is strict. If for all i*

$$\beta_{0\tau} + \beta_{1\tau} x_{1i} + \cdots + \beta_{p\tau} x_{pi} > \beta_{0\tau'} + \beta_{1\tau'} x_{1i} + \cdots + \beta_{p\tau'} x_{pi}, \tag{4}$$

where $\tau < \tau'$, then there exists at least one x_{ji} such that $x_{ji} < 0$.

See the Appendix for a proof. Theorem 1 says that if comonotonicity holds, then the curve-crossing problem may occur anywhere in $R^p = (-\infty, \infty)^p$ except for the first quadrant R^p_+. This is why we require that $\mathbf{x} \in R^p_+ = [0,\ \infty)^p$ in the methodology developed below. The next theorem establishes a relationship between comonotonicity and non-crossing estimates of $q^{\tau}_{y|\mathbf{x}}$.

Theorem 2 *Let τ and τ' be such that $0 \leq \tau < \tau' \leq 1$. Then*

$$\boldsymbol{\alpha}^\top \boldsymbol{\beta}_\tau \leq \boldsymbol{\alpha}^\top \boldsymbol{\beta}_{\tau'}, \quad \textit{for any } \ \boldsymbol{\alpha} \in R^{p+1}_+ \tag{5}$$

if and only if

$$\beta_{j\tau} \leq \beta_{j\tau'}, \quad j = 0, 1, \ldots, p. \tag{6}$$

A proof can be found in the Appendix. Theorem 2 says that condition (5) is equivalent to comonotonicity of the model parameters. Hence if (3) holds, then for any $\boldsymbol{\alpha} = \mathbf{z} = (1, \mathbf{x}) \in R^{p+1}_+$ the monotonicity condition (2) also holds. Hence in our approach $\boldsymbol{\beta}_\tau$ may be estimated by solving the following minimisation problem

$$\min_{\boldsymbol{\beta} \in \Omega} \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k} \left(u_{ik} \right), \tag{7}$$

where Ω is the set of $\boldsymbol{\beta}$ such that $\beta_{j\tau_1} \leq \beta_{j\tau_2} \leq \ldots \leq \beta_{j\tau_K}$ for $j = 0, 1, \ldots, p$.

It is worth mentioning that minimising (7) does not involve any checking procedures for the monotonicity condition (2) at any specific covariate values. All we need to do is to make sure that the covariates take non-negative values, which can be easily achieved through a simple transformation.

Note that for fixed k, u_{ik} can be viewed as a sample from the ALD with a density function defined by $f_k(u) = \tau_k(1-\tau_k)e^{-\rho_{\tau_k}(u)}$. Hence $L_k(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}) = \{\tau_k(1-\tau_k)\}^n e^{-\sum_{i=1}^n \rho_{\tau_k}(u_{ik})}$ can be viewed as the likelihood of u_{ik} $(i = 1, \ldots, n)$. If we let the combined likelihood of the observed data be $L(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}) = \prod_{k=1}^K L_k(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x})$, then

$$L(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}) = \prod_{k=1}^{K} \{\tau_k(1-\tau_k)\}^n e^{-\sum_{i=1}^n \rho_{\tau_k}(u_{ik})}. \tag{8}$$

Consequently minimising (7) is equivalent to maximising (8).

Let $\pi_k(\boldsymbol{\beta}_{\tau_k})$ be a prior density function for $\boldsymbol{\beta}_{\tau_k}$. Then $\pi(\boldsymbol{\beta}) = \prod_{k=1}^K \pi_k(\boldsymbol{\beta}_{\tau_k})$ is the prior density function for $\boldsymbol{\beta}$. The combined posterior density function for $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta} \mid \mathbf{x}, \mathbf{y}) \propto L(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x})\pi(\boldsymbol{\beta}), \quad \boldsymbol{\beta} \in \Omega. \tag{9}$$

Theorem 3 *For any well defined prior density function $\pi(\boldsymbol{\beta})$ on Ω, the posterior density function $\pi(\boldsymbol{\beta} \mid \mathbf{x}, \mathbf{y})$ is also well defined on Ω.*

Theorem 3 holds because $L(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}) \leq 1$. Therefore, $\pi(\boldsymbol{\beta})$ may be chosen from a large class of well defined density functions so that prior knowledge about the model parameters can be properly taken into account. For example, we may let $\pi_k(\boldsymbol{\beta}_{\tau_k})$ be a product of independent density functions of $\beta_{j\tau_k}$. We may also let $\pi_k(\boldsymbol{\beta}_{\tau_k})$ be

$$\beta_{j\tau_1} \sim N(0, \sigma^2), \quad \beta_{j\tau_k} \sim \beta_{j\tau_{k-1}} + \chi^2(df_k), \quad k = 2, 3, \ldots, K,$$

where df_k is the degrees of freedom of a χ^2-distribution. The effects of different prior distributions on the parameter estimation are certainly worthy of investigation in the future.

For illustration purposes, in this paper we let

$$\pi(\boldsymbol{\beta}) = \prod_{k=1}^{K}\prod_{j=0}^{p}\pi_k(\beta_{j\tau_k})\left(\int_{\Omega}\prod_{k=1}^{K}\prod_{j=0}^{p}\pi_k(\beta_{j\tau_k})d\beta_{j\tau_k}\right)^{-1}, \quad \boldsymbol{\beta}\in\Omega,$$

where

$$\pi_k(\beta_{j\tau_k}) = \frac{1}{\sqrt{2\pi}\sigma_{jk}}e^{-\beta_{j\tau_k}^2/2\sigma_{jk}^2}.$$

That is, $\pi(\boldsymbol{\beta})$ is a truncated density function on Ω based on normal densities. Note that large values of σ_{jk} imply weak prior information about $\beta_{j\tau_k}$. Further remarks are given below.

Remark 1. The Bayesian framework used above is to facilitate computation only. This is because expression (8) is clearly not the likelihood of the data. For this reason we call (8) the "quasi-likelihood" of the data. We will use the term "quasi-Bayesian" for the parameter estimation method developed in this paper.

Remark 2. Although Remark 1 means that the resulting posterior is not really a valid posterior in the Bayesian sense, Yang & He (2012) pointed out that work of Yu & Moyeed (2001), Geraci & Bottai (2007) and Yue & Rue (2011) has provided numerical evidence that such a Bayesian approach to quantile regression has merits.

Remark 3. In addition to developing a novel quasi-Bayesian method for (8), we also developed an optimisation method for (7). Extensive simulation studies suggest that the quasi-Bayesian approach outperforms the optimisation approach, which provides further evidence that the quasi-Bayesian approach has merit and may be useful in practice.

Remark 4. In our approach we require non-negative covariates, which can be achieved by a sim-

ple shift transformation. Although such a shift is non-unique, the estimated quantiles of the original response variable will not be affected and can be obtained easily since we are dealing with linear quantile regression. In particular, for cross-sectional data the conditional quantiles of the response variable will not be affected by a shift transformation of the covariates. For time series, if we let $y_t = x_t + c$ and let $q^{\tau}_{y_t}$ be the quantile of y_t, then the quantile of x_t is given by $q^{\tau}_{y_t} - c$.

2.3 Quasi-Bayesian MCMC algorithm

It follows from Theorem 3 that the model parameter $\boldsymbol{\beta}$ may be estimated by a Markov chain Monte Carlo (MCMC) method. The two general steps of the MCMC algorithm are given below, where $\beta^*_{j\tau_k}$ and $\beta_{j\tau_k}$ represent proposed and current parameter values respectively.

1. Obtain a proposed value $\beta^*_{j\tau_k}$ by sampling from $q(\beta_{j\tau_k} \to \beta^*_{j\tau_k})$ such that $\boldsymbol{\beta}^* \in \Omega$, where $q(\beta_{j\tau_k} \to \beta^*_{j\tau_k})$ represents the probability density function of $\beta^*_{j\tau_k}$ conditional on the current value $\beta_{j\tau_k}$, $j = 0, \ldots, p$ and $k = 1, \ldots, K$.

2. Accept the proposed value with probability $\alpha = \min\{AB, 1\}$, where

$$A = \frac{L(\mathbf{y} \mid \boldsymbol{\beta}^*, \mathbf{x})\pi(\boldsymbol{\beta}^*)}{L(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x})\pi(\boldsymbol{\beta})},$$

$$B = \frac{\prod_{j=0}^{p}\prod_{k=1}^{K} q(\beta^*_{j\tau_k} \to \beta_{j\tau_k}) / \int_{\Omega}\prod_{j=0}^{p}\prod_{k=1}^{K} q(\beta^*_{j\tau_k} \to \beta_{j\tau_k}) d\beta_{j\tau_k}}{\prod_{j=0}^{p}\prod_{k=1}^{K} q(\beta_{j\tau_k} \to \beta^*_{j\tau_k}) / \int_{\Omega}\prod_{j=0}^{p}\prod_{k=1}^{K} q(\beta_{j\tau_k} \to \beta^*_{j\tau_k}) d\beta^*_{j\tau_k}},$$

The above general method produces a Markov chain of vectors $\boldsymbol{\beta}$ with the equilibrium distribution given by (9). Let $\boldsymbol{\beta}^{(m)}$ $(m = 1, \ldots, M)$ be values saved from the MCMC algorithm once every L steps after a burn-in period. Then these saved values form a posterior sample of the model parameters, on the basis of which further statistical inferences may be made.

Theorem 4 *Let the quasi-Bayesian estimate of $\boldsymbol{\beta}$ be given by*

$$\hat{\beta}_{j\tau_k} = \frac{1}{M}\sum_{m=1}^{M} \beta_{j\tau_k}^{(m)}, \quad j = 0, 1, \ldots, p,\ k = 1, \ldots, K.$$

Then inequality (5) holds for $\tau = \tau_k$ and $\tau' = \tau_{k+1}$, for $k = 1, \ldots, K-1$.

See the Appendix for a proof. Theorem 4 implies that, conditional on any $\mathbf{x} \in R_+^p$, the quasi-Bayesian estimate $\hat{q}_{y|\mathbf{x}}^{\tau} = \mathbf{z}^\top \hat{\boldsymbol{\beta}}_\tau$ of the conditional quantiles of y is monotone with respect to τ. Consequently, it suffices to design an efficient method for obtaining $\boldsymbol{\beta}^*$ such that (3) holds.

Let $U \sim N(0,1)$, let $\Phi(\cdot)$ be the corresponding probability distribution function, and let V be the random variable representing the truncated version of U over an interval $[a, b]$. The probability distribution function of V is given by $F_V(v) = \frac{\Phi(v)-\Phi(a)}{\Phi(b)-\Phi(a)}$, where $v = \Phi^{-1}\left[r\{\Phi(b) - \Phi(a)\} + \Phi(a)\right]$ and $r \in (0,1)$. This result leads to the following method for obtaining $\boldsymbol{\beta}^*$, where s_{jk} is a standard deviation for a normal distribution, which can be assigned by user.

1. For $j = 0, \ldots, p$ and $k = 1$, obtain a sample $\beta_{j\tau_1}^* \sim N(\beta_{j\tau_1}, s_{j1}^2)$.
2. For $j = 0, \ldots, p$ and $k = 2, \ldots, K$, obtain a sample $\beta_{j\tau_k}^* \sim N(\beta_{j\tau_k}, s_{jk}^2)$ such that $\beta_{j\tau_k}^* \geq \beta_{j\tau_{k-1}}^*$. Specifically,
   (a) Simulate $r \sim U(0,1)$.
   (b) let $a = (\beta_{j\tau_{k-1}}^* - \beta_{j\tau_k})/s_{jk}$.
   (c) Calculate $w = r\{1 - \Phi(a)\} + \Phi(a)$.
   (d) Calculate $v = \Phi^{-1}(w)$.
   (e) Let $\beta_{j\tau_k}^* = \beta_{j\tau_k} + v s_{jk}$.

3. Accept the proposed $\boldsymbol{\beta}^*$ with probability $\alpha = \min\{AB, 1\}$, where A is the same as that given above, while B becomes

$$B = \frac{\prod_{j=0}^{p} q(\beta^*_{j\tau_1} \to \beta_{j\tau_1}) \prod_{k=2}^{K} q(\beta^*_{j\tau_k} \to \beta_{j\tau_k} \mid \beta_{j\tau_{k-1}})}{\prod_{j=0}^{p} q(\beta_{j\tau_1} \to \beta^*_{j\tau_1}) \prod_{k=2}^{K} q(\beta_{j\tau_k} \to \beta^*_{j\tau_k} \mid \beta^*_{j\tau_{k-1}})},$$

where $q(\beta_{j\tau_1} \to \beta^*_{j\tau_1})$ is again a normal density function for $\beta^*_{j\tau_1}$ with mean $\beta_{j\tau_1}$ and variance s^2_{j1}, and

$$q(\beta_{j\tau_k} \to \beta^*_{j\tau_k} \mid \beta^*_{j\tau_{k-1}}) = \frac{1}{C_{jk}\sqrt{2\pi}s_{jk}} e^{-(\beta^*_{j\tau_k} - \beta_{j\tau_k})^2/2s^2_{jk}},$$

where

$$C_{jk} = \int_{\beta^*_{j\tau_{k-1}}}^{\infty} \frac{1}{\sqrt{2\pi}s_{jk}} e^{-(\beta^*_{j\tau_k} - \beta_{j\tau_k})^2/2s^2_{jk}} d\beta^*_{j\tau_k} = 1 - \Phi\left(\frac{\beta^*_{j\tau_{k-1}} - \beta_{j\tau_k}}{s_{jk}}\right).$$

It is worth mentioning that the MCMC method that we have developed can also be applied to time series. This is demonstrated by means of simulation studies which are presented in the next section.

2.4 Further discussion

Note that the marginal distribution of each parameter may be estimated by using the posterior samples. Hence a kind of credible interval for each parameter may also be constructed. However, as pointed out earlier in the paper, this interval is not a credible interval in the conventional Bayesian sense. Therefore we call it a quasi credible interval and refer to it as a QCI. Our extensive simulation studies suggest that these QCIs can be used to measure the performance of the estimated model in an effective manner.

In practice, the problem of how to compare several possible models under the quasi-

Bayesian approach is very challenging. This is because the data do not actually follow the ALD distribution. We suggest checking the overall empirical coverage probabilities of the estimated quantile curves, since well-fitting model should provide good overall coverage probabilities. We also suggest checking the local coverage probabilities using the moving window method developed by Cai et al. (2012).

The basic idea of the moving window method is given below. Since the quasi-Bayesian method is based on the equivalence of (7) and (8), no matter what distribution the error term follows, the τth conditional quantile of the model residual should be approximately zero. On this basis the moving-window method is defined to consist of the following steps: (a) Select a window width h. (b) Calculate the τth sample quantile of the residuals in the ith window. (c) Plot these quantiles against the window index. If the estimated τth quantile model is good, then we should expect that the values of these local sample quantiles should be close to the horizontal line at height 0.

3 Simulation studies

3.1 Data generating processes

In this section, we consider three simulation studies involving both time series and independent data. The first data generating process is given by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + (1 + 0.1y_{t-1} + 0.3y_{t-2})\epsilon_t, \tag{10}$$

where ϵ_t are independently and identically distributed (iid) $N(0, 0.4^2)$, $\beta_0 = 2.2$, $\beta_1 = 0.8$ and $\beta_2 = -0.1$. Process (10) was used to obtain 100 independent time series, each of length 100. The details are as follows: (a) Let $y_1 = y_2 = 0$. (b) For $t = 2, \ldots, 100000$,

simulate $\epsilon_t \sim N(0, 0.4^2)$ and calculate y_t using (10). (c) Save the last 100 values of y_t as the simulated series in order to remove starting value effects. Figure 1 (a) displays one of the simulated time series.

Please note that in this simulation study all the true parameter values were chosen such that the simulated y_t are positive for all t, hence the value of the term $(1+0.1y_{t-1}+0.3y_{t-2})$ is also guaranteed to be positive.

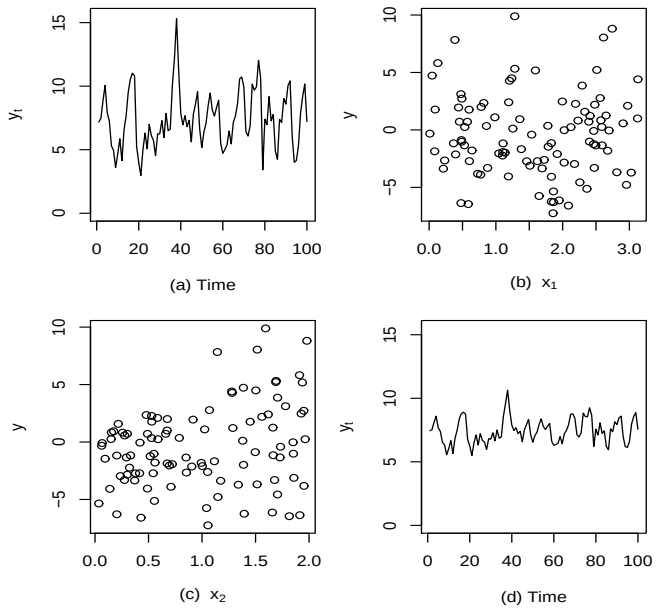

Figure 1: (a) Plot of a simulated time series from (10). (b)-(c) Scatter plots of a data set from (11). (d) Plot of a simulated time series from (12).

The second data generating process is given by

$$y_i = \beta_0 + \beta_1 \sin(x_{1i}) + \beta_2 x_{2i} + (1 + 0.5 \sin(x_{1i}) + 1.8 x_{2i})\epsilon_i, \quad i = 1, \ldots, n, \qquad (11)$$

where $\beta_0 = 1.5$, $\beta_1 = -2.7$, $\beta_2 = -0.5$, $x_1 \in [0, \pi]$ and $x_2 \in [0, 2]$, ϵ_i are iid $N(0, 1)$. Again 100 independent samples, each of length 100, were simulated from process (11) as follows: For $i = 1, \ldots, 100$, we simulated $x_{1i} \sim U(0, \pi)$, $x_{2i} \sim U(0, 2)$ and $\epsilon \sim N(0, 1)$.

Then we used (11) to calculate y_i. One of the simulated data sets is shown in Figure 1 (b) and (c). Note that both $\sin(x_1)$ and x_2 are non-negative.

The third data generating process is similar to (10) and is given by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + (1 + 0.1y_{t-1})\epsilon_t, \tag{12}$$

where ϵ_t are iid $N(0, 0.4^2)$, $\beta_0 = 2.2$, $\beta_1 = 0.8$ and $\beta_2 = -0.1$. Process (12) was used to obtain 100 independent time series, each of length 100.

3.2 Parameter estimation

The quantile autoregressive model corresponding to processes (10) and (12) is

$$q^{\tau}_{y_t|\mathbf{y}_{t-1}} = \beta_{0\tau} + \beta_{1\tau} y_{t-1} + \beta_{2\tau} y_{t-2}, \tag{13}$$

where $\beta_{0\tau} = \beta_0 + q(\tau)$, $\beta_{1\tau} = \beta_1 + 0.1q(\tau)$ for both processes, and for process (10) $\beta_{2\tau} = \beta_2 + 0.3q(\tau)$ while for process (12) $\beta_{2\tau} = \beta_2$ (i.e. a constant!), where $\tau \in (0, 1)$ and $q(\tau)$ is the τth-quantile of $N(0, 0.4^2)$. These true parameter values denoted by $\beta_{j\tau_k}$ are given in Table 1 and Table 3 respectively, where
$\tau \in A_1 = \{\tau_1, \ldots, \tau_9\} = \{0.01, 0.05, 0.25, 0.35, 0.5, 0.65, 0.75, 0.95, 0.99\}$.

Similarly, the quantile regression model corresponding to process (11) is

$$q^{\tau}_{y_i|\mathbf{x}_i} = \beta_{0\tau} + \beta_{1\tau} \sin(x_{1i}) + \beta_{2\tau} x_{2i}, \tag{14}$$

where $\tau \in A_2 = \{\tau_1, \ldots, \tau_{11}\} = \{0.005, 0.1, 0.2, \ldots, 0.9, 0.995\}$. The true parameter values are given in Table 2.

In these simulation studies we used both the quasi-Bayesian method and the optimisa-

tion method to estimate the parameters. We also deliberately chose two different sets of τ values. We hoped that the two methods would produce similar results and that the simulation results would not depend on the chosen quantile levels. We also hoped that the true parameter values of each model would be within the respective 95% QCIs. For process (12) we expected that the estimated $\beta_{2\tau}$ values would be non-decreasing with respect to τ but would still be close to the true value, i.e. -0.1. Finally, we expected that the estimated conditional quantile curves would be non-crossing at all covariate values.

The number of parameters in each simulation study is 27, 33 and 27 respectively. The initial parameter values were obtained by simulating a random sample from the standard normal distribution with the restriction that the results fall in the parameter space Ω. For the quasi-Bayesian approach, we assumed there to be little prior information about the model parameters, hence we deliberately set $\sigma_{jk} = 25$, a relatively large value, for all possible values of j and k.

For each of the 100 independently simulated data sets, a Markov chain of length 500000 was run for the respective models. Testing runs showed that a burn-in period of 50000 values is sufficient. For illustration purposes, Figure 2 shows the time series plots of the parameter values from the MCMC algorithm for the first simulated data set obtained from processes (10). As the saved data file is large, we only plotted the parameter values once every 100 steps. The plots for all other simulated data sets are very similar. We have no concerns about the convergence of the MCMC algorithm.

After the burn-in period, we saved values of the Markov chain once every 100 steps to weaken the autocorrelation between the posterior samples. A 95% QCI for each parameter was constructed from the posterior samples and the estimate of the parameters was calculated. We thus have 100 estimates and 100 associated 95% QCIs for each model parameter. The average of the parameter estimates and the corresponding 95% QCIs are shown in Figure 3, where the first and the third rows correspond to model (13) for process

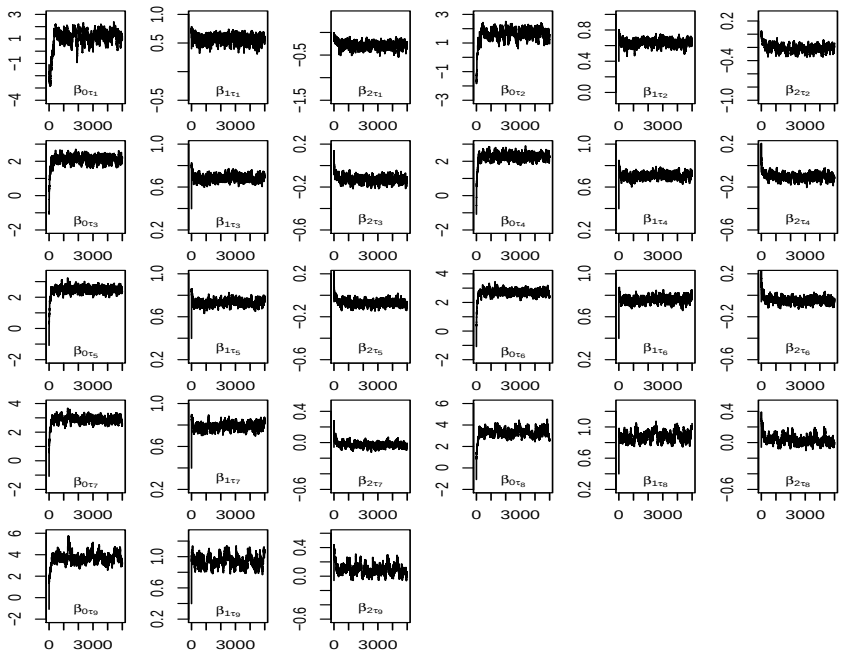

Figure 2: Time series plots of the parameter values from the MCMC algorithm.

(10) and process (12) respectively, while the second row is for model (14). Furthermore, the triangle and the circle symbols correspond to the true and estimated average parameter values respectively, and the continuous curves show the average 95% QCIs of the model parameters. It is seen that on average all the true parameter values are well within the 95% QCIs, suggesting that the performance of the developed methodology is good.

It is also seen that the average performance of the method is good for non-extreme quantile levels, but gets worse at extreme quantile levels in all three simulation studies. The width of the 95% QCIs also becomes wider as τ tends to extremes, which is consistent with what was found by Reich et al. (2010).

It is worth mentioning that in the data generating process (12) we set $\beta_{2\tau} = -0.1$. Although the estimated parameter values are not constant, they are close to the true parameter values for non-extreme τs. However if the true parameter values were decreasing with in-

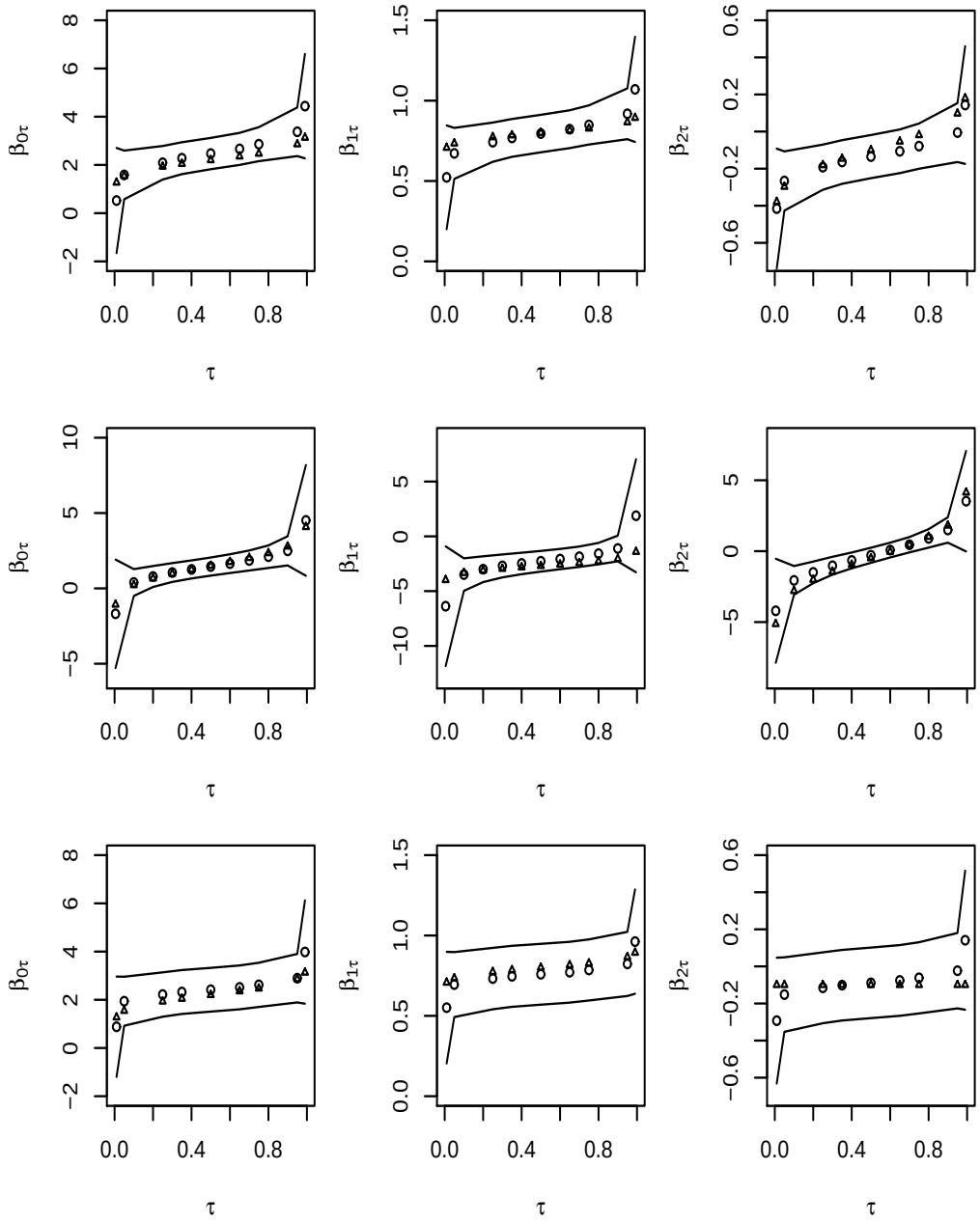

Figure 3: The true parameter values (triangle), the average quasi-Bayesian estimates (circle) and the average 95% QCIs (continuous curves) of the parameters of model (13) (the first and the last rows for processes (10) and (12) respectively) and model (14) (the second row) in the simulation studies.

Table 1: True and the average estimated parameters and the associated standard errors for Simulation study 1.

τ	0.01	0.05	0.25	0.35	0.50	0.65	0.75	0.95	0.99
$\beta_{0\tau_k}$	1.269	1.542	1.930	2.046	2.200	2.354	2.470	2.858	3.131
$\hat{\beta}_{0\tau_k}$	0.523	1.586	2.090	2.279	2.472	2.665	2.861	3.374	4.439
$\tilde{\beta}_{0\tau_k}$	-1.824	-1.653	-1.462	-1.281	-1.041	-0.715	-0.359	0.114	0.724
$SE(\hat{\beta}_{0\tau_k})$	1.114	0.515	0.355	0.338	0.331	0.339	0.360	0.514	1.105
$SE(\tilde{\beta}_{0\tau_k})$	1.765	1.717	1.658	1.577	1.470	1.261	1.088	0.953	0.960
$\beta_{1\tau_k}$	0.707	0.734	0.773	0.785	0.800	0.815	0.827	0.866	0.893
$\hat{\beta}_{1\tau_k}$	0.523	0.672	0.742	0.768	0.795	0.822	0.849	0.918	1.070
$\tilde{\beta}_{1\tau_k}$	0.630	0.653	0.674	0.687	0.704	0.753	0.788	0.984	1.144
$SE(\hat{\beta}_{1\tau_k})$	0.165	0.081	0.062	0.060	0.060	0.060	0.062	0.081	0.167
$SE(\tilde{\beta}_{1\tau_k})$	0.641	0.640	0.634	0.629	0.623	0.658	0.676	0.804	0.855
$\beta_{2\tau_k}$	-0.379	-0.297	-0.181	-0.146	-0.100	-0.054	-0.019	0.097	0.179
$\hat{\beta}_{2\tau_k}$	-0.416	-0.267	-0.192	-0.164	-0.135	-0.106	-0.079	-0.006	0.143
$\tilde{\beta}_{2\tau_k}$	0.615	0.617	0.621	0.625	0.632	0.666	0.721	0.940	1.057
$SE(\hat{\beta}_{2\tau_k})$	0.165	0.081	0.062	0.060	0.059	0.060	0.062	0.081	0.162
$SE(\tilde{\beta}_{2\tau_k})$	0.639	0.639	0.639	0.638	0.637	0.652	0.718	0.879	0.937

MSE	$\beta_{0\tau}$	$\beta_{1\tau}$	$\beta_{2\tau}$
Quasi Bayesian	0.327	0.008	0.002
Optimization	9.288	0.014	0.683

creasing τ then according to Theorem 2 the given method should not used directly. See further discussions on this issue later in the paper.

We now compare the performance of the quasi-Bayesian method and the optimisation method. The average estimated parameter values over the 100 simulated data sets for each of the three models can be found in Tables 1, 2 and 3 respectively, where $\hat{\beta}_{j\tau_k}$ and $\tilde{\beta}_{j\tau_k}$ denote the parameter estimates obtained from the quasi-Bayesian method and the optimisation method respectively, while $SE(\hat{\beta}_{j\tau_k})$ and $SE(\tilde{\beta}_{j\tau_k})$ are the corresponding standard deviations of the estimates. The MSE gives the mean squared errors between the estimated and the true parameter values. It is seen that the average performance of the optimisation method is not very good, which may be explained as follows.

Table 2: True and the average estimated parameters and the associated standard errors for Simulation study 2.

τ	0.005	0.10	0.20	0.30	0.40	0.50	0.600	0.70	0.80	0.90	0.99
$\beta_{0\tau_k}$	-1.076	0.22	0.66	0.98	1.25	1.50	1.753	2.02	2.34	2.78	4.08
$\hat{\beta}_{0\tau_k}$	-1.686	0.39	0.79	1.05	1.26	1.45	1.643	1.85	2.10	2.49	4.51
$\tilde{\beta}_{0\tau_k}$	-0.18	-0.07	0.024	0.11	0.19	0.27	0.37	0.48	0.62	0.85	1.338
$SE(\hat{\beta}_{0\tau_k})$	1.836	0.45	0.35	0.32	0.31	0.31	0.318	0.34	0.38	0.49	1.88
$SE(\tilde{\beta}_{0\tau_k})$	0.73	0.73	0.721	0.71	0.71	0.71	0.70	0.70	0.70	0.73	1.035
$\beta_{1\tau_k}$	-3.988	-3.34	-3.12	-2.96	-2.83	-2.70	-2.573	-2.44	-2.28	-2.06	-1.41
$\hat{\beta}_{1\tau_k}$	-6.376	-3.49	-3.00	-2.70	-2.48	-2.27	-2.076	-1.86	-1.58	-1.11	1.88
$\tilde{\beta}_{1\tau_k}$	-0.53	-0.49	-0.454	-0.42	-0.39	-0.35	-0.30	-0.25	-0.20	-0.14	-0.037
$SE(\hat{\beta}_{1\tau_k})$	2.788	0.76	0.60	0.53	0.50	0.48	0.472	0.47	0.50	0.60	2.63
$SE(\tilde{\beta}_{1\tau_k})$	0.76	0.77	0.787	0.81	0.83	0.88	0.93	0.99	1.04	1.09	1.168
$\beta_{2\tau_k}$	-5.136	-2.81	-2.01	-1.44	-0.96	-0.50	-0.044	0.44	1.01	1.81	4.14
$\hat{\beta}_{2\tau_k}$	-4.205	-2.06	-1.50	-1.03	-0.65	-0.29	0.076	0.45	0.90	1.49	3.53
$\tilde{\beta}_{2\tau_k}$	0.53	0.55	0.559	0.57	0.59	0.63	0.68	0.75	0.85	0.98	1.168
$SE(\hat{\beta}_{2\tau_k})$	1.873	0.52	0.39	0.32	0.28	0.27	0.266	0.28	0.33	0.46	1.81
$SE(\tilde{\beta}_{2\tau_k})$	0.36	0.36	0.357	0.36	0.36	0.40	0.43	0.47	0.54	0.62	0.693

MSE	$\beta_{0\tau}$	$\beta_{1\tau}$	$\beta_{2\tau}$
Quasi Bayesian	0.073	1.72	0.228
Optimization	2.108	5.908	6.171

For illustration purposes, let us consider the first simulation study. We noticed that the estimated constant term $\hat{\beta}_{0k}$ of the model has the largest MSE. To visualize the situation, we arbitrarily chose one of the simulated data sets, let $\tau = 0.5$ and we fixed β_2 at an arbitrarily selected value between -10 and 10. Note that we also checked many other data sets and values of τ and β_2, each time similar results were obtained. Then for a sequence of equally spaced values of β_0 and β_1 between -10 and 10, we calculated the values of the objective function $\sum_{i=1}^{n} \rho_\tau(u_i)$, where $u_i = y_i - \beta_0 - \beta_1 y_{t-1} - \beta_2 y_{t-2}$. Finally we constructed a contour plot of the objective function, shown in Figure 4 (a). Clearly, there exist infinitely many β_0 values that could minimise the objective function. This may explain why the simple optimisation method performs badly in this simulation study.

We also produced the contour plot of the posterior distribution in this case, showed in Figure 4 (b). These contour plots are similar due to the fact that weak prior information was used. However one thing to notice is that the Bayesian posterior estimates of the parameters were obtained based on the mean of the posterior distribution, which may not be the same as the mode of the distribution, while the mode of the distribution is the basis for the conventional optimisation method. This might explain why in such cases the quasi-Bayesian method appears to work better than the conventional optimisation method.

3.3 Some further discussions

It is seen from Tables 1-3 that all the MSEs corresponding to the quasi-Bayesian method are less than those corresponding to the optimisation method, which further confirms that the quasi-Bayesian method is more successful than the optimisation method.

Now let us consider the performance of the methods at extreme levels. Tables 1-3 suggest that larger variations occurred in the parameter estimations since the standard errors of the estimates at extreme levels are generally larger than those at non-extreme levels.

Table 3: True and the average estimated parameters and the associated standard errors for Simulation study 3.

τ	0.01	0.05	0.250	0.350	0.500	0.650	0.750	0.950	0.99
$\beta_{0\tau_k}$	1.27	1.54	1.930	2.046	2.200	2.354	2.470	2.858	3.13
$\hat{\beta}_{0\tau_k}$	0.88	1.94	2.217	2.322	2.417	2.514	2.620	2.895	3.98
$\tilde{\beta}_{0\tau_k}$	-2.12	-1.96	-1.80	-1.62	-1.34	-1.01	-0.59	-0.15	0.31
$SE(\hat{\beta}_{0\tau_k})$	1.06	0.52	0.469	0.464	0.463	0.464	0.468	0.513	1.09
$SE(\tilde{\beta}_{0\tau_k})$	1.52	1.52	1.49	1.41	1.27	1.12	0.94	0.83	0.80
$\beta_{1\tau_k}$	0.71	0.73	0.773	0.785	0.800	0.815	0.827	0.866	0.89
$\hat{\beta}_{1\tau_k}$	0.55	0.69	0.732	0.746	0.759	0.772	0.786	0.823	0.96
$\tilde{\beta}_{1\tau_k}$	0.56	0.59	0.60	0.61	0.62	0.66	0.71	0.81	0.89
$SE(\hat{\beta}_{1\tau_k})$	0.18	0.10	0.097	0.097	0.097	0.097	0.097	0.102	0.17
$SE(\tilde{\beta}_{1\tau_k})$	0.70	0.70	0.70	0.69	0.70	0.71	0.74	0.76	0.79
$\beta_{2\tau_k}$	-0.10	-0.10	-0.100	-0.100	-0.100	-0.100	-0.100	-0.100	-0.10
$\hat{\beta}_{2\tau_k}$	-0.29	-0.15	-0.115	-0.101	-0.089	-0.076	-0.062	-0.023	0.14
$\tilde{\beta}_{2\tau_k}$	0.70	0.70	0.70	0.71	0.71	0.73	0.76	0.84	0.90
$SE(\hat{\beta}_{2\tau_k})$	0.17	0.10	0.098	0.097	0.097	0.097	0.098	0.104	0.19
$SE(\tilde{\beta}_{2\tau_k})$	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.75	0.78

MSE	$\beta_{0\tau}$	$\beta_{1\tau}$	$\beta_{2\tau}$
Quasi Bayesian	0.143	0.005	0.012
Optimization	11.256	0.020	0.728

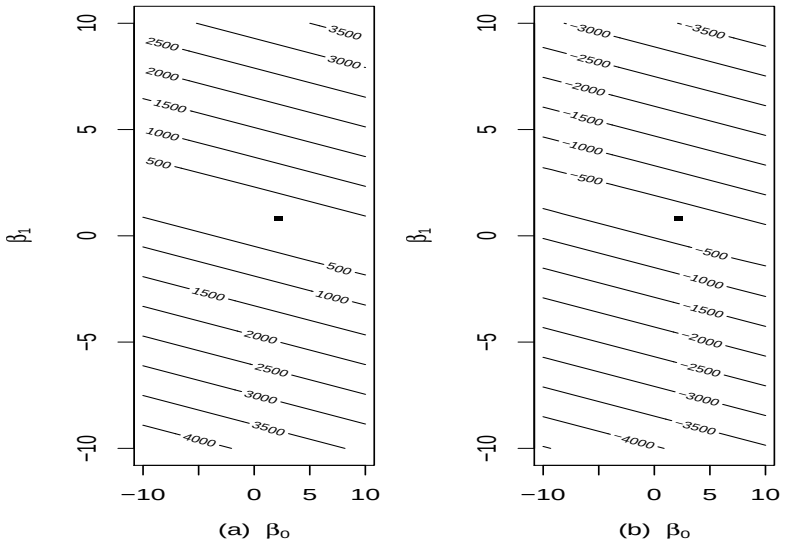

Figure 4: (a) Contour plot of the objective function for the optimisation method. (b) Contour plot of the posterior density function for the quasi-Bayesian method. The square dots correspond to the true values of β_0 and β_1.

Although this is what we expected because less and less information is available as τ goes to extremes, it is worth mentioning that it may not be appropriate to use our methodology to estimate extreme quantiles. Some work about the problem of extreme quantile estimation can be found in the literature. For example Chernozhukov (2005) developed a theory of the large sample properties of extremal quantile regression estimators for the linear quantile regression model. Beirlant et al. (2004) suggested a two-step procedure based on local quantile regression and univariate extreme value theory. Gardes et al. (2010) proposed an estimation method such that the estimation of an extreme quantile is based on observations in a small neighborhood of a given covariate value of interest. Wang et al. (2012) developed two estimation methods for extremal conditional quantiles of heavy-tailed distributions. However none of these methods address the crossing issue. We believe that estimation of non-crossing extreme quantiles is worth investigating in the future.

To check whether the fitted quantile curves are non-crossing, we calculated the differ-

ences $q^{\tau_k}_{y_t|\mathbf{y}_{t-1}} - q^{\tau_{k-1}}_{y_t|\mathbf{y}_{t-1}}$, where $\tau_k \in A_1$ and $t = 3, \ldots, 100$ for model (13) corresponding to processes (10) and (12) respectively, and the differences $q^{\tau_k}_{y_i|\mathbf{x}_i} - q^{\tau_{k-1}}_{y_i|\mathbf{x}_i}$, where $\tau_k \in A_2$ and $i = 1, \ldots, 100$ for model (14). We found that all these differences are non-negative, hence no curve-crossing problem occurred. The first row in Figure 5 displays three of these differences for illustration purposes.

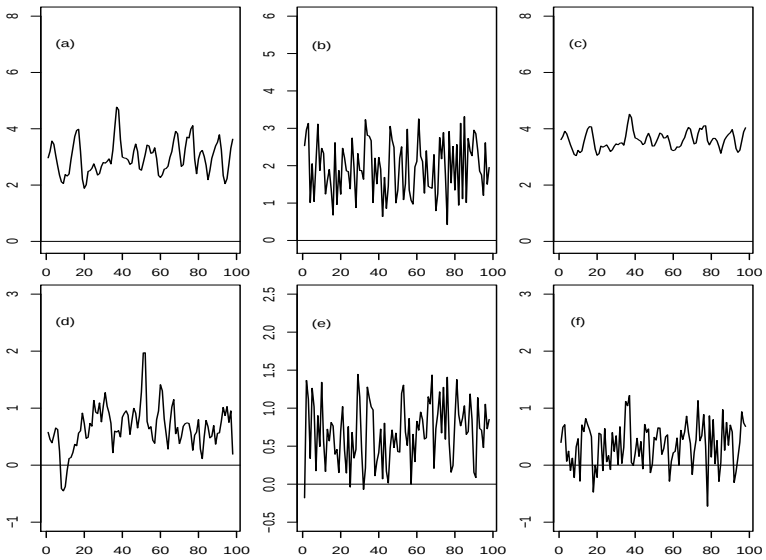

Figure 5: The first row corresponds to our method and the second row corresponds to the conventional method without non-crossing constraints.
(a)(d) Plots of $q^{0.99}_{y_t|\mathbf{y}_{t-1}} - q^{0.95}_{y_t|\mathbf{y}_{t-1}}$ for model (13) corresponding to (10).
(b)(e) Plots of $q^{0.3}_{y_i|\mathbf{x}_i} - q^{0.2}_{y_i|\mathbf{x}_i}$ for model (14).
(c)(f) Plots of $q^{0.05}_{y_t|\mathbf{y}_{t-1}} - q^{0.01}_{y_t|\mathbf{y}_{t-1}}$ for model (13) corresponding to (12).

We also fitted the three models to the simulated data sets (shown in Figure 1) respectively by using the R-package in which no non-crossing constraints are involved. The second row in Figure 5 shows the differences between two fitted conditional quantiles at the same levels as those used in the first row. Clearly, the three data generating processes considered here do result in estimated quantile curves that cross each other if non-crossing constraints are not imposed.

In summary, the simulation studies show that the quasi-Baysian method works well

in parameter estimation and in providing a reasonable QCI for the estimated model parameters. The simulation studies also show that the quasi-Bayesian method appears to outperform the optimisation method with non-crossing constraints. However it may not be appropriate to use our methods when dealing with extreme quantiles.

4 Application to DJIA time series

In this section, we apply the developed methodology to the log-returns of the Dow Jones Industrial Average (DJIA). The data cover the period between 2 January 2004 and 8 October 2010 and are of length 1705. The time series plots of the observed series and its log-returns are displayed in Figure 6. It is seen that the observed DJIA log-returns exhibit the occurrence of extremes and volatility clustering.

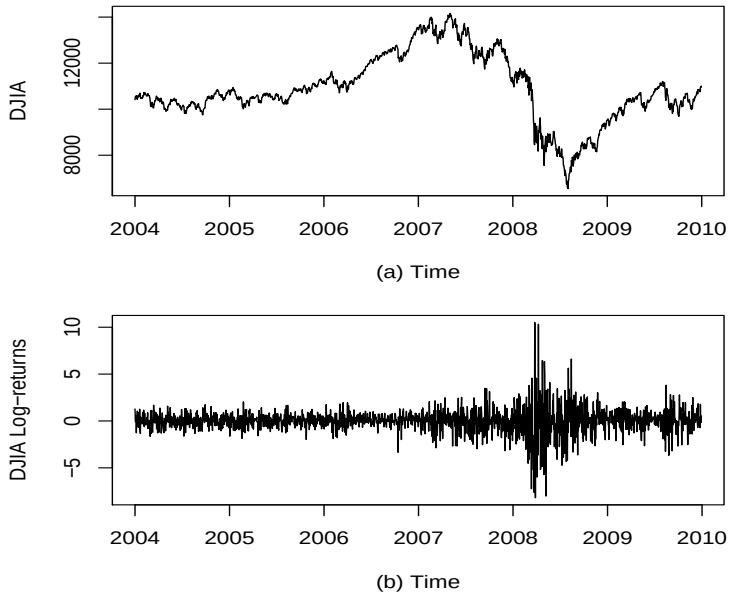

Figure 6: (a) Time series plot of the DJIA between 2 January 2004 and 8 October 2010. (b) Time series plot of the DJIA log-returns.

Note that the smallest log-return is -8.20% during this period of time. To make sure the covariates in our model only take non-negative values, we let $y_t = x_t + C$, where x_t denotes the log return at time t, $t = 1, \ldots, n = 1704$, with the constant C chosen so that $y_t > 0$ for $t = 1, \ldots, 1704$. In this application we let $C = 10$.

We fitted the following models

$$q^{\tau_k}_{y_t|\mathbf{y}_{t-1}} = \beta_{0\tau_k} + \beta_{1\tau_k} y_{t-1} + \cdots + \beta_{p\tau_k} y_{t-p}, \quad \tau_k \in A_1 \tag{15}$$

to the y_t series, where $p = 1, 2, 3$ and 4. Model (15) says that the τ_kth quantile of y_t given $\mathbf{y}_{t-1}$ is $q^{\tau_k}_{y_t|\mathbf{y}_{t-1}}$. It then follows from $y_t = x_t + C$ that the τ_kth quantile of x_t given $\mathbf{x}_{t-1}$ is given by $q^{\tau_k}_{x_t|\mathbf{x}_{t-1}} = q^{\tau_k}_{y_t|\mathbf{y}_{t-1}} - C$.

We checked the overall coverage probabilities of the estimated quantile curves and calculated the MSE between these probabilities and the true quantile levels. The MSE values are 7.5×10^{-4}, 2.2×10^{-4}, 8.9×10^{-5} and 8.5×10^{-2} for $p = 1, 2, 3$ and 4 respectively. We chose the model with the smallest MSE, $p = 3$. Details about fitting this model are given below.

To estimate the parameter values we ran the MCMC algorithm for 10^7 steps with the starting values of $\beta_{j\tau_k}$ randomly simulated from $N(0, 1)$ with the restriction that they fall in the parameter space. After a burn-in period of 2×10^6 steps we saved the parameter values once every 100 steps. Figure 7 provides a summary of the estimates of the parameters and the corresponding 95% QCIs for $\tau_k \in A_1$.

For the purpose of comparison we also fitted model (15) with $p = 3$ to the DJIA log-returns using Bondell et al.'s (2010) method and Schnabel and Eilers' (2013) method. Table 4 lists the estimated parameter values from all methods, where M_1, M_2, M_3 represent our method, Bondell et al.'s method and Schnabel and Eilers' method respectively.

These results clearly show that the estimated parameter values from the three methods

Table 4: Estimated parameter values by using three different methods

	$\hat{\beta}_{0\tau_k}$								
τ	0.01	0.05	0.25	0.35	0.5	0.65	0.75	0.95	0.99
M1	-0.04	0.05	0.14	0.18	0.22	0.24	0.26	0.28	0.32
M2	-4.04	-1.93	-0.48	-0.21	0.06	0.32	0.54	1.74	3.41
M3	-2.70	-1.50	-0.47	-0.26	0.003	0.25	0.45	1.39	2.54
	$\hat{\beta}_{1\tau_k}$								
τ	0.01	0.05	0.25	0.35	0.5	0.65	0.75	0.95	0.99
M1	0.18	0.21	0.23	0.24	0.24	0.25	0.26	0.30	0.37
M2	-0.001	-0.01	-0.05	-0.07	-0.09	-0.09	-0.09	-0.19	-0.19
M3	0.22	0.07	-0.06	-0.09	-0.12	-0.15	-0.18	-0.30	-0.44
	$\hat{\beta}_{2\tau_k}$								
τ	0.01	0.05	0.25	0.35	0.5	0.65	0.75	0.95	0.99
M1	0.16	0.25	0.30	0.31	0.32	0.33	0.35	0.39	0.49
M2	-0.13	-0.02	0.01	-0.01	-0.01	-0.03	-0.05	-0.07	-0.15
M3	0.15	0.04	-0.05	-0.07	-0.10	-0.12	-0.14	-0.22	-0.33
	$\hat{\beta}_{3\tau_k}$								
τ	0.01	0.05	0.25	0.35	0.5	0.65	0.75	0.95	0.99
M1	0.27	0.32	0.39	0.40	0.41	0.42	0.44	0.47	0.53
M2	0.25	0.14	0.07	0.07	0.07	0.07	0.07	0.07	-0.01
M3	0.35	0.22	0.12	0.10	0.07	0.05	0.03	-0.07	-0.19

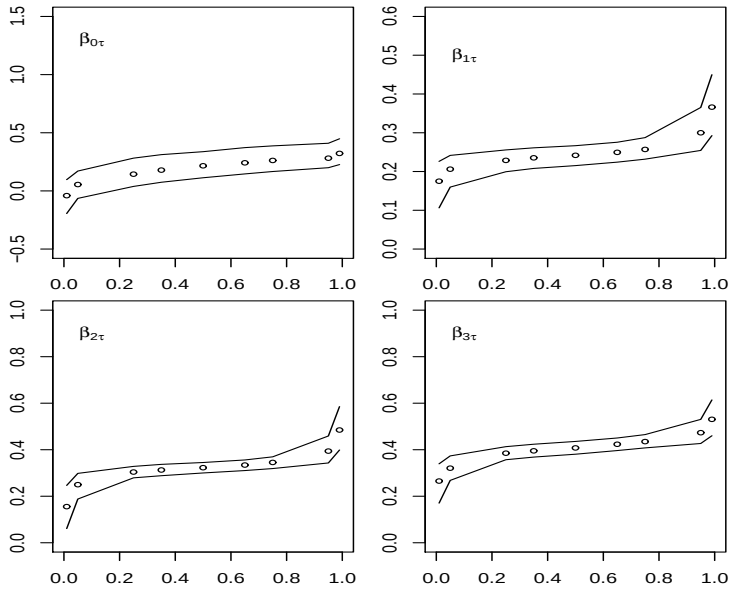

Figure 7: Plots of estimated parameter values and the corresponding 95% QCIs obtained from our method.

are different. Now we compare their performance. We have found that the estimated quantile curves at non-extreme quantile levels are very similar, but differences between them can still be seen visually. Figure 8 displays the estimated quantile curves at three levels, i.e. $\tau = 0.01, 0.5$ and 0.99. The first row corresponds to our method (darker lines) and Bondell et al.'s method (grey lines); the second row corresponds to our method (darker lines) and Schnabel and Eilers' method (grey lines); and the final row corresponds to Bondell et al.'s method (darker lines) and Schnabel and Eilers' method (grey lines). It is seen that the median curves obtained from all methods are very similar. It is also seen that the conditional distribution defined by the estimated quantile curves obtained from our method has the longest tails, while that obtained from the Schnabel and Eilers' method has the shortest. Nevertheless, all methods seem to have produced similar quantile curves despite the fact that the model parameters are different.

Now we check the overall coverage probabilities of the estimated quantile curves from

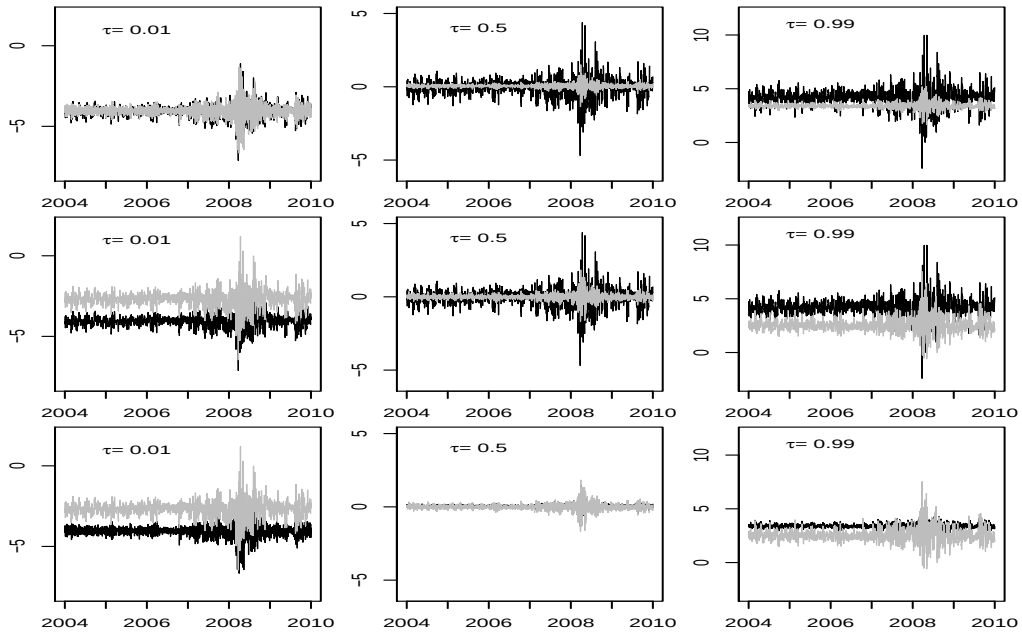

Figure 8: Estimated conditional quantile curves from our method, Bondell et al.'s method and Schnabel and Eilers' method.

each model. Let n_k be the number of the log-returns that are below the τ_kth quantile curve. The proportion at level τ_k may be simply estimated by $n_k/(n-p)$. These estimated values are shown in the second, fifth and sixth columns of Table 5 for the three methods (ours, that of Bondell et al. (2010) and that of Schnabel and Eilers' (2013) respectively). The last row of Table 5 gives the MSE values between the empirical coverage probabilities and the true ones. These MSE results suggest that our method and Bondell's method perform better.

Due to the nature of the quasi-Bayesian approach, we are also able to construct 95% QCIs for each of the coverage probabilities. Specifically, at a level $\tau_k \in A_1$, we used all posterior samples of the model parameters to calculate the proportions. The lower and upper 0.025 quantiles of these proportions then form a 95% QCIs of the true coverage probabilities at this level. The third and the fourth columns of Table 5 show the lower and upper bounds of these QCIs respectively. We noticed that all true probabilities are

Table 5: The coverage probability of the estimated quantile curves and the associated 95% QCI for our method.

	Our method			Bondell et al. method	S & E's method
τ	Proportion	L-bound	U-bound	Proportion	Proportion
0.01	0.008	0.005	0.012	0.010	0.028
0.05	0.048	0.037	0.058	0.052	0.075
0.25	0.236	0.206	0.263	0.249	0.256
0.35	0.340	0.312	0.370	0.350	0.335
0.50	0.487	0.455	0.516	0.499	0.453
0.65	0.638	0.604	0.663	0.651	0.603
0.75	0.740	0.714	0.767	0.751	0.703
0.95	0.943	0.929	0.954	0.951	0.922
0.99	0.985	0.981	0.990	0.991	0.976
MSE	8.94×10^{-5}			9.69×10^{-6}	9.77×10^{-4}

well within these 95% QCIs, which suggest that our estimated coverage probabilities are reasonably good. On the other hand, as it is difficult to use the other two methods to estimate the corresponding confidence intervals, it is difficult to assess the variation of their estimates. Therefore we are not able to compare the fitted models with respect to this criterion.

Now let us consider the local coverage probabilities. Figure 9 shows the local coverage diagnostic plots at five quantile levels, where the grey curves are from our model, continuous darker curves from Bondell's model and the dashed curves from the Schnabel and Eilers' model. It is seen that at non-extreme quantile levels all three models behaves similarly because the majority of the three curves overlapped and are close to the horizontal line, while at extreme quantile levels that the performance of Schnabel and Eilers' model seems slightly better than the other two.

In practice it is of great interest to estimate the conditional distribution of financial returns. This is because once this distribution is available then any quantity of interest about the returns can be obtained easily. For illustration purposes we constructed four conditional distributions at times $t = 1000$, 1200, 1400 and 1700, that were chosen arbitrarily.

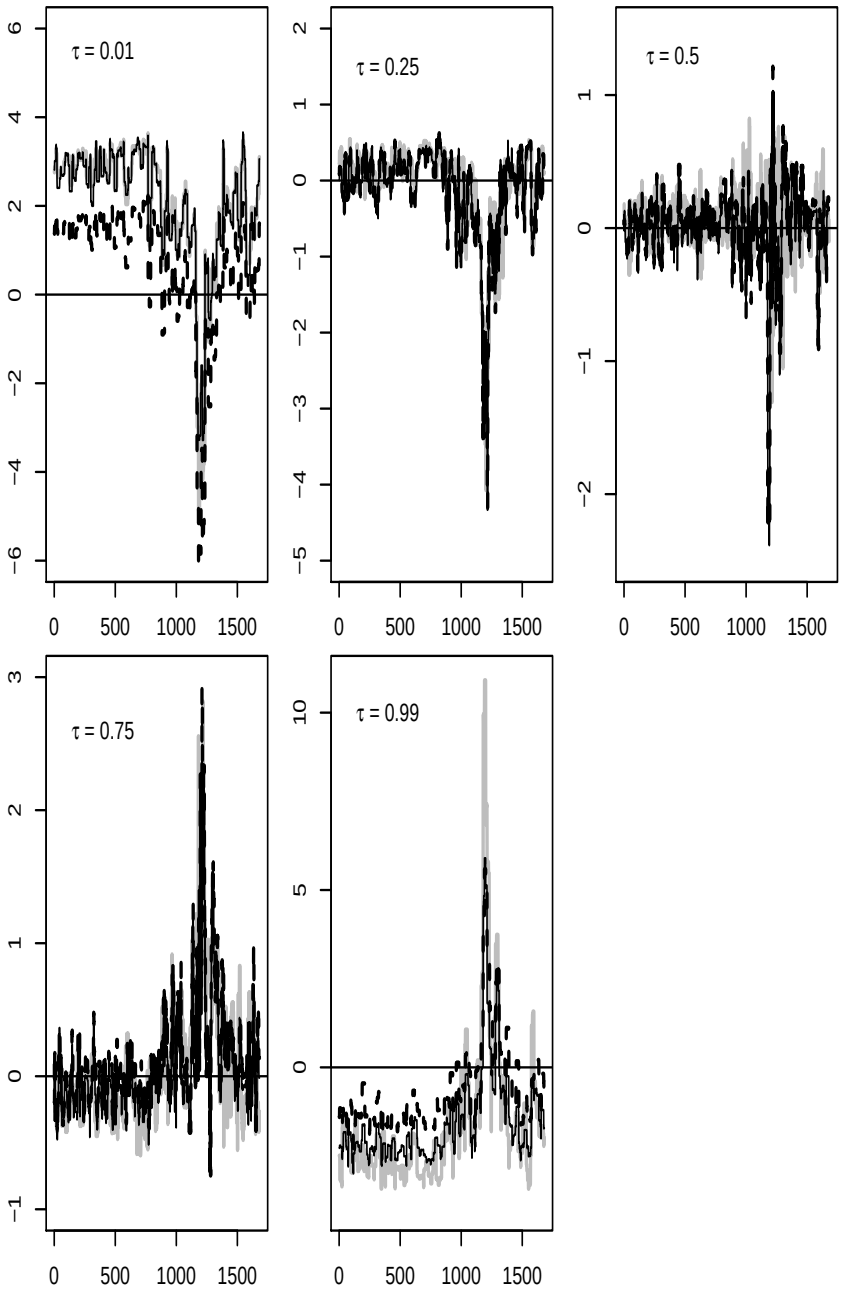

Figure 9: Diagnostic plots of the local probability coverage of the estimated quantile curves from different models.

Figure 10 shows these density functions, where the continuous dark curves are from our model, the dashed curves from Bondell's model, the grey curves from Schnabel and Eilers' model, and the vertical lines correspond to the observed returns at these time points. It is interesting to see that, as shown in Figure 10, although the estimated conditional quantile curves from different methods behave similarly, the estimated density functions conditional on a specific covariate value may not be similar. This may be due to the fact that the parameter spaces of those methods are different. Note that as it is difficult to compare the variation of the estimated conditional distributions in this study, we plan to carry out extensive simulation studies in the future so that the performance of different methods can be compared from a distributional point of view in a systematic way. In practice, we suggest users should try different methods on the same data and choose the most suitable model for their study.

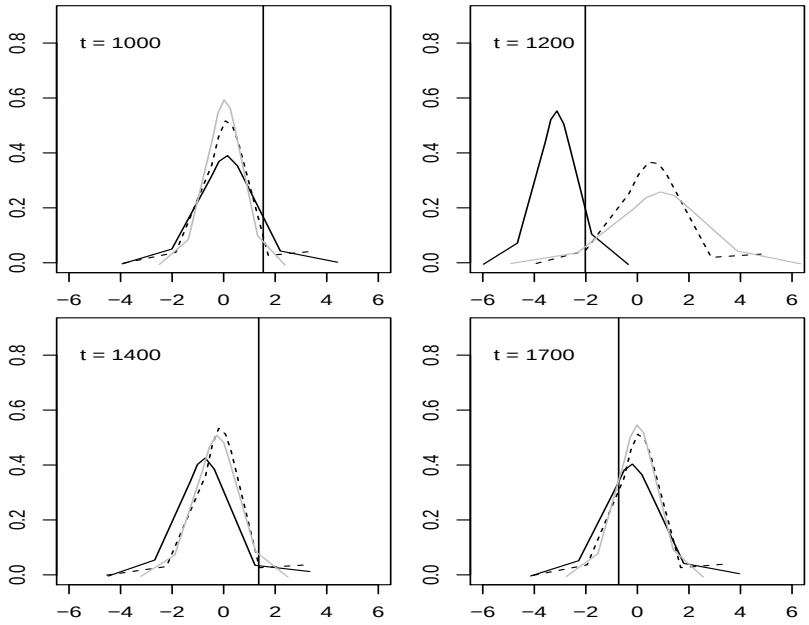

Figure 10: Estimated conditional density functions at different time points.

It is worth mentioning that the estimated conditional density functions (see Figure 10)

are not sufficiently smooth. Smoother conditional density functions might be obtained by increasing the number of quantile levels K. It might also be possible to introduce an additional penalty term to the model to ensure smoothness of the resulting quantile function. We also leave this for future research.

Finally, we would like to point out that, without non-crossing constraints, the conventional quantile regression method will produce invalid estimated conditional quantiles for the data set in this example.

5 Further comments and Conclusions

In this paper we established a relationship between the comonotonicity of the model parameters and the monotonicity of the estimated conditional quantiles for non-negative covariate variables. We also proposed a quasi-Bayesian approach to non-crossing quantile curve estimation.

Our results show that the quasi-Bayesian method not only allows us to estimate non-crossing quantile curves but also allows us to estimate, as by-products of the method, the distribution of any aspect of the response variable at any quantile level. Such information is difficult to obtain from the method of Bondall et al. (2010) or that of Schnabel and Eilers (2013). We noticed that the performance of our method at extreme levels is not as good as it is at non-extreme levels. Consequently careful interpretation of results at extreme levels is required.

In this paper we also developed an optimisation method for parameter estimation. Extensive simulation studies suggest that this quasi-Bayesian method outperforms the more conventional optimisation method.

As we have mentioned earlier in the paper, the main limitation of the proposed method

is that if a true model parameter is monotonically decreasing, it may not be appropriate to use our method directly. However Koenker (2005) showed that it is always possible to reparameterise the model so as to achieve comonotonicity and illustrated this by an example.

The best way to use our method in practice is to reparameterise the model before applying the method. Note that currently techniques for the reparametrisation are model and/or data dependent. There are no unified procedures available for the reparametrisation. Further research on this issue is obviously needed.

Appendix

Proof of Theorem 1:

It follows from (4) that

$$(\beta_{0\tau'} - \beta_{0\tau}) + (\beta_{1\tau'} - \beta_{1\tau})x_{1i} + \cdots + (\beta_{p\tau'} - \beta_{p\tau})x_{pi} < 0. \tag{16}$$

But $\beta_{j\tau'} - \beta_{j\tau} \geq 0$ for all j, so we see that for (16) to hold, we must have at least one x_{ji} such that $x_{ji} < 0$ as required.

Proof of Theorem 2:

First suppose that (5) holds. We need to show that (6) also holds.

Since (5) holds for all $\boldsymbol{\alpha} \in R_+^{p+1}$, it must in particular hold for $\boldsymbol{\alpha} = e_j = (0, \ldots, 1, \ldots, 0)$,

where all the entries of e_j are 0 except for the $j+1$th entry which is equal to 1. That is,

$$(0,\ldots,1,\ldots,0)\begin{pmatrix}\beta_{0\tau}\\ \vdots\\ \beta_{j\tau}\\ \vdots\\ \beta_{p\tau}\end{pmatrix} \leq (0,\ldots,1,\ldots,0)\begin{pmatrix}\beta_{0\tau'}\\ \vdots\\ \beta_{j\tau'}\\ \vdots\\ \beta_{p\tau'}\end{pmatrix}.$$

Therefore, we have $\beta_{j\tau} \leq \beta_{j\tau'}$ for $j = 0,\ldots,p$ as required.

On the other hand, suppose that (6) holds, we need to show that (5) also holds.

First note that for any $\boldsymbol{\alpha} = (\alpha_0,\ldots,\alpha_p) \in R_+^{p+1}$ we have $\alpha_j \geq 0$ and $\boldsymbol{\alpha} = \sum_{j=0}^{p} \alpha_j e_j$. It follows from (6) that $e_j^\top \boldsymbol{\beta}_\tau \leq e_j^\top \boldsymbol{\beta}_{\tau'}$ for $j = 0,\ldots,p$, hence

$$\alpha_j e_j^\top \boldsymbol{\beta}_\tau \leq \alpha_j e_j^\top \boldsymbol{\beta}_{\tau'} \quad \text{implies that} \quad \sum_{j=0}^{p} \alpha_j e_j^\top \boldsymbol{\beta}_\tau \leq \sum_{j=0}^{p} \alpha_j e_j^\top \boldsymbol{\beta}_{\tau'}.$$

So (5) holds as required.

Proof of Theorem 4:

First note that for fixed k and m, and for any $\boldsymbol{\alpha} \in R_+^{p+1}$ we have $\boldsymbol{\alpha}^\top \boldsymbol{\beta}_{\tau_k}^{(m)} \leq \boldsymbol{\alpha}^\top \boldsymbol{\beta}_{\tau_{k+1}}^{(m)}$. Consequently

$$\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\alpha}^\top \boldsymbol{\beta}_{\tau_k}^{(m)} \leq \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\alpha}^\top \boldsymbol{\beta}_{\tau_{k+1}}^{(m)}.$$

That is to say $\boldsymbol{\alpha}^\top \hat{\boldsymbol{\beta}}_{\tau_k} \leq \boldsymbol{\alpha}^\top \hat{\boldsymbol{\beta}}_{\tau_{k+1}}$ as required.

References

[1] Beirlant, J., de Wet, T. & Goegebeur, Y. (2004). Nonparametric Estimation of Extreme Conditional Quantiles. *J. Stat. Comput. Simul.* 74, 567580.

[2] Bondell, H.D., Reich, B.J. & Wang, H. (2010). Non-crossing quantile regression curve estimation. *Biometrika*, 97, 825-838.

[3] Cai, Y., Stander, J. & Davies, N. (2012). A new Bayesian approach to quantile autoregressive time series model estimation and forecasting. *J. Time Series Anal.*, 33, 684-698.

[4] Cai, Y. & Stander, J. (2008). Quantile self-exciting threshold autoregressive time series models. *J. Time Series Anal.*, 29, 186-202.

[5] Chernozhukov, V. (2005). Extremal quantile regression. *Ann. Statist.* 33, 806839

[6] Dette, H. & Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 70, 609-627.

[7] Gardes, L., Girard, S. & Lekina, A. (2010). Functional Nonparametric Estimation of Conditional Extreme Quantiles. *J. Multivariate Anal.* 101, 419433.

[8] Gelfand, A.E. & Dey, D. (1994). Bayesian model choice: Asymptotics and exact calcula- tions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 56, 501-514.

[9] Geraci M. & Bottai M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution, *Biostatistics*, 8, 140-54.

[10] He, X. (1997). Quantile curves without crossing. *Amer. Statist.*, 51, 186-192.

[11] Koenker, R. (1984). A note on L-estimators for linear models. *Statist. Probab. Lett.*, 2, 323-325.

[12] Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

[13] Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33-50.

[14] Koenker, R. & D'Orey, V. (1987). Computing regression quantiles. *Appl. Statist.*, 36, 383-393.

[15] Koenker, R. & D'Orey, V. (1994). A remark on Algorithm AS229: Computing dual regression quantiles and regression rank scores. *Appl. Statist.*, 43, 410-414.

[16] Koenker, R. & Xiao, Z. (2004). Unit root quantile autoregression inference. *J. Amer. Statist. Assoc.*, 99, 775-787.

[17] Koenker, R. & Xiao, Z. (2006). Quantile Autoregression. *J. Amer. Statist. Assoc.*, 101, 980-990.

[18] Kottas, A. & Krnjajic, M. (2009). Bayesian nonparametric modelling in quantile regression. *Scand. J. Stat.*, 36, 297-319.

[19] Lancaster, T. & Jun, S.J. (2010). Bayesian quantile regression methods. *J. Appl. Econometrics*, 25, 287-307.

[20] Lewis, S.M. & Raftery, A.E. (1997). Estimating Bayes' factors via posterior simulation with the Laplace-Metropolis estimator. *J. Amer. Statist. Assoc.*, 92, 648-655.

[21] Newton, M.A. & Raftery, A.E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 56, 3-48.

[22] Reich, B.J., Bondell, H.D. & Wang, H. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, 11, 337-352.

[23] Schnabel, S.K. & Eilers, P.H.C. (2013). Simultaneous estimation of quantile curves using quantile sheets. *AStA Adv. Stat. Anal.*, 97, 77-87.

[24] Thompson, P., Cai, Y., Moyeed, R.A., Reeve, D., & Stander, J. (2010). Bayesian non-parametric quantile regression using splines. *Comput. Statist. Data Anal.*, 54, 1138-1150.

[25] Wang, H.J., Li, D. & He, X. (2012): Estimation of High Conditional Quantiles for Heavy-Tailed Distributions. *J. Amer. Statist. Assoc.*. 107, 1453-1464

[26] Wu, Y. & Liu, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. *Stat. Interface*, 2, 299-310.

[27] Yang Y. & He X. (2012). Bayesian empirical likelihood for quantile regression. *Ann. Statist.* 40, 1102-1131.

[28] Yu, K. & Moyeed, R.A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.*, 54, 437-447.

[29] Yue, Y.R. & Rue, H. (2011). Bayesian inference for additive mixed quantile regression models. *Comput. Statist. Data Anal.* 55, 8496.