Prior elicitation in Bayesian quantile regression for longitudinal data

Rahim Alhamzawi\textsuperscript{1,*}, Keming Yu\textsuperscript{1,**}, and Jianxin Pan\textsuperscript{2,***}

\textsuperscript{1}Department of Mathematics, Brunel University, Uxbridge, UB8 3PH, UK  
\textsuperscript{2}School of Mathematics, University of Manchester, Manchester, M13 9PL, UK

*\textit{email}: Rahim.Al-Hamzawi@brunel.ac.uk

**\textit{email}: Keming.yu@brunel.ac.uk

***\textit{email}: Jianxin.Pan@manchester.ac.uk

\textbf{SUMMARY:} 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.

\textbf{KEY WORDS:} Bayesian quantile regression; Conditional distribution; Gibbs sampling; Longitudinal data; Mixture representation; Random effect.
1. Introduction

Quantile regression models have been widely used for a variety of application (Koenker, 2005; Yu, Lu, and Stander, 2003). It has attracted much interest in recent years because of the flexibility of quantile regression for modeling data with heterogeneous conditional distributions. In addition, a set of quantiles of the response variable (such as the first quartile, the median, the third quartile) may depend on the explanatory variables very differently from the center. Thus, 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 a few literature for quantile regression in longitudinal data and we refer to Koenker (2004), Geraci and Bottai (2007), Reich, Bondell, and Wang (2010) and Yuan and Yin (2010) for an overview.

This paper considers Bayesian quantile regression with random effects. We develop methods for eliciting prior distribution to incorporation of historical data gathered from 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 Geraci and Bottai (2007) and Bayesian MCMC method proposed by Yuan and Yin (2010).

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

\[
y_{ij} = x'_{ij} \beta + z'_{ij} b_i + \epsilon_{ij}, \quad j = 1, ..., n_i, \quad i = 1, ..., N,
\]

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, and \( \epsilon_{ij} \) is the error term. We define the linear mixed quantile functions of the response \( y_{ij} \)

\[
Q_{y_{ij} \mid b}(p \mid x_{ij}, b_i) = x'_{ij} \beta + z'_{ij} b_i, \quad j = 1, ..., n_i, \quad i = 1, ..., N, \tag{1}
\]

where \( Q_{y_{ij} \mid b} \) is the inverse of the cumulative distribution function of \( y_{ij} \) given a vector of unknown subject-specific random effects \( b_i \), \( b_i \sim N_q(0, \Lambda) \), \( \Lambda = \sigma_b^2 \Sigma_b \), \( \Sigma_b \) is the correlation matrix, where the correlation between \( b_{ij} \) and \( b_{ij^*} \) is assumed to be \( \rho |j - j^*| \), with \( 0 < \rho < 1 \), and the \( p \)th quantile of the response \( y_{ij} \) is \( x'_{ij} \beta + z'_{ij} b_i \). We assume the conditional distribution of \( y_{ij} \) given \( b_i \), for \( j = 1, ..., n_i \) and \( i = 1, ..., N \), are independent distribution according to the asymmetric laplace distribution, so that

\[
f(y_{ij} \mid \beta_p, b_i, \sigma) = \frac{p(1-p)}{\sigma} \exp\{-\rho_p(y_{ij} - x'_{ij} \beta - z'_{ij} b_i / \sigma)\}, \tag{2}
\]

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 \( p \)th quantile of this distribution is zero.

Let \( y_i = (y_{i1}, ..., y_{in_i})' \), \( y = (y'_1, ..., y'_N)' \) and \( b = (b'_1, ..., b'_N)' \). The complete-data density of \((y, b)\), for \( i = 1, 2, ..., N \), is then given by

\[
f(y, b \mid \beta, \sigma, \Lambda) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f(y_{ij} \mid \beta, \sigma, b_i) f(b_i \mid \Lambda). \tag{3}
\]

Our interest is on the likelihood function of \( y \) given \( \beta, \sigma \), and \( \Lambda \). Thus, integrate out the
random effects leads to the likelihood

\[ f(y|\beta, \sigma, \Lambda) = \int_{R^q} f(y, b|\beta, \sigma, \Lambda) \, db, \]

\[ = \prod_{i=1}^{N}(\int_{R^q} \prod_{j=1}^{n_i} f(y_{ij}|\beta, b_i, \sigma) f(b_i|\Lambda) \, db_i), \]

\[ = \prod_{i=1}^{N}\left[ \frac{p(1-p)}{\sigma} \right]^{n_i}(2\pi)^{-q/2}|\Lambda|^{-\frac{1}{2}} \times \int_{R^q} \exp\left\{ -\sum_{j=1}^{n_i} \rho_p\left( \frac{y_{ij} - x_{ij}'\beta - z_{ij}'b_i}{\sigma} \right) - \frac{1}{2}b_i'\Lambda^{-1}b_i \right\} \, db_i \]

where \( R^q \) and \( R^q \) denote the \( Nq \) and \( q \)-dimensional Euclidean space, respectively. It is obvious that Equation (5) provides us a degree of shrinkage of the subject-specific regression lines toward the population line.

Model (5) is similar to the \( l_2 \)-penalized check function proposed by Yuan and Yin (2010), which extends the random intercept model proposed by Geraci and Bottai (2007) to a very general case.

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 (PCG) 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.

2. Posterior Inference

2.1 Mixture representation

This section introduce 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.

As provided in Reed and Yu (2009) and Kozumi and Kobayashi (2009) that any variable has 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 \( \varepsilon \) is a random variable follows the asymmetric Laplace distribution with density

\[
f_p(\varepsilon) = \frac{p(1-p)}{\sigma} \exp\{-\rho_p\left(\frac{\varepsilon}{\sigma}\right)\},
\]

where \( \xi \) is a standard normal random variable, and \( t \) is a standard exponential random variable. Then we can represent \( \varepsilon \) as a location mixture of normal given by

\[
\varepsilon = \frac{1-2p}{p(1-p)} \sigma t + \sqrt{\frac{2t}{p(1-p)}} \sigma \xi.
\]

(6)

Now, if the error term \( \varepsilon_{ij} \) is a random variable follows the asymmetric laplace distribution with density

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

where \( \xi_{ij} \) is a standard normal random variable, and \( t_{ij} \) is a standard exponential random variable. Then we can represent \( \varepsilon_{ij} \) as a location mixture of normal given by

\[
\varepsilon_{ij} = \frac{1-2p}{p(1-p)} \sigma t_{ij} + \sqrt{\frac{2t_{ij}}{p(1-p)}} \sigma \xi_{ij}.
\]

(7)

Based on this result, we can equivalently rewrite the response \( y_{ij} \) as

\[
y_{ij} = x'_{ij} \beta + z'_{ij} b_i + \theta \sigma t_{ij} + \phi \sigma \sqrt{t_{ij}} \xi_{ij},
\]

(8)

where

\[
\theta = \frac{1-2p}{p(1-p)} \quad \text{and} \quad \phi^2 = \frac{2}{p(1-p)}.
\]

We assume that \( y_{ij} \) conditionally on \( b_i \) and \( t_{ij} \), for \( j = 1, \ldots, n_i \) and \( i = 1, \ldots, N \), are independently distributed according to the normal with mean \( x'_{ij} \beta + z'_{ij} b_i + \sigma \theta t_{ij} \) and variance
\[ \sigma^2 \phi^2 t_{ij}. \] From now on, it is more convenient with gibbs sampler to work with \( v_{ij} = \sigma t_{ij} \) to avoid appears the scale parameter \( \sigma \) in the conditional mean of \( y_{ij} \). Thus, the density of the response \( y_{ij} \) is given by

\[
f(y_{ij} | \beta, b_i, \sigma, v_{ij}) \propto (\sigma v_{ij})^{-\frac{1}{2}} \exp\left\{ -\frac{(y_{ij} - x_{ij}'\beta - z_{ij}'b_i - \theta v_{ij})^2}{2\sigma \phi^2 v_{ij}} \right\}
\]  

If we let \( v_i = (v_{i1}, \ldots, v_{im})' \) and \( v = (v_1, \ldots, v_N)' \) then the joint density of \( y \) is given by

\[
f(y | \beta, b, \sigma, v) = \prod_{i=1}^{N} f(y_i | \beta, b_i, \sigma, v_i),
\]

\[
f(y_i | \beta, b_i, \sigma, v_i) \propto \sigma^{-\frac{m}{2}} \left( \prod_{j=1}^{m} v_{ij}^{-\frac{1}{2}} \right) \times \exp\left\{ -\sum_{j=1}^{m} \frac{(y_{ij} - x_{ij}'\beta - z_{ij}'b_i - \theta v_{ij})^2}{2\sigma \phi^2 v_{ij}} \right\}.
\]

2.2 Power prior distributions and Gibbs sampler

In this section, we address a quantile dependent prior in Bayesian quantile regression for longitudinal data. Since Yu and Moyeed (2001) Bayesian inference quantile regression has attracted a lot of attention in literature including Hanson and Johnson (2002), Tsionas (2003), Scaccia and Green (2003), Schennach (2005), Dunson and Taylor (2005), Geraci and Bottai (2007), Yu and Stander (2007), Kottas and Krnjajic (2009), Lancaster and Jun (2010), Reed and Yu (2009), Kozumi and Kobayashi (2009), and Yuan and Yin (2010). These Bayesian inference models include Bayesian parametric, Bayesian semiparametric as well as Bayesian nonparametric models. However, almost all these models set priors independent of the values of quantiles, or the prior is the same for modelling different quantiles. This approach may result in 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. It is therefore more reasonable
to set different priors for different quantiles. In this paper, we address a quantile dependent prior for quantile mixed model. Our idea is to set priors based on historical data. Although, one can use improper prior in Bayesian quantile regression, the inference on current data could be more reliable and sensible if there exist historical data gathered from similar previous studies. There are several methods to incorporate the historical data into the analysis of a current study. One of these methods is the power prior proposed by Ibrahim and Chen (2000) which is constructed by raising the likelihood function of the historical data to a power parameter between 0 and 1.

Suppose there exists one historical data from previous study. Let $y_{0ij}$ be the $j$th measurement ($j = 1, \ldots, n_{0i}$) for the $i$th subject ($i = 1, \ldots, N_0$), $x'_{0ij}$ 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 $y_0 = (y'_{01}, \ldots, y'_{0N_0})'$, $y_{0i} = (y_{0i1}, \ldots, y_{0in_{0i}})'$, $X_0 = (X'_{01}, \ldots, X'_{0N_0})'$, and $Z_0 = \text{diag}(Z_{01}, \ldots, Z_{0N_0})$. For the previous study, we assume $v_0 = (v'_{01}, \ldots, v'_{0N_0})$, $v_{0i} = (v_{0i1}, \ldots, v_{0in_{0i}})$ and each $v_{0ij} \sim \text{Exp}(\sigma)$. This can be viewed as the prior distribution on $v_{0ij}$.

For quantile mixed model, we follow Chen et al. (2003) and we propose a prior distribution for $\beta$ taking the form

$$
\pi(\beta_p|D_0, \Lambda, \sigma, a_0, v_0) 
\propto \prod_{i=1}^{N_0} \left( \int_{Rq} \prod_{j=1}^{n_{0i}} [f(y_{0ij}|\beta, b_{0i}, \sigma, v_{0ij})]^{a_0} f(b_{0i}|\Lambda) db_{0i} \right) \pi_0(\beta)
$$

(10)

where $\prod_{j=1}^{n_{0i}} f(y_{0ij}|\beta, b_{0i}, \sigma, v_{0ij}) = f(y_0|\beta, b_{0i}, \sigma, v_{0i})$ be the density for the $i$th subject given a vector of unknown subject specific effects $b_{0i}$, $a_0$ is a fixed parameter, and $0 \leq a_0 \leq 1$. The power parameter $a_0$ represents how much data from the previous study to be used in the current study. There are two special cases for $a_0$, the first case $a_0 = 0$ results no incorporation of the data from the previous study relative to the current study. The second case $a_0 = 1$
Prior elicitation in Bayesian quantile regression for longitudinal data

results full incorporation of the data from the previous study relative to the current study. Therefore, $a_0$ 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 or where there is heterogeneity between two studies (Ibrahim and Chen, 2000). The prior specification is completed by specifying priors for $\sigma, \sigma^2, \rho$ and $\beta_p$. We specify an inverse gamma ($\Gamma$) prior with parameter $(l_{01}, s_{01})$ for $\sigma$, an inverse gamma ($\Gamma$) prior with parameter $(l_{02}, s_{02})$ for $\sigma^2_b$, a scaled beta prior for $\rho$ with parameter $(l_{03}, s_{03})$, and multivariate normal prior with parameter $(\mu_0, B_0)$ for $\beta$. Thus, we propose a joint prior distribution taking the form

$$
\pi(\beta, \sigma, \sigma^2_b, \rho, v_0 | D_0, a_0)
\propto \prod_{i=1}^{N_0} \left( \int_{R^n} \prod_{j=1}^{n_0} [f(y_{0ij}|\beta, b_{0i}, \sigma, v_{0ij})]^{a_0} f(b_{0i}|\Lambda)db_{0i}\pi(v_{0ij}|\sigma) \right)
\times \exp \left\{ -\frac{1}{2} (\beta - \mu_0)'B_0^{-1}(\beta - \mu_0) \right\}^{l_{01}} + 1 \exp \left\{ -\frac{s_{01}}{2\sigma} \right\}
\times \left( \frac{1}{\sigma^2_b} \right)^{l_{02}} + 1 \exp \left\{ -\frac{s_{02}}{2\sigma^2_b} \right\} \times (1 + \rho)^{l_{03} - 1} (1 - \rho)^{s_{03} - 1}
$$

(11)

We see that (11) will not have closed form in general because it depends on the prior that we choose. The joint posterior distribution of $\beta, \sigma, \sigma^2_b$ and $\rho$ is given by

$$
f(\beta, \sigma, \sigma^2_b, \rho, v, v_0 | D, D_0, a_0)
\propto \prod_{i=1}^{N} \int_{R^n} \prod_{j=1}^{n_i} [f(y_{ij}|\beta, b_i, \sigma, v_{ij})] f(b_i|\Lambda)db_i \pi(v_{ij}|\sigma)
\times \pi(\beta, \sigma, \sigma^2_b, \rho, v_0 | D_0, a_0),
$$

(12)

where $D = (N, y, X, Z)$ represent the current study and $v = (v_1', ..., v_N')$, $v_i = (v_{i1}, ..., v_{in_i})$, and each $v_{ij} \sim \text{Exp}(\sigma)$. Gibbs sampler by Geman and Geman (1984) 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 unknowns parameters are needed and each of
these distributions can be obtained by regarding all other parameters in (12) as known. For example, a little algebra shows that the full conditional distribution of $\beta$ is given by

$$\beta|D, D_0, \sigma, b, b_0, v, v_0, a_0 \sim N_{(k+1)}(\hat{\beta}, \hat{B}),$$

where

$$\hat{\beta} = \hat{B}\{\sum_{i=1}^{N} \sum_{j=1}^{n_i} x_{ij}(y_{ij} - z_{ij}' b_i - \theta v_{ij}) \over \sigma \phi^2 v_{ij} \\
+ a_0 \sum_{i=1}^{N_0} \sum_{j=1}^{n_{0i}} x_{0ij}(y_{0ij} - z_{0ij}' b_{0i} - \theta v_{0ij}) \over \sigma \phi^2 v_{0ij} \} + B_0^{-1} \mu_0,$$

$$\hat{B}^{-1} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} x_{ij} x_{ij}' \over \sigma \phi^2 v_{ij} + a_0 \sum_{i=1}^{N_0} \sum_{j=1}^{n_{0i}} x_{0ij} x_{0ij}' \over \sigma \phi^2 v_{0ij} + B_0^{-1}.$$

2.3 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 generalized the the power prior (11) to multiple historical data we assume there are $M$ historical studies denoted by $D_0 = (D_{01}, ..., D_{0M})$ where $D_{0k} = (N_{0k}, y_{0k}, X_{0k}, Z_{0k})$ represent the historical data based on the $k$th study, $k = 1, ..., M$. We define $a_{0k}$ and $v_{0ijk}$ to be the fixed power parameter and an exponential variable, respectively, for the $k$th study, $a_0 = (a_{01}, ..., a_{0M})$. Hence the prior can be generalized as

$$\pi(\beta, \sigma, \sigma_b^2, \rho, v_0|D_0, a_0) \propto \prod_{k=1}^{M} \prod_{i=1}^{N_{0k}} \prod_{j=1}^{n_{0ik}} [f(y_{0ijk}|\beta, b_{0ik}, \sigma, v_{0ijk}) | a_{0k} f(b_{0ik}|\Lambda) db_{0ik} \pi(v_{0ijk}|\sigma)]$$

$$\times \exp\{-{1 \over 2} (\beta - b_0)' B_0^{-1} (\beta - b_0) \}$$

$$\times (1 \over \sigma)^{q_{01} + 1} \exp\{-s_{01} \over 2\sigma \} (1 \over \sigma_b)^{q_{02} + 1} \exp\{-s_{02} \over 2\sigma_b \}$$

$$\times (1 + \rho)^{l_{03} - 1} (1 - \rho)^{s_{03} - 1}.$$
where $y_{0ijk}$ and $b_{0ik}$ denote to the $j$th measurement on the $i$th subject and vector of unknown subject specific random effects for the $k$th study. On the other hand, sometimes the historical data are not available and in this case we put $a_0 = 0$, the historical data are excluded altogether, and the prior (11) reduces to the initial prior. Finally, in case the historical data are not available and the current data are independent, the Bayesian quantile estimates by using prior distribution (11) are closed to Reed and Yu (2009) estimates, and there is code in R (MCMCquantreg) to obtained these estimates.

3. Simulation Study

In this section, we compare the performance of the proposed full Bayesian inference and EM algorithm used by Geraci and Bottai (2007) and Bayesian MCMC method proposed by Yuan and Yin (2010). We used the simulation random intercept model of Geraci and Bottai (2007). Thus, a data set of $N = 100$ subjects in which each subject had 23 scheduled longitudinal measurements was generated from the model.

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

(13)

where $x_{ij} = j, j = 1, 2, \ldots, 23$, 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, $n_0 = n_{0i} = 23, \beta_{00} = 2$ and $\beta_{01} = 2.5$. 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 and $\Gamma(10^{-3}, 10^{-3})$ for the scale parameter. In addition, we take $l_{02} = 7, s_{02} = 0.01, l_{03} = 1$ and $s_{03} = 1$. 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 our Gibbs sampler (GS) with EM algorithm and Bayesian MCMC by using the relative bias averaged over the simulations which is calculated by

$$\hat{\text{bias}}(\hat{\beta}_m) = \frac{1}{1000} \sum_{r=1}^{1000} \frac{\hat{\beta}_m^r - \beta_m}{|\beta_m|},$$

where $\hat{\beta}_m^r, m = 0, 1$ is the estimate quantile for the $r$th replication and $\beta_m$ is the true value.

We used also the estimated relative efficiency to compare our Gibbs sampler where $a_0 = 0$ (no prior data) with EM algorithm and Bayesian MCMC which are calculated by, respectively,

$$\hat{\text{eff}}_{EM}(\hat{\beta}_m) = \frac{S^2_{EM}(\hat{\beta}_m)}{S^2_{GSa_0=0}(\hat{\beta}_m)},$$

$$\hat{\text{eff}}_{MCMC}(\hat{\beta}_m) = \frac{S^2_{MCMC}(\hat{\beta}_m)}{S^2_{GSa_0=0}(\hat{\beta}_m)}.$$  

Furthermore, we conduct sensitive analysis by comparing the Gibbs sampler where $a_0 = 0$ with the Gibbs sampler where $a_0 = 0.50$ and $a_0 = 0.95$ which are calculated, respectively, by

$$\hat{\text{eff}}_{GSa_0=0.50}(\hat{\beta}_m) = \frac{S^2_{GSa_0=0.50}(\hat{\beta}_m)}{S^2_{GSa_0=0}(\hat{\beta}_m)},$$

and

$$\hat{\text{eff}}_{GSa_0=0.95}(\hat{\beta}_m) = \frac{S^2_{GSa_0=0.95}(\hat{\beta}_m)}{S^2_{GSa_0=0}(\hat{\beta}_m)},$$

where

$$S^2(\hat{\beta}_m) = \frac{1}{1000} \sum_{r=1}^{1000} (\hat{\beta}_m^r - \bar{\beta}_m)^2 \quad \text{and} \quad \bar{\beta}_m = \frac{1}{1000} \sum_{r=1}^{1000} \hat{\beta}_m^r.$$

The estimates of the relative bias averaged and relative efficiency for different error term distributions and quantile models are reported in Table 1, along with results of the EM algorithm showed by Geraci and Bottai (2007).

[Table 1 about here.]

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.

Figures 1, 2 and 3 compare the posterior densities of $\beta_0$ and $\beta_1$ estimated by our algorithm for different error distributions: the standard normal, the Students t distribution with 3 degrees of freedom, and the chi-square with 3 degrees of freedom respectively. For each quantile ($p=0.25$ and 0.50), all three posterior densities are almost similar to each other for each $\beta_0$ and $\beta_1$. All posterior densities show that the credible interval (CI) are reasonably narrow and robust in terms of different error distributions.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

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. This is a comforting feature because it is consistent with what we expect from the data. This implies that the posterior mean for the parameters of interest is quite robust with the different weights for power parameter. In addition, we see that as the weights for the historical data increases the 95% credible intervals get narrower. This is also a comforting feature, since it implies that the credible intervals are fairly robust with respect to the increase of the historical data.
4. Analysis Age-Related Macular Degeneration data

We use quantile regression methods to analyze the Age-Related Macular Degeneration data (ARMD) previously analyzed by Chaili (2008). There are 203 patients were randomly selected from three cities 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). In Figure 4, a scatter plot of the change in distance visual acuity against the time in years from entry into the study. It is obvious from Figure 4 that the measurements of the response against the time appear to form four clusters. The change in distance visual acuity for both treated and untreated patients for each visit is shown graphically in Figure 5. It can be seen that the box-plots are very similar with the change in Distance Visual Acuity of the treated patients being slightly lower at each visit than that of untreated patients. In addition, three possible outliers are identified in each group.

[Figure 4 about here.]

[Figure 5 about here.]

We considered 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, sex of the patient, centre
Prior elicitation in Bayesian quantile regression for longitudinal data

(Belfast, Southampton, and London), and treatment (treatment group or control group) (the readers may refer to Chaili (2008) for some details about this experiment). We generated the random effects from the multivariate standard normal distribution. Table 3 summaries the relative bias for our Gibbs sampler and Bayesian MCMC for $\beta$ corresponding to the quantiles $p \in \{0.25, 0.50, 0.75\}$. In addition, Table 3 summaries the relative efficiency comparison between Bayesian MCMC and our Gibbs sampler.

Table 3 about here.

Clearly, 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. However, in general, the bias in our algorithm is very small compared with the bias by Bayesian MCMC. Most noticeably, when $p = 0.50$ the absolute bias generated by our algorithm for all parameters is less than the absolute bias generated by Bayesian MCMC. In addition, the relative efficiency shows that our algorithm is more efficient than Bayesian MCMC. In Figure 6 we present visual summary of Bayesian quantile regression results. Each panel represents one of the six coefficients in the Bayesian quantile regression model. The solid line represents 91 point estimates of the coefficient for $p \in \{0.05, 0.06, ..., 0.94, 0.95\}$. The shaded grey area depicts a 95% credible interval.

Figure 6 about here.

5. 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 behavior of the power prior is clearly and quite robust 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.

References


area. The Statistician. 52, 331-350.

Yu, K. and Stander, J. (2007), Bayesian Analysis of a Tobit Quantile Regression Model. 

Figure 1. plots of posterior densities for $\beta_0(p)$ and $\beta_1(p)$ where $a_0 = 0$ and the error term has a standard normal distribution.
Figure 2. plots of posterior densities for $\beta_0(p)$ and $\beta_1(p)$ where $a_0 = 0$ and the error term has the Students t distribution with 3 degrees of freedom.
Figure 3. plots of posterior densities for $\beta_0(p)$ and $\beta_1(p)$ where $a_0 = 0$ and the error term has the chi-square distribution with 3 degrees of freedom.
Figure 4. scatter plot of the change in distance visual acuity against the time in years from entry into the study.
Figure 5. Boxplots of the change in the distance visual acuity for each visit of the treated and untreated patients.
Figure 6. Quantile regression estimates for the change in distance visual acuity and 95% credible intervals (gray bands) for $p \in \{0.05, 0.06, \ldots, 0.94, 0.95\}$.
Table 1

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

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon \sim N(0, 1)$</th>
<th>$\epsilon \sim t_3$</th>
<th>$\epsilon \sim \chi_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$ bias</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$\beta_0$ efficiency</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>$\beta_1$ bias</td>
<td>0.00009</td>
<td>0.00010</td>
<td>0.00000</td>
</tr>
<tr>
<td>$\beta_1$ efficiency</td>
<td>1.00000</td>
<td>0.87130</td>
<td>0.73221</td>
</tr>
<tr>
<td>$\epsilon \sim N(0, 1)$</td>
<td>GS $a_0=0.00$</td>
<td>GS $a_0=0.50$</td>
<td>GS $a_0=0.95$</td>
</tr>
<tr>
<td>GS $a_0=0.00$</td>
<td>0.25</td>
<td>0.00013</td>
<td>0.00007</td>
</tr>
<tr>
<td>GS $a_0=0.50$</td>
<td>0.25</td>
<td>0.00011</td>
<td>0.00002</td>
</tr>
<tr>
<td>GS $a_0=0.95$</td>
<td>0.25</td>
<td>0.00517</td>
<td>0.00053</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.25</td>
<td>0.00240</td>
<td>4.93114</td>
</tr>
<tr>
<td>EM</td>
<td>0.25</td>
<td>0.00006</td>
<td>1.00000</td>
</tr>
<tr>
<td>$\epsilon \sim t_3$</td>
<td>GS $a_0=0.00$</td>
<td>GS $a_0=0.50$</td>
<td>GS $a_0=0.95$</td>
</tr>
<tr>
<td>GS $a_0=0.00$</td>
<td>0.50</td>
<td>0.00005</td>
<td>0.96027</td>
</tr>
<tr>
<td>GS $a_0=0.50$</td>
<td>0.50</td>
<td>0.00002</td>
<td>0.90345</td>
</tr>
<tr>
<td>GS $a_0=0.95$</td>
<td>0.50</td>
<td>0.00736</td>
<td>4.09938</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.50</td>
<td>0.00820</td>
<td>3.99210</td>
</tr>
<tr>
<td>EM</td>
<td>0.50</td>
<td>0.27739</td>
<td>3.56667</td>
</tr>
<tr>
<td>$\epsilon \sim \chi_3$</td>
<td>GS $a_0=0.00$</td>
<td>GS $a_0=0.50$</td>
<td>GS $a_0=0.95$</td>
</tr>
<tr>
<td>GS $a_0=0.00$</td>
<td>0.50</td>
<td>0.00012</td>
<td>0.95731</td>
</tr>
<tr>
<td>GS $a_0=0.50$</td>
<td>0.50</td>
<td>0.00006</td>
<td>0.94618</td>
</tr>
<tr>
<td>GS $a_0=0.95$</td>
<td>0.50</td>
<td>0.00391</td>
<td>1.20034</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.50</td>
<td>0.00430</td>
<td>3.54281</td>
</tr>
<tr>
<td>EM</td>
<td>0.50</td>
<td>-0.13920</td>
<td>3.64298</td>
</tr>
<tr>
<td>$\epsilon \sim \chi_3$</td>
<td>GS $a_0=0.00$</td>
<td>GS $a_0=0.50$</td>
<td>GS $a_0=0.95$</td>
</tr>
<tr>
<td>GS $a_0=0.00$</td>
<td>0.50</td>
<td>0.00000</td>
<td>0.99589</td>
</tr>
<tr>
<td>GS $a_0=0.50$</td>
<td>0.50</td>
<td>0.00005</td>
<td>0.92721</td>
</tr>
<tr>
<td>GS $a_0=0.95$</td>
<td>0.50</td>
<td>-0.00771</td>
<td>2.47111</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.50</td>
<td>0.02680</td>
<td>4.59660</td>
</tr>
<tr>
<td>EM</td>
<td>0.50</td>
<td>0.09013</td>
<td>1.00000</td>
</tr>
<tr>
<td>$\epsilon \sim \chi_3$</td>
<td>GS $a_0=0.00$</td>
<td>GS $a_0=0.50$</td>
<td>GS $a_0=0.95$</td>
</tr>
<tr>
<td>GS $a_0=0.00$</td>
<td>0.50</td>
<td>0.00002</td>
<td>0.92774</td>
</tr>
<tr>
<td>GS $a_0=0.50$</td>
<td>0.50</td>
<td>0.00018</td>
<td>0.89447</td>
</tr>
<tr>
<td>GS $a_0=0.95$</td>
<td>0.50</td>
<td>0.00835</td>
<td>2.31421</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.50</td>
<td>0.01920</td>
<td>3.75119</td>
</tr>
<tr>
<td>EM</td>
<td>0.50</td>
<td>0.00000</td>
<td>2.15328</td>
</tr>
</tbody>
</table>
Table 2
Posterior means and 95% credible intervals (CrI) for $\beta_0$ and $\beta_1$ for the simulated data.

<table>
<thead>
<tr>
<th>$a_0$</th>
<th>$p$</th>
<th>mean $\beta_0$ (95% CrI)</th>
<th>mean $\beta_1$ (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.25</td>
<td>0.3361 (0.2246, 0.4721)</td>
<td>2.0000 (1.9763, 2.0033)</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>0.5895 (0.4631, 0.7007)</td>
<td>2.0010 (1.9968, 2.0137)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.25</td>
<td>0.6478 (0.5301, 0.7642)</td>
<td>2.0032 (1.9994, 2.0161)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.25</td>
<td>0.8391 (0.7258, 0.9435)</td>
<td>2.0073 (1.9930, 2.0093)</td>
</tr>
<tr>
<td>1.00</td>
<td>0.25</td>
<td>0.8660 (0.7568, 0.9544)</td>
<td>2.0120 (1.9939, 2.0101)</td>
</tr>
<tr>
<td>0.00</td>
<td>0.50</td>
<td>1.1035 (0.8732, 1.3551)</td>
<td>1.9987 (1.8973, 2.1814)</td>
</tr>
<tr>
<td>0.25</td>
<td>0.50</td>
<td>1.5101 (1.2353, 1.6460)</td>
<td>2.0880 (1.9391, 2.2060)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>1.5657 (1.3439, 1.7164)</td>
<td>2.1290 (1.9399, 2.2061)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.50</td>
<td>1.6722 (1.6347, 1.7409)</td>
<td>2.1421 (1.9677, 2.2171)</td>
</tr>
<tr>
<td>1.00</td>
<td>0.50</td>
<td>1.7371 (1.6944, 1.7749)</td>
<td>2.1558 (1.9934, 2.2276)</td>
</tr>
</tbody>
</table>
### Table 3

*Estimated bias and relative efficiency by using Gibbs sampler (GS) and Bayesian MCMC.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>p</th>
<th>Parameter</th>
<th>bias (GS)</th>
<th>bias (MCMC)</th>
<th>efficiency (MCMC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.25</td>
<td>$\beta_0$</td>
<td>0.0031</td>
<td>0.0022</td>
<td>2.1089</td>
</tr>
<tr>
<td>Time</td>
<td>0.25</td>
<td>$\beta_1$</td>
<td>0.0017</td>
<td>0.0012</td>
<td>4.7826</td>
</tr>
<tr>
<td>Age</td>
<td>0.25</td>
<td>$\beta_2$</td>
<td>0.0011</td>
<td>0.0041</td>
<td>3.9628</td>
</tr>
<tr>
<td>Sex</td>
<td>0.25</td>
<td>$\beta_3$</td>
<td>0.0047</td>
<td>0.0035</td>
<td>4.8733</td>
</tr>
<tr>
<td>Centre</td>
<td>0.25</td>
<td>$\beta_4$</td>
<td>0.0019</td>
<td>0.0024</td>
<td>2.6667</td>
</tr>
<tr>
<td>Treat</td>
<td>0.25</td>
<td>$\beta_5$</td>
<td>0.0016</td>
<td>-0.0043</td>
<td>3.9032</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.50</td>
<td>$\beta_0$</td>
<td>0.0004</td>
<td>0.0091</td>
<td>1.5680</td>
</tr>
<tr>
<td>Time</td>
<td>0.50</td>
<td>$\beta_1$</td>
<td>0.0001</td>
<td>0.0053</td>
<td>2.4902</td>
</tr>
<tr>
<td>Age</td>
<td>0.50</td>
<td>$\beta_2$</td>
<td>0.0002</td>
<td>-0.0011</td>
<td>1.3395</td>
</tr>
<tr>
<td>Sex</td>
<td>0.50</td>
<td>$\beta_3$</td>
<td>0.0011</td>
<td>0.0016</td>
<td>4.1753</td>
</tr>
<tr>
<td>Centre</td>
<td>0.50</td>
<td>$\beta_4$</td>
<td>0.0002</td>
<td>-0.0027</td>
<td>1.5439</td>
</tr>
<tr>
<td>Treat</td>
<td>0.50</td>
<td>$\beta_5$</td>
<td>0.0013</td>
<td>-0.0039</td>
<td>2.7641</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.75</td>
<td>$\beta_0$</td>
<td>0.0021</td>
<td>0.0015</td>
<td>2.5392</td>
</tr>
<tr>
<td>Time</td>
<td>0.75</td>
<td>$\beta_1$</td>
<td>0.0025</td>
<td>0.0038</td>
<td>3.4053</td>
</tr>
<tr>
<td>Age</td>
<td>0.75</td>
<td>$\beta_2$</td>
<td>0.0011</td>
<td>0.0019</td>
<td>1.0003</td>
</tr>
<tr>
<td>Sex</td>
<td>0.75</td>
<td>$\beta_3$</td>
<td>0.0023</td>
<td>0.0021</td>
<td>5.2953</td>
</tr>
<tr>
<td>Centre</td>
<td>0.75</td>
<td>$\beta_4$</td>
<td>0.0008</td>
<td>-0.0013</td>
<td>2.1193</td>
</tr>
<tr>
<td>Treat</td>
<td>0.75</td>
<td>$\beta_5$</td>
<td>0.0033</td>
<td>0.0028</td>
<td>2.3395</td>
</tr>
</tbody>
</table>