Bayesian Lasso mixed quantile regression

Rahim Alhamzawi*, Keming Yu

Department of Mathematics, Brunel University, Uxbridge UB8 3PH, UK

Abstract

In this paper, we discuss the regularization in linear mixed quantile regression. A hierarchical Bayesian model is used to shrink the fixed and random effects toward the common population values by introducing an l_1 penalty in the mixed quantile regression check function. A Gibbs sampler is developed to simulate the parameters from the posterior distributions. Through simulation studies and analysis of an age-related macular degeneration data, we assess the performance of the proposed method. The simulation studies and the age-related macular degeneration data analysis indicate that the proposed method performs well in comparison to the other approaches.

Keywords: Asymmetric Laplace distribution; Gibbs sampler; Random effects; Longitudinal data; Quantile regression.

1. Introduction

Clustered data are encountered in a wide variety of applications including agriculture, economics, educational, ecology, geology, medicine and social repeated measures studies. The linear mixed model with random effects (Laird and Ware, 1982) has been widely used to describe the clustered data, due to the flexibility for modeling fixed and random effects. In this model, the fixed effects give the population intercept and slopes, while the random effects account for the heterogeneity among the clusters. One of the serious challenges in the

*Corresponding author.

Email address: rahim.al-hamzawi@brunel.ac.uk (Rahim Alhamzawi)

linear mixed model lies in selecting both random and fixed effects. To solve this problem, model selection criteria such as AIC (Akaike, 1973) and BIC (Schwarz, 1978) have been used over the years to select the mixed effects by comparing a list of models. Recently, Bayesian methods have been proposed for selecting the mixed effects (see, Chen and Dunson, 2003; Kinney and Dunson, 2008; Saville and Herring, 2009; Bondell et al., 2010; Ibrahim et al., 2011). These approaches focus on the traditional least square regression. Compared to the traditional least square regression, quantile regression is more robust to non-normal errors and outliers (Koenker, 2005). The median is the best-known example of the quantile and plays the central role (Koenker, 2005; Yu et al., 2003). Quantile regression has gradually emerged as a comprehensive extension to the least square regression (Koenker, 2005). It is insensitive to heteroscedastic data and outliers, and thus is able to accommodate non-normal errors, which are common in many real world applications (Koenker and Bassett, 1978; Koenker, 2005). Quantile regression has been the subject of great theoretical interest as well as numerous practical applications in a number of fields such as finance, social science and medicine. A comprehensive account of these recent applications can be found in (Koenker, 2005).

Variable selection for fixed effects in quantile regression has attracted much research interest recently (see for example, Zou and Yuan, 2008; Li and Zhu, 2008; Wu and Liu, 2009; Bradic et al., 2010; Li et al, 2010). In this paper, we present a Bayesian approach to select the random effects, together with the fixed effects in the quantile regression models.

Our motivating example is an analysis of age-related macular degeneration data which is previously analyzed by Chaili (2008). This study had a total of 203 patients which were randomly selected from three cities (centers) in the United Kingdom to measure the treatment effects of teletherapy on the loss of vision associated with the progress of age related macular degeneration. The objective of this study is to investigate the relationship between the distance visual acuity (DVA) and a set of covariates. The change in distance visual acuity of each patient was measured four times over a two year period, on the 3th, 6th, 12th and 24th months. In this paper we are interested in selecting the most significant predictors as well as random effects for the quantile regression model, relating to the change in distance visual acuity. The selection of random effects is important in this application, in order to know which predictors have coefficients that vary among subjects. Our goal in this study is to shrink the fixed and random effects toward the common population values by introducing an l_1 penalty in the mixed quantile regression check function.

The rest of the paper is organized as follows. In Section 2, we review linear mixed-effects (LME) models and its re-parameterization . In Section 3, we present the penalized linear mixed quantile regression. We outline the Bayesian MCMC estimation procedure in Section 4. In Section 5, we carry out simulation examples to examine the performance of the method proposed and in Section 6, we illustrate the performance of our method via analysis of the age-related macular degeneration data. We conclude with a brief discussion in Section 7.

2. Linear Mixed Models

Suppose there are N subjects under study so that y_{ij} denote the jth response for subject i, for $i = 1, ..., N$ and $j = 1, ..., n_i$, $\boldsymbol{x}'_{ij}$ and $\boldsymbol{z}'_{ij}$ are rows of the $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ matrices, $\boldsymbol{X}_i$ is $n_i \times k$ and $\boldsymbol{Z}_i$ is $n_i \times q$. Then the linear mixed model is defined as:

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\alpha}_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2), \tag{1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i$ are k and q-dimensional unknown parameters and random effects, respectively, and ε_{ij} is the error term. Here $\boldsymbol{\alpha}_i \sim N(0, \boldsymbol{\Sigma})$.

A Markov chain Monte Carlo (MCMC) method has been suggested by Chen and Dunson (2003) to identify any random effect with zero variance. Their approach is built under the re-parameterized random effect model and it is based on a prior with mass at zero for the random effect variances. Recently, this reparametrization appeared in papers by Kinney and Dunson (2008), Bondell et al. (2010), Saville and Herring (2009) and Ibrahim et al. (2011) to contact Bayesian variable selection in mixed models. Chen and Dunson (2003) reparametrized the linear mixed model (1) as

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\Lambda\Gamma b}_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2), \tag{2}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(d_1, d_2, ..., d_q)$ is a diagonal matrix, $\boldsymbol{\Gamma}$ is a lower triangular matrix of dimension $q \times q$, whose $(l, r)^{th}$ element is denoted by γ_{lr}, $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{q(q-1)/2})'$, $\boldsymbol{b}_i = (b_{i1}, ..., b_{iq})'$ and $\boldsymbol{b}_i \sim \mathrm{N}(0, I)$. Under this reparametrization, the diagonal elements of $\boldsymbol{\Sigma}$ are $\sigma_{ll} = d_l^2 + \sum_{r=1}^{l-1} \gamma_{lr}^2 d_l^2$, $l = 1, \cdots, q$. Given $\boldsymbol{\Gamma}$, $\boldsymbol{b}_i$ and rearranging terms, the authors showed that the diagonal elements of $\boldsymbol{\Lambda}$ can be expressed as regression coefficients in a normal linear regression model. They also showed that, given $\boldsymbol{\Lambda}$ and $\boldsymbol{b}_i$, the parameters in $\boldsymbol{\Gamma}$ can be expressed as regression coefficients in a normal linear regression model. Under this reparametrization, the parameters in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ have the conditional conjugacy property that allows for a simple and efficient Gibbs sampling algorithm for fitting the linear mixed model. Recently, Bondell et al. (2010) developed this idea by using a penalized joint log-likelihood function with an adaptive penalty for the selection and estimation of both the fixed and random effects.

3. The penalized linear mixed quantile regression

Our approach is to set up the problem as a Bayesian quantile regression problem under the l_1 penalty. For the pth regression quantile, we can define a joint penalized criterion under the l_1 penalty as

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \rho_p(y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta}_p - \boldsymbol{z}'_{ij}\boldsymbol{\Lambda\Gamma b}_{pi}) + \lambda(\sum_{s=1}^{k} |\beta_{ps}| + \sum_{s=1}^{q} |d_{ps}|), \tag{3}$$

where $\rho_p(u) = u\{p - I(u < 0)\}$ is the so-called check function. For simplicity of notation, we will omit the subscript p in the remainder of the paper. Thus, Equation (3) becomes

$$\min_{\beta} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \rho_p(y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i) + \lambda(\sum_{s=1}^{k} |\beta_s| + \sum_{s=1}^{q} |d_s|), \tag{4}$$

A parametric connection between the minimization problem in (4) and maximum likelihood theory is given by assuming the error distribution is the asymmetric Laplace distribution; see Koenker and Machado (1999) and Yu and Moyeed (2001). This error distribution also connects the quantile regression model with normal regression models; see Kozumi and Kobayashi (2009) and Reed and Yu (2009). If we employ a Laplace prior $p(\beta_s|\sigma, \lambda) = \sigma\lambda/2 \exp\{-\sigma\lambda|\beta_s|\}$ on β_s, a Laplace prior $p(d_s|\sigma, \lambda) = \sigma\lambda/2 \exp\{-\sigma\lambda|d_s|\}$ on d_s and assume that the residuals ε_{ij} follow an asymmetric Laplace distribution $\mathrm{AL}(0, \sigma, p)$, where the parameters are the location, scale, and skewness, respectively, then the posterior distribution of $\boldsymbol{\beta}$ is

$$f(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \sigma, \lambda, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = \exp\{-\sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{|\varepsilon_{ij}| + (2p-1)\varepsilon_{ij}}{2\sigma} - \sigma\lambda \sum_{s=1}^{k} |\beta_s| - \sigma\lambda \sum_{s=1}^{q} |d_s|\}, \tag{5}$$

where $\varepsilon_{ij} = y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i$. In order to make the model inference tractable, the likelihood function of the asymmetric laplace distribution can be represented as a scale mixture of normals with an exponential mixing density (see , Kozumi and Kobayashi, 2009; Reed and Yu, 2009). This representation can be written as

$$\prod_{i=1}^{N} \prod_{j=1}^{n_i} \sigma^{-1} \exp\{-\frac{|\varepsilon_{ij}| + (2p-1)\varepsilon_{ij}}{2\sigma}\} = \prod_{i=1}^{N} \prod_{j=1}^{n_i} \int_0^{\infty} \frac{1}{\sigma\sqrt{4\pi\sigma v_{ij}}} \exp\{-\frac{(\varepsilon_{ij} - \xi v_{ij})^2}{4\sigma v_{ij}} - \zeta v_{ij}\} dv_{ij}, \tag{6}$$

where $\xi = (1 - 2p)$ and $\zeta = p(1 - p)/\sigma$ (see also, Alhamzawi and Yu, 2012, for some detials). The Laplace prior can also be represented as a scale mixture of normals (see, Andrews and Mallows, 1974; Park and Casella, 2008)

$$
\begin{aligned}
p(\boldsymbol{\beta}|\lambda_1) &= \prod_{s=1}^{k} \frac{\lambda_1}{2} \exp\{-\lambda_1|\beta_s|\}, \\
&= \prod_{s=1}^{k} \int_0^{\infty} \frac{1}{\sqrt{2\pi t_s}} \exp\{-\frac{\beta_s^2}{2t_s}\} \frac{\lambda_1^2}{2} \exp\{-\frac{\lambda_1^2}{2} t_s\} dt_s,
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
p(\boldsymbol{d}|\lambda_1) &= \prod_{s=1}^{q} \frac{\lambda_1}{2} \exp\{-\lambda_1|d_s|\}, \\
&= \prod_{s=1}^{q} \int_0^{\infty} \frac{1}{\sqrt{2\pi \eta_s}} \exp\{-\frac{d_s^2}{2\eta_s}\} \frac{\lambda_1^2}{2} \exp\{-\frac{\lambda_1^2}{2} \eta_s\} d\eta_s,
\end{aligned} \tag{8}
$$

where $\lambda_1 = \sigma\lambda$. The prior specification for the model in (4) is completed by specifying priors for $(\sigma, \lambda_1^2, \boldsymbol{\gamma})$. We specify a conjugate Inverse Gamma prior InverseGamma(g_1, g_2) to the scale parameter σ, a conjugate gamma prior Gamma(c_1, c_2) for λ_1^2 and a multivariate normal prior $\mathrm{N}(0, \boldsymbol{\Gamma}_0)$ for $\boldsymbol{\gamma}$.

4. Bayesian sampler for variable selection

A Gibbs sampling algorithm for the Bayesian quantile regression is constructed by sampling the parameters from their full conditional distributions.

- Full conditional distribution of $\boldsymbol{\beta}$.

  Let $\boldsymbol{T} = \mathrm{diag}(t_1^{-1}, ..., t_k^{-1})$. Then, the full conditional distribution of $\boldsymbol{\beta}$ is a multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ where

$$
\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = (\sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}'}{2\sigma v_{ij}} + T)^{-1}, \qquad \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{\boldsymbol{x}_{ij}(y_{ij} - \boldsymbol{z}_{ij}'\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{b}_i - \xi v_{ij})}{2\sigma v_{ij}} \tag{9}
$$

- Full conditional distribution of $\boldsymbol{b}_i$.

The full conditional distribution of $\boldsymbol{b}_i$ is a multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{b}_i}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{b}_i}$ where

$$\boldsymbol{\Sigma}_{\boldsymbol{b}_i} = (\sum_{j=1}^{n_i} \frac{(\boldsymbol{z}'_{ij}\boldsymbol{\Lambda\Gamma})(\boldsymbol{z}'_{ij}\boldsymbol{\Lambda\Gamma})'}{2\sigma v_{ij}} + \boldsymbol{I})^{-1}, \quad \boldsymbol{\mu}_{\boldsymbol{b}_i} = \boldsymbol{\Sigma}_{\boldsymbol{b}_i} \sum_{j=1}^{n_i} \frac{(\boldsymbol{z}'_{ij}\boldsymbol{\Lambda\Gamma})(y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \xi v_{ij})}{2\sigma v_{ij}} \quad (10)$$

- Full conditional distribution of σ.

$$\sigma|\boldsymbol{\beta}, \boldsymbol{v} \quad \sim \text{InverseGamma}(\frac{3n}{2} + g_1, \frac{(\varepsilon_{ij} - \xi v_{ij})^2}{4v_{ij}} + \zeta v_{ij} + g_2).$$

- Full conditional distribution of $\boldsymbol{v}$.

The full conditional posterior distribution of each v_{ij}^{-1} is Inverse Gaussian(μ', λ') where

$$\mu' = \sqrt{\frac{1}{(y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{\Lambda\Gamma b}_i)^2}} \quad \text{and} \quad \lambda' = \frac{1}{2\sigma}, \quad (11)$$

in the parameterization of inverse Gaussian density given by

$$f(x|\lambda', \mu') = \sqrt{\frac{\lambda'}{2\pi}} x^{-3/2} \exp\{\frac{-\lambda'(x-\mu')^2}{2(\mu')^2 x}\}, \quad x > 0; \quad (12)$$

(Chhikara and Folks, 1989, see,). We use the `rinvGauss()` function in the R package `SuppDists` (Bob, 2009) to sample from generalized inverse Gaussian distribution.

- Full conditional distribution of $\boldsymbol{t}$.

The full conditional posterior distribution of each t_s is Inverse Gaussian(μ', λ') where

$$\mu' = \sqrt{\frac{\lambda_1^2}{\beta_s^2}} \quad \text{and} \quad \lambda' = \lambda_1^2. \quad (13)$$

- Full conditional distribution of $\boldsymbol{\eta}$.

The full conditional posterior distribution of each η_s is Inverse Gaussian(μ', λ') where

$$\mu' = \sqrt{\frac{\lambda_1^2}{d_s^2}} \quad \text{and} \quad \lambda' = \lambda_1^2. \tag{14}$$

- Full conditional distribution of λ_1^2.

The full conditional posterior distribution of λ_1^2 is Gamma(f_1, f_2) where

$$f_1 = k + q + c_1 \quad \text{and} \quad f_2 = \sum_{s=1}^{k} \frac{t_s}{2} + \sum_{s=1}^{q} \frac{\eta_s}{2} + c_2. \tag{15}$$

- Full conditional distribution of $\boldsymbol{\gamma}$.

The full conditional posterior distribution of $\boldsymbol{\gamma}$ is a multivariate normal distribution with mean $\boldsymbol{\mu_\gamma}$ and variance-covariance matrix $\boldsymbol{\Sigma_\gamma}$ where

$$\boldsymbol{\Sigma_\gamma} = (\sum_{i=1}^{N}\sum_{j=1}^{n_i} \frac{\boldsymbol{u}_{ij}\boldsymbol{u}'_{ij}}{2\sigma v_{ij}} + \boldsymbol{\Gamma_0}^{-1})^{-1}, \qquad \boldsymbol{\mu_\gamma} = \boldsymbol{\Sigma_\gamma}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} \frac{\boldsymbol{u}_{ij}(y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \xi v_{ij})}{2\sigma v_{ij}}), \tag{16}$$

where $\boldsymbol{u}_{ij} = (b_{il} d_m z_{ijm} : l = 1, ..., q, m = l+1, ..., q)'$

- Full conditional distribution of $\boldsymbol{d}$.

The full conditional posterior distribution of $\boldsymbol{d}$ is multivariate normal distribution with mean $\boldsymbol{\mu_d}$ and variance-covariance matrix $\boldsymbol{\Sigma_d}$ where

$$\boldsymbol{\Sigma_d} = (\sum_{i=1}^{N}\sum_{j=1}^{n_i} \frac{\boldsymbol{w}_{ij}\boldsymbol{w}'_{ij}}{2\sigma v_{ij}} + \boldsymbol{\Omega})^{-1}, \qquad \boldsymbol{\mu_d} = \boldsymbol{\Sigma_d}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} \frac{\boldsymbol{w}_{ij}(y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \xi v_{ij})}{2\sigma v_{ij}}), \tag{17}$$

where $\boldsymbol{\Omega} = \text{diag}(\eta_1^{-1}, \cdots, \eta_q^{-1})$ and $\boldsymbol{w}_{ij} = (b_{il}\gamma_m z_{ijm} : l = 1, ..., q, m = l+1, ..., q)'$.

5. Simulation studies

We use simulation studies to examine the proposed approach. We compare four models: the Bayesian quantile regression for longitudinal data models (BQRGS) reported in Luo et al. (2011), the Bayesian adaptive lasso quantile regression (BALQR) for the fixed effects as described in Alhamzawi et al. (2011), the standard frequentist quantile regression for the fixed effects (QR) in the R package quantreg (Koenker, 2012), and our approach is referred to as "PMQ". The methods are evaluated based on the median of mean absolute deviations (MMAD), i.e. median $(\sum_{i=1}^{N}\sum_{j=1}^{n_i}|y_{ij}-\hat{y}_{ij}|/\sum_{i=1}^{N}n_i)$, where $\hat{y}_{ij}$ is the predicted value of y_{ij} and the median is taken over 100 simulations. MMAD is a good way of providing information on how well a method performs, where a lower MMAD indicates a better performance.

We report the standard deviation of MAD (SD) for each method. Correlation coefficients between $\boldsymbol{y}=(y_{ij},...,y_{N_{n_N}})'$ and $\hat{\boldsymbol{y}}=(\hat{y}_{ij},...,\hat{y}_{N_{n_N}})'$ for each simulated data set and for each method are also reported. A model with a higher correlation coefficient between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$, indicates a better performance. We consider 4 model designs:

Design 1 (sparse case): We generate data from the model

$$y_{ij}=\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{z}'_{ij}\boldsymbol{\alpha}_i+\varepsilon_{ij}, \tag{18}$$

for $i=1,\cdots,50$, $j=1,\cdots,5$ and $\boldsymbol{x}_{ij}=(x_{ij1},x_{ij2},x_{ij3},x_{ij4},x_{ij5},x_{ij6},x_{ij7},x_{ij8})'$, where x_{ijg} is generated from a uniform distribution over $(-2,2)$ for $g=1,\cdots,8$. We set $\boldsymbol{z}_{ij}=\boldsymbol{x}_{ij}$. We consider the true model

$$y_{ij}=(\beta_1+\alpha_{i1})x_{ij1}+(\beta_2+\alpha_{i2})x_{ij2}+(\beta_3+\alpha_{i3})x_{ij3}+\varepsilon_{ij} \tag{19}$$

where $(\beta_1, \beta_2, \beta_3) = (1, 1, 1)$, $\boldsymbol{\alpha}_i \sim N(0, \boldsymbol{\Sigma})$, and true variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.75 & 0.10 & 0.05 \\ 0.10 & 0.55 & 0.15 \\ 0.05 & 0.15 & 0.10 \end{pmatrix}$$

We simulated the error ε_{ij} from three possible error distributions: a standard normal distribution, a t_3 distribution with three degrees of freedom, and a Chi-squared distribution with three degrees of freedom (χ^2_3). For each $p \in (0.50, 0.75, 0.95)$, we run 100 simulations. The prior specifications are the same as those in Section (3), and we set $c_1 = c_2 = g_1 = g_2 = 0.1$. We run our Gibbs sampler for 25,000 iterations with an initial burn-in of 5000 iterations.

Design 2 (sparse case): The setup in this design is the same as Design 1, except we set $\boldsymbol{z}_{ij} = \boldsymbol{x}_{ij}$ plus a random intercept term. We generate data from the model

$$y_{ij} = \alpha_{i0} + (\beta_1 + \alpha_{i1})x_{ij1} + (\beta_2 + \alpha_{i2})x_{ij2} + (\beta_3 + \alpha_{i3})x_{ij3} + \varepsilon_{ij}, \tag{20}$$

where $(\beta_1, \beta_2, \beta_3) = (3, 1.5, 2)$, $\boldsymbol{\alpha}_i \sim N(0, \boldsymbol{\Sigma})$, and true variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.95 & 0.45 & 0.03 & 0.05 \\ 0.45 & 0.75 & 0.10 & 0.05 \\ 0.03 & 0.10 & 0.55 & 0.15 \\ 0.05 & 0.05 & 0.15 & 0.10 \end{pmatrix}$$

Design 3 (dense case): The setup in this design is the same as Design 2, except we

Table 1: MMADs with standard deviations of MADs in their corresponding parentheses for Design 1.

p	Method	Error Distribution		
		normal	t_3	χ^2_3
0.50	PMQ	0.903 (0.014)	0.918 (0.011)	1.289 (0.017)
	BQRGS	1.023 (0.055)	1.327 (0.062)	1.725 (0.038)
	BALQR	1.221 (0.092)	1.321 (0.075)	2.051 (0.095)
	QR	1.287 (0.083)	1.704 (0.084)	2.193 (0.090)
0.75	PMQ	0.936 (0.015)	0.985 (0.021)	1.226 (0.025)
	BQRGS	1.045 (0.057)	1.363 (0.035)	1.633 (0.059)
	BALQR	1.212 (0.091)	1.459 (0.091)	1.925 (0.111)
	QR	1.295 (0.097)	1.697 (0.106)	2.196 (0.115)
0.95	PMQ	1.001 (0.003)	1.251 (0.016)	1.634 (0.035)
	BQRGS	1.103 (0.026)	1.434 (0.126)	2.014 (0.084)
	BALQR	1.222 (0.096)	1.741 (0.112)	2.116 (0.109)
	QR	1.320 (0.084)	1.744 (0.102)	2.215 (0.094)

set $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8) = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)$.

Design 4 (sparse recovery problem where the number of predictors exceeds the sample size). The setup in this design is the same as Design 2, except we set N=30, n_i=3, $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2}, \cdots, x_{ij100})'$ and $\boldsymbol{\beta} = (\underbrace{1, 1, 1}_{3}, \underbrace{0, \cdots, 0}_{97})$, which corresponds to the case that the number of variables is greater than the sample size.

Table 2: MMADs with standard deviations of MADs in their corresponding parentheses for Design 2.

p	Method	Error Distribution		
		normal	t_3	χ^2_3
0.50	PMQ	0.981 (0.009)	1.126 (0.013)	1.121 (0.005)
	BQRGS	1.041 (0.048)	1.329 (0.072)	1.331 (0.021)
	BALQR	1.281 (0.106)	1.351 (0.096)	1.827 (0.011)
	QR	1.476 (0.082)	1.881 (0.094)	2.281 (0.124)
0.75	PMQ	1.083 (0.011)	1.075 (0.001)	0.987 (0.003)
	BQRGS	1.329 (0.032)	1.512 (0.031)	1.329 (0.021)
	BALQR	1.425 (0.102)	1.924 (0.099)	1.855 (0.065)
	QR	1.509 (0.096)	1.865 (0.106)	2.290 (0.138)
0.95	PMQ	1.092 (0.008)	1.167 (0.019)	1.248 (0.011)
	BQRGS	1.331 (0.026)	1.308 (0.103)	1.403 (0.048)
	BALQR	1.429 (0.092)	1.812 (0.101)	2.071 (0.094)
	QR	1.516 (0.095)	1.904 (0.112)	2.361 (0.127)

Design 1, 2, 3 and 4 show that, in terms of the MMAD and the standard deviation of MAD (SD), our proposed method performs better than the other methods in general. The results of the MMAD and standard deviations for Designs 1-4 are reported in Table 1, 2 and Figure 1, 2, respectively. We see that the proposed method (PMQ) tends to give lower MMAD and standard deviations compared with the other approaches reviewed for all distributions under considerations, suggesting good performance of the algorithm. As expected, the results of MMAD and the standard deviations criteria show that the BALQR and QR do not perform well because they ignore the random effects entirely.

Instead of looking at the MMAD and the standard deviations criteria, we may also look at the correlation coefficient between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$. The results of the correlation coefficients over 100 simulations for Designs 1-3 are reported in Figure 3, 4 and 5, respectively. The figures are plotted using the R package `vioplot` (Adler, 2005). We see that the proposed

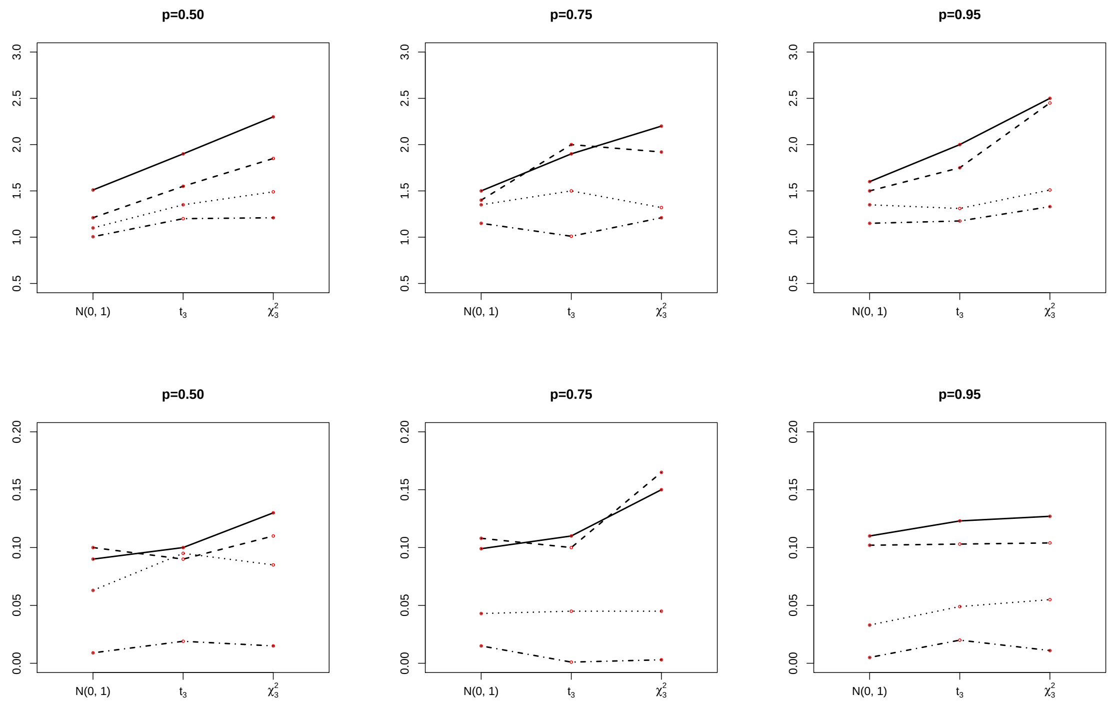

Figure 1: The upper panels represent the MMADs and the lower panels represent the standard deviations for the simulated data in Design 3. The solid line denotes the QR method, the dashed line denotes the BALQR, the dotted line denotes the BQRGS and the dotted-dashed line denotes our Gibbs sampler (PMQ).

method (PMQ) tends to give higher correlation coefficient between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ compared with the other approaches reviewed, suggesting good performance of the proposed algorithm. We also observe that BQRGS gives a higher correlation coefficient compared with BALQR and QR. Similar conclusions are also observed for Design 4 of which, the correlation coefficient results are not shown here.

We may also look at the estimation of the parameter vector $\boldsymbol{\beta}$. From Table 3 we observe that, in general, the proposed method performs well when comparing the estimates of β_j with the true values of β_j. Table 4 shows the posterior means, standard deviations and 95% credible intervals for the random effect variances in the simulated data of Design 1 when the error is normal by using our proposed method. As we can observe from Table 4, all the

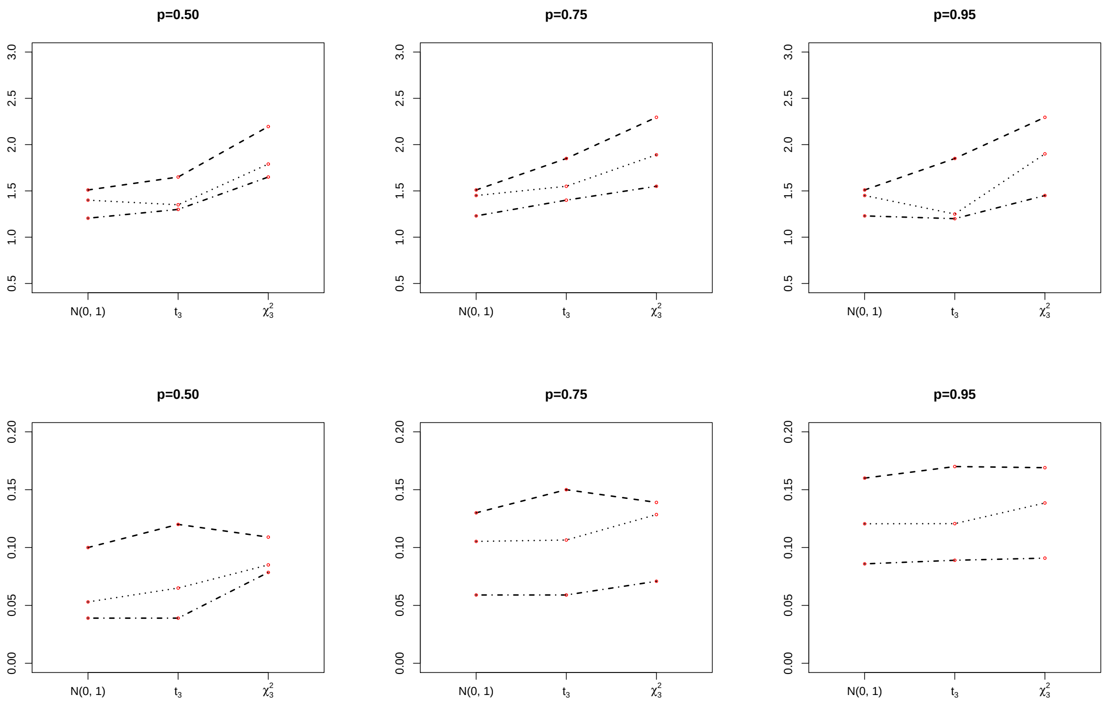

Figure 2: The upper panels represent the MMADs and the lower panels represent the standard deviations for the simulated data in Design 4. The dashed line denotes the BALQR, the dotted line denotes the BQRGS and the dotted-dashed line denotes our Gibbs sampler (PMQ). Because the number of variables is greater than the sample size, the design matrix is singular. As a result, the standard QR whose results are not shown here fails in this Design.

credible intervals contain the true parameter values, indicating that our algorithm performs well.

Table 3: Posterior means for the simulated data in Design 1 when the error is normal.

p	Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
	$\boldsymbol{\beta}^{true}$	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000
0.50	PMQ	1.004	1.001	0.998	0.004	0.000	0.005	0.011	0.002
	BQRGS	1.021	0.992	1.002	0.012	0.043	0.048	0.019	-0.057
	BALQR	1.052	0.988	1.013	0.002	0.002	-0.003	0.009	-0.019
	QR	0.983	0.991	0.992	0.003	-0.012	-0.015	-0.011	0.031
0.75	PMQ	1.009	0.994	0.997	0.000	0.001	-0.008	-0.001	-0.013
	BQRGS	0.911	0.937	1.112	-0.032	0.045	0.037	0.013	0.023
	BALQR	1.013	0.978	1.012	0.007	0.003	-0.012	-0.009	0.002
	QR	1.015	0.953	0.991	0.016	0.026	-0.014	-0.021	0.014
0.95	PMQ	1.012	0.987	1.019	0.007	-0.014	0.021	0.004	0.009
	BQRGS	0.944	0.891	0.965	0.045	0.035	-0.002	0.015	0.021
	BALQR	0.967	0.969	0.989	0.013	0.005	0.003	-0.001	-0.008
	QR	0.958	0.977	1.022	0.016	-0.019	0.002	-0.013	0.011

6. Age-related macular degeneration data

We illustrate the proposed method with the age-related Macular Degeneration data (ARMD) previously analyzed by Chaili (2008). There are 203 patients which were randomly selected from three cities (centers) in the United Kingdom to measure the treatment effects of teletherapy on the loss of vision associated with the progress of age related macular degeneration. The sample consists of 70 patients from London, 84 from Belfast and 49 from Southampton. Of which, 101 patients were randomly assigned to a treatment medication group and 102 to a control group. The response variable, the change in Distance Visual Acuity (DVA), of each patient was measured four times over a two year period, on the 3th, 6th, 12th and 24th months (Chaili, 2008). Owing to the possible heterogeneity among the subjects, this dataset is of specific interest to us. In this section, our quantile regression model contains seven covariates. The seven covariates are time (x_1), age (x_2), sex (x_3), centre at which the examination took place (x_4), treatment (x_5), the index eye of the patient (x_6)

Table 4: Posterior means (Mean), standard deviations (SD) and 95% credible intervals (CrI) for the random effect variances in the simulated data of Design 1 when the error is normal by using our proposed method.

p		σ_{11}	σ_{22}	σ_{33}	σ_{44}	σ_{55}	σ_{66}	σ_{77}	σ_{88}
0.50	σ^{true}	0.75	0.55	0.10	0.00	0.00	0.00	0.00	0.00
	Mean	0.83	0.41	0.13	0.07	0.01	0.06	0.05	0.02
	SD	0.31	0.27	0.24	0.11	0.07	0.04	0.03	0.05
	CrI	(0.45, 1.13)	(0.31, 0.76)	(0.03, 0.21)	(0.00, 0.17)	(0.00, 0.03)	(0.00, 0.14)	(0.00, 0.09)	(0.00, 0.04)
0.75	Mean	0.93	0.49	0.18	0.03	0.00	0.08	0.06	0.05
	SD	0.26	0.30	0.16	0.04	0.02	0.09	0.08	0.01
	CrI	(0.53, 1.27)	(0.26, 0.71)	(0.04, 0.30)	(0.00, 0.08)	(0.00, 0.04)	(0.03, 0.17)	(0.00, 0.11)	(0.02, 0.09)
0.95	Mean	0.99	0.62	0.08	0.05	0.03	0.2	0.00	0.02
	SD	0.43	0.35	0.09	0.02	0.01	0.05	0.01	0.03
	CrI	(0.61, 1.32)	(0.33, 0.81)	(0.02, 0.17)	(0.01, 0.11)	(0.00, 0.07)	(0.00, 0.06)	(0.00, 0.02)	(0.00, 0.05)

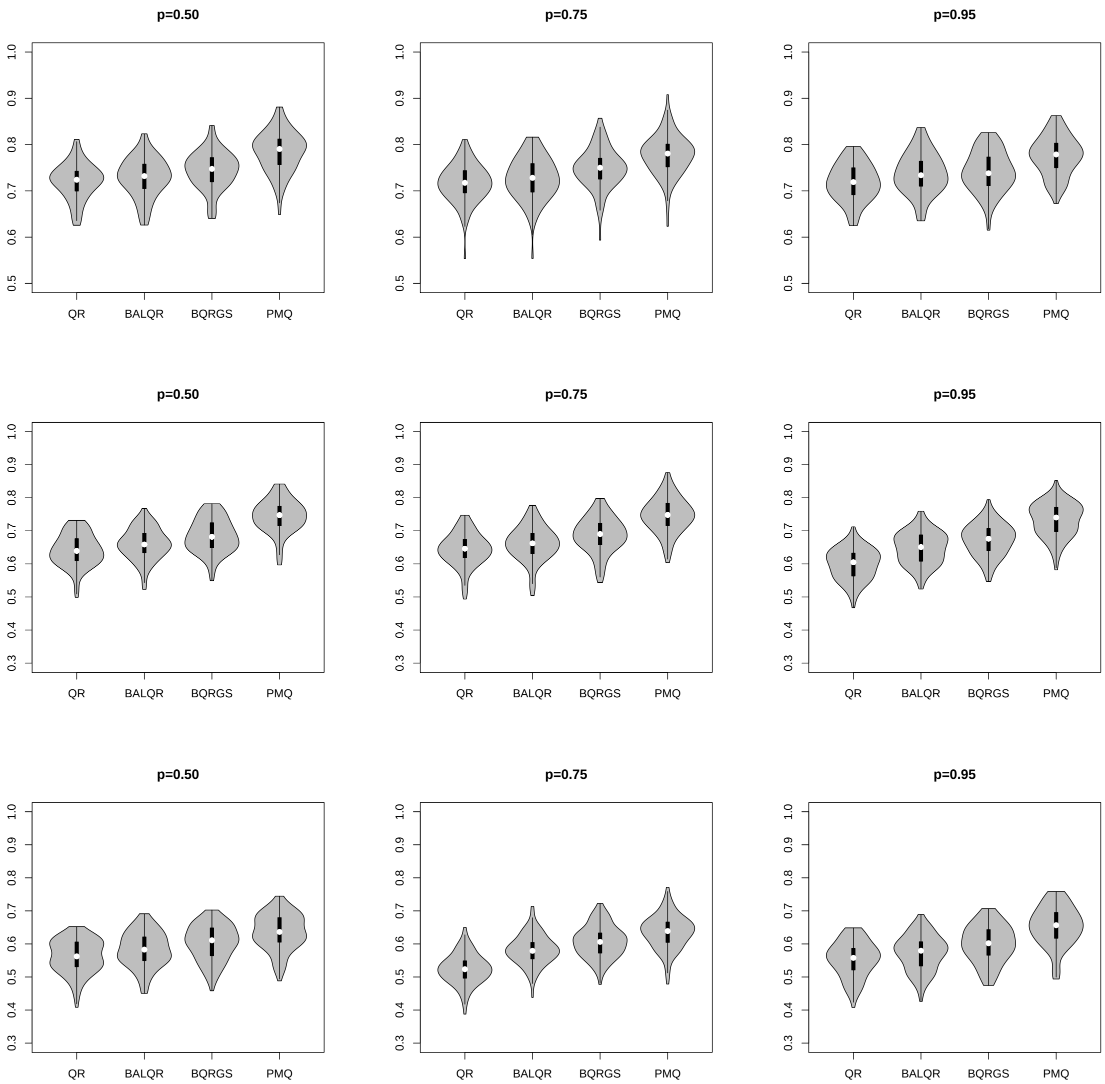

Figure 3: Correlation coefficients between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ for Design 1 over 100 simulations. The panels from the top to the bottom correspond to the standard normal distribution, $t_{(3)}$ distribution and $\chi^2_{(3)}$ distribution for the errors, respectively.

and either both or one eye affected by the condition (x_7). In this example we set $\boldsymbol{z}_{ij} = \boldsymbol{x}_{ij}$. and we assume the random effects follow the multivariate normal distribution $\mathrm{N}_q(0, \boldsymbol{I})$. The results of mean absolute deviation (MAD) and the standard deviations of AD are reported in Table 5. Clearly, we can see from Table 5 that the results of MMAD and the standard deviations criteria show that the proposed method perform well compared with the other

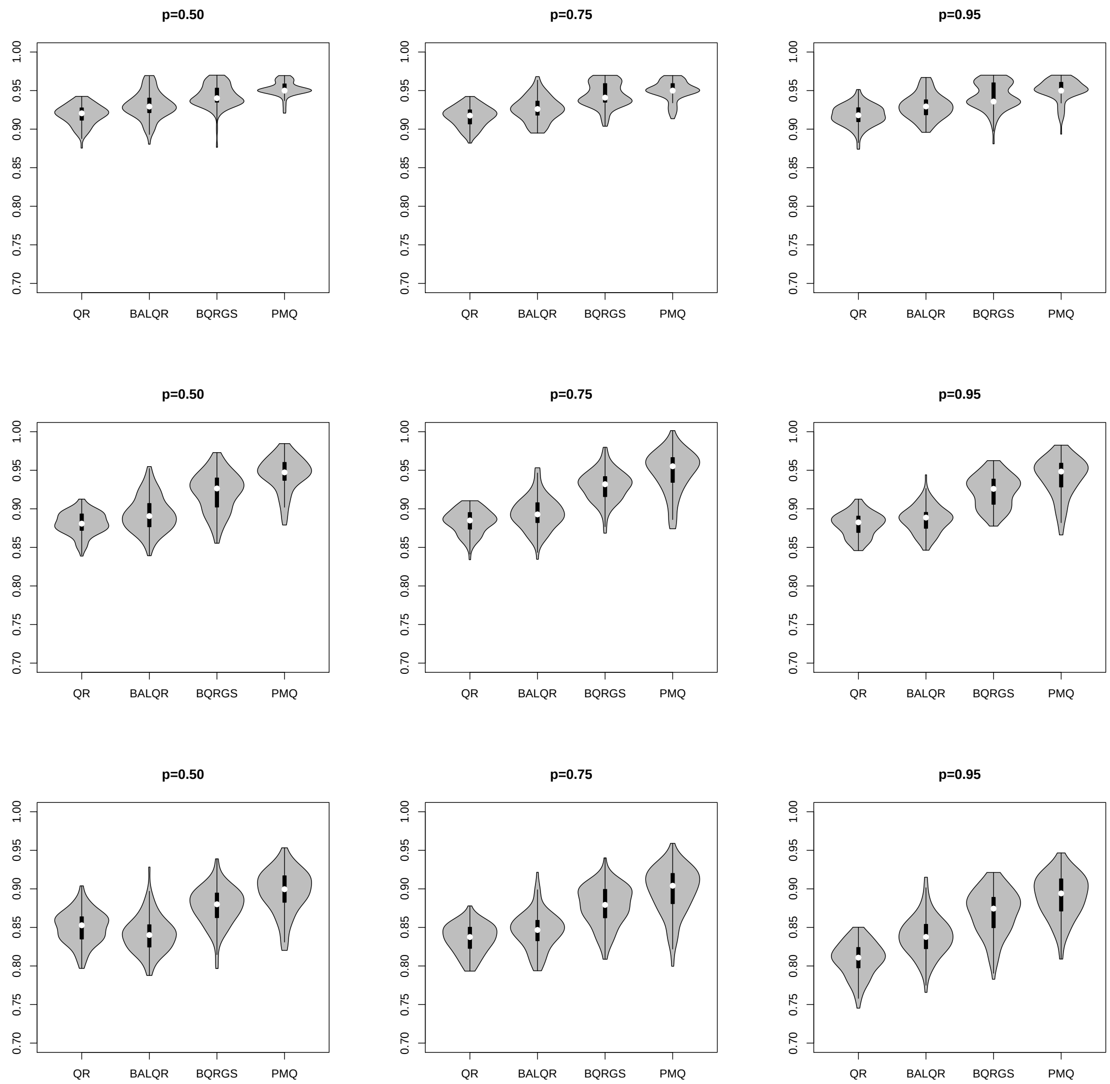

Figure 4: Correlation coefficients between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ for Design 2 over 100 simulations. The panels from the top to the bottom correspond to the standard normal distribution, $t_{(3)}$ distribution and $\chi^2_{(3)}$ distribution for the errors, respectively.

approaches.

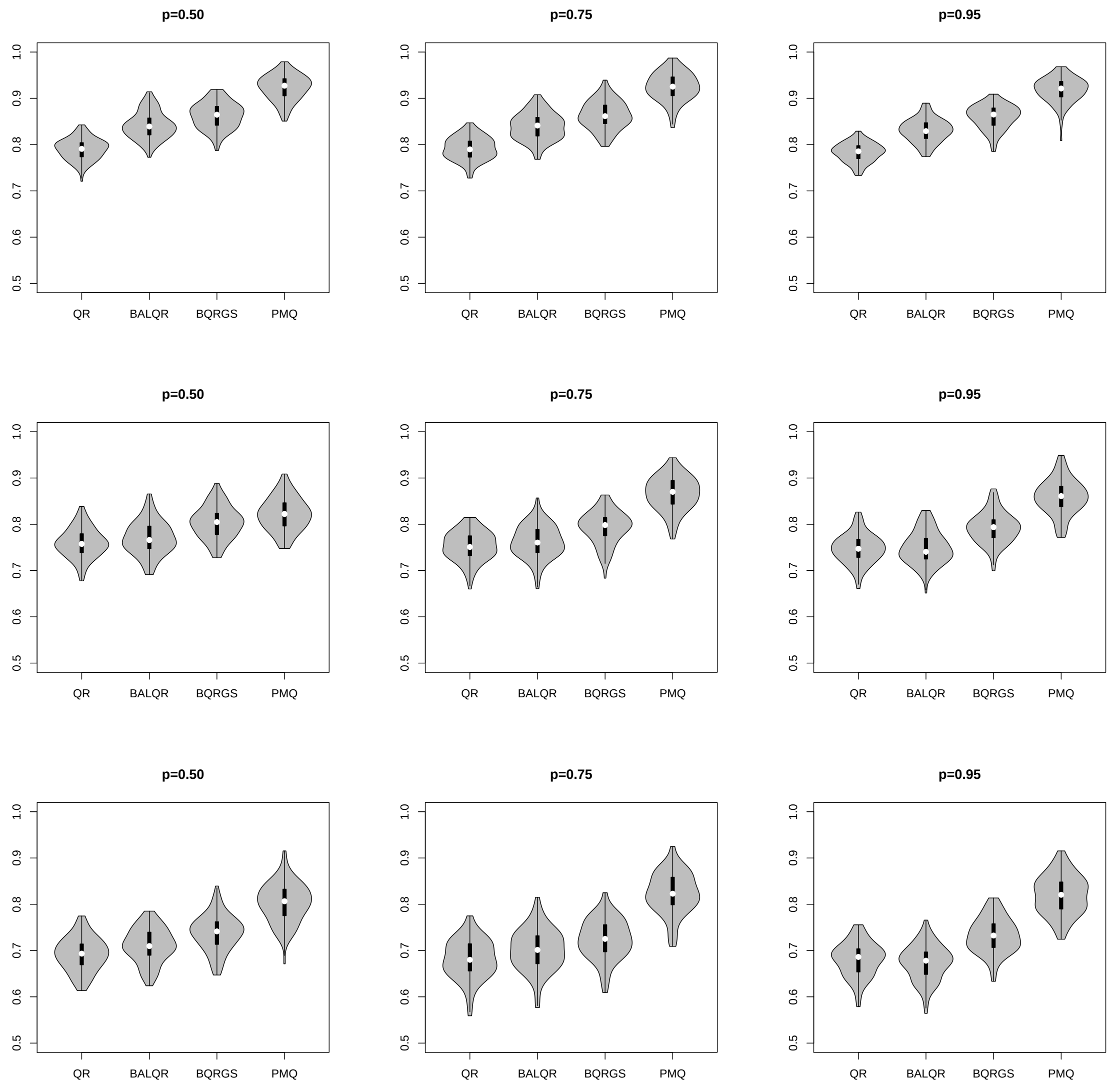

Figure 5: Correlation coefficients between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ for Design 3 over 100 simulations. The panels from the top to the bottom correspond to the standard normal distribution, $t_{(3)}$ distribution and $\chi^2_{(3)}$ distribution for the errors, respectively.

7. Conclusion

In this paper, we have proposed a Bayesian hierarchical model for variable selection and estimation in mixed quantile regression models. Similar to Chen and Dunson (2003) we re-parameterized linear mixed quantile regression model so that functions of the covariance parameters of the random effects distribution are incorporated as regression coefficients.

Table 5: MADs with standard deviations of ADs in their corresponding parentheses for the age-related macular degeneration data.

Method	$p = 0.50$	$p = 0.75$	$p = 0.95$
PMQ	0.043 (0.014)	0.056 (0.026)	0.153 (0.105)
BQRGS	0.213 (0.053)	0.287 (0.051)	0.325 (0.203)
BALQR	0.316 (0.067)	0.331 (0.147)	0.369 (0.133)
QR	0.203 (0.179)	0.251 (0.196)	0.471 (0.255)

We have introduced a Gibbs sampler for Bayesian mixed quantile regression with the joint Lasso penalty for fixed effects and random effect variances. This Gibbs sampler is based on a theoretic derivation of the skewed Laplace distribution as a scale mixture of normal distributions. By using simulated and age-related macular degeneration data we have shown that the proposed method can outperform the commonly used methods with respect to estimation.

Bibliography

Adler, D. (2005). *`vioplot`: A violin plot is a combination of a box plot and a kernel density plot*. R package version 0.2. url = http://cran.r-project.org/web/packages/vioplot/index.html.

Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. In Second international symposium on information theory, eds. Petrov, B. N. and Csaki, F.

Alhamzawi, R. and Yu, K. (2012). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics and Data Analysis*. In press. available online 24 January 2012.

Alhamzawi, R., Yu, K. and Benoit, D.F. (2011). Bayesian adaptive LASSO quantile regression, *Statistical Modelling* **12**, 279-297.

Andrews, D. F. and Mallows, C. L. (1974). Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society, Ser. B* **36**, 99–102.

Bob Wheeler (2009. SuppDists: Supplementary distributions. R package version 1.1-8.

Bondell, H. D., Krishna, A., and Ghosh, S. K. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*. **66**, 69–77.

Bradic, J., Fan, J. and Wang, W. (2010). Penalized composite quasi-likelihood for ultrahigh-dimensional variable selection. *Journal of the Royal Statistical Society, Ser. B* **73**, 325–349.

Chaili, PX. (2008). Statistical modelling of age-related macular degeneration. A dissertation submitted to The University of Manchester for the degree of Master of Science in the Faculty of Engineering and Physical Sciences.

Chen, Z. and Dunson, D. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762-769.

Chhikara, R.S. and Folks, L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications.* Marcel Dekker, New York.

Ibrahim, J. G., Zhu, H., Garcia, R. I. and Guo, R. (2011). Fixed and Random Effects Selection in Mixed Effects Models. *Biometrics* **67**, 495-503.

Kinney, S. K. and Dunson, D. B. (2008). Fixed and random effects selection in linear and logistic models. *Biometrics* **63**, 690–698.

Koenker, R. (2005). *Quantile Regression.* Cambridge U. Press.

Koenker, R. (2012). *quantreg: Quantile Regression.* R package version 4.79. url = http://CRAN.R-project.org/package=quantreg.

Koenker, R. and Bassett, G. Jr. (1978). Regression Quantiles. *Econometrica* **46,** 33–50.

Koenker, R. and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression *Journal of the American Statistical Association* **94,** 1296-1310.

Kozumi, H. and Kobayashi, G. (2009). Gibbs Sampling Methods for Bayesian Quantile Regression. *Journal of Statistical Computation and Simulation* **81,** 1565-1578.

Laird, N. M. and Ware, J. L. (1982). Random-effects models for longitudinaldata. *Biometrics* **38**, 963-974.

Li, Y. and Zhu, J. (2008) L1-norm quantile regression. *J. Computnl Graph. Statist.* **17**, 163-185.

Li, Q., Xi, R. and Lin, N. (2010). Bayesian Regularized Quantile Regression. *Bayesian Analysis* **5**, 1-24.

Luo, Y., Lian, H., Tian, M. (2011). Bayesian quantile regression for longitudinal data model. *Journal of Statistical Computation and Simulation*. To appear.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103**, 681–686.

Reed, C. and Yu, K. (2009). A Partially collapsed Gibbs sampler for Bayesian quantile regression. Technical report. Department of Mathematical Sciences, Brunel University.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.

Saville, B. R. and Herring, A. H. (2009). Testing Random Effects in the Linear Mixed Model Using Approximate Bayes Factors. *Biometrics* **65**, 369-376. doi: 10.1111/j.1541-0420.2008.01107.x

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica* **19**, 801-817.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile Regression: Applications and Current Research Areas. *The Statistician* **52**, 331-350.

Yu, K. and Moyeed, R.A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* **54,** 437– 447.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics* **36**, 1108-1126.