



## Research Article

## Open Access

## Prior Elicitation in Bayesian Quantile Regression for Longitudinal Data

Rahim Alhamzawi<sup>1\*</sup>, Keming Yu<sup>1</sup> and Jianxin Pan<sup>2</sup><sup>1</sup>Department of Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, UK<sup>2</sup>School of Mathematics, University of Manchester, Manchester, M13 9PL, UK**Abstract**

In this paper, we introduce Bayesian quantile regression for longitudinal data in terms of informative priors and Gibbs sampling. We develop methods for eliciting prior distribution to incorporate historical data gathered from similar previous studies. The methods can be used either with no prior data or with complete prior data. The advantage of the methods is that the prior distribution is changing automatically when we change the quantile. We propose Gibbs sampling methods which are computationally efficient and easy to implement. The methods are illustrated with both simulation and real data.

**Keywords:** Bayesian quantile regression; Conditional distribution; Gibbs sampling; Longitudinal data; Mixture representation; Random effect

**Introduction**

Quantile regression models have been widely used for a variety of application [1,2]. It has attracted much interest in recent years because of the flexibility of quantile regression for modelling data with heterogeneous conditional distributions. In addition, a set of quantiles of the response variable (such as the first quartile, the median and the third quartile) may depend on the explanatory variables very differently from the center. Thus, a set of quantiles may give a more complete picture of the relation between the explanatory variables and the response variable than the mean regression. Furthermore, quantile regression makes very minimal assumptions on the error term distribution and thus its estimators may be more robust than the mean regression when the error term is non normal. Consequently, quantile regression has emerged as a useful supplement to standard mean regression models.

One of the serious challenges in quantile regression lies in analysis of longitudinal data in which repeated measurements are made on the same subject over time as well as in specification of quantile dependent prior distributions in quantile regression. There exists little literature for quantile regression in longitudinal data and we refer to [3-6].

This paper considers Bayesian quantile regression with random effects. We develop methods for eliciting prior distribution to incorporate historical data gathered from similar previous studies. The methods can be used either with no prior data or with complete prior data. The advantage of the methods is that the prior distribution is changing automatically when we change the quantile as well as precision. In addition, we propose Gibbs sampler for Bayesian quantile regression with random effects which is computationally efficient and easy to implement compared with expectation maximization algorithm proposed by [4] and Bayesian MCMC method proposed by [6].

The rest of this article is organized as follows. Section 2 introduces asymmetric Laplace as scale mixtures of normal distributions, we elicit power prior distribution, and describe Gibbs sampler (GS) for Bayesian quantile regression. In Section 3, we illustrate the Gibbs sampler by analyzing simulated data and compare our results with Bayesian MCMC and EM algorithm. Section 4 analyzes an age-related macular degeneration data set. We conclude with a brief discussion in Section 5.

**Model and Methods****Random effects model for longitudinal data**

Suppose there are  $N$  subjects under study so that  $y_{ij}$  denote the  $j^{\text{th}}$  measurement on the  $i^{\text{th}}$  subject, for  $i=1, \dots, N$  and  $j=1, \dots, n_i$ . We start with the following latent regression model:

$$y_{ij} = x'_{ij}\beta + z'_{ij}b_i + \varepsilon_{ij}, \quad j=1, \dots, n_i, \quad i=1, \dots, N,$$

where  $x'_{ij}$  and  $z'_{ij}$  are rows of the  $X_i$  and  $Z_i$  matrices,  $X_i$  is  $n_i \times (k+1)$  and  $Z_i$  is  $n_i \times q$ ,  $\beta$  and  $b_i$  are  $(k+1)$  and  $q$ -dimensional unknown parameters and random effects, respectively, and  $\varepsilon_{ij}$  is the error term. We define the linear mixed quantile functions of the response  $y_{ij}$ .

$$Q_{y_{ij}|b_i}(p | x_{ij}, b_i) = x'_{ij}\beta_p + z'_{ij}b_i, \quad j=1, \dots, n_i, \quad i=1, \dots, N, \quad (1)$$

where  $Q_{y_{ij}|b_i}$  is the inverse of the cumulative distribution function of  $y_{ij}$  given a vector of unknown subject-specific random effects  $b_i$ ,  $b_i \sim N_q(0, \Lambda)$ ,  $\Lambda$  is a symmetric nonsingular matrix, and the  $p^{\text{th}}$  quantile of the response  $y_{ij}$  is  $x'_{ij}\beta_p + z'_{ij}b_i$ . We assume the conditional distribution of  $y_{ij}$  given  $b_i$ , for  $j=1, \dots, n_i$  and  $i=1, \dots, N$ , is an independent distribution according to the asymmetric laplace distribution, so that

$$f(y_{ij} | \beta_p, b_i, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\rho_p\left(\frac{y_{ij} - x'_{ij}\beta_p - z'_{ij}b_i}{\sigma}\right)\right\}, \quad (2)$$

where  $\sigma > 0$  is a scale parameter, and  $\rho_p(w) = wp - w1_{(-\infty, 0)}(w)$ . The parameter  $p$  determines the skewness of the distribution and the  $p^{\text{th}}$  quantile of this distribution is zero.

**\*Corresponding author:** Rahim Alhamzawi, Department of Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, UK, E-mail: [Rahim.Al-hamzawi@brunel.ac.uk](mailto:Rahim.Al-hamzawi@brunel.ac.uk)

**Received** December 02, 2010; **Accepted** April 18, 2011; **Published** September 25, 2011

**Citation:** Alhamzawi R, Yu K, Pan J (2011) Prior Elicitation in Bayesian Quantile Regression for Longitudinal Data. J Biomet Biostat 2:115. doi:10.4172/2155-6180.1000115

**Copyright:** © 2011 Alhamzawi R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Let  $y_i = (y_{i1}, \dots, y_{in_i})'$ ,  $y = (y_1', \dots, y_N')'$  and  $b = (b_1', \dots, b_N')'$ . The complete data density of  $(y, b)$ , for  $i = 1, 2, \dots, N$ , is then given by

$$f(y, b | \beta_p, \sigma, \Lambda) = f(y | b, \beta_p, \sigma) f(b | \Lambda) \tag{3}$$

There are several attractive properties of (3). First, it takes within subject correlation into account while allowing each individual to have a unique correlation subject. Second, it provides us a degree of shrinkage of the subject-specific regression lines toward the population line. Finally, a nice property of the asymmetric laplace distributions is that it can be represented as a scale mixture of normal distributions [7-10].

Our interest lies in the likelihood function of  $y$  given  $\beta_p, \sigma$ , and  $\Lambda$ . To induce the correlation structure on the responses, we integrate out the random effects

$$\begin{aligned} f(y | \beta_p, \sigma, \Lambda) &= \int_{R^{Nq}} f(y, b | \beta_p, \sigma, \Lambda) db, \\ &= \prod_{i=1}^N \left( \int_{R^q} \prod_{j=1}^{n_i} f(y_{ij} | \beta_p, b_i, \sigma) f(b_i | \Lambda) db_i \right), \\ &= \prod_{i=1}^N \left\{ \left[ \frac{p(1-p)}{\sigma} \right]^{n_i} (2\pi)^{-q/2} |\Lambda|^{-\frac{1}{2}} \right. \\ &\quad \left. \times \int_{R^q} \exp\left\{ -\sum_{j=1}^{n_i} \rho_p \left( \frac{y_{ij} - x'_{ij}\beta_p - z'_{ij}b_i}{\sigma} \right) - \frac{1}{2} b_i' \Lambda^{-1} b_i \right\} db_i \right\} \end{aligned} \tag{5}$$

where  $R^{Nq}$  and  $R^q$  denote the  $Nq$  and  $q$ -dimensional Euclidean space, respectively.

Model (2.5) is similar to the  $L_2$ -penalized check function proposed by [6] which extends the random intercept model proposed by [4] to a very general case.

### Mixture representation

This section introduces Gibbs sampler as procedure to estimate the parameters of interest. We adopt a full Bayesian approach to quantile regression for longitudinal data. Consider the linear mixed quantile functions of the response  $y_{ij}$  (1), where the error term  $\varepsilon_{ij}$  has asymmetric Laplace distribution with the  $p$ th quantile equal to zero.

Recently, [7-10] proved that the asymmetric Laplace distribution can be viewed as a mixture of an exponential and a scaled normal distribution. This can be recognized in the following lemma.

Lemma Suppose  $\xi_{ij}$  is a standard normal random variable,  $t_{ij}$  is a standard exponential random variable and  $\varepsilon_{ij}$  is a random variable follows the asymmetric laplace distribution with density

$$f_p(\varepsilon_{ij}) = \frac{p(1-p)}{\sigma} \exp\left\{ -\rho_p \left( \frac{\varepsilon_{ij}}{\sigma} \right) \right\},$$

Then we can represent  $\varepsilon_{ij}$  as a location mixture of normal given by  $\varepsilon_{ij} = \theta + \sigma t_{ij} + \sigma \varphi \xi_{ij} \sqrt{t_{ij}}$ ,  $\tag{6}$

where

$$\theta = \frac{1-2p}{p(1-p)} \quad \text{and} \quad \varphi^2 = \frac{2}{p(1-p)}.$$

We assume  $\xi_{ij}$  and  $t_{ij}$  are mutually independent. This mixture representation allow us to express a quantile regression model as a normal regression model. In addition, it provides an easy way to construct Gibbs sampler as well as saving time over sampling random

effect.

We assume that  $y_{ij}$  conditionally on  $b_i$  and  $t_{ij}$ , for  $j = 1, \dots, n_i$  and  $i = 1, \dots, N$ , are independently distributed according to the normal with mean  $x'_{ij}\beta_p + z'_{ij}b_i + \sigma\theta t_{ij}$  and variance  $\sigma^2\varphi^2 t_{ij}$ . From now on, it is more convenient with Gibbs sampler to work with  $v_{ij} = \sigma t_{ij}$  to avoid the scale parameter  $\sigma$  appearing in the conditional mean of  $y_{ij}$  [10].

### Power prior distributions and Gibbs sampler

In this section, we address a quantile dependent prior in Bayesian quantile regression for longitudinal data. Since [11]. Bayesian inference quantile regression has attracted a lot of attention in literature including [4-7,9,10,12-21]. However, almost all these models set priors independent of the values of quantiles, or the prior is the same for modelling different quantiles [27]. This approach may result in an inflexibility in quantile modelling. For example, a 95% quantile regression model should have different parameter values from the median quantile, and thus the priors used for modelling the quantiles should be different [22]. In this paper, we address a quantile dependent prior for quantile mixed model. Our idea is to set priors based on historical data. The power prior by [23] is one of several methods to incorporate historical data into the analysis of currnt data. We adapt this prior distribution to be used in quantile regression.

Suppose there exists one historical data from a previous study. Let  $y_{0ij}$  be the  $j$ th measurement ( $j = 1, \dots, n_{0i}$ ) for the  $i$ th subject ( $i = 1, \dots, N_0$ ),  $x'_{0ij}$  and  $z'_{0ij}$  are rows of the  $X_{0i}$  and  $Z_{0i}$  matrices,  $X_{0i}$  is  $n_{0i} \times (k+1)$  and  $Z_{0i}$  is  $n_{0i} \times q$ , and let  $b_{0i}$  is a vector of unknown subject-specific effects for the previous study. Denoting by  $D_0 = (N_0, y_0, X_0, Z_0)$  a historical data with  $N_0$  subjects measuring the same response variable and covariates as the current study where  $y_0 = (y'_{01}, \dots, y'_{0N_0})'$ ,  $y_{0i} = (y_{0i1}, \dots, y_{0in_{0i}})'$ ,  $X_0 = (X'_{01}, \dots, X'_{0N_0})'$  and  $Z_0 = \text{diag}(Z_{01}, \dots, Z_{0N_0})$ . For the previous study, we assume  $v_0 = (v'_{01}, \dots, v'_{0N_0})'$ ,  $v_{0i} = (v_{0i1}, \dots, v_{0in_{0i}})$  and each  $v_{0ij} \sim \text{Exp}(\sigma)$ . For quantile mixed model, we define a prior distribution for  $\beta_p$  taking the form

$$\begin{aligned} \pi(\beta_p | D_0, \Lambda, \sigma, a_0, v_0) \\ \propto \prod_{i=1}^{N_0} \left( \int_{R^q} \prod_{j=1}^{n_{0i}} [f(y_{0ij} | \beta_p, b_{0i}, \sigma, v_{0ij})]^{a_0} f(b_{0i} | \Lambda) db_{0i} \right) \end{aligned} \tag{7}$$

where  $\prod_{j=1}^{n_{0i}} f(y_{0ij} | \beta_p, b_{0i}, \sigma, v_{0ij}) = f(y_{0i} | \beta_p, b_{0i}, \sigma, v_{0i})$  be the density for the  $i$ th subject given a vector of unknown subject specific effects  $b_{0i}$ , and  $0 \leq a_0 \leq 1$ . There are several attractive properties of (7). First, the prior distribution (7) is dependnt on the quantile. Second, it provides an easy way to construct Gibbs sampler via the mixture representation. We assume  $a_0$  as known parameter. The power parameter  $a_0$  represents how much data from the previous study should be used in the current study. The main role of  $a_0$  is that  $a_0$  controls the influence of the data gathered from previous studies that is similar to the current study. Such control is important when the sample size of the current data is quite different from the sample size of historical data [23]. Chen et al. [24] choose  $a_0$  as random to allow the flexibility in expressing our uncertainty about the power parameter via a prior mean and variance. The prior mean and variance is determined by the investigator. Recently [25,26] recognized that using  $a_0$  as random is inappropriate.

Neuenschwander et al. [25] shown that elicitation of specific values

of  $a_0$  can be done via expert opinions, or via a meta-analytic. Neelon et al. [26] recommend choosing  $a_0$  based on expert opinion about the commensurability of current and historical studies. The authors assign the power parameter a range of fixed values as part of a sensitivity analysis. If, for example, investigators are concerned about the similarity between the studies, they may choose  $a_0 \leq 0.50$ . On the other hand, if there is strong belief about the similarity between the studies, they may choose  $a_0 > 0.50$ . They also recommended conducting reference analyses in which  $a_0$  is set to 0 as lower bound and 1 as upper bound. In this paper, following [26] we take a range of values for  $a_0$  between 0 and 1 based on expert opinion about the commensurability of current and historical studies.

In this paper, we assume  $\Lambda = c^{-1}I_q$ , where  $C$  is precision parameter. The prior specification is completed by specifying priors for  $\sigma, c$  and  $\beta_p$ . We specify an inverse gamma (IG) prior with parameter  $(l_{01}, s_{01})$  for  $\sigma$ , a gamma ( $\Gamma$ ) prior with parameter  $(l_{02}, s_{02})$  for  $c$ , and multivariate normal prior with parameter  $(\mu_0, B_0)$  for  $\beta_p$ . Thus, we define a joint prior distribution taking the form

$$\begin{aligned} &\pi(\beta_p, \sigma, c, v_0 | D_0, a_0) \\ &\propto \prod_{i=1}^{N_0} \left( \int_{R^q} \prod_{j=1}^{n_{0i}} [f(y_{0ij} | \beta_p, b_{0i}, \sigma, v_{0ij})]^{a_0} f(b_{0i} | \Lambda) db_{0i} \pi(v_{0ij} | \sigma) \right) \\ &\times \exp\left\{-\frac{1}{2}(\beta_p - \mu_0)' B_0^{-1} (\beta_p - \mu_0)\right\} \left(\frac{1}{\sigma}\right)^{l_{01}+1} \exp\left\{-\frac{s_{01}}{\sigma}\right\} \times c^{l_{02}-1} \exp\{-cs_{02}\} \end{aligned} \tag{8}$$

We see that (8) will not have closed form in general because it depends on the prior that we choose. The joint posterior distribution of  $\beta_p, \sigma, c, v$  and  $v_0$  is given by

$$\begin{aligned} &f(\beta_p, \sigma, c, v, v_0 | D, D_0, a_0) \\ &\propto \left( \prod_{i=1}^N \prod_{j=1}^{n_i} f(y_{ij} | \beta_p, b_i, \sigma, v_{ij}) f(b_i | \Lambda) db_i \pi(v_{ij} | \sigma) \right) \\ &\times \pi(\beta_p, \sigma, c, v_0 | D_0, a_0), \end{aligned} \tag{9}$$

where  $D = (N, y, X, Z)$  represent the current study and  $v = (v'_N, \dots, v'_1)$ ,  $v_i = (v_{i1}, \dots, v_{in_i})$ , and each  $v_{ij} \sim \text{Exp}(\sigma)$ .

Incorporating historical information into the analysis of new information through a prior distribution provides a natural framework for updating information across studies [26]. The most common way of combining the historical information into the analysis of new information is through hierarchical modeling. There are times, however, when investigators want to control the influence of the historical information on the new information. In this paper, we use the power prior distribution, because it introduces a power parameter that explicitly controls the amount of weight assigned to the historical data. Such control is important when the sample size of the current data is quite different from the sample size of the historical data or where there is heterogeneity between two studies [23]. In addition, this prior has an attractive property in Bayesian quantile regression as it is dependent upon the quantile level [27].

Gibbs sampler by [28] is very popular method for constructing a Markov chain in Bayesian inference, and it is used to generate a sequence of samples from the full conditional distribution. We use Gibbs sampling in Bayesian quantile regression to estimate the parameters of interest from our mixture representation. Thus, the fully

conditional posterior distributions of all unknown parameters are needed and each of these distributions can be obtained by regarding all other parameters in (9) as known. The efficient Gibbs sampler works as follows:

1. Fix the value of  $p$  and  $a_0$  so that the  $p^{\text{th}}$  quantile and the weight of the historical data are modelled.

2. Simulate the scale parameter  $\sigma : \sigma | \cdot \sim$  Inverse Gamma  $(d_1, f_1)$  where

$$d_1 = (l_{01} + \sum_{i=1}^N n_i + a_0 \sum_{i=1}^{N_0} n_{0i}), \tag{10}$$

$$\begin{aligned} &\text{and} \\ &f_1 = s_{01} + \sum_{i=1}^N \sum_{j=1}^{n_i} \rho_p (y_{ij} - x'_{ij} \beta_p - z'_{ij} b_i) \\ &+ a_0 \sum_{i=1}^{N_0} \sum_{j=1}^{n_{0i}} \rho_p (y_{0ij} - x'_{0ij} \beta_p - z'_{0ij} b_{0i}). \end{aligned}$$

3. Simulate  $v_{ij} | \cdot \sim$  Generalized Inverse Gaussian  $(\frac{1}{2}, d_2, f_2), i = 1, \dots, N, j = 1, \dots, n_i$ , where

$$d_2 = \sqrt{\frac{(y_{ij} - x'_{ij} \beta_p - z'_{ij} b_i)^2}{\sigma \phi^2}} \text{ and } f_2 = \sqrt{\frac{\theta^2}{\sigma \phi^2} + \frac{2}{\sigma}}, \tag{11}$$

where the probability density function of Generalized Inverse Gaussian  $(r, f_2, d_2)$  is given by

$$\begin{aligned} &f(x | r, f_2, d_2) = \frac{(\frac{d_2}{2})^r}{2C_r(f_2 d_2)} x^{r-1} \exp\left\{-\frac{1}{2}(f_2^2 x^{-1} + d_2^2 x)\right\}, \\ &x > 0, -\infty < r < \infty, f_2, d_2 \geq 0, \\ &\text{and } C_r(\cdot) \text{ is a modified Bessel function of the third kind [29]} \end{aligned}$$

4. Simulate  $v_{0ij} | \cdot \sim$  Generalized Inverse Gaussian  $(\frac{2-a_0}{2}, d_3, f_3), i = 1, \dots, N_0, j = 1, \dots, n_{0i}$ , where

$$d_3 = \sqrt{\frac{a_0 (y_{0ij} - x'_{0ij} \beta_p - z'_{0ij} b_{0i})^2}{\sigma \phi^2}} \text{ and } f_3 = \sqrt{\frac{a_0 \theta^2}{\sigma \phi^2} + \frac{2}{\sigma}}.$$

5. Simulate the fixed effect parameters  $\beta_p | \cdot \sim N(\mu_{\beta_p}, \Sigma_{\beta_p})$ , where

$$\begin{aligned} \mu_{\beta_p} &= \Sigma_{\beta_p}^{-1} \left\{ \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{x_{ij} (y_{ij} - z'_{ij} b_i - \theta v_{ij})}{\sigma \phi^2 v_{ij}} \right. \\ &+ a_0 \sum_{i=1}^{N_0} \sum_{j=1}^{n_{0i}} \frac{x_{0ij} (y_{0ij} - z'_{0ij} b_{0i} - \theta v_{0ij})}{\sigma \phi^2 v_{0ij}} + B_0^{-1} \mu_0 \left. \right\}, \\ \Sigma_{\beta_p}^{-1} &= \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{x_{ij} x'_{ij}}{\sigma \phi^2 v_{ij}} + a_0 \sum_{i=1}^{N_0} \sum_{j=1}^{n_{0i}} \frac{x_{0ij} x'_{0ij}}{\sigma \phi^2 v_{0ij}} + B_0^{-1}. \end{aligned}$$

6. Simulate the random effect parameters  $b | \cdot \sim N(\mu_b, \Sigma_b)$ , where

$$\begin{aligned} \mu_b &= \Sigma_b^{-1} \left\{ \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{z_{ij} (y_{ij} - x'_{ij} \beta_p - \theta v_{ij})}{\sigma \phi^2 v_{ij}} \right\}, \\ \text{and } \Sigma_b^{-1} &= \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{z_{ij} z'_{ij}}{\sigma \phi^2 v_{ij}} + cI. \end{aligned}$$

7. Simulate the random effect parameters  $b_0 | \cdot \sim N(\mu_{b_0}, \Sigma_{b_0})$ , where

$$\mu_{b_0} = a_0 \sum_{i=1}^{N_0} \sum_{j=1}^{n_{0i}} \frac{z_{0ij}(y_{0ij} - x'_{0ij}\beta_p - \theta v_{0ij})}{\sigma \varphi^2 v_{0ij}}, \text{ and}$$

$$\Sigma_{b_0}^{-1} = a_0 \sum_{i=1}^{N_0} \sum_{j=1}^{n_{0i}} \frac{z_{0ij} z'_{0ij}}{\sigma \varphi^2 v_{0ij}} + cI.$$

8. Simulate the random effect precision parameter

$$c | \cdot \sim \text{Gamma}(l_{02} + \frac{q + q_0}{2}, s_{02} + \frac{b'b + b'_0 b_0}{2}). \tag{12}$$

### Some extensions

Using the power prior distributions depend on the availability of the historical data. In the previous part we elicited power prior distribution from one historical data and this prior can be easily generalized to multiple historical data. To generalize the power prior (8) to multiple historical data we assume there are  $M$  historical studies denoted by  $D_0 = (D_{01}, \dots, D_{0M})$  where  $D_{0k} = (N_{0k}, y_{0k}, X_{0k}, Z_{0k})$  represent the historical data based on the  $k^{\text{th}}$  study,  $k = 1, \dots, M$ . We define  $a_{0k}$  and  $v_{0ijk}$  to be the fixed power parameter and an exponential variable, respectively, for the  $k$  th study,  $a_0 = (a_{01}, \dots, a_{0M})$ . Hence the prior can be generalized as

$$\pi(\beta_p, \sigma, \sigma_b^2, \rho, v_0 | D_0, a_0)$$

$$\propto \prod_{k=1}^M \left( \prod_{i=1}^{N_{0k}} \prod_{j=1}^{n_{0ik}} [f(y_{0ijk} | \beta_p, b_{0ik}, \sigma, v_{0ijk})]^{a_{0k}} f(b_{0ik} | \Lambda) db_{0ik} \pi(v_{0ijk} | \sigma) \right)$$

$$\times \exp\left\{-\frac{1}{2}(\beta_p - b_0)' B_0^{-1}(\beta_p - b_0)\right\} \left(\frac{1}{\sigma}\right)^{01+1} \exp\left\{-\frac{s_{01}}{\sigma}\right\}$$

$$\times c^{l_{02}-1} \exp\{-cs_{02}\} \tag{13}$$

where  $y_{0ijk}$  and  $b_{0ik}$  denote the  $j^{\text{th}}$  measurement on the  $i^{\text{th}}$  subject and vector of unknown subject specific random effects for the  $k^{\text{th}}$  study. On the other hand, sometimes the historical data is not available and in this case we put  $a_0$ , the historical data is excluded altogether, and the prior (8) reduces to the initial prior. Finally, in case the historical data is not available and the current data is independent, the Bayesian quantile estimates by using prior distribution [8] are closed to [9] estimates, and there is code in R (MCMCquantreg) to obtain these estimates.

### Simulation study

In this section, we compare the performance of the proposed Bayesian inference and EM algorithm used by [4] and Bayesian MCMC method proposed by [6]. We use the simulation random intercept model of [4]. Thus, a data set of  $N=100$  subjects in which each subject had 23 scheduled longitudinal measurements was generated from the model.

$$y_{ij} = 1 + 2x_{ij} + b_i + \varepsilon_{ij}, \quad j = 1, \dots, 23, \quad i = 1, \dots, 100, \tag{14}$$

where  $x_{ij} = j, j = 1, 2, \dots, 23$ , is constant throughout the simulation study and  $q = 1$ . We generated the random effects from the standard normal distribution. For the historical data we use the same model with  $N_0 = 25$  subjects,  $n_0 = n_{0i} = 23$ ,  $\beta_{00(p)} = 2$  and  $\beta_{01(p)} = 2.5$ , where  $\beta_{00(p)}$  and  $\beta_{01(p)}$  represent the intercept and the slope for the historical study. Like Geraci and Bottai (2007), we simulated the error term from three different distributions: the standard normal, the Student's  $t$  distribution with three degrees of freedom, and the chi-square with

3 degrees of freedom. We use initial prior  $N(0, 10^6)$  for each regression parameters,  $\Gamma(10^{-3}, 10^{-3})$  for the scale and precision parameter. We estimate the parameters of interest by using our Gibbs sampler and we simulated 1000 replications for each distribution for the error term. We run our Gibbs sampler for 5000 iterations with an initial burn-in of 1000 iterations. We conduct sensitive analyses with respect to three different choices for  $a_0$ . We assumed a range of values for  $a_0$  ( $a_0 = 0, 0.50$  and  $0.95$ ). We compared the estimates of parameters in terms of relative bias averaged and relative efficiency by different algorithm: Gibbs sampler (GS), EM algorithm and Bayesian MCMC. The relative bias averaged over the simulations and the relative efficiency for EM and Bayesian MCMC were calculated by, respectively,

$$\text{bias}(\hat{\beta}_{pm}) = \frac{1}{1000} \sum_{r=1}^{1000} \frac{\hat{\beta}_{pm}^r - \beta_{pm}}{|\beta_{pm}|}$$

$$\text{eff}_{EM}(\hat{\beta}_{pm}) = \frac{S_{EM}^2(\hat{\beta}_{pm})}{S_{GS, a_0=0}^2(\hat{\beta}_{pm})},$$

$$\text{eff}_{MCMC}(\hat{\beta}_{pm}) = \frac{S_{MCMC}^2(\hat{\beta}_{pm})}{S_{GS, a_0=0}^2(\hat{\beta}_{pm})}.$$

where  $\hat{\beta}_{pm}^r, m = 0, 1$  is the estimate quantile for the  $r^{\text{th}}$  replication and  $\beta_{pm}$  is the true value.  $S^2(\hat{\beta}_{pm}) = \frac{1}{1000} \sum_{r=1}^{1000} (\hat{\beta}_{pm}^r - \bar{\beta}_{pm})^2$  and  $\bar{\beta}_{pm} = \frac{1}{1000} \sum_{r=1}^{1000} \hat{\beta}_{pm}^r$ . Furthermore, we conduct sensitive analysis by comparing the Gibbs sampler where  $a_0 = 0$  with the Gibbs sampler where  $a_0 = 0.50$  and  $a_0 = 0.95$

The estimates of the relative bias averaged and relative efficiency for different error term distributions and quantile models are reported in Table 1.

Clearly, the relative biases due to the three approaches are more or less the same, while the GS yields positive biases more than the EM and Bayesian MCMC which have sometimes negative biases. However, in general, the absolute bias in our algorithm is very small compared with the absolute bias by EM and Bayesian MCMC. In addition, the relative efficiency shows that our algorithm is more efficient than EM and Bayesian MCMC. Moreover, we see that as the weight of historical data increases the bias becomes smaller and the efficiency increases too.

Table 2 summaries the posterior means for  $\beta_0$  and  $\beta_1$  where the error is normal at two different quantiles, namely 25% and 50% with respect to five different choices for  $a_0$ . We assumed a range of values for  $a_0$  ( $a_0 = 0, 0.25, 0.50, 0.75$  and  $1$ ). We see that as the weights for the historical data increases the posterior means for  $\beta_{(0.25)}$  and  $\beta_{(0.50)}$  increase and the 95% credible intervals get narrower.

### Analysis age-related macular degeneration data

We use quantile regression methods to analyze the Age-Related Macular Degeneration data (ARMD) previously analyzed by [30]. There are 203 patients 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.

Model	$p$	bias $\beta_0$	Efficiency $\beta_0$	Bias $\beta_1$	Efficiency $\beta_1$
$\varepsilon \sim N(0,1)$					
GS $a_0=0.00$	0.25	0.00027	1.00000	0.00009	1.00000
GS $a_0=0.50$	0.25	0.00013	0.87132	0.00007	0.73221
GS $a_0=0.95$	0.25	0.00011	0.82339	0.00002	0.71291
MCMC	0.25	0.00517	6.55291	0.00053	4.81135
EM	0.25	0.00240	4.93114	0.00000	3.19351
GS $a_0=0.00$	0.50	0.00006	1.00000	0.00012	1.00000
GS $a_0=0.50$	0.50	0.00005	0.96027	0.00010	0.87130
GS $a_0=0.95$	0.50	0.00002	0.90345	0.00013	0.65734
MCMC	0.50	0.00736	4.09938	-0.00024	3.89316
EM	0.50	0.00820	3.99210	0.00000	2.31730
$\varepsilon \sim t_3$					
GS $a_0=0.00$	0.25	0.00008	1.00000	0.00017	1.00000
GS $a_0=0.50$	0.25	0.00003	0.91247	0.00011	0.82553
GS $a_0=0.95$	0.25	0.00005	0.91136	0.00008	0.79735
MCMC	0.25	0.27739	3.56667	-0.00195	2.95862
EM	0.25	-0.13920	3.64298	0.00010	2.76704
GS $a_0=0.00$	0.50	0.00012	1.00000	0.00007	1.00000
GS $a_0=0.50$	0.50	0.00002	0.95731	0.00002	0.91337
GS $a_0=0.95$	0.50	0.00006	0.94618	0.00003	0.86935
MCMC	0.50	0.00391	1.20034	0.00029	1.33942
EM	0.50	0.00430	3.54281	0.00010	2.70529
$\varepsilon \sim \chi_3$					
GS $a_0=0.00$	0.25	0.00000	1.00000	0.00006	1.00000
GS $a_0=0.50$	0.25	0.00005	0.99589	0.00009	0.88395
GS $a_0=0.95$	0.25	0.00008	0.92721	0.00004	0.83557
MCMC	0.25	-0.00771	2.24711	0.00032	2.41503
EM	0.25	0.02680	4.59660	-0.00010	3.99103
GS $a_0=0.00$	0.50	0.00013	1.00000	0.00000	1.00000
GS $a_0=0.50$	0.50	0.00029	0.92774	0.00007	0.73221
GS $a_0=0.95$	0.50	0.00018	0.89447	0.00002	0.63291
MCMC	0.50	0.00835	2.31421	0.00017	2.53098
EM	0.50	0.01920	3.75119	0.00000	2.15328

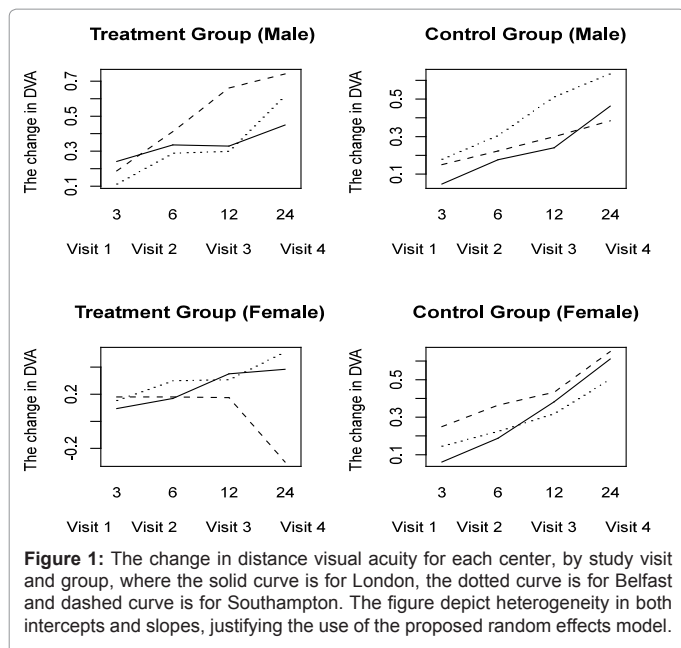
Table 1: Estimated bias and relative efficiency for different error distribution by using Gibbs sampler (GS), Bayesian MCMC, and EM algorithm.

$a_0$	$P$	mean $\beta_0$ (95%CrI)	mean $\beta_1$ (95% CrI)
0.00	0.25	0.3361 (0.2246, 0.4721)	2.0000 (1.9763, 2.0033)
0.25	0.25	0.5895 (0.4631, 0.7007)	2.0010 (1.9968, 2.0137)
0.50	0.25	0.6478 (0.5301, 0.7642)	2.0032 (1.9994, 2.0161)
0.75	0.25	0.8391 (0.7258, 0.9435)	2.0073 (1.9930, 2.0093)
1.00	0.25	0.8660 (0.7568, 0.9544)	2.0120 (1.9939, 2.0101)
0.00	0.50	1.1035 (0.8732, 1.3551)	1.9987 (1.8973, 2.1814)
0.25	0.50	1.5101 (1.2353, 1.6460)	2.0880 (1.9391, 2.2060)
0.50	0.50	1.5657 (1.3439, 1.7164)	2.1290 (1.9399, 2.2061)
0.75	0.50	1.6722 (1.6347, 1.7409)	2.1421 (1.9677, 2.2171)
1.00	0.50	1.7371 (1.6944, 1.7749)	2.1558 (1.9934, 2.2276)

Table 2: Posterior means and 95% credible intervals (CrI) for  $\beta_0$  and  $\beta_1$  for the simulated data.

The response variable, the change in Distance Visual Acuity (DVA), of each patients was measured four times over a two year period, on the 3th, 6th, 12th and 24th months [30]. The data set contains 84 male and 115 female. For the purposes of illustration, we use the female as the historical data, from which we will construct our prior, and we use the data for the male as the current data. Figure 1 displays the change in distance visual acuity for each center, by studying visit and group (treatment group or control group) for the male and female. Since the plots indicated extensive between groups heterogeneity, both at baseline and over visit, we used the linear mixed quantile model (1) to show how the distance visual acuity is affected by five covariates and an intercept term: the actual time of the visits of each patient on the 3th, 6th, 12th and 24th months, age of the patient, centre (Belfast,

Southampton, and London), treatment (treatment group or control group) and the visit  $\times$  group interaction (the readers may refer to [30] for some details about this experiment). We take  $b_i \sim N_2(0, I)$  is a  $2 \times 1$  vector comprising specific intercept and slope parameters and  $z_{ij} = (1, \text{visit}_{ij})'$ . We proposed a similar regression model for the historical study (female). We use initial prior  $N(0, 10^6)$  for each regression parameters and  $\Gamma(10^{-3}, 10^{-3})$  for the scale and precision parameter. Under specific quantile level, we estimate the power parameter that calibrate the power prior distribution in quantile regression to a corresponding hierarchical model. We obtain  $\hat{a}_0 = 0.23, 0.31$  and  $0.21$  for  $p = 0.25, 0.50$  and  $0.75$ , respectively. Table 3 summaries the relative bias for Bayesian MCMC and Gibbs sampler with several values of  $a_0$ , including  $a_0 = \hat{a}_0$ . In addition, Table 3 summaries the relative



efficiency comparison between Bayesian MCMC and our Gibbs sampler.

Again, the relative biases due to both approaches are more or less the same, while the GS yields more positive biases than Bayesian MCMC which sometimes gives negative biases. Most noticeably, when  $p=0.50$  the absolute bias generated by GS for all parameters is less than the absolute bias generated by Bayesian MCMC and the gap in terms of mean squared error between GS and Bayesian MCMC

decreased compared with first and third quartile. When estimating the first quartile, the loss of efficiency of the Bayesian MCMC, with respect to GS, was 118% for the intercept and more than this value for the slopes. It can be argued that the performance of Gibbs sampler is better than Bayesian MCMC may be due to the fact that Bayesian MCMC assumed improper prior distribution (uniform) for all regression and scale parameters. This reason becomes clear with informative prior, we see that as the weight of historical data increases the bias becomes smaller and the efficiency increases too.

### Discussion

In this paper, we have introduced quantile regression for longitudinal data using the asymmetric Laplace distribution from a Bayesian point of view. The methods for eliciting prior distributions can be used either with no prior data or with complete prior data. The methods have been outlined for unknown parameters. We compared our Gibbs sampler with Bayesian MCMC and EM algorithm by using the relative bias averaged and the estimated relative efficiency. We have found that the bias in our algorithm is very small compare with the bias by Bayesian MCMC and EM algorithm. In addition, the relative efficiency showed that our algorithm is more efficiency than Bayesian MCMC and EM algorithm. Finally, we have showed that the behaviour of the power prior is clearly with different weights of the power parameter. Incorporating historical information gathered from similar previous studies into the analysis of current study through a prior distribution provides a natural framework for updating information to yield better results. In particular, the power prior can be used to incorporate the historical data for updating information across studies.

Variable	$p$	Parameter	Bias MCMC	efficiency MCMC	bias GS $a_0=0.0$	Efficiency GS $a_0=0.0$	Bias GS $a_0=\hat{a}$	Efficiency GS $a_0=\hat{a}$	Bias GS $a_0=0.50$	Efficiency GS $a_0=0.50$	Bias GS $a_0=0.95$	Efficiency GS $a_0=0.95$
Intercept	0.25	$\beta_0$	0.0018	2.1771	0.0015	0.9713	0.0011	0.9548	0.0003	0.9263	0.0007	0.7511
Time	0.25	$\beta_1$	0.0015	4.2531	0.0019	0.9910	0.0015	0.9625	0.0013	0.9226	0.0004	0.8162
Age	0.25	$\beta_2$	0.0052	3.6621	0.0008	1.0007	0.0013	0.9162	0.0003	0.8541	0.0000	0.7381
Center	0.25	$\beta_3$	0.0031	4.1954	0.0037	0.9568	0.0025	0.9417	0.0011	0.9771	0.0007	0.8572
Treat	0.25	$\beta_4$	0.0028	2.3718	0.0021	0.9936	0.0017	0.8611	0.0009	0.8819	0.0003	0.8338
Visit x Group	0.25	$\beta_5$	-0.0033	3.6931	0.0013	1.0125	0.0009	0.9352	0.0008	0.8926	0.0011	0.8811
Intercept	0.50	$\beta_0$	0.0082	1.6416	0.0005	1.0173	0.0003	0.9812	0.0001	0.8115	0.0004	0.7739
Time	0.50	$\beta_1$	0.0071	1.4358	0.0012	0.9972	0.0012	0.8714	0.0009	0.8909	0.0000	0.8162
Age	0.50	$\beta_2$	-0.0018	1.8871	0.0007	0.9615	0.0012	0.8759	0.0000	0.7791	0.0005	0.7150
Center	0.50	$\beta_3$	-0.0021	2.2701	0.0009	0.9926	0.0004	0.9157	0.0008	0.8183	0.0001	0.7962
Treat	0.50	$\beta_4$	0.0033	1.3277	0.0013	0.9732	0.0010	0.9456	0.0010	0.8317	0.0006	0.8110
Visit x Group	0.50	$\beta_5$	-0.0041	2.7140	0.0011	1.1704	0.0011	0.9627	0.0007	0.9122	0.0003	0.7155
Intercept	0.75	$\beta_0$	0.0015	3.1147	0.0009	1.0063	0.0011	0.8931	0.0004	0.8102	0.0009	0.7107
Time	0.75	$\beta_1$	0.0027	3.8031	0.0016	1.126	0.0010	0.8825	0.0010	0.8339	0.0006	0.7263
Age	0.75	$\beta_2$	0.0017	2.5731	0.0021	1.0286	0.0023	0.9345	0.0006	0.8513	0.0015	0.6610
Center	0.75	$\beta_3$	-0.0019	5.1703	0.0024	1.1351	0.0019	0.9132	0.0017	0.7959	0.0008	0.7289
Treat	0.75	$\beta_4$	-0.0014	2.3319	0.0009	0.9938	0.0012	0.8627	0.0005	0.8208	0.0002	0.7338
Visit x Group	0.75	$\beta_5$	0.0026	3.4265	0.0014	1.0346	0.0016	0.8236	0.0009	0.7107	0.0005	0.6619

Table 3: Estimated bias and relative efficiency by using Gibbs sampler GS and Bayesian MCMC for the real data.

## References

1. Koenker R (2005) Quantile Regression. New York: Cambridge University Press.
2. Yu K, Lu Z, Stander, et al (2003) Quantile regression: applications and current research area. The Statistician 52: 331-350.
3. Koenker R (2004) Quantile regression for longitudinal data. Journal of Multivariate Analysis 91: 74-89.
4. Geraci M, Bottai M (2007) Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics 8: 140-154.
5. Reich BJ, Bondell H, Wang HJ (2010) Flexible Bayesian quantile regression for independent and clustered data. Biostatistics 2: 337-352.
6. Yuan Y, Yin G (2010) Bayesian quantile regression for longitudinal studies with nonignorable missing data. Biometrics 66: 105-114.
7. Tsonas EG (2003) Bayesian Quantile Inference. Journal of statistical computation and simulation 73: 659-674.
8. Rue H, Held L (2005) Gaussian Markov Random Fields: Theory and Applications. In: Monographs on Statistics and Applied Probability Chapman and Hall, London.
9. Reed C, Yu K (2009) A Partially collapsed Gibbs sampler for Bayesian quantile regression. Technical report, Department of Mathematical Sciences, Brunel University
10. Kozumi H, Kobayashi G (2011) Gibbs Sampling Methods for Bayesian Quantile Regression. Journal of Statistical Computation and Simulation 1-14.
11. Yu K, Moyeed RA (2001) Bayesian quantile regression. Statistics Probability Letters 54: 437- 447.
12. Hanson T, Johnson WO (2002) Modeling regression error with a mixture of Polya trees. Journal of the American Statistical Association 97: 1020-1033.
13. Scaccia L, Green PJ (2003) Bayesian growth curves using normal mixtures with nonparametric weights. Journal of Computational and Graphical Statistics 12: 208-331.
14. Schennach SM (2005) Bayesian exponentially tilted empirical likelihood. Biometrika 92: 31-46.
15. Dunson DB, Taylor JA (2005) Approximate Bayesian inference for quantiles. Nonparametric Statistics 17: 385-400.
16. Yu K, Stander J (2007) Bayesian Analysis of a Tobit Quantile Regression Model. Journal of Econometrics 137: 260-276.
17. Kottas A, Krnjajic M (2009) Bayesian semiparametric modeling in quantile regression. Scandinavian Journal of Statistics. 36: 297-319.
18. Lancaster T, Jun SJ (2010). Bayesian quantile regression methods. Journal of Applied Econometrics. 25: 287-307.
19. Li Q, Xi R, Lin N (2010) Bayesian Regularized Quantile Regression. Bayesian Analysis 5: 1-24.
20. Benoit DF, Poel D (2010) Binary quantile regression: A Bayesian approach based on the asymmetric laplace distribution. J Appl Econ 54: 437-447.
21. Reich BJ, Fuentes M, Dunson DB (2011) Bayesian spatial quantile regression. Accepted. Journal of the American Statistical Association 106: 6-20.
22. Alhamzawi R, Yu K, Vinciotti V, Tucker A (2011) Prior elicitation for mixed quantile regression with an allometric model.
23. Ibrahim JG, Chen MH (2000) Power prior distributions for regression models. Statistical Science 15: 46-60.
24. Chen, MH, Ibrahim JG, Shao QM, Weiss RE (2003) Prior elicitation for model selection and estimation in generalized linear mixed models. Journal of Statistical Planning and Inference 111: 57-76.
25. Neuenschwander B, Branson M, Spiegelhalter D (2009) A note on the power prior. Statistics in Medicine 28: 3562-3566.
26. Neelon B, O'Malley AJ (2010) The use of power prior distributions for incorporating historical data into a Bayesian analysis. Duke University, Durham.
27. Alhamzawi R, Yu K (2011) Power Prior Elicitation in Bayesian Quantile Regression. Journal of Probability and Statistics 1-16.
28. Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Journal of Applied Statistics 20: 25-62.
29. Barndorff-Nielsen OE, Shephard N (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. Journal of the Royal Statistical Society 63: 167-241.
30. Chaili PX (2008) Statistical modelling of age-related macular degeneration. University of Manchester.
31. Chen MH, Ibrahim JG (2006) The relationship between the power prior and hierarchical models. Bayesian Anal 1: 551-574.

### Submit your next manuscript and get advantages of OMICS Group submissions

#### Unique features:

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

#### Special features:

- 100 Open Access Journals
- 10,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, DOAJ, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: [www.omicsonline.org/submission](http://www.omicsonline.org/submission)