

# SEMIPARAMETRIC BAYESIAN LINEAR QUANTILE MIXED MODEL AND ITS APPLICATION

by

Ming Yang, MS

APPROVED:

---

Stacia DeSantis, PHD

---

Sheng Luo, PHD

---

David R. Lairson, PHD

---

Xiaoming Liu, PHD

Copyright  
by  
Ming Yang, MS  
2014

## DEDICATION

xxx

xxx

# SEMIPARAMETRIC BAYESIAN LINEAR MIXED MODEL AND ITS APPLICATION

by

Ming Yang, MS

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS

SCHOOL OF PUBLIC HEALTH

Houston, Texas

month, 2014

## ACKNOWLEDGEMENTS

text text text

# SEMIPARAMETRIC BAYESIAN LINEAR MIXED MODEL AND ITS APPLICATION

Ming Yang, MS  
The University of Texas  
School of Public Health, 2014

Thesis Chair, Stacia DeSantis, PhD

text text text

# Contents

<b>1</b>	<b>Background</b>	<b>9</b>
<b>2</b>	<b>Specific Aims and Hypotheses</b>	<b>12</b>
<b>3</b>	<b>Data</b>	<b>12</b>
<b>4</b>	<b>Methods</b>	<b>13</b>
4.1	Nonparametric Bayesian . . . . .	13
4.1.1	Dirichlet Process and Dirichlet Process Mixture Models . . . . .	13
4.1.2	Centered Dirichlet Process Mixture . . . . .	15
4.1.3	Truncation . . . . .	15
4.2	Linear Quantile Mixed Models . . . . .	16
4.2.1	Quantile Regression and Asymmetric Laplace Distribution . . . . .	16
4.2.2	Bayesian Linear Quantile Mixed Model . . . . .	18
<b>5</b>	<b>Plan for Simulation Studies</b>	<b>19</b>
5.1	Bayesian Quantile Regression Method . . . . .	19
5.2	Dirichlet Process Mixture Prior for Random Effects . . . . .	21
5.2.1	Estimation of Fixed Effects . . . . .	22
5.2.2	Prediction of Random Effects . . . . .	22
<b>6</b>	<b>References</b>	<b>25</b>

## List of Tables

1	Simulation study with 500 datasets, $N(0, 1)$ for error term . . . . .	21
2	Sample summary table: estimation of fixed effects when $u_i \sim$ mixture Normal	22
3	Sample summary table: statistics of prediction of random effects (RAD is based on 30 posterior samples) . . . . .	25



# List of Figures

1	Predicted densities of random effects that follow beta distribution . . . . .	23
2	Predicted densities of random effects that follow mixture of normals . . . . .	23
3	Predicted densities of random effects that follow $N(0.6, 0.5^2)$ . . . . .	23

# 1 Background

Mixed effects model methodology, first introduced by R.A. Fisher [6], is a statistical tool that is used across a wide variety of disciplines including biostatistical contexts. Mixed effects models are especially popular in researches involving repeated measurements or observations from multilevel (or hierarchical) structure where the correlation between observations is not negligible. In the context of Bayesian data analysis for linear mixed models (LMM) or generalized linear models (GLMM), normal (or multivariate normal) distribution is a conventional prior for the random effects. However, the choice of normal prior is simply for mathematical convenience but without any theoretical justification. Several studies have revealed that estimation for the fixed effects is relatively robust to the prior choice of the random effects [26] [2] [12], which suggests that the estimates for the fixed effects are asymptotically valid even when the normal prior is not appropriate. However, in practice, inference on random effects may be also of interest, for example, the block random effects in randomized block designs. Accurate estimation of the random effects is especially important when we are trying to make predictions for some specific subject in the future. Since the estimate of the random effects is very sensitive to the choice of the prior and different priors will lead to totally different posterior distributions. In such cases, simply imposing normal prior to random effects is criticized for causing potential consequences, such as the results are sensitive to outliers [3], less efficiency for non-normal data [24], and the validity of the normality assumption is hard to assess since they are latent variables. In addition, we always would like to express our prior belief more accurately by choosing other distributions rather than blindly using normal in modeling the random effect.

Some authors have proposed to use more flexible parametric alternatives than normal such as the heavier-tailed  $t$  distribution and its skew extension [20]. But these are still restrictive because they don't allow multi-modality in the data [3]. In handling all these problems at once, a Bayesian nonparametric method provides an ideal solution by using the Dirichlet

process (DP) to replace the normal distribution as a much more flexible prior for random effects. In general, by using DP prior we model the distribution of the random effects with infinite mixture of distribution functions so that we capture the uncertainty and irregularity (outliers, skewness and clustering, etc.) in the data. More details about DP will be introduced in Section 4.1.2. The performance of DP prior and normal prior will be compared using simulation study.

Quantile regression models are becoming more and more popular in the statistical community in recent years. Compared with the ubiquitous mean regression (a.n.a. linear regression) models, quantile regression models provide a much more comprehensive and focused insight into the association between the variables by studying the conditional quantiles functions of the outcome, which may not be observed by looking only at conditional mean models [13]. In quantile regression, the regression coefficients ( $\beta$ ) are functions of the quantile ( $\tau$ ), and their estimates vary according to different quantiles. Thus quantile regression provides a way to studying the heterogeneity of the outcome that is associated with the covariates [13]. Moreover, quantile regression is more robust against outliers in the outcome compared with the mean regression, which is an immediate extension from the property of quantiles.

For a linear quantile regression model, [14] introduced a method in estimating the conditional quantiles. As an introductory material, Koenker [15] briefly covers the fundamentals of quantile regression, parameter estimation techniques, inference, asymptotic theory, etc., his book [13] provides more comprehensive and deeper introduction on quantile regression related topics. Yu and Moyeed [33] introduced the idea of Bayesian quantile regression by modeling the error term using asymmetric Laplace distribution (ALD) followed the idea proposed in [14]. [19] developed a Gibbs sampling algorithm for Bayesian quantile regression models, in which they used a location-scale mixture representation of the ALD. Many works have been done to extend the quantile regression method to accommodate longitudinal data. [11] developed a quasi-likelihood method for median regression model for longitudinal data. [8] proposed to fit the quantile regression for longitudinal data based on ALD and the

estimation is made by using Monte Carlo EM algorithm, later on [21] followed the idea of [8] and extended the model from random intercept to include random regression coefficients as well. [7] proposed a working correlation model for quantile regression for longitudinal data, a induced smoothing method was used to make the inference of the estimators. Fully Bayesian techniques were also developed for fitting linear quantile mixed models, for example [22] used the similar idea as [19] did in decomposing the error term as a location-scale mixture and developed a Gibbs sampling algorithm for quantile linear mixed model. The fully Bayesian method is appealing because it is easy to implement and to make inference, the uncertainty of the unknowns is taken into account, and it is flexible in the distribution of random effects. The detailed background about Bayesian quantile linear mixed model will be introduced in Section 4.2.2. In addition, temporal

A novel idea that combines the linear quantile mixed model and nonparametric Bayesian methodology in studying the longitudinal data can be an important supplement to existing methods. Longitudinal or correlated data commonly arise in biostatistical context. For example, they can be measurements that are taken repeatedly over time, or outcomes from grouped patients in clinical trials. Moreover, real data can be far from ideal situation, which means extreme observations can be possibly included. It is crucial that a proper method to be used in order to make a better investigation of the relationship between the outcome of interest and potential risk factors. Traditional mean regression model is easy to grasp and implement, but it also suffers from its fragility in facing of the outliers. Ignoring the outliers in analysis could lead to misleading conclusion [15]. The quantile regression method can be a sensible supplement where the mean regression shouldn't be used in those situations.

*[Add background for topic two here.]*

To sum up, this research project focuses on developing and applying new statistical methods in analysis longitudinal data and will cover the following topics. First, a statistical method that combines nonparametric Bayesian technique and the quartile regression model in an-

alyzing longitudinal data will be introduced. In this part, the linear quantile mixed model is used so that the within subject correlation is accounted and the results also will be more robust to potential outlying observations. Since we used quantile regress instead of mean regression, a more comprehensive insight about the association between the outcome and covariates will be learned. As mentioned, in this part we will use DP to model the random effects in linear quantile mixed model, which is robust to the true underlying distribution of random effects and supposed to provide better prediction of the random effects. Second, as studying the temporal data, we will focus on the prediction of future outcomes.

## 2 Specific Aims and Hypotheses

1. To incorporate nonparametric Bayesian method with quantile linear mixed model to analyze longitudinal data.
2. To obtain accurate prediction of the random effects in mixed model so that accurate prediction for future outcomes can be made. When given new covariates  $\mathbf{x}_0$ , the posterior predictive density of future outcome, given observed values, is formulated as

$$r(\tilde{y}|\mathbf{y}) = \int g(\tilde{y}, \boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} = \int q(\tilde{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \quad (1)$$

where  $\boldsymbol{\theta}$  are the posterior samples from MCMC iterations.

## 3 Data

Traumatic brain injury (TBI) data.

## 4 Methods

### 4.1 Nonparametric Bayesian

#### 4.1.1 Dirichlet Process and Dirichlet Process Mixture Models

A *Dirichlet process* (DP), first introduced by Ferguson in 1973 [5], is a random probability model (RPM), is a probability distribution whose domain is itself a set of probability distributions. A Dirichlet process, denoted by  $DP(\alpha, H)$ , is specified by two parameters: the concentration parameter  $\alpha$ , which is a positive real number and the base distribution (or base measure)  $H$  that is an arbitrary distribution. A draw from DP will return a random distribution (the output distribution) over some some probability space  $\Theta$  that can be drawn from  $H$ . Let  $A_1, \dots, A_r$  be any finite partition of the probability space  $\Theta$ , then we say the RPM  $G$  is a a Dirichlet process with base measure  $H$  and concentration parameter  $\alpha$  if

$$(G(A_1), \dots, G(A_r)) \sim Dir(\alpha H(A_1), \dots, \alpha H(A_r)) \quad (2)$$

An intuitive understanding of Equation (2) is that the concentration parameter  $\alpha$  defines the inverse variance of DP as  $Var(G(A)) = H(A)(1 - H(A))/(\alpha + 1)$  while the base distribution defines the mean of DP:  $E(G(A)) = H(A)$ . Larger  $\alpha$  leads to smaller variance of DP and as a result, it will concentrate more around its mean. Consequently, as  $\alpha \rightarrow \infty$ , we have  $G(A) \rightarrow H(A)$  [5].

Regarding the computational purpose, there are different realization methods for DP and here we only mention two of them. The first is the Pólya urn scheme [1], which is formulated

as follows

$$\begin{aligned}
& \theta_i | G \stackrel{iid}{\sim} G, \text{ for } i = 1, 2, \dots, \text{ where} \\
& G \sim DP(\alpha, H) \\
& \theta_1 \sim H, \text{ and } \theta_{n+1} | \theta_1, \dots, \theta_n \sim \frac{1}{\alpha + n} \left( \alpha H + \sum_{i=1}^n \delta_{\theta_i} \right).
\end{aligned} \tag{3}$$

The other is the stick-breaking representation [31]:

$$\begin{aligned}
& \theta_i | G \stackrel{iid}{\sim} G \text{ for } i = 1, 2, \dots, \text{ where} \\
& G \sim DP(\alpha, H) \\
& G = \sum_{i=1}^{\infty} V_i \prod_{l < i} (1 - V_l) \delta_{\theta_i} \text{ with } V_i \stackrel{iid}{\sim} \text{Beta}(1, \alpha) \text{ and } \theta_i \stackrel{iid}{\sim} H.
\end{aligned} \tag{4}$$

where in both representations  $\delta_{\theta_i}$  is the point mass at  $\theta_i$ . Due to its simplicity, using the stick-breaking structure, with a selection of appropriate priors for the hyper-parameters for parameters in (4), a fully Bayesian algorithm is easy to develop and implement, and inference for DPM models can be made from MCMC simulation.

From (4) we can see that the DP is just a infinite mixtures of point masses, which include infinitely many parameters. And although the base distribution  $H$  is continuous, the DP is almost surely discrete. The discreteness property of DP limits its applications to many problems where the distribution to model is believed to be continuous. To overcome this constraint, a simple extension was developed by introducing additional convolution:

$$\begin{aligned}
& \theta_i | \boldsymbol{\eta}_k \sim f(\cdot | \boldsymbol{\eta}_k), \text{ } k = 1, 2, \dots \\
& \boldsymbol{\eta}_k | G \stackrel{iid}{\sim} G, \text{ } G \sim DP(\alpha, H)
\end{aligned} \tag{5}$$

where  $f(\cdot)$  is a continuous density function (usually is normal distribution) for  $\theta_i$  with parameter  $\boldsymbol{\eta}$  and as a result,  $\theta_i \sim \sum_{k=1}^{\infty} w_k f(\cdot | \boldsymbol{\eta}_k)$ , where  $w_k = V_k \prod_{l < k} (1 - V_l)$ . This model is referred as Dirichlet process mixture (DPM) model[4].

#### 4.1.2 Centered Dirichlet Process Mixture

In statistical modeling, the parameter  $\theta$  (in some parameter space  $\Theta$ ) in the probability distribution  $\mathcal{P}_\theta$  is not identifiable if and only if there is another parameter  $\theta^* \in \Theta$ , with  $\theta^* \neq \theta$  and  $\mathcal{P}_\theta = \mathcal{P}_{\theta^*}$ . In linear mixed models, i.e.  $y_{ij} = \beta_0 + \boldsymbol{\beta}'\mathbf{X}_{ij} + \theta_i + \varepsilon_{ij}$ , with  $E(y_{ij}) = \beta_0 + \boldsymbol{\beta}'\mathbf{X}_{ij} + \mu_{\theta_i}$ , to avoid the identifiability problem of the regression intercept and mean parameter of the random effects, people usually constrain the normal prior to have mean zero with some variance parameter. Similar problem also occur when we use the DPM prior for the random effects, since mean of the mixture distribute is not zero by default in (5). To work around the identifiability problem, [32] proposed a revised stick-breaking procedure that standardizes the random samples from DPM to make its mean be zero and the resulting DPM model is named centered DPM (CDPM) model. Take the mixtures of normal model as an example, the procedure is done as follows:

$$\begin{aligned}
(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) | G &\stackrel{iid}{\sim} G, \quad k = 1, 2, \dots \\
\boldsymbol{\mu}_k^* &= \boldsymbol{\mu}_k - \sum_{i=k}^{\infty} w_k \boldsymbol{\mu}_k, \quad \text{with } w_k = V_k \prod_{l < k} (1 - V_l), \quad V_k \stackrel{iid}{\sim} \text{Beta}(1, \alpha) \\
\theta_i | \boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k &\sim N(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k)
\end{aligned} \tag{6}$$

#### 4.1.3 Truncation

However, it is impossible to get exact samples from a DP since it needs to perform draws from an infinitely many mixtures of densities. By noticing that the random weights  $w_i$  assigned to the densities decrease as the number of mixture index  $i$  increases, some authors proposed



a strategy to bypass this issue by choosing a large enough  $N$  for mixture to approximate the full DPM [25] [10] [17] [27], i.e.

$$G^* = \sum_{k=1}^N w_k f(\cdot) \quad (7)$$

In [9], Ishwaran gives some theoretical comparison to assess the difference between truncated DP and full DP under the setting of generalized linear mixed models. He also proposed a blocked Gibbs sampler based on the truncation of the stick-breaking representation. Based on Ishwaran's theory, if we choose  $N = 200$  and  $\alpha = 10$ , the loss of the approximation by truncation is negligible ( $\leq 9.1e^{-7}$ ). In our application, we will compare the performance of truncated DPM for different values of  $N$ .

## 4.2 Linear Quantile Mixed Models

### 4.2.1 Quantile Regression and Asymmetric Laplace Distribution

Let  $Y$  be a real valued random variable with cumulative distribution function  $F_Y(y) = P(Y \leq y)$ . The  $\tau^{th}$  quantile of  $Y$ , where  $\tau \in [0, 1]$ , is given by

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf \{y : F_Y(y) \geq \tau\} \quad (8)$$

In quantile regression context, similar as in linear regression, the conditional quantile of  $Y$  is model as a linear function of other covariate, which is defined as

$$Q_{y|\mathbf{x}}(\tau) = \mathbf{x}^T \boldsymbol{\beta}_\tau \quad (9)$$

Given the data sample, the estimates of the regression coefficients at  $\tau^{th}$  quantile can be obtained by solving

$$\hat{\boldsymbol{\beta}}_\tau = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n [\rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta})], \quad (10)$$

where the loss function  $\rho_\tau(\cdot)$  is defined as  $\rho_\tau(y) = y(\tau - I(y < 0))$ .

There is not direct solution to solve (10) rather the minimization problem can be reformulated as a linear programming problem, where simplex methods or interior point methods can be applied to solve for the estimates [13]. The minimization problem can also be rephrased as a maximum-likelihood problem by using the asymmetrical Laplace distribution (ALD). [16] [33].

Suppose a random variable  $Y$  follows  $\text{ALD}(\mu, \sigma, \tau)$ , then the probability density function of  $Y$  can be written as

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left[ -\rho_\tau \left( \frac{y-\mu}{\sigma} \right) \right], \quad (11)$$

where  $\mu \in (-\infty, \infty)$  is the location parameter,  $\sigma$  is the scale parameter and  $\tau \in (0, 1)$  is the skewness parameter. Thus, in a standard linear model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad (12)$$

if we assume the random error  $\varepsilon_i \sim \text{ALD}(0, \sigma, \tau)$ , then  $y_i|\mathbf{x}_i \sim \text{ALD}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma, \tau)$ , that is the likelihood function can be written as

$$L(\boldsymbol{\beta}, \sigma; \mathbf{y}, \tau) \propto \frac{1}{\sigma^n} \exp \left[ -\sum_{i=1}^n \rho_\tau \left( \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right]. \quad (13)$$

If we treat  $\sigma$  in (13) as nuisance then the maximization of (13) with respect to  $\boldsymbol{\beta}$  is exactly the same as that in (10).

### 4.2.2 Bayesian Linear Quantile Mixed Model

Linear quantile mixed model (LQMM) is defined as

$$Q_{y_{ij}|\mathbf{x}_{ij}}(\tau) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\alpha}_i, \quad i = 1, \dots, N, \quad j = 1, \dots, n_i. \quad (14)$$

where  $y_{ij}$  is the response variable for subject  $i$  at time  $j$ ,  $\mathbf{x}_{ij}$  is the  $p$ -dimensional fixed effects covariates and  $\boldsymbol{\beta}_\tau$  is the corresponding  $p \times 1$  vector of fixed effects, while  $\mathbf{z}_{ij}$  is the  $k$ -dimensional random effects covariates and  $\boldsymbol{\alpha}_i$  is the corresponding  $k \times 1$  vector of random effects. Under the assumption that the random error follows  $\text{ALD}(0, \sigma, \tau)$  distribution, conditional on the random effects  $\boldsymbol{\alpha}_i$ ,  $y_{ij}$ 's are independently and identically distributed as  $\text{ALD}(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\alpha}_i, \sigma, \tau)$ :

$$f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp \left[ -\rho_\tau \left( \frac{y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta} - \mathbf{z}_{ij}'\boldsymbol{\alpha}_i}{\sigma} \right) \right] \quad (15)$$

To develop a Gibbs sampler for model (15), we need to assume a location-scale mixture representation of the ALD [18]. Under this assumption the random error is represented as

$$\varepsilon_{ij} = \gamma_1 e_{ij} + \gamma_2 \sqrt{\sigma e_{ij}} u_{ij},$$

where

$$\gamma_1 = \frac{1-2\tau}{\tau(1-\tau)}, \quad \text{and} \quad \gamma_2^2 = \frac{2}{\tau(1-\tau)},$$

and

$$u_{ij} \sim N(0, 1), \quad \text{and} \quad e_{ij} \sim \exp(\sigma).$$

The linear mixed model is then reparameterized as

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\alpha}_i + \gamma_1 e_{ij} + \gamma_2 \sqrt{\sigma e_{ij}} u_{ij}, \quad (16)$$

or equivalently

$$f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, e_{ij}, \sigma) = \frac{1}{\sqrt{2\pi\gamma_2^2\sigma e_{ij}}} \exp \left[ -\frac{1}{2\gamma_2^2\sigma e_{ij}} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta} - \mathbf{z}_{ij}^T \boldsymbol{\alpha}_i - \gamma_1 e_{ij})^2 \right]. \quad (17)$$

Following (16), a fully specified Bayesian model would include follows:

$$\begin{aligned} \mathbf{u} &\sim \prod_{i=1}^N \prod_{j=1}^{n_i} \exp \left( -\frac{u_{ij}^2}{2} \right) \\ \mathbf{e} &\sim \prod_{i=1}^N \prod_{j=1}^{n_i} \frac{1}{\sigma} \exp \left( -\frac{e_{ij}}{\sigma} \right) \\ \boldsymbol{\beta} &\sim MVN_p(\mathbf{0}, \boldsymbol{\Sigma}) \\ \boldsymbol{\alpha}_i|v &\sim MVN_p(\mathbf{0}, v^2 \mathbf{I}) \text{ or } \boldsymbol{\alpha}_i|G \sim DP(\alpha_0, H) \\ \sigma &\sim IG(a_0, b_0) \\ v &\sim IG(c_0, d_0) \text{ or } \alpha_0 \sim IG(e_0, f_0), H \sim N(\mu_0, \sigma_0). \end{aligned}$$

## 5 Plan for Simulation Studies

### 5.1 Bayesian Quantile Regression Method

In this section, the simulation study will focus on fitting the Bayesian quantile linear mixed model (BLQMM) by using the ALD representation for the random errors. The focus will be checking the performance of the BLQMM method in estimating the regression coefficients. Here we will use the normal distribution as the prior for the random effects.

We simulate data from the model

$$y_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta} + \mathbf{z}_{ij}' \boldsymbol{\alpha}_i + (1 + \gamma x_{ij}) \varepsilon_{ij} \quad (18)$$

where the covariates  $x_{ij}$  not only affects the location of  $y_{ij}$  but also its scale, which is one of the reasons why we prefer to use quantile regression for such data. Based on equation (18), the settings are similar to strategy one for the fixed and random effects and  $\gamma$  is a parameter chosen for the heteroscedasticity of the data [28]. When  $\gamma = 0$ , (18) reduces to

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\alpha}_i + \varepsilon_{ij}. \quad (19)$$

Based on equation (18), for a simplified model

$$y_{ij} = \beta_0 + \beta_1 * x_{ij} + \alpha_i + (1 + \gamma x_{ij})\varepsilon_{ij} \quad (20)$$

The simulation study is carried out as follows: generate data from equation (20), where

$$\beta_0 = 10, \beta_1 = -1, \gamma = 1/11$$

$$x_{ij} \sim \text{unif}(0, 10), \alpha_i \sim N(0, 1), \varepsilon_{ij} \sim N(0, 1)$$

Given  $\boldsymbol{\beta}$  and  $\tau$ , the true values of  $\tau^{th}$  quantile regression coefficients are given by

$$\beta_0(\tau) = \beta_0 + q_\tau(\varepsilon_i),$$

$$\beta_1(\tau) = \beta_1 + \gamma q_\tau(\varepsilon_i).$$

The Gibbs developed in Section 4.2.2 will be implemented using **Jags** and the **inferno** will be made based on the MCMC samples. The simulation results will be shown as follows

Table 1: Simulation study with 500 datasets,  $N(0, 1)$  for error term

$\tau = 0.05$	<i>bias</i>	<i>RMSE</i>
$\beta_0(8.355)$	-0.0387	2.089
$\beta_1(-1.150)$	0.0754	0.118
<hr/>		
$\tau = 0.25$		
$\beta_0(9.326)$	-0.0300	1.049
$\beta_1(-1.061)$	0.0214	0.085
<hr/>		
$\tau = 0.5$		
$\beta_0(10)$	-0.0039	0.354
$\beta_1(-1)$	0.0078	0.071
<hr/>		
$\tau = 0.75$		
$\beta_0(10.674)$	0.0243	0.992
$\beta_1(-0.939)$	-0.0096	0.089
<hr/>		
$\tau = 0.95$		
$\beta_0(11.644)$	0.0374	2.070
$\beta_1(-0.850)$	-0.0707	0.117

## 5.2 Dirichlet Process Mixture Prior for Random Effects

In this section of simulation study we aim to compare the performance of different prior distributions in modeling the random effects. We will investigate the effects of various priors on estimating the fixed and predicting the random effect with emphasis on random effects prediction. Estimation bias and credible interval will be calculated for the fixed effects for different prior assignments. Regarding comparing the predicted random effects we will compare with respect to (a) estimating the true underlying distribution of the random effects, (b) the sums of squared difference between predicted versus generated random effects and

(c) the posterior predictive values of the outcome based on the predicted random effects from different models.

The prior distributions under comparison include Gaussian (conventional one),  $t$  distribution (heavier-tailed), and nonparametric Bayesian ones (DPM, CDPM). And the true underlying random effects are generated from Gaussian or non-Gaussian distributions like  $t$  distribution (heavy tail),  $Beta$  distribution (positively skewed) , and mixture of normals (multimodal).

### 5.2.1 Estimation of Fixed Effects

Datasets will be simulated from the following formula with constant intercept

$$y_{ij} = \beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \varepsilon_{ij} \quad (21)$$

Table 2: Sample summary table: estimation of fixed effects when  $u_i \sim$  mixture Normal

<i>par</i>	<i>truth</i>	<i>Normal</i>		<i>DPM</i>		<i>CDPM</i>	
		<i>est</i>	<i>95% CI</i>	<i>est.</i>	<i>95% CI</i>	<i>est.</i>	<i>95% CI</i>
$\beta_1$	1	0.929	(0.775, 1.089)	0.793	(-0.357,1.803)	0.928	(0.820,1.037)
$\beta_2$	1	1.016	(0.963, 1.067)	1.017	(0.965,1.069)	1.017	(0.966,1.067)
$\beta_3$	-1	-0.957	(-1.007, -0.907)	-0.957	(-1.010,-0.906)	-0.956	(-1.008,-0.905)

### 5.2.2 Prediction of Random Effects

#### Prediction of the distribution of random effects

To compare the performance of DPM and conventional normal prior we simulated data based on Equation (19), specifically,

$$\boldsymbol{\beta} = (1, 1, -1)',$$

$$\mathbf{x}_{ij} = (x_{2ij}, x_{3ij})' \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{I} \right),$$

$$\varepsilon_{ij} \sim N(0, 1).$$

$\mathbf{x}_{1ij}$  is a vector of 1's and simulate  $\alpha_i$  from following different distributions respectively. We fit the simulated data with normal prior and DPPM for the random effect and compare the results by plotting them against the empirical data density of the random effect.

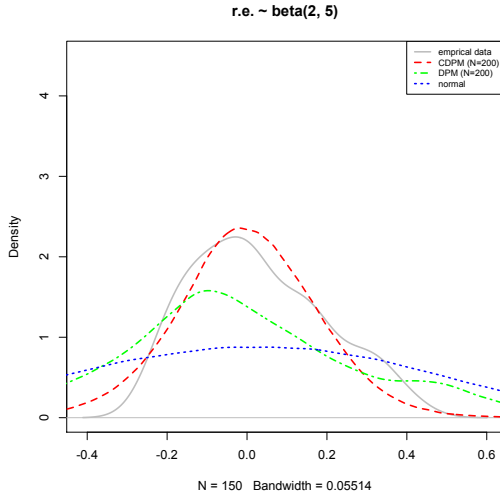


Figure 1: Predicted densities of random effects that follow beta distribution

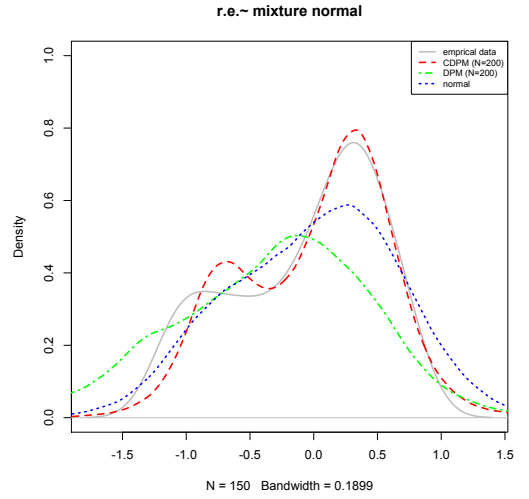


Figure 2: Predicted densities of random effects that follow mixture of normals

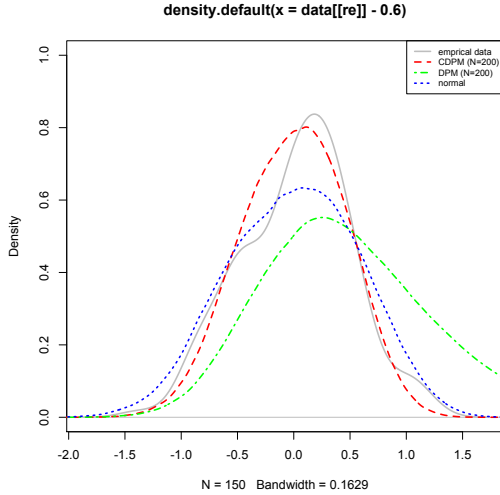


Figure 3: Predicted densities of random effects that follow  $N(0.6, 0.5^2)$



### Sum of squared difference

The best predictor (BP) is defined as the one that minimize the mean squared error of the prediction [23], that is to minimize

$$E[(\tilde{\mathbf{u}} - \mathbf{u})^T \mathbf{A}(\tilde{\mathbf{u}} - \mathbf{u})] = \int \int (\tilde{\mathbf{u}} - \mathbf{u})^T \mathbf{A}(\tilde{\mathbf{u}} - \mathbf{u}) f(\mathbf{u}, \mathbf{y}) d\mathbf{y} d\mathbf{u}, \quad (22)$$

where  $\mathbf{A}$  is a positive definite symmetrical matrix.

[30] shows that the to minimize (22), the BP of  $\mathbf{u}$  is given by

$$\tilde{\mathbf{u}} = E(\mathbf{u}|\mathbf{y}), \quad (23)$$

that is the posterior mean of the random effects  $\mathbf{u}$ . Thus, based on the result of (23), to compare the performance of different models in predicting the random effects, one criteria that we can use is the sum of squared difference (SSD) between the predicted and generated random effects is calculated as follows:

$$SSD = \sum_{i=1}^N (\tilde{u}_i - u_i)^2, \quad (24)$$

where  $\tilde{u}_i$  is the posterior mean of the predicted random effects for subject  $i$  and the model with smaller SSD is preferred.

### Posterior predictive values

We can also sample the response variable based on the predicted distribution of the random effects, as a result, we can compare the posterior predictive values with the observed ones by using  $Q - Q$  plot or relative absolute deviation (RAD) [29], which is defined as

$$RAD = \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{\tilde{y}_{ij} - y_{ij}}{y_{ij}}, \quad (25)$$

where  $\tilde{y}_{ij}$  is the posterior predictive values sampled based on (1). Similar as SSD criteria, model with smaller RAD is preferred.

Table 3: Sample summary table: statistics of prediction of random effects (RAD is based on 30 posterior samples)

	<i>dist. of <math>u_i</math></i>	<i>Normal</i>	<i>DPM</i>	<i>CDPM</i>
	$N(0.6, 0.25)$	1.00 (70.52)	0.22 (15.33)	0.98 (69.33)
<b>SSD</b>	$beta(2,5)$	1.00 (11.96)	0.29 (3.50)	0.27 (3.24)
	$MN$	1.00 (13.48)	1.08 (14.56)	0.80 (10.84)
	$N(0.6, 0.25)$	1.00 (-2685.27)	1.04 (-2787.77)	1.02 (-2739.31)
<b>RAD</b>	$beta(2,5)$	1.00 (1939.24)	0.81 (1563.85)	0.84 (1634.97)
	$MN$	1.00 (264.39)	0.86 (226.64)	0.87 (228.96)

## References

- [1] David Blackwell and James B MacQueen. Ferguson distributions via pólya urn schemes. *The annals of statistics*, pages 353–355, 1973.
- [2] Steven M Butler and Thomas A Louis. Random effects models with non-parametric priors. *Statistics in medicine*, 11(14-15):1981–2000, 1992.
- [3] David B Dunson. Nonparametric bayes applications to biostatistics. *Bayesian nonparametrics*, 28:223, 2010.
- [4] Michael D Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588, 1995.

- [5] Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.
- [6] Ronald A Fisher. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the royal society of Edinburgh*, 52(02):399–433, 1919.
- [7] Liya Fu and You-Gan Wang. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8):2526–2538, 2012.
- [8] Marco Geraci and Matteo Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154, 2007.
- [9] Hemant Ishwaran. Inference for the random effects in bayesian generalized linear mixed models.
- [10] Hemant Ishwaran and Lancelot F James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453), 2001.
- [11] Sin-Ho Jung. Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, 91(433):251–257, 1996.
- [12] Ken P Kleinman and Joseph G Ibrahim. A semiparametric bayesian approach to the random effects model. *Biometrics*, pages 921–938, 1998.
- [13] Roger Koenker. *Quantile regression*. Cambridge university press, 2005.
- [14] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.
- [15] Roger Koenker and Kevin Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

- [16] Roger Koenker and Jose AF Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310, 1999.
- [17] Athanasios Kottas and Alan E Gelfand. Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96(456):1458–1468, 2001.
- [18] Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgorski. *The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Exonomics, Engineering, and Finance*. Number 183. Springer, 2001.
- [19] Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation*, 81(11):1565–1578, 2011.
- [20] Katherine J Lee and Simon G Thompson. Flexible parametric models for random-effects distributions. *Statistics in medicine*, 27(3):418–434, 2008.
- [21] Yuan Liu and Matteo Bottai. Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics*, 5(1), 2009.
- [22] Youxi Luo, Heng Lian, and Maozai Tian. Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 82(11):1635–1649, 2012.
- [23] Charles E McCulloch and John M Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001.
- [24] Devan V Mehrotra. Non-iterative robust estimators of variance components in within-subject designs. *Statistics in medicine*, 16(13):1465–1479, 1997.
- [25] Pietro Muliere and Luca Tardella. Approximating distributions of random functionals of ferguson-dirichlet priors. *Canadian Journal of Statistics*, 26(2):283–297, 1998.

- [26] John M Neuhaus, Walter W Hauck, and John D Kalbfleisch. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79(4):755–762, 1992.
- [27] David I Ohlssen, Linda D Sharples, and David J Spiegelhalter. Flexible random-effects models using bayesian semi-parametric models: applications to institutional comparisons. *Statistics in medicine*, 26(9):2088–2112, 2007.
- [28] Craig Reed and Keming Yu. A partially collapsed gibbs sampler for bayesian quantile regression. 2009.
- [29] Reyhaneh Rikhtehgaran and Iraj Kazemi. Semi-parametric bayesian estimation of mixed-effects models using the multivariate skew-normal distribution. *Computational Statistics*, 28(5):2007–2027, 2013.
- [30] Shayle R Searle, George Casella, and Charles E McCulloch. *Variance components*, volume 391. John Wiley & Sons, 2009.
- [31] Jayaram Sethuraman. A constructive definition of dirichlet priors. Technical report, DTIC Document, 1991.
- [32] Mingan Yang, David B Dunson, and Donna Baird. Semiparametric bayes hierarchical models with mean and variance constraints. *Computational statistics & data analysis*, 54(9):2172–2186, 2010.
- [33] Keming Yu and Rana A Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.