6-30-2016

Bayesian Semiparametric Quantile Regression for Clustered Data

Xin Tong
University of South Carolina

Bayesian Semiparametric Quantile Regression for Clustered Data

by

Xin Tong

Bachelor of Science
Nanjing University 2007

Master of Science in Public Health
University of South Carolina 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Biostatistics

The Norman J. Arnold School of Public Health

University of South Carolina

2016

Accepted by:

Bo Cai, Major Professor

Marco Geraci, Committee Member

Alexander McLain, Committee Member

Lianming Wang, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Acknowledgments

I would like to express my deep and sincere gratitude to everyone who supported and helped me over the past few years. They have made my graduate study possible.

First and foremost, I would like to thank my advisor Bo Cai for his insightful guidance, endless patience and invaluable advice throughout my doctoral study. This thesis would not be possible without his constant support and inspiration make my doctoral research productive and stimulating.

I am especially thankful to my committee members, Dr. Marco Geraci, Dr. Alex McLain and Dr. Lianming Wang for joining my committee. I greatly appreciate constructive feedback they have provided. I have learned a lot about the concept and modeling methodology of quantile regression for longitudinal data by talking with Dr. Geraci. I am very grateful to Dr. Wang for advising me with his expertise on interval-censored data. I would like to express my gratitude to Dr. McLain for his perpetual encouragement along the way.

The faculty of this department are very resourceful and have provided me with a tremendous graduate education. Further thanks go to Dr. James Hussey, Dr. Robert Moran, Dr. Jiajia Zhang for their enlightenment, which has provided me a solid foundation to progress upon my graduate study. I also would like to thank Dr. Suzanne McDermott for offering me a research assistantship during my PhD study.

Most of all, I thank my parents, Lianfang Tong, Yahui Wang and my husband, Dr. Diyang Chu, for their love and support. To them I dedicate this dissertation.

Abstract

Traditional frequentist quantile regression makes few assumptions on the form of the error distribution and thus is able to accommodate non-normal errors. However, inference on the quantile regression models could be challenging for the unknown error distribution, though asymptotic or resampling methods were developed. Bayesian literature on quantile regression with random effects is relatively limited. The quantile regression approach proposed in this dissertation is founded on Bayesian probabilistic modeling for the underlying unknown distributions. By adopting the error density with a nonparametric scale mixture models, we developed Bayesian semiparametric models to make an inference on any quantile of interest and to allow for flexible shapes of the error densities.

In this dissertation, we aimed to develop Bayesian semiparametric quantile regressions for both longitudinal data and clustered interval-censored data. We first proposed a semiparametric quantile mixed effect regression for clustered data, which relaxed normality assumption for both random effects and the error term. We then developed a semiparametric accelerated failure time quantile regression for the clustered interval-censored data. Both of the methods allow for estimates for the subgroup specific parameters and the detection of heterogeneity in the random effects population under nonparametric settings. Markov chain Monte Carlo (MCMC) methods provide computationally feasible implementations of Bayesian inference and learning. However, the speed of convergence can be challenging for highly complex and non-conjugate models. Specifically, Gibbs sampling algorithm that employs the addition of auxiliary parameters was used to speed up posterior sampling in our study. Sev-

eral variations of the proposed model were considered and compared via the deviance information criterion. The performance of the proposed methods was evaluated by extensive simulation studies, and examples using data from Orthodontic clinics and lymphatic filariasis drug studies were presented as illustration.

Table of Contents

List of Tables

List of Figures

Introduction

Linear mixed models are frequently used to describe longitudinal and repeated measurements data where random effects serve to model the between-individual correlation structure. One of the basic model assumptions of mixed model is the normality of the random effects, which is chosen mainly for computational convenience. However, the normality assumption is typically made for the sake of convenience, rather than from some theoretical justification, and might be inappropriate in some situations. Another troublesome fact, as noted by [Burr and Doss, 2005], is that random effects, unlike error terms, cannot be checked (there are no residuals). Thus we are totally dependent on this uncheckable model assumption. It is of potential interest to model the error nonparametrically while accounting for the correlation of observations within the same subject. Extensive work have been done on a more flexible, possibly nonparametric distribution for random effects in both frequentist and Bayesian methods, such as Dirichlet Process (DP) prior to model random effects [Bush and MacEachern, 1996, Kleinman and Ibrahim, 1998b].

Most linear mixed models also assume that random errors are normally distributed with constant variances. However, in some applications, the covariates may have different impacts at different locations of the response distribution. For instance, in a birth weight study, Abrevaya and Dahl [2008] found that gender of the baby and the mother's prenatal-care visits have different effects at the lower and upper quantiles of the infant birth weight distribution. In such scenarios, modeling the mean is limiting and thus cannot accommodate population heterogeneity. Quantile regression extends regression for the mean to the analysis of the entire conditional distribution of the

outcome variable, by allowing the tail quantiles of the conditional distribution of the response variable as functions of a set of covariates [Koenker and Bassett Jr, 1978]. By focusing on the conditional quantiles, quantile regression offers an alternative analytic tool that can automatically capture the heterogeneity in covariate effects at different quantiles of the response distribution without modeling the heteroscedasticity. Quantile regression is free from the distributional assumptions and robust to the presence of outliers and therefore quantile regression is appropriate with skewed distributions.

The dissertation is organized as follows. We begin with an introduction of mixed models, quantile regression and quantile regression for longitudinal data in chapter 1. The use of asymmetric Laplace distribution (ALD) for the error distribution provides a natural way to deal with the Bayesian quantile regression [Yu and Moyeed, 2001], but lacks flexibility in error distribution. [Kottas and Krnjanić, 2009] proposed an approach to inference for nonparametric quantile regression founded on probabilistic modeling for the underlying unknown (random) distributions. The flexibility of such inference under nonparametric prior models is attractive and can be incorporated into mixed models for longitudinal data. Chapter 2 presents a Bayesian semiparametric approach to random effects model using Dirichlet process mixtures for the error distribution. Two simulation results are presented and indicate that the presented approach works well for practical situations. In chapter 3, we provide an overview of interval-censored data and two common regression models, proportional hazard model and accelerated failure time model, and discuss some existing approaches for regression analysis of clustered interval-censored failure time data. We are interested to explore a flexible quantile regression framework for analysing time-to-event data that are randomly interval-censored in chapter 4. Quantile regression models the conditional quantiles of the survival time directly and allows the covariates to have different effects at different tails of the survival distribution and thus is able to cap-

ture important population heterogeneity [Koenker, 2005]. We adopt an accelerated failure time model with random effects to analyze clustered interval-censored data in the nonparametric quantile regression framework. A nonparametric Dirichlet process prior random effects component has been used to provide flexible distributional forms to the random effects. In addition, a Bayesian quantile regression model for clustered interval-censored data with parametric error distribution is also constructed for comparison. Lastly, chapter 5 provides some discussion and offer some recommendations on how to proceed in this Bayesian nonparametric quantile regression with random effects.

Chapter 1

Background of Quantile Regression for Longitudinal Data

Inference of quantile analysis has received increasing attention in recent years. Quantile regression seeks to model each quantile of the response distribution, whether separately or jointly, conditional upon covariates. It extends regression for the mean to the analysis of the entire conditional distribution of the outcome variable. Regression quantiles are robust against the influence of outliers [Huber, 1981], and in many regression examples, we might expect a different structural relationship for the higher (or lower) responses than the average responses. In such applications, mean (or median) regression approaches would likely overlook important features that could be uncovered by a more general quantile regression analysis.

The construction of infant and adolescent growth charts provides a motivating application of applying quantile regression to longitudinal data [Wei et al., 2006]. It is of importance to construct reference growth charts that accurately represent the conditional quantiles of the growth distribution without unduly constraining the estimation process by unverifiable distributional assumptions.

1.1 Mixed Models with Heterogeneity in Random Effects

Longitudinal data are characterized by repeated measurements on the same subject over time, as may be collected in clinical trials, panel research, epidemiological studies, etc. Linear mixed models are commonly used for longitudinal data analysis,

where covariate effects are modeled parametrically and with-in subject correlations are modeled using random effects [Laird and Ware, 1982]. Let y_{ij} denote the response for ith subject and jth repeated measurements, the random effects model for y_{ij} is given by

$$y_{ij} = X_{ij}\boldsymbol{\beta} + Z_{ij}\boldsymbol{b_i} + \epsilon_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i, \tag{1.1}$$

where the subscript i would index individual subjects, and the subscript j would index the n_i distinct measurements made on the ith subject. y_{ij} is the outcome for the jth measure on the ith subject, X_{ij} is a $1 \times p$ vector of fixed covariates; $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector of regression coefficients, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, referred to as fixed effects in the models; Z_{ij} is typically a subset of X_{ij}, $1 \times q$ vector of covariates for the $q \times 1$ vector of random effects $\boldsymbol{b_i}$ with $q \leq p$, $\boldsymbol{b_i} = (b_{i1}, \ldots, b_{iq})^T$ is assumed to be normally distributed as $N_q(0, \boldsymbol{D})$. Here $\boldsymbol{D}$ is a $q \times q$ positive-definite covariance matrix. ϵ_{ij} is usually assumed to be independent and normally distributed as $N_{n_i}(0, \sigma^2 I)$. The random effects component $Z_{ij}\boldsymbol{b_i}$ can be considered as the deviation from the population mean $X_{ij}\boldsymbol{\beta}$. $\boldsymbol{b_i}$ is a vector of individual specific regression coefficients, which is assumed to be independently distributed from the error terms $\boldsymbol{\epsilon}_i$. Marginalizing the random effects $\boldsymbol{b_i}$, we have

$$y_{ij}|\boldsymbol{\beta}, D, \sigma^2 \sim N(X_{ij}\boldsymbol{\beta}, Z_{ij}\boldsymbol{D}Z_{ij}^T + \sigma^2\boldsymbol{I}) \tag{1.2}$$

Inference for regression parameters can be done by generalized least squares, maximum likelihood methods or empirical Bayesian method.

The dominant paradigm in the random effects literature has been a Gaussian structure in which covariates exert a pure location shift effect on the response variable. McCulloch and Neuhaus [2011] argued that misspecifying the shape of a random effects distribution does not matter as long as the assumed distributions are not quite far from non-normality, not as extreme as a two-point, discrete distribution, indicating a high degree of robustness. But it is also of interest to explore various forms

of heterogeneity associated with the covariates under less stringent distributional assumptions, while still accounting for individual specific effects. Frequentist inference regarding fixed effects should not be much affected asymptotically by changing the distribution of the random effects since the first two moments of the marginal distribution of the response variable do not depend on the normality of the distribution of the random effects. In equation 1.2, the individual regression coefficients simply result in a complex covariance structure. On the other hand, bayesian inference for fixed effects will depend on (y, σ^2, D), and this dependence will be sensitive to the distributional form ascribed to the $\boldsymbol{b_i}$ [Kleinman and Ibrahim, 1998b]. Verbeke and Lesaffre [1996] shows that random effects may be badly estimated under normal error while the true error distribution is not normally distributed, and the current methods for inspecting the appropriateness of the model assumptions are not sound. A more flexible and robust approach to solve the Gaussian restricted random effect distribution problem is to use a nonparametric distribution, such as Dirichlet process which will be discussed in section 1.3.1. Such model has the potential to capture more types of variability in those effects with the possible end result of more precise estimates of the fixed effects.

1.2 Quantile Regression

Quantile regression has been used as an attractive analytic technique to examine many situations including risk management, portfolio optimization and asset pricing in econometrics. The use of quantile regressions is versatile with many advantages. It allows us to study the impact of predictors on different quantiles of the response distribution, and thus provides a complete picture of the relationship between the response variable and the predictors. The entire conditional distribution of the dependent variable can be characterised by using different values of τ. Median regression methods can be more efficient than mean regression estimators in the presence of

heteroscedasticity. Quantiles are invariant to monotonic transformations and are not sensitive to outlier observations on the dependent variable. Finally, when the error term is non-normal, quantile regression estimators may be more efficient than least squares estimators since estimation and inference are distribution-free.

Basics of Quantile Regression

By extending the standard additive regression formulation, the τth quantile regression model for (continuous) response observations y_i, with associated covariate vectors x_i, $i = 1, \ldots, n$, can be written as

$$y_i = h(x_i) + \epsilon_i \tag{1.3}$$

where the ϵ_i is the error term whose distribution (with density $f_\tau(\cdot)$) is restricted to guarantee the τth quantile of ϵ_i is zero, that is $\int_{-\infty}^{0} f_\tau(\epsilon_i) d\epsilon_i = \tau$. The linear model for the τth quantile is

$$y_i = x_i^T \boldsymbol{\beta}_\tau + \epsilon_i \tag{1.4}$$

where the τth quantile of ϵ_i is zero. Koenker and Bassett Jr [1978] specify the τ-th conditional quantile function as

$$Q_\tau(y_i|\boldsymbol{\beta}, \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_\tau, i = 1, ..., n$$

where $\tau \in (0, 1)$, y_i is a scalar response variable with conditional cumulative distribution F_y, $Q_{y_i}(\cdot) \equiv F_{y_i}^{-1}(\cdot)$, $\boldsymbol{\beta}_\tau \in R^p$ is a column vector of unknown fixed parameters with length p. One can trace the entire conditional distribution of the dependent variable, conditional on the set of predictors, by increasing τ from 0 to 1. Conditional quantile model specifies a different model for each quantile of the outcome distribution. Interpretation of $\boldsymbol{\beta}_\tau$ is straightforward: the intercept term simply represents the baseline predicted quantile, while each slope can be interpreted as the rate of change of the τth response quantile per unit change in the value of the corresponding covariate (the others being fixed). It is flexible for modeling data with heterogeneous

conditional distributions and makes no distributional assumption about the error term in the model. However, there is no probability model for the response distribution in classical quantile regression, point estimation for $\hat{\boldsymbol{\beta}}_\tau$ proceeds by solving an optimization problem of the loss function

$$\text{argmin}_{\beta \in \mathbb{R}^\tau} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \boldsymbol{\beta}) \tag{1.5}$$

where $\rho_\tau(\upsilon) = \upsilon(\tau - I(\upsilon \leq 0))$ is the quantile loss function; see [Koenker and Bassett Jr, 1978]. $I(\cdot)$ is an indicator function such that $I(\upsilon \leq 0) = 1$ if $\upsilon \leq 0$ and $I(\upsilon \leq 0) = 0$, otherwise.

Asymmetric Laplace distribution provides a natural link between minimization of the quantile loss function and maximum likelihood theory by assuming the error term in equation (1.4) follows ALD. See [Koenker and Machado, 1999, Yu and Moyeed, 2001, Yu and Zhang, 2005] for more details. A random variable Y follows an ALD if its corresponding probability density is

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau(\frac{y-\mu}{\sigma})\right\} \tag{1.6}$$

where $\rho_\tau(\cdot)$ is the loss function, $\sigma > 0$ is the scale parameter and $-\infty < \mu < +\infty$ is the location parameter.

Assume $y_i \sim ALD(\mu_i, \sigma, \tau), \mu_i = x_i^T \boldsymbol{\beta}_\tau$. The likelihood for N independent observations is

$$L(\boldsymbol{\beta}, \sigma, \boldsymbol{y}, \tau) = \left[\frac{\tau(1-\tau)}{\sigma}\right]^N \exp\left\{-\sum_{i=1}^{N} \rho_\tau(\frac{y_i - x_i^T \boldsymbol{\beta}_\tau}{\sigma})\right\} \tag{1.7}$$

Considering σ a nuisance parameter, the maximization of the likelihood in (1.7) with respect to the parameter $\boldsymbol{\beta}_\tau$ is equivalent to the minimization of the loss function in (1.5). The relationship between the loss function and ALD can be used to reformulate the quantile regression method in the likelihood framework. By utilizing this property, [Koenker and Machado, 1999] proposed a likelihood-based goodness-of-fit test for quantile regression. Yu and Moyeed [2001] proposed a Bayesian approximate

inference about the quantile regression by using asymmetric Laplace distribution to form likelihood function, and Yu et al. [2003] studied the Bayesian estimation procedure for the Tobit quantile regression model with censored data. Geraci and Bottai [2007] and Yuan and Yin [2010] extend the asymmetric Laplace distribution idea for longitudinal studies, which will be explained in the next section.

An Example

An example of the low birth weight study that was carried out by Koenker [2005] will be used in this section. The data set used for this example is a subset of 50,000 observations. We are interested in investigating the relationship between infants?birth weight (LBW: in grams) and only four predictors of interest, such as: the gender of the infant, marital status of the mother, prenatal care, and smoking status of the mother during pregnancy. The linear regression model for this example is:

$$\text{LBW} = \alpha + \beta_1 \times \text{Gender} + \beta_2 \times \text{Married} + \beta_3 \times \text{Prenatal care} + \beta_4 \times \text{Smoke}$$

Linear regression is used to model the relationship between a set of predictor variable and a response variable. It estimates the mean value of the response variable for given levels of the predictor variables. The least square result and quantile regression results are summarized in Table 1.1.

In each plot of Figure 1.1, the regression coefficient at a given quantile indicates the effect on birth weight of a unit change in that variable, assuming that the other variables are fixed, with 95% confidence interval bands.

The intercept can be interpreted as the estimated conditional quantile function of the birth weight distribution of a girl born to an unmarried mother who didn't smoke, and had her first prenatal visit in the first trimester of the pregnancy [Koenker, 2005]. The least-square estimate present a birth weight of 3224 grams for a girl born to an

Table 1.1 Coefficients estimates for quantile regression and linear regression for the birth weight example

		Quantile regression				
Characteristic	Linear regression	5th	10th	50th	90th	95th
Intercept	3224	2353	2608	3252	3856	4031
Married	161.1	227	171	149	141	165
Boy	115.9	28	84	121	142	142
Prenatal Care	-227.0	-536	-418	-164	-111	-57
Smoke	-200.9	-255	-226	-190	-177	-199

unmarried mother who didn't smoke, and had her first prenatal visit in the first trimester of the pregnancy.

According to the linear regression model,the mean weight of boys are 115.9 grams

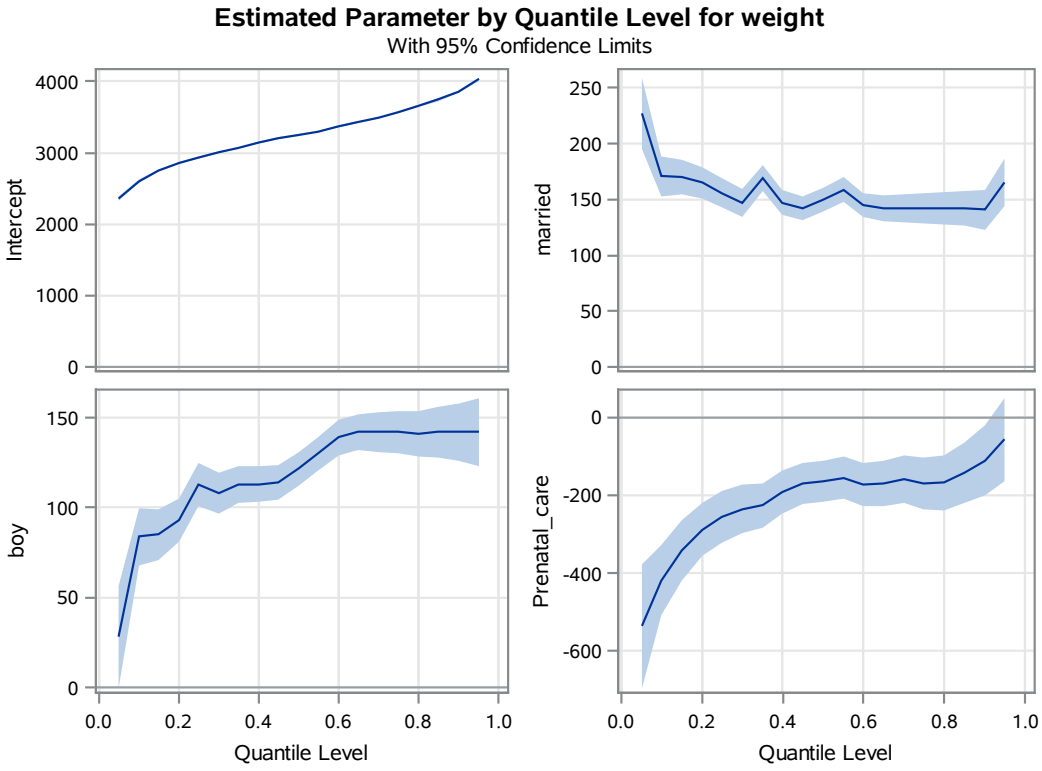


Figure 1.1 Estimated parameters by quantile level

heavier than girls. But the disparity is much smaller than 100 grams in the lower tail and larger than 120 grams in the upper tail of the distribution. The marital status of the mother seems to be associated with a rather large positive effect on birth weight, especially in the lower tail of the distribution. The mean weight of babies born to mothers with no prenatal care is -227 grams lower than that of babies born to mothers who had a prenatal visit in the first trimester. The quantile regression results indicate that the effect of no prenatal care has a larger negative impact on the lower quantiles of birth weight. The 5-th quantile of birth weight for infants born to mothers who had no prenatal care is 536 grams lower than for infants born to mothers had a prenatal visit in the first trimester. The linear regression model underestimates this effect at the 5-th quantile. The deleterious effect of smoking is

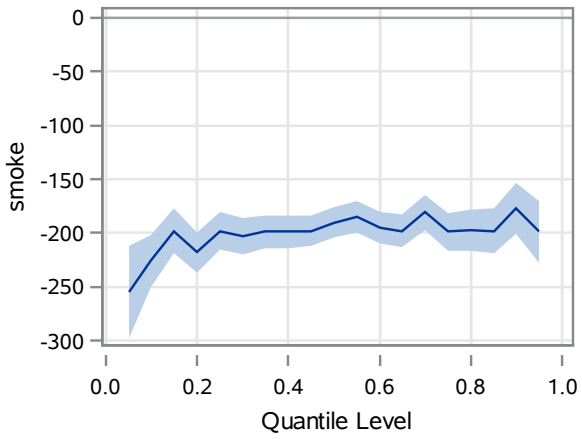


Figure 1.2 Estimated parameter smoking by quantile level

shown in Figure 1.2. Smoking during the pregnancy is associated with a decrease of about $160 - 180$ grams in birth weight. The effect of smoking is quite stable over the entire distribution, as the least-squares estimates of the mother smoking effect are mainly overlapped with the quantile regression confidence band.

1.3 Quantile Regression for Longitudinal Data

Koenker [2004] first considered the penalized interpretation of the classical random effects estimator in order to estimate quantile functions with subject specific fixed effects. For a conditional quantile regression model with a random intercept, a ℓ_1 penalty is employed to shrink the random effects towards a common value. Koenker [2004] propose to estimate α_i and β_{τ_k} at multiple quantile levels $\tau_k, k = 1, \ldots, q$ by minimizing the following penalized objective function:

$$\sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_k \rho_{\tau_k}(y_{ij} - \alpha_i - X_{ij}^T\beta_{\tau_k}) + \lambda\sum_{i=1}^{n}|\alpha_i|,$$

where w_k is the weight on the quantile level τ_k and λ is the regularization parameter that controls the variation of α_i and helps shrink b_i towards a common value. As in most regularization problems, the choice of the penalty parameter λ is crucial. For $\lambda \to 0$ we obtain the fixed effect estimator by solving $\min_{\alpha,\beta}\sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_k\rho_{\tau_k}(y_{ij} - \alpha_i - X_{ij}^T\beta_{\tau_k})$; while for $\lambda \to \infty$ the $\hat{\alpha}_i \to 0$ for all $i = 1, \ldots, n$ and we obtain an estimate of the model with only the fixed effects.

As mentioned in previous section, Geraci and Bottai [2007] and Yuan and Yin [2010] extend the asymmetric Laplace distribution idea in Yu and Moyeed [2001] for quantile regression with independent data to clustered data, where the conditional distribution of the response is assumed to follow an asymmetric Laplace distribution with the mean depending on the covariates and the skewness parameter depending on the quantile level of interest. The conditional quantile regression model includes subject specific random intercepts to account for within−subject dependence in longitudinal studies. But a model with random intercepts only cannot obviously account for between−clusters heterogeneity associated with given explanatory variables. Geraci and Bottai [2014] propose a class of models, called linear quantile mixed models (LQMMs), which include both random intercepts and random slopes.

Consider longitudinal data with repeated measurements in the form $(y_{ij}, \boldsymbol{x}_{ij}^T)$, for

$j = 1, \ldots, n_i, i = 1, \ldots, N$, where y_{ij} is the jth measurement on the ith subject. For a given $\tau \in (0, 1)$,a conditional quantile mixed model for continuous response y_{ij} is defined by

$$Q_{y_{ij}|\boldsymbol{b}_i}(\tau|\boldsymbol{x}_{ij}, \boldsymbol{b}_i) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i \tag{1.8}$$

where $\boldsymbol{z}_{ij}$ denotes a $q \times 1$ subset of covariates $\boldsymbol{x}_{ij}$, $\boldsymbol{b}_i$ is a $q \times 1$ vector of random effects. $Q_{y_{ij}|b_i}(\cdot)$ is the inverse of the cumulative distribution function of the response conditional on a location-shift random effect b_i. The expression 1.8 is equivalent to

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i + \epsilon_{ij} \tag{1.9}$$

with $Q_{\epsilon_{ij}}(\tau|\boldsymbol{x}_{ij}, \boldsymbol{b}_i) = 0$. The error term ϵ_{ij} are assumed to be independently distributed as ALD and the random effect $\boldsymbol{b}_i$ are also independent and follow a multivariate distribution. Usually $\boldsymbol{b}_i$ and ϵ_{ij} are independent of each other. The conditional density function can be written as

$$f(y_{ij}|\beta, \boldsymbol{b}_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau(\frac{y_{ij} - \mu_{ij}}{\sigma})\right\},$$

where $\mu_{ij} = x_{ij}^T\beta + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i$ is the linear predictor of the τth quantile. The random structure above allows to account for between-individual heterogeneity associated with given explanatory variables. A recent review of linear quantile regression models for longitudinal observations is provided by Marino and Farcomeni [2015]. The majority of the econometrics literature that studies quantile regression models for panel data with fixed effects propose inference procedures based on the assumption that the number of periods goes to infinity when the sample size goes to infinity. This assumption allows to estimate unobservable fixed effects. Canay [2011] introduces a two-step estimator for panel data quantile regression models which gets rid of the fixed effects under the assumption that these effects are location shifters. Lamarche [2010] elaborates the asymptotic theory of penalized quantile regression estimators for panel data by deriving an asymptotic approximation for the optimal value of the

tuning parameter. Galvao [2011] studies a quantile regression dynamic panel model with fixed effects.

Jang and Wang [2015] approximate the central density by linearly interpolating the conditional quantile functions of the response at multiple quantiles and estimate the tail densities by adopting extreme value theory. Through joint-quantile modeling, their method can yield the joint posterior distribution of quantile coefficients at multiple quantiles and meanwhile avoid the quantile crossing issue.

1.4 Bayesian Nonparametric Modeling for Quantile Regression

Recently Bayesian nonparametric methods have been studied and developed extensively, Müller and Quintana [2004] provide an overview of the respective methodologies. By modeling the response distribution and the regression function nonparametrically, Bayesian nonparametric modeling provides a flexible framework for the general regression problem. Instead of defining priors as probability distributions for parameters in parametric Bayesian analysis, nonparametric Bayesian's prior beliefs are expressed as random probability measures assigned directly to the set of probability distributions. Dirichlet process (DP), which can be understood as an infinite dimensional probability distribution over probability distributions, was originally formalized by Ferguson [1973] and Antoniak [1974]. And Dirichlet process mixtures are used to model the joint distribution of response and covariates, from which full inference is obtained for the desired conditional distribution for response given covariates. Many papers have been devoted to developing the practicality of using DP priors [Escobar and West, 1995, MacEachern and Müller, 1998, Neal, 2000].

Yu and Moyeed [2001] propose a Bayesian approach by noting that minimizing equation (1.5) is equivalent to maximizing a likelihood function under the asymmetric Laplace error distribution. However, it has limitations regarding modeling skewness and tail behavior because the same parameter determines both skewness and quantile

level. Nonparametric error distributions based on Dirichlet process mixture models will be discussed later in this section.

Dirichlet Process Mixture Models (DPMM)

Consider a model for joint distribution of the response, y, and the vector of covariates, $\boldsymbol{x}$, which in general comprises both continuous and categorical covariates. We use a nonparametric mixture model for the joint density $\boldsymbol{z} = (y, \boldsymbol{x})$,

$$f(\boldsymbol{z}; G) = \int k(\boldsymbol{z}; \boldsymbol{\theta}) dG(\boldsymbol{\theta})$$

with a parametric kernel density $k(\boldsymbol{z}; \boldsymbol{\theta})$ and a random mixing distribution G that is modeled non-parametrically by using an infinite dimensional Bayesian model. Following Ferguson [1973], a distribution G on Φ follows a Dirichlet process $DP(\alpha G_0)$ if, given an finite measurable partition, $A_1, ..., A_k$ of Φ, the joint distribution of $(G(A_1), ..., G(A_k))$ is Dirichlet $(\alpha G_0(A_1), ..., \alpha G_0(A_k))$, where $G(A_i)$ and $\alpha G_0(A_i)$ denote the probability of set A_i under G and G_0, respectively. The base distribution G_0 has a parametric form on Φ and $E(G(A)) = G_0(A)$. G_0 acts as a prior distribution for component parameters in the DPMM discussed later in this section. The concentration or precision parameter $\alpha > 0$ controls the variance of the DP. As α increases, a sample G is more likely to be close to base G_0. A draw from the Dirichlet process is a probability distribution. The most commonly used DP definition is its constructive definition by Sethuraman [1994], the stick-breaking construction, which characterized DP realizations as countable mixtures of point masses. A random distribution $G \sim DP(\alpha, G_0)$, then the stick-breaking representation of G is as the following

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k},$$

where δ_{θ_k} is a point mass at θ_k. The locations of the point passes θ_k arise iid from G_0. The weight π_k arise from a stick breaking mechanism based on iid draws from a

Beta$(1, \alpha)$ distribution

$$\pi_k = \beta_k \prod_{j=1}^{k-1} (1 - \beta_j)$$

This construction emphasizes that the samples from a DP are discrete with probability one. The term stick-breaking comes from the interpretation of mixing proportion π_k; which is given by successively breaking a unit-length stick into infinitely many pieces.

Suppose in a hierarchical model, the data $\boldsymbol{z}_i = (\boldsymbol{x}_i, y_i : i = 1, \ldots, n)$ are from a parametric distribution f with latent mixing parameters θ_i, a DP prior can be placed on the parameter distribution to ensure greater model flexibility and robustness. The general Bayesian hierarchical model by linking DP to nonparametric Bayesian modeling can then be written as:

$$\begin{aligned} z_i|\theta_i &\overset{\text{ind}}{\sim} k(z_i|\theta_i), \qquad i = 1, \ldots, n \\ \theta_i &\overset{\text{iid}}{\sim} G, \\ G &\sim DP(\alpha, G_0) \end{aligned}$$

This is often called Dirichlet process mixture model (DPMM), that we use a DP prior on the parameter and then complete the model by introducing a likelihood. According to the discreteness property of the DP and its stick-breaking representation, this model implies $z_i \sim \sum_{k=1}^{\infty} \pi_k k(z_i|\theta_k)$, where θ_k are infinite samples from G_0. Under this formulation, the DPMM is interpreted as a flexible mixture model in which the number of components is infinite. The assumption of exchangeability [De Finetti, 1935] makes sure that the joint distribution is invariant to the order in which observations are assigned to clusters. Given the heterogeneity in random effects, we are also particularly interested in classifying individuals in clusters, a nonparametric Dirichlet Process prior seems to be a reasonable alternative to place on random effects.

Subgroup Identification through Dirichlet Process

There are two schemes in clustering: one produces a hierarchical sequence of partitions, and the other allocates observations into proposed clusters. The latter are based on the concept that the observations come from a heterogeneous population consisting of several clusters. Each cluster can be modeled by a distinct parametric distribution and a mixture of these clusters, a finite mixture model, is used to model the heterogeneity of the overall population. More explicitly, given observations $\boldsymbol{y} = \{y_1, ..., y_n\}$, assume that there are K clusters differentiated by θ_k, which can be either a scaler or a vector. Let $f(\cdot|\theta_k)$ be the density of cluster k; the finite mixture model is of the form

$$f(x|\theta) = \sum_{k=1}^{K} \pi_k f(x|\theta_k)$$

with $\pi_k \geq 0$, for $k = 1, ..., K$, and $\sum_{k=1}^{K} \pi_k = 1$. The probability π_k also represents the prior probability that an observation comes from each cluster k. One of the challenges of using finite mixture model in clustering analysis is in determining the number of clusters K.

The number of clusters is automatically learnt from data through DP mixture models. DP mixtures models became popular in recent years because of the development of simple and efficient MCMC algorithms for posterior computation [MacEachern, 1994, Escobar and West, 1998, MacEachern and Müller, 1998]. It is a generalization of mixture models to infinite components in a Bayesian framework. The Pólya Urn scheme from Blackwell and MacQueen [1973], which most popular algorithms rely on, integrates out the infinite dimensional G to obtain the conditional prior of φ_i given $\boldsymbol{\varphi}^{(i)} = (\varphi_1, \ldots, \varphi_{i-1}, \varphi_{i+1}, \ldots, \varphi_n)'$:

$$\varphi_i|\boldsymbol{\varphi}^{(i)} \sim \frac{\alpha}{\alpha + n - 1} G_0 + \frac{1}{\alpha + n - 1} \sum_{j \neq i} \delta_{\varphi_j},$$

which generates new values from $\varphi_i \sim G_0$ with probability $\frac{\alpha}{\alpha+n-1}$ and otherwise sets

φ_i equal to an element of $\boldsymbol{\varphi}^{(i)}$ that is chosen by sampling from a discrete uniform distribution.

The tendency of the DP to cluster subjects into groups have identical coefficients is quite apparent from this structure. In particular, the n subjects are allocated to $K \leq n$ distinct values, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)'$, with the induced prior on k stochastically increasing with α and n [Antoniak, 1974]. Hence, although the expression is infinite, we obtain a finite number of K clusters on integrating out the random weights and random atoms. For this reason, DP methods are commonly used to allow uncertainty in the number of mixture components and for clustering.

Flexible Error Distribution

Another focus in Bayesian nonparametric modeling is to find flexible error distribution. It is natural to extend a parametric class of distributions to a nonparametric model through mixing. Kottas and Krnjanić [2009] develope two families of nonparametric prior distributions based on Dirichlet process mixture models. The first model is a general scale mixture of ALD to allow more flexible tail behavior but does not affect the skewness of the kernel of the mixture. The second is a flexible scale mixture of uniform densities which can capture the shape (skewness, tail behavior) of general unimodal error densities $f_p(\cdot)$, which is a representation of non-increasing densities on the positive real line.

Specifically, a density $f_p(\cdot)$ on positive real line (R^+) is non-increasing if and only if there exists a distribution function G on R^+ such that $f(t) \equiv f(t, G) = \int \theta^{-1} 1_{0,\theta}(t) dG(\theta)$. This result can be employed to provide a mixture representation for any unimodal density on the real line with pth quantile (mode) equal to zero, $\int \int k_p(\epsilon; \sigma_1, \sigma_2) dG_1(\sigma_1) dG_2(\sigma_2)$. Here G_1 and G_2 are general mixing distributions,

supported on R^+, and

$$k_p(\epsilon;\sigma_1,\sigma_2) = \frac{p}{\sigma_1}1_{(-\sigma_1,0)}(\epsilon) + \frac{1-p}{\sigma_2}1_{[0,\sigma_2)}(\epsilon) \tag{1.10}$$

with $0 < p < 1$, and $\sigma_r > 0$, $r = 1, 2$. Assuming independent DP priors for G_1 and G_2, we obtain the model for the error density $\int\int k_p(\epsilon;\sigma_1,\sigma_2)dG_1(\sigma_1)dG_2(\sigma_2)$, $G_r \sim DP(\alpha_r G_{r0}), r = 1, 2$. In the context of quantile regression, this model is sufficiently flexible to capture general forms of skewness and tail behavior [Kottas and Krnjajić, 2009].

Instead of posing a parametric likelihood, Reich et al. [2010] propose to model the likelihood nonparametrically by an infinite mixture of quantile-restricted two-component Gaussian mixtures and to accommodate error heteroscedasticity by specifying its form parametrically. They followed location-shift model with random subject effects b_i,

$$y_{ij} = x_{ij}^T\beta + b_i + x_{ij}^T\gamma e_{ij},$$

where $x_{ij}^T\gamma > 0$ for all x_{ij}. They defined the random error distribution as the infinite mixture

$$h(e|\boldsymbol{\mu},\boldsymbol{\sigma}^2) = \sum_{l=1}^{\infty} p_l f(e|\boldsymbol{\mu}_l,\boldsymbol{\sigma}_l^2,q_l),$$

where p_l is the mixing proportion with $\sum_{l=1}^{\infty} p_l = 1$, and the base density $f(e|\boldsymbol{\mu}_l,\boldsymbol{\sigma}_l^2,q_l)$ is the two component normal mixture

$$f(e|\boldsymbol{\mu}_l,\boldsymbol{\sigma}_l^2,q_l) = q_l\phi(\mu_{1l},\sigma_{1l}^2) + (1-q_l)\phi(\mu_{2l},\sigma_{2l}^2)$$

where $\phi(\mu,\sigma^2)$ is the normal density with mean μ and variance σ^2. The parameters are chosen to ensure $\int_{-\infty}^{0} f(e|\boldsymbol{\mu}_l,\boldsymbol{\sigma}_l^2,q_l)de = \tau$, that is, to make the τ-th quantile of e to be zero.

Learning about the Precision Parameters

The precision parameter, α, of the Dirichlet process is extremely important for the model. When α is small, then G tends to concentrate on a few atoms of probability.

When α is large, then G is a distribution with many support points and the nonparametric model is 'closer' to G_0, the baseline model. These features are to be borne in mind when considering priors for α. Escobar and West [1998] discuss various effects of the parameter α and then issues arising in learning about α within the MCMC analysis in their book. They developed a Gibbs sampler wherein draws from the conditional posterior distribution of α are computed by drawing successive samples from relatively familiar distributions. In Escobar and West [1995]'s approach, prior uncertainty in α is specified using a Gamma(a, b) distribution with fixed hyperparameters or a mixture of gamma distributions. In many cases the values of a and b are simply varied over a range that is thought to be reasonable, Escobar and West [1995] mentioned the possibility of using a Gamma(a, b) reference prior by letting $a \to 0, b \to 0$. There are at two advantages of choosing a gamma prior, 1)the distribution is proper so there is no need to prove posterior propriety. This is especially important in models with many parameters where propriety may be difficult or impossible to prove; 2)the gamma prior leads to full conditional distributions that are easy to sample. Gamma prior can sometimes be unstable, so a grid search should be used to find a proper value for α.

Some researchers adopt an empirical Bayes approach wherein the maximum likelihood estimate (MLE) of α is computed and inferences about all other model parameters are made conditional on the MLE [Liu, 1996]. MLE of α is computed by alternating between inference and estimation steps until convergence [Dorazio et al., 2008]. The shortcomings are, 1) it fails to account for errors in estimating α; 2) the empirical Bayes approach is computationally demanding, usually requiring repeated applications of the Gibbs sampler or other Markov chain Monte Carlo algorithm; 3) MLE of α can be unstable in some situations [Kyung et al., 2010].

Chapter 2

A Bayesian Nonparametric Quantile Regression for Repeated Measures with a Flexible Error Distribution

In longitudinal studies, several repeated measurements or contaminated replicates are often taken on the same subject with errors. Bayesian approach for the analysis of repeated measures offers a flexible way of combining data with prior information and inference can be always provided without the need for approximations using modern computation methods. However, quantile regression is not equipped with a parametric likelihood, and therefore, Bayesian quantile regression for repeated measures is challenging. As mentioned in 1.2.3, there exist several likelihood estimation methods in parametric and semiparametric frameworks in literature. Approaches extended from Yu and Moyeed [2001] are based on the asymmetric Laplace distribution for the errors. Note that in ALD, the same parameter determines both skewness and quantile level, hence limiting its flexibility in modeling skewness and tail behavior. In this chapter, we propose a Bayesian nonparametric quantile regression for repeated measures with error using a flexible distribution, which can capture the shape (e.g., skewness, tail behavior) of any unimodal error density. DP mixture models are fit by marginalizing the random mixing distribution over its DP prior and using the resulting Pólya urn representation for the latent mixing parameters in sampling from the posterior.

The outline of this chapter is as follows. Section 2.1 describes the proposed method

and prior specification. Section 2.2 discusses the computational details and posterior inference. Simulation studies to assess the performance of the proposed method are presented in Section 2.3. An application to a data set from orthodontic clinic is illustrated in Section 2.4. Finally, conclusions and discussion are presented in Section 2.5. The technical details are given in Appendix A.

2.1 Model Setup

Suppose we observe the data in the form $(y_{ij}, X_{ij}, Z_{ij}), i = 1, \ldots, n; j = 1, \ldots, n_i$. y_{ij} is the jth observation of the continuous response variable on the ith subject, X_{ij} is the $p \times 1$ covariate vector for fixed effects, and the first covariate is one corresponding to the intercept. For a given subject, the covariates in X_i, for instance, gender, are assumed to remain constant across the repeated measurements. Z_{ij} is the $q \times 1$ covariate vector for random effects. A linear mixed model has the form

$$y_{ij} = X_{ij}\boldsymbol{\beta} + Z_{ij}\boldsymbol{b_i} + \epsilon_{ij}, i = 1, \ldots, n; j = 1, \ldots, n_i, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the unknown regression coefficient vector associated with p model covariates, $\boldsymbol{b_i} = (b_{i,1}, \ldots, b_{i,q})^T$ are i.i.d. random effects vectors associated with q model covariates.

To allow more flexibility in Model 2.1 we assume that the true density of the regression errors $\boldsymbol{\epsilon}$ belongs to a non-parametric scale mixture of uniform densities. Specifically, A density $f_p(\cdot)$ on positive real line (R^+) is non-increasing if and only if there exists a distribution function G on R^+ such that $f(t) \equiv f(t, G) = \int \theta^{-1} 1_{0,\theta}(t) dG(\theta)$. The result on the real line with pth quantile (mode) equal to zero,

$$\int \int k_p(\epsilon; \sigma_1, \sigma_2) dG_1(\sigma_1) dG_2(\sigma_2).$$

Here G_1 and G_2 are general mixing distributions, supported on R^+, and

$$k_p(\epsilon; \sigma_1, \sigma_2) = \frac{p}{\sigma_1} 1_{(-\sigma_1, 0)}(\epsilon) + \frac{1-p}{\sigma_2} 1_{[0, \sigma_2)}(\epsilon) \tag{2.2}$$

with $0 < p < 1$, and $\sigma_r > 0$, $r = 1, 2$. With latent mixing parameters σ_{1i} and σ_{2i} for each response, the model can be represented by

$$\boldsymbol{y}_i, y_{ij}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i} \overset{\text{ind}}{\sim} k_p(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}; \sigma_{1i}, \sigma_{2i}), i = 1, \ldots, n, j = 1, \ldots, n_i$$

The likelihood becomes

$$\prod_{i=1}^{n}\prod_{j=1}^{n_i} \frac{p}{\sigma_{1i}} 1_{(-\sigma_{1i},0)}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) + \frac{1-p}{\sigma_{2i}} 1_{[0,\sigma_{2i})}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) \qquad (2.3)$$

2.2 Prior Specification

Independent normal priors are given to fixed effect $\boldsymbol{\beta}$. σ_{1i}, σ_{2i} are assigned independent DP priors.

$$\begin{aligned}
\sigma_{ri}|G_r &\overset{iid}{\sim} G_r, r = 1, 2, i = 1, \ldots, n \\
G_r|\alpha_r, d_r &\sim DP(\alpha_r G_{r0}), r = 1, 2, \\
\boldsymbol{b_i}|G_b &\sim G_b, i = 1, \ldots, n \\
G_b|\alpha_b, G_{0b} &\sim DP(\alpha_b G_{0b})
\end{aligned}$$

Various noninformative prior distributions for σ have been suggested in Bayesian literature and software, including an improper uniform density on σ_{ri} [Gelman, 2006], proper distributions such as $p(\sigma_{ri}^2) \sim$ inverse-gamma. Many Bayesians have preferred the inverse-gamma prior family, possibly because its conditional conjugacy suggested clean mathematical properties. However, by writing the hierarchical model in the above form, we see conditional conjugacy in the wider class of half-t distributions on σ_{ri} , which include the uniform and half-Cauchy densities on σ_{ri} (as well as inverse-gamma on σ_{ri}^2 as special cases. From this perspective, the inverse-gamma family has nothing special to offer, and we prefer to work on the scale of the standard deviation parameter σ_{ri}, which is typically directly interpretable in the original model. We consider uniform distributions Uniform(0, A) for G_{r0}. As illustrated in Gelman

[2006], the choice of "noninformative" prior distribution can have a big effect on inferences, especially for problems where the number of clusters (repeated measures) is small or the cluster-level variance is close to zero. For a finite but sufficiently large A, inferences are not sensitive to the choice of A. Random effects $\boldsymbol{b}_i$ are assigned independent DP priors. A normal base measure with zero mean is specified to G_{0b}. An assumed DP prior with zero mean may still imply a non-zero mean for random effect distribution, to ensure that $E(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$, the additional constraint $E(\boldsymbol{b}_i = \boldsymbol{0})$ is needed. DP precision parameters play an important role and the optimal values of α_r, $r = 1, 2, b$ are decided by the grid search.

2.3 Posterior Computation

Bayesian approach provides a flexible framework for representing the intricate nature of the word and our knowledge of it, and the Monte Carlo methods provide a corresponding flexible mechanism for inference within this framework. MCMC integration methods, especially the Metropolis-Hastings algorithm and the Gibbs sampler have emerged as extremely popular tools for the analysis of complex statistical models in a short period of computers' development. Properly defined and implemented, MCMC methods enable users to successively sample values from values from a convergent Markov chain, and reduce complex high-dimensional problems to a sequence of much lower-dimensional ones. Details of MCMC methods can be found in [Geyer, 2011].

Markov Chain Sampling Methods for Dirichlet Process Mixture Models

Many different Markov chain Monte Carlo sampling techniques have been developed for making posterior inferences from DP mixture models; Neal (2000) is a good reference. The posterior inference methods for DP mixture models we use in this study is a combination of MCMC methods from Escobar and West [1995] and Neal [2000]. They are based on a marginalization of the random mixing distributions over their DP

priors [Blackwell and MacQueen, 1973]. The discreteness of the DP priors [Blackwell and MacQueen, 1973, Sethuraman, 1994] induces a clustering. We use $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$ as illustrations. The most direct approach to sampling for DP mixture models is to repeatedly draw values for each θ_i from its conditional distribution given both the data and the θ_j for $j \neq i$ (written as θ_{-i}). This conditional distribution is obtained by combining the likelihood for θ_i that results from y_i having distribution $F(\theta_i)$, which will be written as $F(y_i, \theta_i)$ and since the observations are exchangeable, the prior conditional on θ_{-i}, which is

$$\theta_i|\theta_{-i} \sim \frac{\alpha}{\alpha + n - 1} G_0 + \frac{1}{\alpha + n - 1} \sum_{j \neq i} \delta(\theta_j)$$

When combined with the likelihood, this yields the following conditional distribution for use in Gibbs sampling

$$\theta_i|\theta_{-i}, y_i \sim r_i H_i + \sum_{j \neq i} q_{i,j} \delta(\theta_j)$$

Here, H_i is the posterior distribution for θ based on the prior G_0 and the single observation y_i with likelihood $F(y_i, \theta)$. The values of the $q_{i,j}$ and of r_i are defined by

$$\begin{aligned} q_{i,j} &= bF(y_i, \theta_j) \\ r_i &= b\alpha \int F(y_i, \theta) dG_0(\theta) \end{aligned}$$

where b is such that $\sum_{j \neq i} q_{i,j} + r_i = 1$.

There are two main approaches for Dirichlet Process mixtures, MCMC and variational inference [Escobar and West, 1995, MacEachern and Müller, 1998, Blei and Jordan, 2005]. Use of Dirichlet process mixture models has become computationally feasible as the development of Markov chain methods for sampling from the posterior distribution of the component distribution or of the associations of mixture components with observations. In the Chinese restaurant prior, we can easily swap customer i to the last customer to arrive by taking advantage of exchangeability, which yields a straightforward formula for the conditional for Gibbs sampling. Methods based

on Gibbs sampling can be easily implemented for models based on conjugate prior distribution. Lavine and West [1992] use the Gibbs sampling approach to calculate normal mixture models. Such iterative resampling method is applied to the mixture model by introducing the classification variables which identify data points with specific components. However, when non-conjugate priors are used, it is often difficult to perform numerical integration in straightforward Gibbs sampling. Later, a Markov chain method for sampling from the posterior distribution of a Dirichlet process mixture model was presented, extended Gibbs sampling for the indicators specifying which mixture component is associated with each observation by using a set of auxiliary parameters [MacEachern and Müller, 1998, Neal, 2000]. The method is simple to implement and more efficient than previous ways of handling general Dirichlet process mixture models with non-conjugate priors. Blei and Jordan [2005] presented a variational inference algorithm, which is a class of deterministic algorithms that convert inference problems into optimization problems.

Gibbs Samplig Algorithm with Auxiliary Parameter

MacEachern and Müller [1998] devised an approach to handle non-conjugate priors that uses a mapping from a set of auxiliary parameter to parameters in use. Models with non-conjugate priors can be handled by applying Gibbs sampling to a state that has been extended by the addition of auxiliary parameters [Neal, 2000]. In this approach, the auxiliary parameters are regarded as existing only temporarily; this allows more flexibility in constructing algorithms. The permanent state of the Markov chain will be x, but a variable y will be introduced temporarily and discarded during Markov chain simulation. The implementation of our DP is performed using a modified algorithm proposed in Neal [2000] (Algorithm 8), which is appropriate for models with no closed form solutions and therefore applies Gibbs sampling with the inclusion of auxiliary parameters.

Let c_i indicates the "latent class" associated with observation y_i, with the numbering of the c_i being of no significance. We can use this technique to update c_i for a Dirichlet process mixture model without having to integrate with respect G_0. The permanent state of the Markov chain consist of $\boldsymbol{c} = (c_1, ..., c_n)$ and $\boldsymbol{\phi} = (\phi_c : c \in \{c_1, .., c_2\})$. For each class, c, the parameters ϕ_c determine the distribution of observations from that class; all such ϕ_c is denoted by ϕ When c_i is updated, we will introduce temporarily auxiliary parameters that represent potential values for θ that are not associated with any other observations. We then update c_i by Gibbs sampling with respect to the distribution that includes these auxiliary parameters. Figure 2.1 represents the conditional prior distribution for a new observation using auxiliary parameters. In this setup, the number of auxiliary parameters $m = 3$. The component for the new observation is chosen from among the four components associated with other observations plus three possible new components ($m = 3$), with parameters, ϕ_5, ϕ_6, ϕ_7, drawn independently from G_0. The probabilities used for this choice are shown at the top. The probability of c_i being equal to a c in $\{1, \ldots, k^-\}$ will be $n_{-i,c}/(n-1+\alpha)$, where $n_{-i,c}$ is the number of times c occurs among the c_j for $j \neq i$. The probability of c_i having some other value will be $\alpha/(n-1+\alpha)$, which is split equally among $m = 3$ components introduced. The dashed arrows illustrate the possibilities of choosing an existing component, or a new component that uses one of the auxiliary parameters in the figure.

To perform a Gibbs sampling update for c_i in this representation of the posterior distribution, c_i must be either one of the components associated with other observations or one of the auxiliary components that were introduced, we can easily do Gibbs sampling by evaluating the relative probabilities of these possibilities. Once a new value for c_i has been chosen, we discard all ϕ values that are not now associated with an observation.

Let the state of the Markov chain consist of $\boldsymbol{c} = (c_1, \ldots, c_n)$ and $\boldsymbol{\phi} = (\phi_c : c \in$

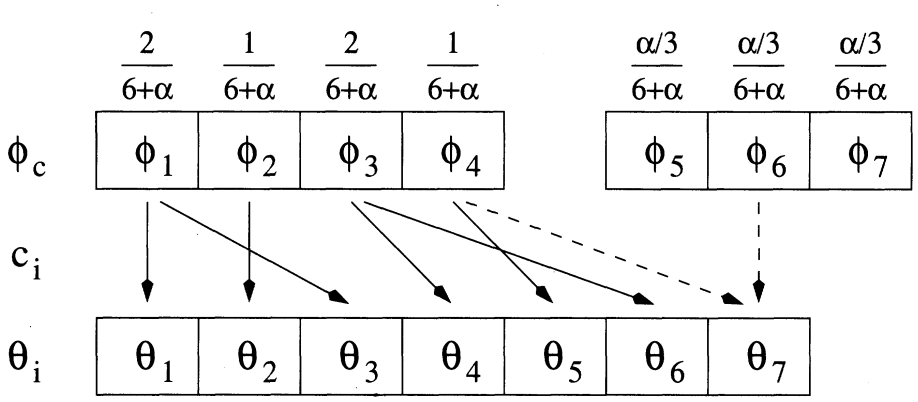


Figure 2.1 Using auxiliary parameters to represent conditional prior distribution for a new observation.

$\{c_1, \ldots, c_n\}$). The algorithms can be summarized by repeated sampling as follows:

- For $i = 1, \ldots, n$:, let k^- be the number of distinct c_j for $j \neq i$, and let $h = k^- + m$. Label these c_j with values in $\{1, \ldots, k^-\}$. If $c_i = c_j$ for some $j \neq i$, draw values independently from G_0 for those ϕ_c for which $k^- < c \leq h$. If $c_i \neq c_j$ for all $j \neq i$, let c_i have the label $k^- + 1$, and draw values independetnly from G_0 for those G_0 for those ϕ_c for which $k^- + 1 < c \leq h$. Draw a new value for c_i from $\{1, \ldots, h\}$ using the following probabilities:

$$P(c_i = c | c_{-i}, y_i, \phi_1, \ldots, \phi_h) = \begin{cases} b\frac{n_{-i,c}}{n-1+\alpha} F(y_i, \phi_c) & \text{for } 1 \leq c \leq k^-, \\ b\frac{\alpha/m}{n-1+\alpha} F(y_i, \phi_c) & \text{for } k^- \leq c \leq h \end{cases} \tag{2.4}$$

  where $n_{-i,c}$ is the number of c_j for $j \neq i$ that are equal to c, and b is the appropriate normalizing constant. Change the state to contain only those ϕ_c that are now associated with one or more observation.

- For all $c \in \{c_1, \ldots, c_n\}$: Draw a new value from $\phi_c | y_i$ such that $c_i = c$ by using the Metropolis-Hastings algorithm to ϕ_c that leaves this distribution invariant.

The posterior distributions for σ_{ri} and b_i can be handled by applying Gibbs sampling with auxilliary parameters described in this section. The likelihood of i-th observation is $F(\boldsymbol{y_i}, \boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{ri}) = \prod_{j=1}^{n_i} p(\boldsymbol{y_i} | \boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i})$. When updating σ_{ri}, we

assume both $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ are "known". By placing a Uniform distribution to base distribution G_{r0}, the conditional distribution for σ_{ri} is

$$p(\sigma_{ri}|\boldsymbol{y_i}, G_{r0}) \propto \frac{\alpha_r}{\alpha_r + n - 1}\text{Unif}(d_r)F(\boldsymbol{y_i}, \sigma_{ri}) + \frac{1}{\alpha_r + n - 1}\sum_{g \neq i}\delta(\sigma_{rg})F(\boldsymbol{y_i}, \sigma_{ri})$$

For each $\boldsymbol{b_i}$, suppose a Multivariate normal hyperprior for base distribution G_{0b} is $N_p(\mathbf{0}, D)$, the conditional distribution

$$\begin{aligned} p(\boldsymbol{b_i}|\boldsymbol{y_i}) &\propto \text{prior}(\boldsymbol{b_i}|D)L(\boldsymbol{y_i}|\boldsymbol{b_i}) \\ &\propto N_p(\mathbf{0}, D)p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i}) \\ &\propto \frac{1}{(2\pi)^{1/2} \mid D \mid^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{b_i} - \mathbf{0})^T D^{-1}(\boldsymbol{b_i} - \mathbf{0})\Big) \\ &\times \prod_{j=1}^{n_i}\frac{p}{\sigma_{1i}}1_{(-\sigma_{1i},0)}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) + \frac{1-p}{\sigma_{2i}}1_{[0,\sigma_{2i})}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) \end{aligned}$$

Bayesian Computing for Other Parameters

Updating fixed effects $\boldsymbol{\beta}$ requires Metropolis-Hastings algorithms as full conditionals distributions are not recognizable. Suppose the prior is a multivariate normal distribution $N_p(\mathbf{0}, \Sigma_0)$, the conditional distribution of $\boldsymbol{\beta}$ is

$$\begin{aligned} p(\boldsymbol{\beta}|\boldsymbol{y_i}) &\propto N_p(\mathbf{0}, \Sigma_0)\prod_{i=1}^{n} p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i}) \\ &\propto \frac{1}{(2\pi)^{1/2} \mid \Sigma_0 \mid^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{0})^T \Sigma_0^{-1}(\boldsymbol{\beta} - \mathbf{0})\Big) \\ &\times \prod_{i=1}^{n}\prod_{j=1}^{n_i}\frac{p}{\sigma_{1i}}1_{(-\sigma_{1i},0)}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) + \frac{1-p}{\sigma_{2i}}1_{[0,\sigma_{2i})}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) \end{aligned}$$

For the (t)-th iteration, update $\boldsymbol{\beta}$ by using the following Metropolis-Hastings algorithm. Draw $\boldsymbol{\beta}^*$ from the proposal distribution $N_p(\boldsymbol{\beta}^{t-1}, T_0)$, and take

$$\boldsymbol{\beta}^t = \begin{cases} \boldsymbol{\beta}^* & \text{with probability } \min(r, 1), \\ \boldsymbol{\beta}^{t-1} & \text{with probability1-min}(r, 1), \end{cases} \tag{2.5}$$

where

$$r = \frac{p(\boldsymbol{\beta}^*|\boldsymbol{y_i})/J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1})}{p(\boldsymbol{\beta}^{t-1}|\boldsymbol{y_i})/J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)}$$

since the jump distribution is symmetric, $J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1}) = J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)$,

$$J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1}) = \frac{1}{(2\pi)^{1/2} \mid T_0 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{t-1})^T T_0^{-1} (\boldsymbol{\beta}^* - \boldsymbol{\beta}^{t-1}) \right)$$

$$J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*) = \frac{1}{(2\pi)^{1/2} \mid T_0 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}^{t-1} - \boldsymbol{\beta}^*)^T T_0^{-1} (\boldsymbol{\beta}^{t-1} - \boldsymbol{\beta}^*) \right)$$

Lastly, we can directly sample the covariance matrix D using normal-inverse Wishart conjugacy. Given the mean vector, the conjugate prior for D^- is the Wishart distribution, a generalization of the gamma distribution to p dimensions.

Wishart$_p(\nu_0, \Phi_0)$, with two parameters ν_0 and Φ_0, is

$$p(D^{-1}|\nu_0, \Phi_0) \propto \mid D^{-1} \mid^{(\nu_0 - p - 1)/2} \exp\left(-\frac{1}{2}\text{tr}(\Phi_0^{-1} D^{-1}) \right)$$

The posterior distribution of D^{-1} can be derived from Bayes Theorem, Wishart$_p(n + \nu_0, \left(\Phi_0^{-1} + \sum_{i=1}^n \boldsymbol{b_i}\boldsymbol{b_i}^T\right)^{-1})$, with more details described in the Appendix.A.

$$\begin{aligned} p(D^{-1}|\boldsymbol{b_i}) &\propto \prod_{i=1}^n p(\boldsymbol{b_i}|D) p(D^{-1}|\nu_0, \Phi_0) \\ D^{-1}|\boldsymbol{b_i} &\sim \text{Wishart}_p(n + \nu_0, \left(\Phi_0^{-1} + \sum_{i=1}^n \boldsymbol{b_i}\boldsymbol{b_i}^T\right)^{-1}) \end{aligned}$$

2.4 Simulation Studies

We next access the performance of the proposed method through simulation studies, under which the model's superiority is reflected compared with traditional mixture models and frequentist method.

Simulation Study I

We generate the simulation data from the model,

$$y_{ij} = \beta_0 + u_i + (\beta_1 + \nu_i)x_{ij} + \beta_2 z_{ij} + (1 + \gamma x_{ij})\epsilon_{ij} \tag{2.6}$$

Table 2.1 Simulation study scenarios

Model description	(n,M)	μ	ν	ϵ	Γ
(1) location shift symmetric	(5,50)	N(0,5)	-	N(0,5)	0
(3) location shift symmetric	(5,50)	N(0,5)	-	t(3)	0
(5) location shift asymmetric	(5,50)	N(0,5)	-	χ^2_2	0
(6) location shift asymmetric	(5,50)	t(3)	-	χ^2_2	0
(14) heteroscedastic symmetric	(10,100)	N(0,5)	-	N(0,5)	0.25
(16) heteroscedastic asymmetric	(10,100)	t(3)	-	χ^2_2	0.25
(19) location shift heavy-tailed $\text{cor}(\mu,\nu)>0$	(10,100)	t(3)	t(3)	t(3)	0
(23) location shift with 5% contamination	(10,100)	N(0,5) + N(0,50)	-	χ^2_2	0

where $i = 1, \ldots, M, j = 1, \ldots, n$, with a sample size of M subjects with n repeated measures. $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (100, 2, 1)'$, u_i and ν_i are cluster-specific random effects, $x_{ij} = \delta_i + \eta_{ij}, \delta_i \sim N(0,1), \eta_{ij} \sim N(0,1)$ and $z_{ij} \sim \text{Binom}(1, 0.5)$. The model setup follows an simulation study in Geraci and Bottai [2014] and we select eight model scenarios from the paper to evaluate our proposed estimator under different assumptions. We consider a sample size of $M = 50$ and 100 subjects, and include both symmetric and skewed measurement error scenarios. For symmetric errors, we use Normal and heavy tailed (Student t) errors. For skewed cases, we let errors follow χ^2_2 distribution. Models with both homogenous errors and heteroscedasticity are considered. The true random effects distribution is assumed to be symmetric, either normal or t distribution. A summary for different data generating distributions and sample sizes is given in Table 2.1.

In all cases, datasets were generated independently 300 times. Posterior can be sampled by MCMC techniques for DP mixtures described before. For proposed method BNQM, we do grid search to find the optimal combination of three precision

parameters ($\alpha \in \{0.1, 1, 5, 20\}$) and optimal values are chosen with the smallest DIC (Deviance Information of Criterion). DIC is a Bayesian model comparison criterion based on trade-off between the goodness of fit and complexity of the model. In Gelman et al. [2014], for a likelihood $p(y|\theta)$

$$DIC = 2\hat{D}_{avg}(y) - D_{\hat{\theta}}(y)$$

Based on the lowest DIC value, we set $\alpha_1 = 0.1$, $\alpha_2 = 5$ and $\alpha_b = 20$ for $\tau = 0.1$ and 0.5, and for $\tau = 0.9, \alpha_1 = 20$, $\alpha_2 = 1$ and $\alpha_b = 20$.

We evaluate the methods by using sample standard deviation (SSD), estimated standard errors (ESE), the 95% coverage probabilities (CP) and mean squared errors (MSE) of each model at the quantile coefficients. The SSD is calculated as the standard deviation of the posterior means of the parameter estimates. On the other hand, the ESE is the mean of the standard deviations of the parameter estimates. The 95% CP is the posterior probability of the 95% CIs to include the true value. The mean squared errors can be obtained by

$$\frac{1}{N}\sum_{m=1}^{N}(\hat{\beta}_j^m - \beta_j)^2$$

where $\hat{\beta}_j^m, j = 0, 1, 2$ are the parameter estimates in the mth simulation run. The posterior means were used as the estimates.

To compare the results of our Bayesian nonparametric quantile mixed model (BNQM) with other approaches, we also fit LQMMs that assumes asymmetric Laplace error. The package 'lqmm' is available from the Comprehensive R Archive Network (CRAN) to fit LQMMs, which were estimated with the gradient search algorithm described in Geraci and Bottai [2014]. The default normal random effects assumption leads to a Gauss-Hermite quadrature, while robust (Gauss-Laguerre) random effects estimation is also provided. K = 11 nodes was used in model scenarios 1, 3, 5, and 6. For data generated under models 14, 16, 19 and 23, the number of nodes was set to K = 17. The tolerance parameter and the maximum number of iterations were

set to, respectively, 10^{-3} and 500 for the likelihood, and to 10^{-5} and 10 for the scale parameter.

For each case, the simulation is run 300 times at quantile levels $\tau \in \{0.5, 0.75, 0.9\}$. In Table 2.2, the results are summarized for scenario (1), assuming homoscedastic normal random errors with mean 0 and variance 5, $\epsilon_{ij} \sim N(0, 5)$ and normal random effects, $u_i \sim N(0, 5)$. Relative bias, averaged over the simulation realizations, was low at all three quantiles. The parameters were estimated using the proposed model fairly accurately with high coverage probabilities including the true values. The results are very similar comparing to the LQMM estimator which indicates both methods work well under normal assumptions.

The convergence was evaluated by a critical examination of the trace plots, and by using Geweke's method [Geweke, 1992]. With 4,000 MCMC samples, we discard the first 2,000 samples as burn-in, and estimate parameters by averaging the posterior means. A trace plot can be used to determine whether the chain is mixing well and reached its stationary distribution. The trace plots for fixed effect coefficients β_0, β_1 and β_2 and the random effects coefficients b_0, b_1, and b_2 are shown in Figure 2.2. The trace plots indicate very good convergence of the sample chains of the parameters. The chains appear to be centered on their true parameter values with little variation. Thus, we could conclude the chains have reached their stationary distributions and inference can be drawn from them.

Geweke [1992]'s convergence diagnostic for Markov chain is based on a test of equality of the means of the first and last part of a Markov chain (the first 10% and the last 50% has been used in this dissertation). The Geweke's statistic has an asymptotically standard normal distribution and the corresponding P-value can be obtained. If the P-value is greater than a certain significance level, say 0.05, we conclude that the means of the first and last parts of the chain are equal and hence the chain has reached its stationary distribution. Table 2.3 shows the Geweke's

Table 2.2 Simulation Results for Scenario (1)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	0.000	0.496	0.489	0.93	0.247
		β_1	-0.002	0.177	0.196	0.98	0.031
		β_2	0.041	0.348	0.370	0.96	0.122
0.5	BNQM	β_0	-0.001	0.435	0.186	0.88	0.202
		β_1	-0.007	0.215	0.143	0.99	0.046
		β_2	0.095	0.425	0.277	0.94	0.189
0.75	LQMM	β_0	0.001	0.529	0.560	0.95	0.301
		β_1	-0.000	0.189	0.215	0.98	0.036
		β_2	0.048	0.372	0.410	0.96	0.140
0.75	BNQM	β_0	0.000	0.424	0.209	0.92	0.180
		β_1	-0.004	0.196	0.146	0.98	0.039
		β_2	0.023	0.383	0.272	0.91	0.147
0.9	LQMM	β_0	0.002	0.580	0.659	0.97	0.382
		β_1	-0.001	0.222	0.272	0.97	0.049
		β_2	0.016	0.424	0.515	0.98	0.179
0.9	BNQM	β_0	-0.003	0.455	0.306	0.90	0.305
		β_1	-0.006	0.203	0.172	0.98	0.041
		β_2	0.000	0.367	0.297	0.94	0.134

diagnostics for the fixed effect coefficients and the random effects. All the P-values corresponding to the Geweke's statistics are greater than significance level (0.05). Therefore, we could conclude that there is no difference between the sample means of the first 10% and the last 50% of the chains and the chains have converged well.

Results are summarized for scenario (3), (5), (6) in Table 2.4, 2.5, 2.6. Scenario (3) assumes random errors to be t distribution with 3 degrees of freedom, and normal random effect, $N(0, 5)$. Scenario (5) assumes random errors to be χ^2 distribution with 2 degrees of freedom, and normal random effect, $N(0, 5)$. Scenario (6) assumes

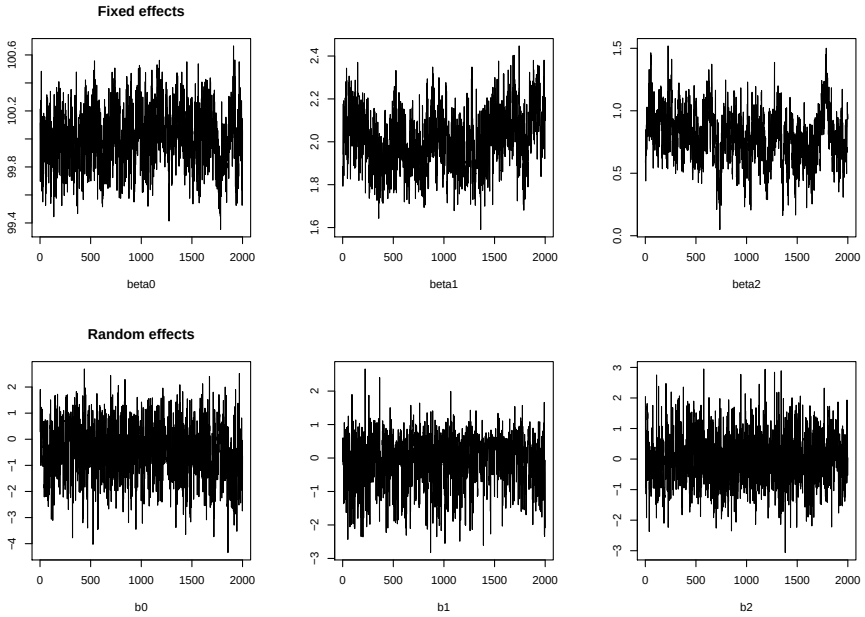


Figure 2.2 Trace plot at quantile level 0.5 in Scenario (1)

Table 2.3 Geweke's Statistics for Markov Chains of the Proposed Model in Scenario (1)

Parameter	Z score	P-value
β_0	-0.4607	0.3588
β_1	0.2242	0.3890
β_2	1.9751	0.0567
b_{0i}	0.8343	0.2817
b_{1i}	0.3882	0.3882
b_{2i}	0.2026	0.2026

random errors to be χ^2 distribution with 2 degrees of freedom, and random effect, $t(df = 3)$. All three scenario models have homoscedastic random errors. The relative bias, averaged over the simulation realizations, sample standard deviations (SSD), estimated standard deviations (ESD) and the corresponding 95% coverage probabilities (CPs) for the fixed effect coefficients using the proposed model and LQMM are presented. Estimation bias was low at all considered quantiles for one dimensional

random-effects models. As expected, the performance of our proposed method was satisfactory under symmetric and skewed measurement errors. Mixture distribution of two uniform densities is sufficiently flexible to capture general forms of skewness and tail behavior.

Scenario (14) and (16) have heteroscedastic random errors. Results are summarized in Table 2.7 and Table 2.8. Scenario (14) assumes random errors to be $N(0, 5)$, and random effect, $N(0, 5)$. Scenario (16) assumes random errors to be χ^2 distribution with 2 degrees of freedom, and random effect to be $t(3)$. Overall, the performance of both methods was satisfactory. In the context of quantile regression, our proposed scale mixtures of uniform densities can capture the shape (e.g., skewness, tail behavior) of any unimodal error density.

Scenario (19) assumes heavy tailed random error and correlated random effects (Table 2.9). We observed that an increase of bias in the slope of x for LQMM estimator that was also mentioned in Geraci and Bottai [2014], specifically the relative bias is -0.213 when $\tau = 0.75$, -0.182 when $\tau = 0.9$; their corresponding variances are inflated. According to the authors, random effects distributions characterized by heavy tails may require using a larger number of nodes to reduce the estimation bias. In our proposed method, flexible nonparametric prior models for the random effects demonstrated the superiority of nonparametric mixture models over parametric models. Typically, this normality assumption for the random effects is made for the sake of convenience, rather than from some theoretical justification, and may not actually hold. Potentially, the parametric random effect assumption could be relaxed by using a DP mixture for the unknown random effects distribution in generalized linear mixed model [Kleinman and Ibrahim, 1998a]. Due to data sparsity in the tail areas, estimates from quantile regression are often unstable at tails, especially for heavy-tailed distributions.

Outlier contamination in the random intercept (scenario 23, Table 2.9) did not

impact the estimation bias in both methods.

Table 2.4 Simulation Results for Scenario (3)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	0.000	0.527	0.478	0.91	0.278
		β_1	0.000	0.135	0.146	0.98	0.018
		β_2	0.018	0.216	240	0.97	0.047
0.5	BNQM	β_0	-0.003	0.374	0.176	0.94	0.170
		β_1	-0.011	0.152	0.145	0.94	0.024
		β_2	0.095	0.282	0.225	0.90	0.079
0.75	LQMM	β_0	0.004	0.581	0.555	0.89	0.484
		β_1	0.005	0.142	0.167	0.97	0.020
		β_2	0.044	0.249	0.276	0.96	0.064
0.75	BNQM	β_0	0.004	0.792	0.193	0.91	0.842
		β_1	-0.005	0.138	0.120	0.91	0.019
		β_2	-0.046	0.227	0.212	0.93	0.054
0.9	LQMM	β_0	0.008	0.598	0.670	0.78	0.999
		β_1	0.004	0.188	0.219	0.97	0.035
		β_2	-0.007	0.321	0.377	0.98	0.103
0.9	BNQM	β_0	0.005	0.819	0.203	0.93	0.241
		β_1	-0.006	0.128	0.109	0.92	0.017
		β_2	-0.059	0.218	0.197	0.89	0.051

Simulation Study II

In this section, we compare our proposed method to Reich's marginal random effects model [Reich et al., 2010]. We consider a model 2.7 with one random effect to coordinate with the marginal model.

$$\text{Model 1} : y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_{0i} + \epsilon_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i \qquad (2.7)$$

Table 2.5 Simulation Results for Scenario (5)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	0.003	0.535	0.496	0.87	0.407
		β_1	-0.004	0.136	0.159	0.97	0.019
		β_2	-0.035	0.258	0.281	0.96	0.068
0.5	BNQM	β_0	0.003	0.363	0.193	0.88	0.222
		β_1	-0.009	0.105	0.104	0.95	0.011
		β_2	0.061	0.174	0.184	0.96	0.034
0.75	LQMM	β_0	0.005	0.561	0.561	0.87	0.539
		β_1	-0.001	0.191	0.205	0.96	0.036
		β_2	-0.007	0.375	0.381	0.94	0.140
0.75	BNQM	β_0	-0.002	0.510	0.182	0.91	0.291
		β_1	-0.010	0.129	0.112	0.91	0.017
		β_2	-0.050	0.194	0.189	0.94	0.040
0.9	LQMM	β_0	0.005	0.683	0.831	0.96	0.699
		β_1	-0.004	0.265	0.302	0.98	0.070
		β_2	-0.037	0.531	0.582	0.96	0.282
0.9	BNQM	β_0	-0.005	0.464	0.290	0.90	0.449
		β_1	-0.013	0.112	0.104	0.93	0.013
		β_2	-0.022	0.199	0.196	0.94	0.040

where $x_{1ij} \sim N(5,1)$, $x_{2ij} \sim$ Binomial distribution (1, 0.5), a normal random effect distribution for $b_{0i} \sim$ Normal(0,1); ϵ_{ij} is the random error from Gamma distribution Gamma(2,2). The coefficients are $(\beta_0, \beta_1, \beta_2) = (1, 1, 1)$. A sample size of 100 with 10 repeated measures is applied, and the simulation is repeated 200 times at three quantile levels, $(0.1, 0.5, 0.9)$. We evaluate the methods by using sample standard deviation (SSD), estimated standard errors (ESE), the 95% coverage probabilities (CP) and mean squared errors (MSE) of each model at the quantile coefficients.

The results are summarized in Table 2.11. We notice that the coverage proba-

Table 2.6 Simulation Results for Scenario (6)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	0.003	0.318	0.332	0.90	0.167
		β_1	0.006	0.121	0.127	0.95	0.015
		β_2	-0.022	0.233	0.251	0.96	0.054
0.5	BNQM	β_0	0.002	0.321	0.173	0.97	0.148
		β_1	-0.004	0.095	0.091	0.95	0.009
		β_2	-0.028	0.161	0.165	0.95	0.027
0.75	LQMM	β_0	0.005	0.561	0.561	0.87	0.539
		β_1	0.001	0.191	0.205	0.96	0.036
		β_2	-0.007	0.375	0.381	0.94	0.140
0.75	BNQM	β_0	-0.002	0.388	0.180	0.96	0.183
		β_1	-0.011	0.117	0.100	0.91	0.014
		β_2	-0.004	0.187	0.177	0.94	0.035
0.9	LQMM	β_0	0.003	0.731	0.787	0.97	0.629
		β_1	0.005	0.263	0.280	0.96	0.069
		β_2	0.036	0.520	0.570	0.98	0.271
0.9	BNQM	β_0	-0.003	0.771	0.284	0.86	0.685
		β_1	-0.006	0.099	0.094	0.93	0.010
		β_2	0.002	0.198	0.183	0.92	0.039

bilities at $\tau = 0.9$ for Reich's method is relatively low, 0.49 for β_0 and 0.52 for β_1. Coverage probabilities of 95% credible intervals are comparable for the two methods at quantile $\tau = 0.1$ and $\tau = 0.5$. At quantile level $\tau = 0.1$ and 0.9, SSD for β_0 is much higher for Reich's method , which indicates the posterior means from simulated data sets are spread apart. And it is not surprising to see their corresponding MSEs are high, MSE = 0.330 when $\tau = 0.1$ and MSE = 0.255 when $\tau = 0.9$. There is little difference when $\tau = 0.5$. These imply Reich's marginal random effects model does not have a stable estimation in tail quantiles. Later, in [Smith et al., 2015], a

Table 2.7 Simulation Results for Scenario (14)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	0.000	0.370	0.381	0.95	0.136
		β_1	0.003	0.085	0.098	0.97	0.007
		β_2	-0.007	0.156	0.167	0.97	0.024
0.5	BNQM	β_0	-0.001	0.506	0.173	0.91	0.278
		β_1	-0.042	0.256	0.141	0.95	0.073
		β_2	0.082	0.392	0.240	0.93	0.160
0.75	LQMM	β_0	0.001	0.398	0.435	0.96	0.163
		β_1	-0.03	0.098	0.110	0.92	0.015
		β_2	-0.007	0.184	0.186	0.96	0.034
0.75	BNQM	β_0	-0.004	0.439	0.289	0.91	0.349
		β_1	0.004	0.369	0.244	0.97	0.136
		β_2	0.020	0.298	0.302	0.95	0.089
0.9	LQMM	β_0	0.003	0.368	0.476	0.95	0.205
		β_1	-0.065	0.122	0.128	0.71	0.046
		β_2	0.001	0.227	0.229	0.95	0.052
0.9	BNQM	β_0	-0.004	0.460	0.306	0.92	0.356
		β_1	-0.065	0.327	0.370	0.91	0.138
		β_2	0.022	0.331	0.270	0.91	0.110

simulation study on clustered data were conducted and it showed low coverages in some quantiles. But it's not clear whether difficulty in estimating tail quantile is the cause from their averaged results. They used a copula model to resemble the usual mixed model in that covariates affect both the marginal population distribution via fixed effects and subject-specific distributions via random slopes. The copula approach maintains the marginal distributions of the population-level quantile effects while accounting for within-subject dependence, enabling inference at the population and subject levels. The use of marginal models for repeated measurements is often

Table 2.8 Simulation Results for Scenario (16)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	0.002	0.263	0.242	0.82	0.117
		β_1	-0.007	0.065	0.066	0.94	0.005
		β_2	-0.021	0.108	0.111	0.96	0.012
0.5	BNQM	β_0	0.001	0.244	0.142	0.88	0.074
		β_1	-0.069	0.099	0.095	0.98	0.036
		β_2	-0.067	0.173	0.171	0.92	0.034
0.75	LQMM	β_0	0.002	0.294	0.279	0.89	0.125
		β_1	-0.032	0.090	0.093	0.85	0.016
		β_2	-0.002	0.172	0.171	0.95	0.030
0.75	BNQM	β_0	-0.003	0.272	0.171	0.93	0.154
		β_1	0.070	0.248	0.161	0.87	0.097
		β_2	-0.035	0.185	0.175	0.92	0.035
0.9	LQMM	β_0	0.002	0.390	0.413	0.94	0.185
		β_1	-0.069	0.133	0.147	0.71	0.065
		β_2	-0.009	0.284	0.284	0.95	0.080
0.9	BNQM	β_0	-0.004	0.476	0.324	0.88	0.409
		β_1	0.015	0.240	0.291	0.98	0.060
		β_2	-0.020	0.202	0.184	0.92	0.041

discouraged [Lindsey and Lambert, 1998, Crouchley and Davies, 1999, Lee et al., 2004] as predictions correspond to hypothetical individuals only. Lee et al. [2004] stated that differences are mainly caused by the choice of unidentifiable constraints on the random effects and they discuss the advantages of conditional models over marginal models. Vaida and Blanchard [2005] argued that the conditional likelihood should be used when the focus is on clusters and marginal likelihood should be used if the research aim is population focused. Marino and Farcomeni [2015] provides an overview distinguishing between these two approaches. For example, the interpre-

Table 2.9 Simulation Results for Scenario (19)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	-0.002	1.907	1.939	0.98	3.659
		β_1	-0.096	4.848	5.224	0.97	23.46
		β_2	-0.003	0.211	0.243	0.98	0.044
0.5	BNQM	β_0	-0.003	0.417	0.232	0.95	0.239
		β_1	-0.022	0.260	0.128	0.95	0.070
		β_2	0.061	0.204	0.213	0.93	0.045
0.75	LQMM	β_0	0.017	5.047	5.290	0.94	28.39
		β_1	-0.213	3.864	7.801	0.97	15.05
		β_2	-0.001	0.228	0.287	0.99	0.052
0.75	BNQM	β_0	0.006	0.216	0.125	0.91	0.455
		β_1	-0.000	0.239	0.077	0.98	0.057
		β_2	-0.059	0.178	0.157	0.87	0.035
0.9	LQMM	β_0	0.082	9.430	11.43	0.97	157.4
		β_1	-0.182	3.204	9.922	0.99	10.35
		β_2	0.035	0.323	0.398	0.99	0.105
0.9	BNQM	β_0	-0.003	0.383	0.518	0.91	0.241
		β_1	-0.004	0.270	0.474	0.94	0.073
		β_2	0.084	0.214	0.184	0.92	0.053

tation is different. In the marginal formulation, parameters describe the effect of covariates on the τ-th population response quantile. On the other hand, in the conditional framework, regression coefficients have an individual-specific interpretation allow to describe sources of unobserved heterogeneity that influence the dependence between longitudinal observations.

Trace plots for fixed effect coefficients for our proposed method are illustrated in Figure 2.3. These trace plots indicate very good convergence of the sample chains of the parameters and parameters estimates appear to be centered on their true values

Table 2.10 Simulation Results for Scenario (23)

p	Method	Coef	Relative Bias	SSD	ESE	CPs	MSE
0.5	LQMM	β_0	0.003	0.486	0.500	0.93	0.316
		β_1	0.005	0.083	0.099	0.99	0.007
		β_2	-0.001	0.122	0.141	0.98	0.015
0.5	BNQM	β_0	0.003	0.477	0.196	0.92	0.299
		β_1	0.010	0.113	0.102	0.95	0.013
		β_2	-0.051	0.183	0.181	0.92	0.036
0.75	LQMM	β_0	0.003	0.539	0.534	0.89	0.416
		β_1	0.003	0.109	0.132	0.98	0.012
		β_2	0.023	0.196	0.215	0.98	0.039
0.75	BNQM	β_0	-0.001	0.523	0.193	0.91	0.280
		β_1	-0.014	0.116	0.110	0.94	0.014
		β_2	-0.056	0.195	0.190	0.90	0.041
0.9	LQMM	β_0	0.004	0.730	0.640	0.84	0.725
		β_1	0.003	0.162	0.195	0.98	0.026
		β_2	0.039	0.309	0.359	0.97	0.097
0.9	BNQM	β_0	-0.003	0.828	0.294	0.87	0.771
		β_1	-0.013	0.108	0.103	0.95	0.012
		β_2	-0.044	0.206	0.195	0.92	0.044

with little variation. Comparisons between estimated and true errors are presented in Figure 2.4. The solid black line is simulated error (from Gamma distribution), and the colored dash curves are estimated errors at different quantile levels. We can see how close the simulation results are to the true errors in Figure 2.5 by centering errors at each quantile.

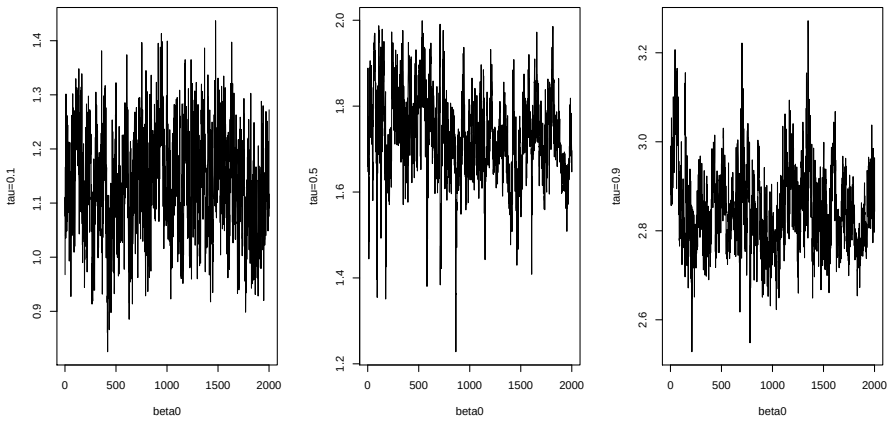


Figure 2.3 Trace plots of Proposed method at quantiles: 0.1, 0.5, 0.9

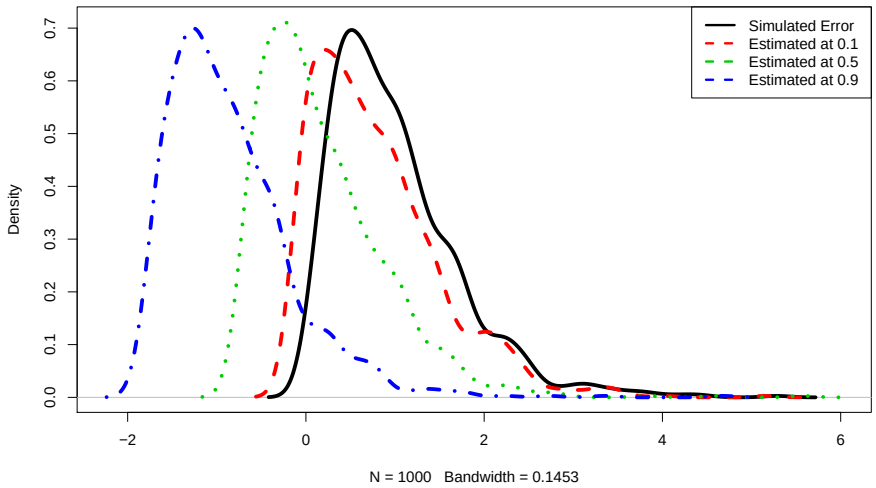


Figure 2.4 Errors of Proposed method at quantiles: 0.1, 0.5, 0.9

Table 2.11 Summaries and coverage probabilities of 95% intervals: comparing Reich's marginal random effects model and the proposed method

p	Method	Coef	Bias	SSD	ESE	CPs	MSE
0.1	Reich	β_0	0.101	0.551	0.149	0.87	0.330
		β_1	0.007	0.031	0.028	0.92	0.001
		β_2	0.003	0.035	0.032	0.96	0.001
0.1	BNQM	β_0	-0.059	0.113	0.083	0.85	0.018
		β_1	0.005	0.022	0.016	0.88	0.000
		β_2	0.017	0.035	0.033	0.93	0.002
0.5	Reich	β_0	0.041	0.210	0.145	0.82	0.046
		β_1	-0.007	0.041	0.028	0.85	0.002
		β_2	-0.005	0.044	0.035	0.90	0.002
0.5	BNQM	β_0	-0.009	0.171	0.189	0.95	0.029
		β_1	0.002	0.035	0.037	0.99	0.001
		β_2	0.006	0.078	0.079	0.98	0.006
0.9	Reich	β_0	0.090	0.498	0.166	0.49	0.255
		β_1	-0.011	0.099	0.032	0.52	0.010
		β_2	-0.020	0.103	0.077	0.84	0.011
0.9	BNQM	β_0	-0.011	0.129	0.173	0.97	0.018
		β_1	0.004	0.023	0.033	0.99	0.001
		β_2	0.001	0.041	0.069	0.99	0.002

Simulation Study III

We then investigate the performance of the proposed method under scenarios with possible skewness and multimodality, particularly when normality assumption of random effects distribution does not hold. Again, we consider the model (2.6) for generating the simulation data

$$y_{ij} = \beta_0 + u_i + (\beta_1 + \nu_i)x_{ij} + \beta_2 z_{ij} + (1 + \gamma x_{ij})\epsilon_{ij}$$

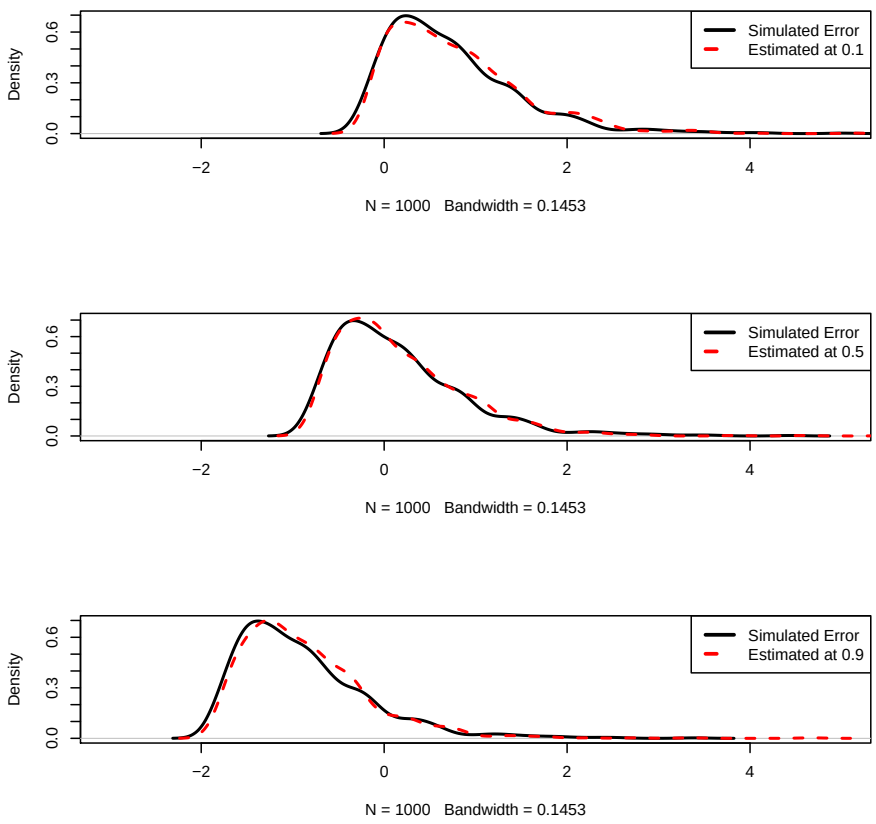


Figure 2.5 Errors comparison of Proposed method at quantiles: 0.1, 0.5, 0.9

where $i = 1, \ldots, m, j = 1, \ldots, n$, $(\beta_0, \beta_1, \beta_2) = (100, 2, 1)$, $z_{ij} \sim \text{Binom}(1, 0.5)$, $x_{ij} = \delta_i + \eta_{ij}, \delta_i \sim N(0, 1), \eta_{ij} \sim N(0, 1)$. A sample size of $n = 100$ subjects with $m = 10$ repeated measures for each is applied for each scenario. The design is similar as the previous scenarios except normal mixture random effects or random errors.

(a) $u_i \sim \frac{11}{18}N(-1.05, 0.2) + \frac{7}{18}N(1.65, 0.15), \nu = 0, \epsilon_{ij} \sim N(0, 5), \gamma = 0.25$. Intraclass correlation of (a) is 0.28.

(b) $u_i \sim \frac{11}{18}N(-1.05, 0.2) + \frac{7}{18}N(1.65, 0.15), \nu = 0, \epsilon_{ij} \sim \chi^2_2, \gamma = 0.25$. Intraclass correlation of (b) is 0.49.

(c) $u_i \sim \frac{11}{18}N(-0.35, 0.05) + \frac{7}{18}N(0.55, 0.10), \nu = 0, \epsilon_{ij} \sim 0.7 \times N(-1, 0.1) + 0.3 \times N(1, 0.1), \gamma = 0.25$. Intraclass correlation of model (c) is 0.22.

The relative bias of posterior means, sample standard deviations (SSD), estimated standard deviations (ESD) and the corresponding 95% coverage probabilities (CPs) for the fixed effect coefficients using the proposed model are summarized in Table 2.12. In case (a) and case (b), we assume a binormal mixture distribution for random intercept, u_i. Overall, the performance of BNQM is satisfied. Estimation bias of slope x of LQMM estimator is high, which may be caused by the violation of normal random effect assumption. The bias still exists using Gauss-Laguerre (type = 'robust') quadrature.

Model (c) assumes normal mixture errors, we use the random effect with smaller variance to ensure a moderate ICC. Overall, the performance is satisfied. The bias of tail quantiles is slightly higher than median for BNQM, which might be caused by the fact that nonparametric approaches restrict the error densities to necessarily have their modes at the quantile of interest. In Figure 2.6, we can see the estimated errors is very close to the binormal mixture errors after centering at each quantile.

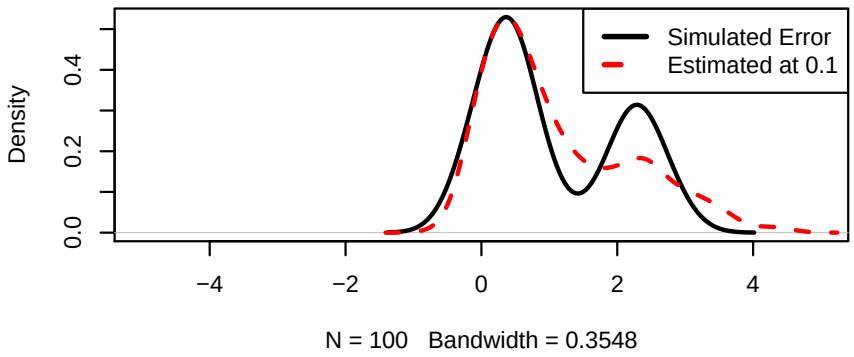
2.5 Orthodontic Growth Data

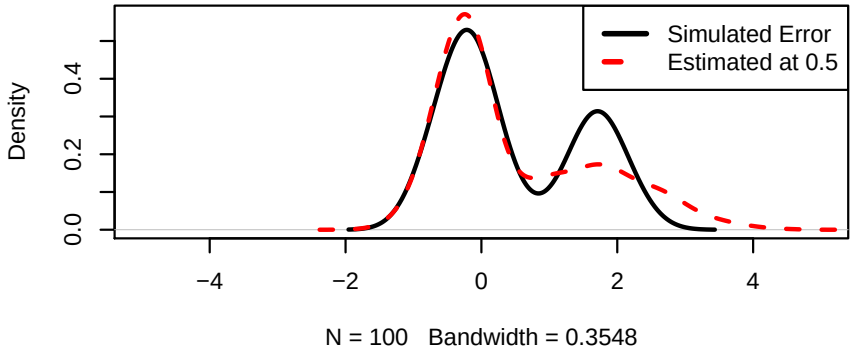
In this section, we analyze the Orthodontic data first reported in Potthoff and Roy (1964). Investigators at University of North Carolina Dental School followed the teeth growth of 27 children (16 males, 11 females) from age 8 until age 14. Every two years they measured the distance between the pituitary gland and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head. The data set is available in the package *nlme* and has been analyzed by several random effect models [Sheather, 2009, Geraci, 2014]. We are interested in applying our method to a subset (i.e. girls) of the data. Figure 2.7 shows a plot of distance against age for each of the 11 girls. The model we consider for subject $i(i = 1, 2, \ldots, 11)$ at age $j(j = 1, 2, 3, 4)$ is as follows:

$$\text{Distance}_{ij} = \beta_0 + b_{0i} + \beta_1 \text{Age}_j + e_{ij}$$

Table 2.12 Simulation Results (a), (b), (c)

p	Model	Coef	Bias	SSD	ESE	CPs	MSE	LQMM (Bias)
0.1	a	β_0	-0.007	0.482	0.524	0.70	0.735	-0.001
		β_1	-0.065	0.321	0.359	0.92	0.109	0.112
		β_2	-0.026	0.547	0.429	0.86	0.298	0.019
0.5	a	β_0	-0.001	0.520	0.445	0.83	0.282	0.000
		β_1	-0.014	0.238	0.180	0.85	0.057	0.001
		β_2	0.016	0.409	0.292	0.85	0.167	-0.020
0.9	a	β_0	-0.002	0.494	0.501	0.90	0.310	0.001
		β_1	0.049	0.266	0.384	0.99	0.088	-0.055
		β_2	-0.015	0.288	0.300	0.96	0.083	-0.020
0.1	b	β_0	-0.005	0.344	0.491	0.85	0.380	-0.002
		β_1	0.070	0.094	0.334	0.99	0.029	0.028
		β_2	0.072	0.211	0.402	0.90	0.050	-0.002
0.5	b	β_0	-0.001	0.262	0.126	0.84	0.080	0.002
		β_1	-0.027	0.128	0.085	0.99	0.020	-0.008
		β_2	-0.074	0.144	0.158	0.92	0.026	-0.020
0.9	b	β_0	-0.003	0.466	0.314	0.93	0.307	0.002
		β_1	-0.003	0.229	0.285	0.97	0.052	-0.066
		β_2	-0.049	0.186	0.176	0.92	0.037	-0.015
0.1	c	β_0	-0.001	0.116	0.042	0.91	0.032	-0.001
		β_1	0.088	0.046	0.031	0.89	0.024	0.015
		β_2	-0.018	0.113	0.052	0.94	0.013	0.005
0.5	c	β_0	0.001	0.144	0.059	0.96	0.040	0.001
		β_1	0.002	0.101	0.055	0.89	0.010	-0.004
		β_2	-0.034	0.111	0.086	0.92	0.014	-0.008
0.9	c	β_0	0.002	0.189	0.058	0.98	0.072	-0.001
		β_1	-0.095	0.191	0.055	0.88	0.083	-0.018
		β_2	-0.010	0.185	0.085	0.94	0.034	-0.001





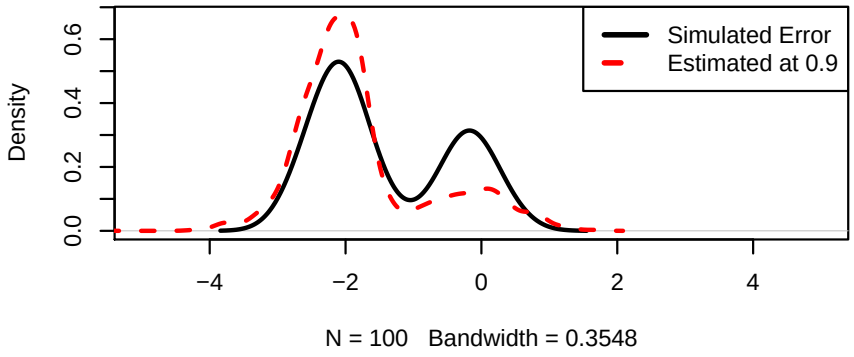


Figure 2.6 Errors in Scenario (c)

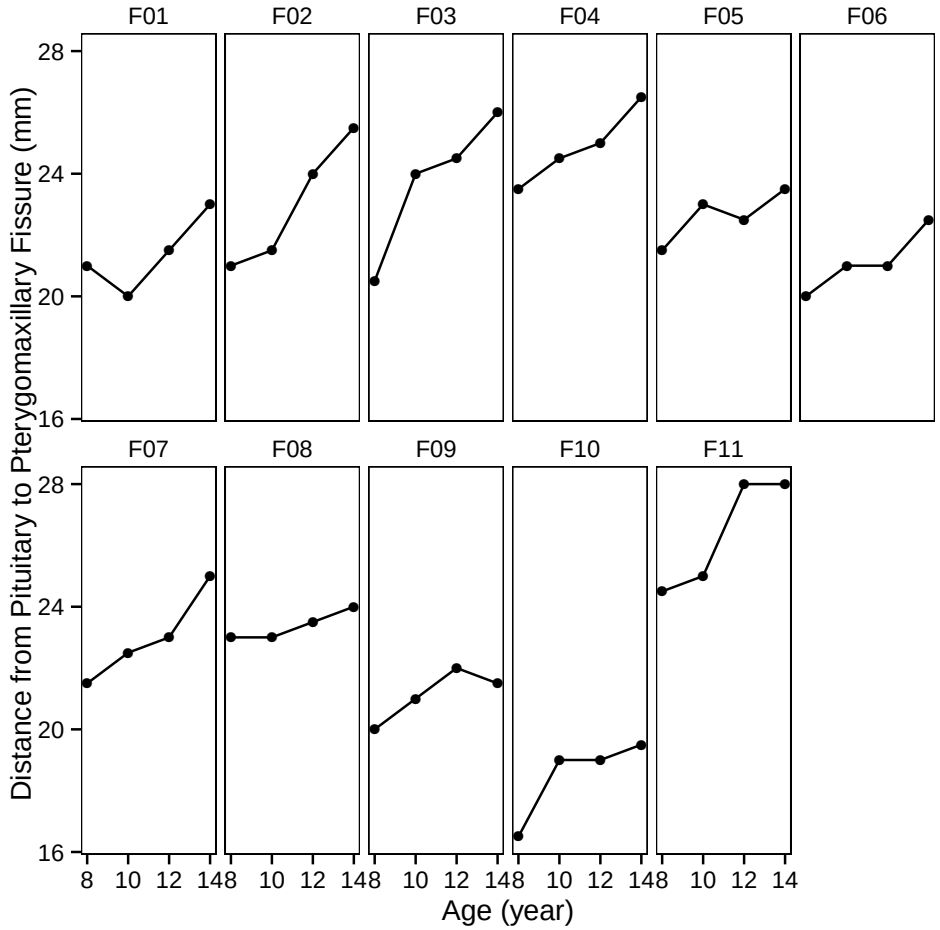


Figure 2.7 Plots of Distance against Age for each female

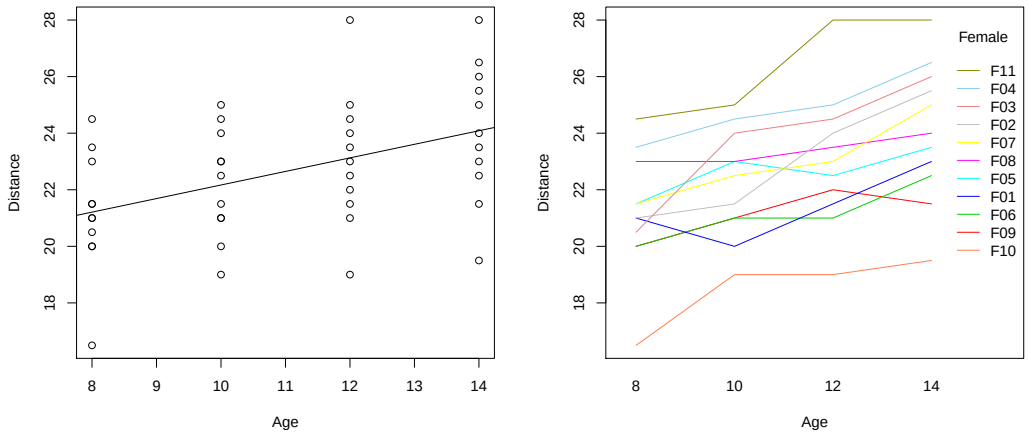


Figure 2.8 Relationship between Distance and Age for each female in one plot

where the distribution of random effect b_{0i} is assumed to have a Dirichlet Process prior. The error term e_{ij} unspecified and we only know its τth quantile equal to 0. The model is a random intercepts model since it assumes that the intercepts differ randomly across the 11 female subjects but distance increases linearly with age at the same fixed rate, and age is modeled as a fixed effect. Figure 2.8 shows the varying intercepts among subjects. The model assumes correlation between two distance measurements for the same child is constant no matter which ages they were taken. For example the ith child, two measurements $D_{i,j}$ and $D_{i,k}$ are taken at age j and k.

$$\begin{aligned}
\text{Corr}(\text{Distance}_{i,j}, \text{Distance}_{i,k}) &= \frac{\text{cov}(D_{ij}, D_{ik})}{\sigma_{ij}\sigma_{ik}} \\
&= \frac{\text{cov}(b_i + e_{ij}, b_i + e_{ik})}{\text{sd}(b_i + e_{ij})\text{sd}(b_i + e_{ik})} \\
&= \frac{\text{cov}(b_i, b_i)}{\sqrt{\sigma_b^2 + \sigma_e^2}\sqrt{\sigma_b^2 + \sigma_e^2}} \\
&= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}
\end{aligned}$$

The correlations between two distances measurements for the same female subject over each time interval range from 0.830 to 0.948. Thus, it seems the constant correlation over time assumption is reasonable for female subjects.

We first find the optimal precision parameters and summarize the fixed effect results at $\tau = 0.1, \ldots, 0.9$ in Table 2.13. Figure 2.9 presents a concise summary of the quantile regression results. Figure 2.10 presents a concise summary of the quantile regression results from LQMM. The solid line with filled dots represents the point estimates, $\hat{\beta}_0(\tau)$, with the shaded gray area depicting a 95% pointwise confidence band. Superimposed on the plot is a dashed line representing the ordinary least squares estimate of the mean effect, with two dotted lines again representing a 95% confidence interval for this coefficient. The intercept of the model may be interpreted as the estimated conditional quantile function of the distance distribution of girl at age 11 (Following the example in *lqmm* package, age is centered at 11 years so that

Table 2.13 Coefficient estimates for Orthodontic data set at $\tau = 0.1, \ldots, 0.9$

	τ								
Coeff	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	BNQR estimates								
$\beta_0(\tau)$	20.69	20.99	21.04	21.79	22.09	22.08	23.02	23.28	23.49
$\beta_1(\tau)$	0.45	0.46	0.45	0.43	0.45	0.49	0.49	0.49	0.48
	LQMM estimates								
$\beta_0(\tau)$	20.76	21.27	21.50	22.74	22.94	23.01	23.11	23.25	23.53
$\beta_1(\tau)$	0.25	0.42	0.50	0.41	0.44	0.42	0.46	0.50	0.54

Table 2.14 Quantile coefficient $\beta_0(\tau = 0.5)$: fixed effect and random effect

Subjects	F01	F02	F03	F04	F05	F06
LQMM	21.43	22.95	23.66	24.71	22.60	21.20
BNQM	20.97	22.47	23.10	24.30	22.35	20.88
Subjects	F07	F08	F09	F10	F11	
LQMM	22.95	23.30	21.20	18.74	26.11	
BNQM	22.68	23.33	20.88	18.27	23.99	

the cross-product between intercept and slope is zero). The rate of growth in girls is around 0.43 to 0.49 mm per year at all nine quantiles, which is almost unchanged as we assumed. Figure 2.11 and figure 2.12 are the trace plots and error comparisons, respectively. The simulations converge well and the errors are comparable between our proposed method and *lqmm*, and both have τth quantile equal to 0. We conclude the chains have reached their stationary distributions and inference can be drawn from them.

Combining with random effect, the intercept effect for each girl in two methods are compared in 2.14. Both method show that F10 has the smallest quantile intercept coefficient while F11 has the largest value.

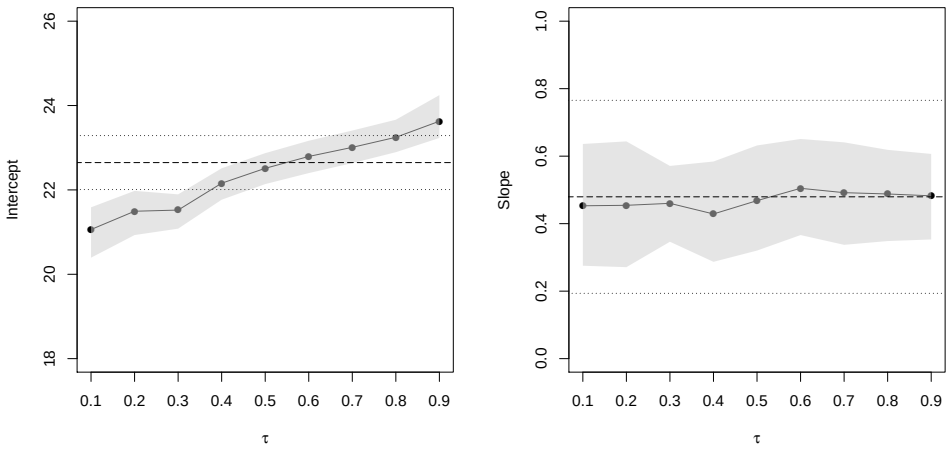


Figure 2.9 Regression quantiles for distance and 95% confidence intervals (shaded area). Least squares estimates (dashed lines) and 95% confidence intervals (dotted lines) are reported.

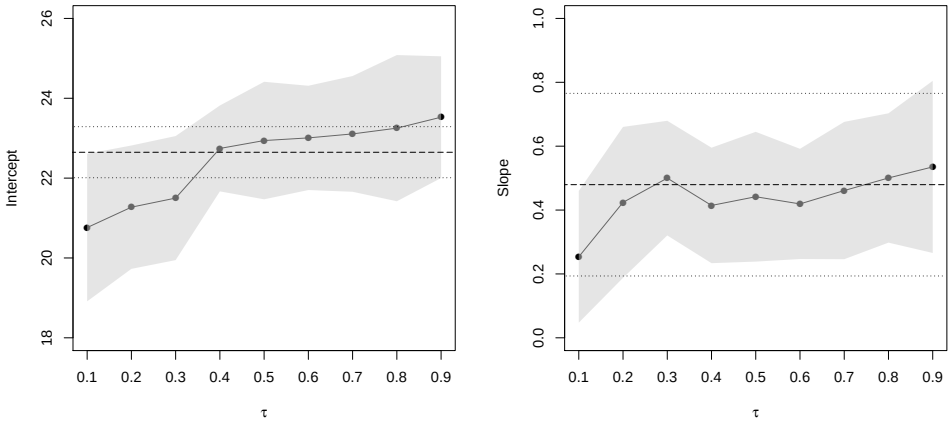


Figure 2.10 LQMM: Regression quantiles for distance and 95% confidence intervals (shaded area). Least squares estimates (dashed lines) and 95% confidence intervals (dotted lines) are reported.

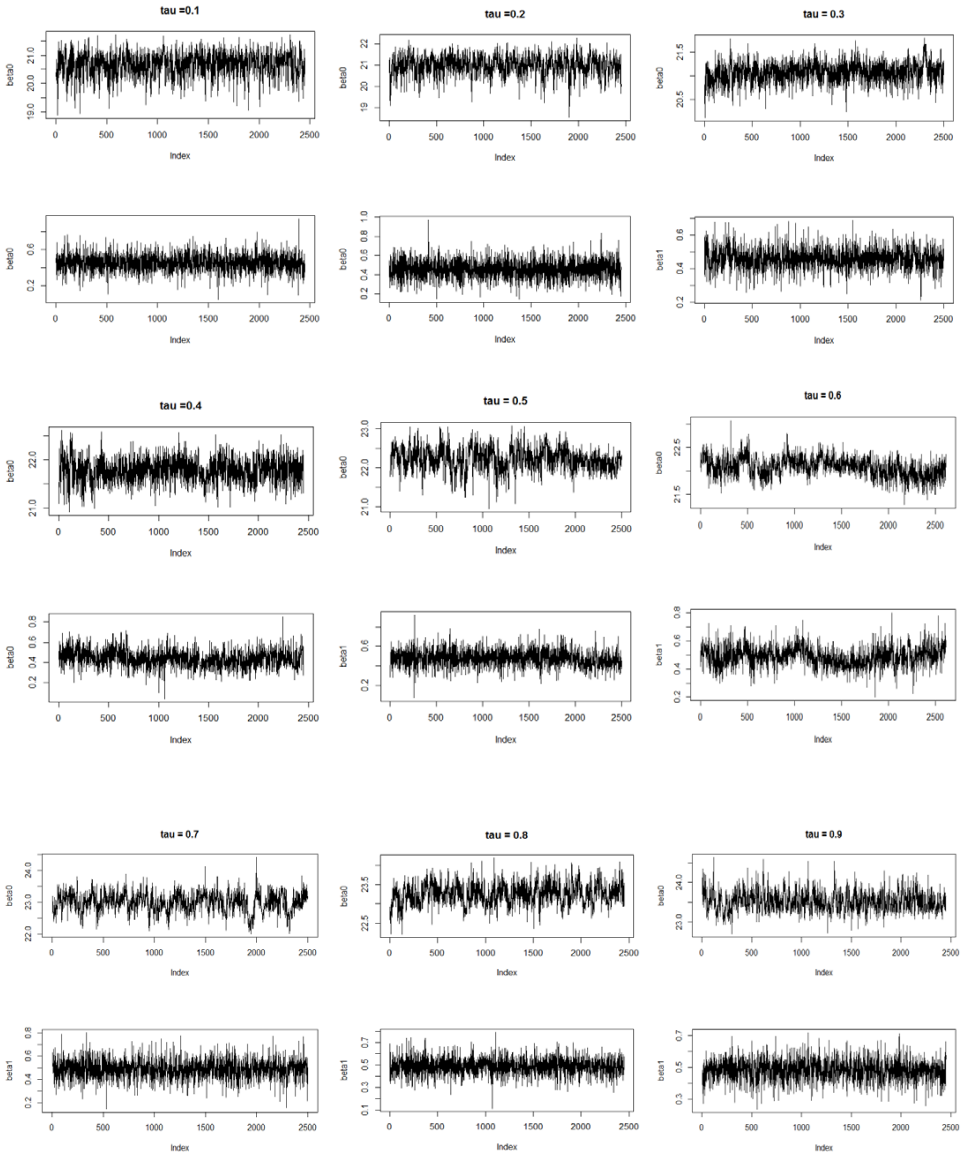


Figure 2.11 Trace plot for each quantile level

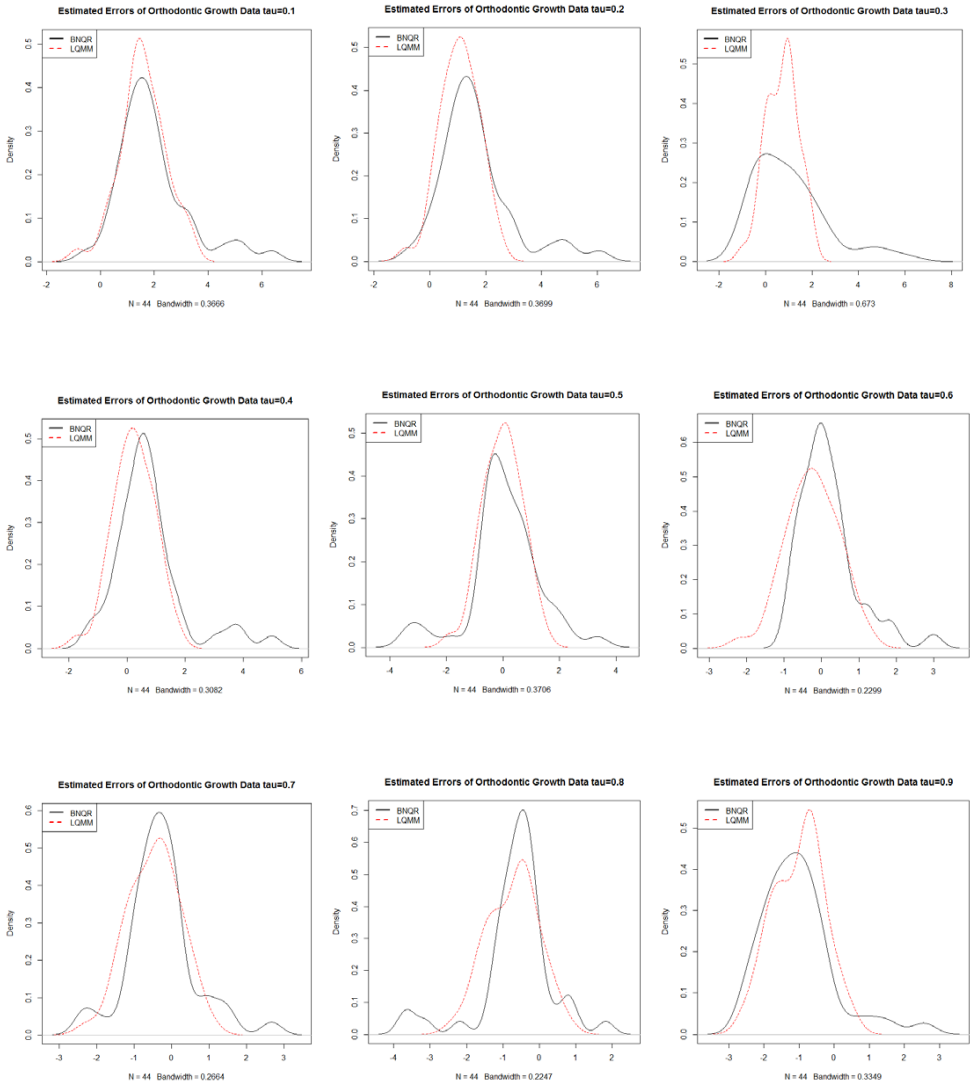


Figure 2.12 Error plot for each quantile level

Chapter 3

Clustered Interval-Censored Data and Regression Methods

In survival data, subjects are followed over time for a event of particular interest. One feature of survival data is censoring, which means the data is incomplete. But unlike missing data, censored data provide partial information because an observed failure time may fall into a certain range. There are usually three types of censored data, right-, left-, and interval-censored data. The right-censored or left-censored data mean that the failure times of interest are observed greater or less than censoring times, respectively. The third type of censoring is interval-censoring [Finkelstein, 1986, Sun, 2007], which is commonly used to describe a type of sampling scheme or incomplete data in medical studies that entail periodical follow-up. For instance in the acquired immune deficiency syndrome (AIDS) studies that concern the human immunodeficiency virus (HIV) infection and the AIDS incubation time (the time from HIV infection to AIDS diagnosis), an individual due for the scheduled visits may miss some visits and may return with a changed disease status, which gives rise to "interval-censored data". The time to emergence of a tooth in children is a typical example. We only know the true emergence time is greater than the last observation time at which the tooth has not emerged and less than or equal to the first observation time at which the tooth has emerged, thus giving an interval that contains the true (but unobserved) time of emergence of tooth. In such cases, the observed failure time falls into a certain interval with the exact occurrence times being unknown.

In multicenter clinical trials for HIV, some types of cancer, or other infectious diseases, clustered interval-censoring has become increasingly common. Collecting multiple observations on an experimental unit results in clustered data. The observations within a cluster are typically correlated. Mixed models can account for within-cluster correlation by including cluster-specific random effects in regression models. In this chapter, the context of the interval-censored data and statistical inferences approaches is reviewed, with particular focus on modeling approach to clustered interval-censored data.

3.1 Interval-Censored Data

We introduce three data-collection schemes of interval-censored data in this section. To proceed, some notions are established first. Let T be within some random time interval, denote U and V as the two examination times satisfying $U \leq V$ with probability 1. Assuming that each subject is observed twice, the observed data consists of

$$\{U, V, \delta_1 = I(T \leq U), \delta_2 = I(U < T \leq V), \delta_3 = I(V < T < \infty)\}$$

Note that by taking $U = V$, case I interval-censored data can be described by this formulation too. In this situation, each study subject is observed only once and the failure time of interest is known only to be smaller or greater than the observation time. In other words, the failure time of interest is either left- or right-censored and such data are also called case I interval-censored data.

One example is the lung tumor data, there is of great interest in knowing whether the environment accelerates the time to lung tumor [Dinse and Lagakos, 1983]. In this paper, the 114 RFM mice with nonlethal lung tumors are assigned to two treatments, conventional environment (CE) and germ-free environment (GE). The time to tumor occurrence is only known to be smaller or greater than the observed time of death.

That is, we only observe left- or right-censored failure times. As defined above, this type of data is called current status data or case I interval-censored data.

In correspondence, general interval-censored data that are not current status data are usually referred to as case II interval-censored data. Until the mid-1980s, many articles about general interval-censored data began to appear. A study conducted between 1976 and 1980 in Boston on early breast cancer patients was usually considered to be a good example [Finkelstein and Wolfe, 1985]. Physicians evaluated cosmetic appearance including breast retraction of the patients after treatment every four or six months. In the data set, patients who did not experience breast retraction are right censored observations denoted by intervals with no finite right end points; intervals are given by the last clinic visit at which breast retraction had not yet occurred and the first clinic visit time at which breast retraction was detected.

A natural generalization of case II interval-censoring is case k interval-censoring where there are $k > 2$ examination times for each individual. For each individual i there is a sequence of examination time

$$0 < Y_{i1} < Y_{i2}, \ldots, < Y_{in_i} < \infty$$

$i = 1, \ldots, n$. $\boldsymbol{Y_i} = (Y_{i1}, \ldots, Y_{in_i})$. Let T_i be the ith individual's unobserved event time. The observations become

$$\{U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, i = 1, \ldots, n\}$$

In practice, it is convenient to reduce case k interval-censoring to case II interval-censoring by considering the following three possibilities:

1. If the event has occurred by the first examination. Denote $U_i = Y_{i1}$, and let $V_i = Y_{i2}$. Furthermore, let $\delta_{1i} = I(T_i \leq U_i)$, $\delta_{2i} = I(U_i \leq T_i \leq V_i)$, and $\delta_{3i} = I(V_i < T_i)$. Then $\delta_{1i} = 1, \delta_{2i} = 0, \delta_{3i} = 0$;

2. T_i is known to be between a range (Y_{iL}, Y_{iR}), where Y_{iL} is the last examination time proceeding T_i, and Y_{iR} is the first examination time following T_i. Denote

$U_i = Y_{iL}$, and let $V_i = Y_{iR}$. Similarly, let $\delta_{1i} = I(T_i \leq U_i)$, $\delta_{2i} = I(U_i \leq T_i \leq V_i)$, and $\delta_{3i} = I(V_i < T_i)$. Then $\delta_{1i} = 0, \delta_{2i} = 1, \delta_{3i} = 0$;

3. At the last examination, the event has not occurred. Denote $V_i = Y_{in_i}$, then $\delta_{1i} = 0, \delta_{2i} = 0, \delta_{3i} = 1$.

Another generalization of the formulation is to assume that there exists a set of observation time points, say $U_1 \leq U_2 \cdots \leq U_k$, for each study subject, where k is a random integer. The observed information then has the form

$$\{(k, U_j, \delta_j = I(U_{j-1} < T \leq U_j)), j = 1, \ldots, k\},$$

Both representations give rise to the same likelihood function [Sun, 2007]. There are some other types of censoring data such as doubly interval-censored data which is rare compared to others [Sun, 1995].

3.2 Regression Analysis for Interval-Censored Data

For the analysis of interval-censored failure time data in medical and health studies, estimation of the cumulative distribution function (cdf) of survival time or the survival function is perhaps the most important and common task.

Let T denote the survival time of interest in a survival study and $F = Pr(T \leq t)$ its cdf. Suppose that observed data can be represented by $\{I_i\}_{i=1}^n$, where $I_i = (L_i, R_i]$ is the interval known to contain the unobserved survival time associated with the ith subject. If $L_i = 0$, we have a left-censored observation and if $R_i = \infty$, we have a right-censored observation. In this section, we will discuss two common regression analysis of interval-censored data. Other models such as the proportional odds regression model, additive hazards model and the logistic model for discrete interval-censored failure time data are discussed in [Sun, 2007].

Proportional Hazards (PH) Model

In survival analysis, the most commonly used regression model is perhaps the Cox proportional hazards (PH) model, which has the form

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t)\exp(\boldsymbol{X}^{'}\boldsymbol{\beta}),$$

where $\lambda_0(t)$ denotes the unknown baseline hazard function and $\boldsymbol{\beta}$ the p-dimensional vector of regression efficients. The coditional density and the survival function of T given $\boldsymbol{X}$ has the forms

$$f(t;\boldsymbol{X}) = \lambda_0(t)\exp(\boldsymbol{X}^{'}\boldsymbol{\beta})\exp(-\Lambda_0(t)\exp(\boldsymbol{X}^{'}\boldsymbol{\beta}))$$

$$S(t;\boldsymbol{X}) = \exp(-\Lambda_0(t)\exp(\boldsymbol{X}^{'}\boldsymbol{\beta})) = S_0(t)^{\exp(\boldsymbol{X}^{'}\boldsymbol{\beta})}$$

For inference about $\boldsymbol{\beta}$ and the cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, full likelihood approach maximizes L over $\boldsymbol{\beta}$ and $\Lambda_0(t)$ simultaneously [Finkelstein, 1986].

Consider a survival study that consists of n independent subjects and gives to interval-censored data

$$\{(L_i, R_i], \boldsymbol{X_i}; i = 1, \ldots, n\}$$

for the survival times of interest. $(L_i, R_i]$ is the interval within which the unobserved survival event associated with the ith subject to occur. If $L_i = 0$, we have a left-censored observation and if $R_i = \infty$, we have a right-censored observation. $\boldsymbol{X_i}$ represents the p dimensional vector of covariates associated with the ith subject and assume that the censoring mechanism is independent of the covariates. let $S(t;\boldsymbol{X})$ denote the survival function for a subject with covariates $\boldsymbol{X}$. Then the likelihood contribution has the form

$$L = \prod_{i=1}^{n}[S(L_i;\boldsymbol{X_i}) - S(R_i;\boldsymbol{X_i})]$$

assuming that $L_i < R_i$ for all $i = 1, \ldots, n$.

Under PH model, the log of likelihood function has the form

$$l(\boldsymbol{\beta}, S_0) = \sum_{i=1}^{n} \log\Big\{ [S_0(L_i)]^{\exp(\boldsymbol{X_i}'\boldsymbol{\beta})} - [S_0(R_i)]^{\exp(\boldsymbol{X_i}'\boldsymbol{\beta})} \Big\}$$

in terms of the regression parameter $\boldsymbol{\beta}$ and the baseline survival function $S_0(t)$.

For inference, the maximum likelihood approach first studied in Finkelstein [1986]. The likelihood depends on S_0 only through its values at the different observation time points. Thus one only needs to focus on estimating the values of S_0 at these time points. Let $s_0 = 0 < s_1 < \cdots < s_{m+1} = \infty$ denote the ordered distinct time points of all observed interval end points $\{L_i, R_i; i = 1, \ldots, n\}$ and $\alpha_{ij} = I(s_j \in (L_i, R_i]), j = 1, \ldots, m, i = 1, \ldots, n$. $S_0(s_j)$ can be written as

$$S_0(s_j) = \prod_{k=1}^{j} \exp(-\exp(\alpha_k)) = \exp(-\sum_{k=1}^{j} \exp(\alpha_k)), j = 1, \ldots, m.$$

Then the log likelihood function $l()$ can be rewritten as

$$l(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \log\left\{ \sum_{j=1}^{m+1} \alpha_{ij} \Big[\exp(-\alpha_{j-1}\exp(\boldsymbol{X_i}'\boldsymbol{\beta})) - \exp(-\alpha_j \exp(\boldsymbol{X_i}'\boldsymbol{\beta})) \Big] \right\}$$

where $\alpha_j = \sum_{k=0}^{j} \exp(\alpha_k), \alpha_0 = -\infty$ and $\alpha_{m+1} = \infty$. To maximize the log likelihood function above, the Newton-Raphson algorithm can be used by treating it as a log likelihood function arising from a parametric model. Then the maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can be determined by solving the score equations of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

An alternative to the above full likelihood approach is the marginal likelihood approach. This approach defines a marginal likelihood as the summation of the probabilities of the ranking of the underlying and unobserved failure times that are consistent with observed interval-censored data [Satten, 1996]. It does not require estimation of the baseline cumulative hazard function, but it requires solving complicated score equations and involves massive computational effort. The disadvantage is that it does not have a simple and easily manageable form, resulting in the need for great computational effort. [Satten, 1996] investigated this approach and proposed a

Gibbs sampling procedure for generating underlying rankings and the use of stochastic approximation for solving the score functions. Goggins et al. [1998] developed a Monte Carlo EM algorithm for determination of regression parameter estimates under the model. The marginal likelihood approach focuses only on the finite-dimensional regression parameters, while full likelihood approach directly estimates the finite dimensional regression parameters and the infinite-dimensional nuisance parameter simultaneously.

Other semeparametric approaches have been developed for regression analysis of interval-censored data under the proportional hazards (PH) model. Pan [1999] proposed a generalized gradient projection method to estimate the regression coefficients besides estimating the baseline hazard using the nonparametric maximum likelihood estimation method. Applying the local likelihood and a penalized spline-based approach are proposed [Betensky et al., 2002, Cai and Betensky, 2003]. In these approaches, some finite-dimensional functions are used to approximate the log baseline hazard function in the PH model. Cai and Betensky [2003] developed a penalized likelihood approach and modeled the logarithm of the baseline hazard function with a linear spline. Zhang et al. [2010] proposed a sieve maximum likelihood method adopting B-splines for the logarithm of the cumulative baseline hazard function. From a Bayesian perspective, Sinha et al. [1999] used piecewise constant function for the baseline hazard function and Çetinyürek Yavuz and Lambert [2011] modeled the baseline density function with penalized B-splines, and Lin et al. [2014] approximated the baseline cumulative hazard function with monotone splines. A comprehensive review was given by Zhang and Sun [2010].

Accelerated Failure Time (AFT) Model

Cox proportional hazard model has been extensively used in medical research. However, the assumption of proportional hazard functions is strong and may be violated.

One important alternative is the accelerated failure time (AFT) model. The AFT approach is analogous to the classical linear regression approach. In such representation, the natural logarithm of the survival time $Y = \log(T)$ is modeled. This is the natural transformation made in linear models to convert positive variables to observations on the entire real line. A linear model is assumed for $Y_i, i = 1, \ldots, n$, namely,

$$Y_i = log(T_i) = \mu + \boldsymbol{\beta}^{'} X_i + W_i \tag{3.1}$$

T_i denote the time to the event, $\boldsymbol{X}$ a vector of fixed time explanatory covariates and $W_i, i = 1, \ldots, n$ is the error distribution. Common choices for the error distribution include the standard normal distribution which yields a log normal regression model, the extreme value distribution , which yields a Weibull regression model, or a logistic distribution, which yields a log logistic regression model. One can also formulate a nonparametric form for the errors using a mixture distribution.

The concept that covariates affect the failure time in the AFT model and the PH model is similar but the way to affect survival function is different. Let $S_0(t)$ denote the survival function of $T = e^Y$ when X is zero, that is, $S_0(t)$ is the survival function of $\exp(\mu + W)$.

$$Pr(T > t|\boldsymbol{X}) = Pr(T > log(t)|\boldsymbol{X}) = Pr(\mu + W > log(t) - \boldsymbol{\beta}^T \boldsymbol{X}|\boldsymbol{X})$$

$$= Pr(\exp(\mu + W) > t\exp(-\boldsymbol{\beta}^{'} \boldsymbol{X})|\boldsymbol{X}) = S(t\exp(-\boldsymbol{\beta}^{'} \boldsymbol{X}))$$

Notice that the effect of the explanatory variables in the original time scale is to change the time scale by a factor $\exp(-\boldsymbol{\beta}^{'} X)$. Depending on the sign of $\boldsymbol{\beta}^{'} X$, the time is either accelerated by a constant factor or degraded by a constant factor. Note that the hazard rate of an individual with a covariate value Z for this class of models is related to a baseline hazard rate h_0 by

$$h(t|\boldsymbol{X}) = h_0(t\exp(-\boldsymbol{\beta}^{'} \boldsymbol{X}))\exp(-\boldsymbol{\beta}^{'} \boldsymbol{X})$$

The underlying assumption for AFT models is that the effect of covariates is multiplicative (proportional) with respect to survival time. In other words, the effect is to change the timescale and therefore to accelerate or decelerate the time to failure. Although the PH model specifies that the effect of covariates on the hazard is multiplicative, it does not give a direct relationship between $\boldsymbol{X}$ and T because $h_0(t)$ is arbitrary. In contrast, the model 3.1 specifies a linear relationship between log T and $\boldsymbol{X}$.

Maximum likelihood approach is relatively hard on AFT model. The main difficulty is that the regression parameter and the nuisance function are tangled in the likelihood function [Sun, 2007], though Huang and Wellner [1997] prove that the consistency of the maximum likelihood estimators of those two, no asymptotic distribution theory for the estimators is available yet, even for the case of current status data. Linear rank statistics can be used to estimate regression parameters. For AFT models, inferences about the regression parameter are based on approximate likelihood or estimating equations. For these approximate likelihood or estimating equation approaches, one needs to investigate their efficiency, which is usually more challenging than developing the approximate likelihood or estimating equations [Sun, 2007]. References include Betensky et al. [2001], Pu and Li [1999], Li and Pu [2003], Rabinowitz et al. [1995], Xue et al. [2006].

3.3 Accelerated Failure Time Model with Random Effects

The concept of frailty was originally introduced by Vaupel [1976] to characterize the heterogeneity among clusters in survival analysis. The frailty component is often assumed to have a parametric distribution such as gamma or log-normal distributions. Goethals et al. [2009] modele interval-censored clustered cow udder quarter infection times under the Gamma frailty PH model. Zhang and Sun [2010] propose weighted score functions where the weight depends on cluster size or a within-cluster

resampling procedure to avoid the correlation issue within clusters. Kim [2010] model failure times and cluster size jointly using the PH model with normal random effect and a mixed ordinal regression model. Li et al. [2012] propose frailty additive hazards model for clustered interval-censored data. Pan et al. [2015] propose a multiple frailty proportional hazards model, not only accounting for the baseline heterogeneity and effect variation across clusters for predictors, but also quantifying the probabilities of the existence of such frailties. Most of those methods have focused on one shared frailty or equivalently one random intercept term to account for variation in baseline risk across clusters. However, the frailty PH model has some important drawbacks. First, the implied correlation structure is too simple, e.g., in the analysis of the multi-center clinical trials only the center effect and not the center by treatment interaction can be controlled for. Second, the choice of the frailty distribution can have a crucial impact on the results for the regression parameters of interest [Hougaard, 2012].

In contrast to the PH model, neglected covariates in the AFT model do not cause bias in estimating the regression parameters for the included covariates [Hougaard, 1999]. Komárek and Lesaffre [2007] propose a Bayesian accelerated failure time model whose error distribution is modelled in a flexible way as a finite normal mixture. An advantage of the full Bayesian approach is the fact that a general random effect vector can be easily included in the model. An extension of the AFT model, the mixed effects accelerated failure time (MEAFT) model, takes into account the within-cluster correlations explicitly by including random effects in the regression expression, as in a classical linear mixed model of Laird and Ware [1982].

Let T_{ij} be the exact event time of interest that cannot be observed, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the unknown regression coefficient vector, X_{ij} is the covariate vector for fixed effects, $\boldsymbol{b_i} = (b_{i,1}, \ldots, b_{i,q})^T$ are i.i.d. random effects vectors. Z_{ij} is the covariate vector for random effects, ϵ_{ij} are i.i.d. error random variables with a density $k_\tau(\epsilon_{ij})$ as in

equation (2.4), $i = 1, \ldots, n, j = 1, \ldots, n_i$. For the jth subject in the ith cluster,

$$\log(T_{ij}) = y_{ij} = \boldsymbol{\beta}^T X_{ij} + \boldsymbol{b}_i^T Z_{ij} + \epsilon_{ij} \tag{3.2}$$

The (i, j)th true log-event time y_{ij} is only known to lie in the interval $(y_{ij}^L, y_{ij}^R]$. Generally, a single observation $(\delta_{ij1}, \delta_{ij2}, \delta_{ij3}, y_{ij}^L, y_{ij}^R, X_{ij})$ has the conditional likelihood of jth observation in ith cluster given by

$$L_{ij} = \left\{F(y_{ij}^R)\right\}^{\delta_{ij1}} \left\{F(y_{ij}^R) - F(y_{ij}^L)\right\}^{\delta_{ij2}} \left\{1 - F(y_{ij}^L)\right\}^{\delta_{ij3}} \tag{3.3}$$

where $F(\cdot)$ is the conditional cumulative distribution function of T_{ij}. We define $F_\tau(x) = \int_{-\infty}^{x} f_\tau^2(x; G_1, G_2)dx$. The likelihood contribution of the ith cluster is given by

$$L_i = \int_{R^q} \{\prod_{j=1}^{n_i} \int_{y_{ij}^L}^{y_{ij}^U} f(y - \beta^T x_{ij} - \boldsymbol{b}^T z_{ij})dy\} g(\boldsymbol{b})d\boldsymbol{b},$$

Due to multiple integration in the likelihood above, it is rather cumbersome to use maximum-likelihood based methods for the model (4.2) with interval-censored observations, even with error and random effects being parametrically specified. Stochastic versions of standard estimation techniques can be used; Komárek and Lesaffre [2007]'s Bayesian approach avoids full parametric assumptions (like normality) concerning the error density, while still being computationally tractable for both interval-censored data and a general covariate vector z_{ij}. They specify the density of the error term ϵ_{ij} in (4.2) as

$$f(\epsilon) = \sum_{j=1}^{k} w_j \psi(\epsilon | \mu_j, \sigma_j^2),$$

with $\psi(\cdot|\mu_j, \sigma_j^2)$ is density of $N(\mu_j, \sigma_j^2)$. Note that k is the number of mixture components and unknown, as well as the mixture weights $w = (w_1, ..., w_k)^T$, means $\mu = (\mu_1, ..., \mu_k)^T$, and variances $\sigma^2 = (\sigma_1^2, ..., \sigma_k^2)^T$. For the actual implementation of the reversible jump MCMC algorithm, Gibbs sampler was conducted and the auxiliary variable (AV) method was employed.

Chapter 4

Bayesian Quantile Analysis of Clustered Interval-Censored Data

We are interested to explore a flexible quantile regression framework for analysing time-to-event data that are randomly interval-censored. Unlike hazard regression, quantile regression models the conditional quantiles of the survival time directly. Therefore, the results are easier to interpret. More importantly, quantile regression allows the covariates to have different effects at different tails of the survival distribution and thus is able to capture important population heterogeneity Koenker [2005]. In Section 5.1, a quantile regression with parametric error distribution is constructed to analyze clustered interval-censored data. Then in Section 5.2, a nonparametric error distribution is applied to the same Bayesian AFT model with random effects. We run simulations to compare these two methods and apply our method to a real data set.

4.1 Quantile AFT Model for Clustered Interval-Censored Data

The accelerated failure time (AFT) model with random effects can be used to model correlated survival data in the mixed model framework. The asymmetric Laplace distribution is commonly used for parametric Bayesian quantile regression since its τ-th quantile is zero; the double exponential distribution is a special case of the asymmetric Laplace distribution with $\tau = 0.5$.

Let T_{ij} be the exact event time that cannot be observed, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is

the unknown regression coefficient vector, X_{ij} is the covariate vector for fixed effects, $\boldsymbol{b_i} = (b_{i,1}, \ldots, b_{i,q})^T$ are i.i.d. random effects vectors. Z_{ij} is the covariate vector for random effects, ϵ_{ij} are i.i.d. error random variables with a density $k_\tau(\epsilon_{ij})$ as in equation (2.4), $i = 1, \ldots, n, j = 1, \ldots, n_i$. For the jth subject in the ith cluster,

$$\log(T_{ij}) = y_{ij} = \boldsymbol{\beta}^T X_{ij} + \boldsymbol{b}_i^T Z_{ij} + \epsilon_{ij} \tag{4.1}$$

The (i, j)th true log-event time y_{ij} is only known to lie in the interval $(y_{ij}^L, y_{ij}^R]$. Assuming $y_{ij} \sim$ ALD, covariate vectors $X_{ij} = \{X_{ij1}, ..., X_{ijp}\}$, $i = 1, ..., n$, $j = 1, ..., n_i$, $\boldsymbol{\beta} = \{\beta_1, ..., \beta_p\}$ are fixed effects and $\boldsymbol{b}_i$ are random effects. The response distribution can be written as

$$f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma_i, \tau) = \frac{\tau(1-\tau)}{\sigma_i} \exp\left\{-\rho_\tau(\frac{y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b}_i}{\sigma_i})\right\}$$

The likelihood function for $\boldsymbol{\beta}, \boldsymbol{b}_i$ and σ_i is as follows

$$L(\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma_i; \boldsymbol{y}_i, \tau) = \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma_i, \tau)$$

$$\propto \frac{1}{\sigma_i^{n_i}} \exp\left\{-\sum_{j=1}^{n_i}\left(\frac{y_{ij} - X_{ij}^T\boldsymbol{\beta} - Z_{ij}\boldsymbol{b}_i}{\sigma_i}\right)\left(\tau - I(\frac{y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b}_i}{\sigma_i} \leq 0)\right)\right\}$$

The prior distributions for σ_i, $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$

$$\begin{aligned}
\sigma_i|G &\overset{iid}{\sim} G, i = 1, \ldots, n \\
G|\alpha, d &\sim DP(\alpha G_0) \\
\boldsymbol{\beta} &\sim N_p(\mu_0, \Sigma_0) \\
\boldsymbol{b_i}|G_b &\sim G_b, i = 1, \ldots, n \\
G_b|\alpha_b, G_{0b} &\sim DP(\alpha_b G_{0b})
\end{aligned}$$

where μ_0 and Σ_0 are hyper parameters assumed to be known. G_0 has a uniform prior, and $G_{0b} \sim MVN(\boldsymbol{0}, \boldsymbol{D})$.

We want to generate data from a truncated distribution over interval $(y_{ij}^L, y_{ij}^R]$. $f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma_i, \tau) = \frac{\tau(1-\tau)}{\sigma_i} \exp\left\{-\rho_\tau(\frac{y_{ij}-X_{ij}\beta-Z_{ij}\boldsymbol{b}_i}{\sigma_i})\right\}, \epsilon_{ij} = y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b}_i,$

$$f_\tau(\epsilon_{ij};\sigma_i) = \begin{cases} \frac{\tau(1-\tau)}{\sigma_i} \exp\left\{\frac{1-\tau}{\sigma_i}\epsilon_{ij}\right\} & \text{if } \epsilon_{ij} < 0 \\ \frac{\tau(1-\tau)}{\sigma_i} \exp\left\{-\frac{\tau}{\sigma_i}\epsilon_{ij}\right\} & \text{if } \epsilon_{ij} \geq 0 \end{cases}$$

Taking integral of the density, the CDF is $F_\tau(\epsilon_{ij};\sigma_i) = \int_{-\infty}^{\infty} f_\tau(\epsilon_{ij};\sigma_i)$ is as follows

$$F_\tau(\epsilon_{ij};\sigma_i) = \begin{cases} \tau e^{\frac{1-\tau}{\sigma_i}\epsilon_{ij}} & \text{if } \epsilon_{ij} < 0 \\ 1-(1-\tau)e^{-\frac{\tau}{\sigma_i}\epsilon_{ij}} & \text{if } \epsilon_{ij} \geq 0 \end{cases}$$

We can use the inverse CDF algorithm to generate ϵ_{ij} from a truncated distribution over interval $(y_{ij}^L - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}, y_{ij}^R - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}]$. For a distribution $F_\tau(\epsilon_{ij};\sigma_i)$, we generate uniform random variate U on the interval $(F_\tau(y_{ij}^L - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i};\sigma_i), F_\tau(y_{ij}^R - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i};\sigma_i)]$ and then apply the inverse CDF,

$$F_\tau^{-1}(U;\sigma_i) = \frac{\sigma}{1-\tau}\log\frac{U}{\tau}1_{(0,\tau)}U - \frac{\sigma}{\tau}\log\frac{U-1}{\tau-1}1_{[\tau,1)}U$$

the resulting value ϵ_{ij} follow the $F_\tau(\epsilon_{ij};\sigma_i)$ distribution truncated to $(y_{ij}^L - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}, y_{ij}^R - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}]$, which is the same as y_{ij} truncated on $(y_{ij}^L, y_{ij}^U]$.

4.2 Nonparametric Quantile AFT model

Here, we develop the Bayesian nonparametric model of DP mixture implied conditional regression to the context of quantile AFT model with random effects. The data consist of a set of covariate vectors X and corresponding responses that cannot be observed. Let T_{ij} be the exact event time of interest that cannot be observed, X_{ij} is the covariate vector for fixed effects, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the unknown regression coefficient vector, $\boldsymbol{b_i} = (b_{i,1}, \ldots, b_{i,q})^T$ are i.i.d. random effects vectors. Z_{ij} is the covariate vector for random effects, ϵ_{ij} are i.i.d. error random variables with a density $k_\tau(\epsilon_{ij})$ as in equation (2.4), $i = 1, \ldots, n, j = 1, \ldots, n_i$. For the jth subject in the ith cluster,

$$\log(T_{ij}) = y_{ij} = \boldsymbol{\beta}^T X_{ij} + \boldsymbol{b}_i^T Z_{ij} + \epsilon_{ij} \tag{4.2}$$

The (i,j)th true log-event time y_{ij} is only known to lie in the interval $(y_{ij}^L, y_{ij}^R]$.

The density of the error term ϵ_{ij} is specified as

$$k_\tau(\epsilon_{ij};\sigma_{1i},\sigma_{2i}) = \frac{\tau}{\sigma_{1i}} 1_{(-\sigma_{1i},0)}(\epsilon_{ij}) + \frac{1-\tau}{\sigma_{2i}} 1_{[0,\sigma_{2i})}(\epsilon_{ij})$$

Assuming independent DP priors for the mixing distributions G_1 and G_2, we obtain the error density as $f_\tau^2(\epsilon; G_1, G_2) = \int\int k_\tau(\epsilon;\sigma_1,\sigma_2) dG_1(\sigma_1) dG_2(\sigma_2)$, and it is defined that $\int_{-\infty}^0 f_\tau^2(\epsilon; G_1, G_2) d\epsilon = \tau$.

In the presence of interval-censoring, the unobserved (log) event time is within an interval: $y_{ij} \in (y_{ij}^L, y_{ij}^R]$. Because $-\sigma_{1i} < y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i} < \sigma_{2i}$, y_{ij} can be randomly sampled from $(\max(y_{ij}^L, X_{ij}\boldsymbol{\beta} + Z_{ij}\boldsymbol{b_i} - \sigma_{1i}), \min(y_{ij}^R, X_{ij}\boldsymbol{\beta} + Z_{ij}\boldsymbol{b_i} + \sigma_{2i})]$. Thus, the likelihood contribution of subject j in cluster i is given by

$$\int_{y_{ij}^L}^{y_{ij}^R} k_\tau(y - X_{i,j}\boldsymbol{\beta} - Z_{i,j}\boldsymbol{b_i}; \sigma_{1,i}, \sigma_{2,i}) dy$$

Similarly, we need to generate data from a truncated distribution over interval $(y_{ij}^L, y_{ij}^R]$. Taking integral of the density $k_\tau(\epsilon_{ij};\sigma_{1i},\sigma_{2i}) = \frac{\tau}{\sigma_{1i}} 1_{(-\sigma_{1i},0)}(\epsilon_{ij}) + \frac{1-\tau}{\sigma_{2i}} 1_{[0,\sigma_{2i})}(\epsilon_{ij})$, the CDF $F_\tau(\epsilon_{ij};\sigma_{1i},\sigma_{2i}) = \int_{-\infty}^{\infty} k_p(\epsilon_{ij};\sigma_{1i},\sigma_{2i})$ is as follows

$$F_\tau(\epsilon_{ij};\sigma_{1i},\sigma_{2i}) = \begin{cases} 0 & \text{if } \epsilon_{ij} \le -\sigma_{1i} \\ \frac{\tau(\epsilon_{ij}+\sigma_{1i})}{\sigma_{1i}} & \text{if } -\sigma_{1i} < \epsilon_{ij} < 0 \\ \tau + \frac{(1-\tau)\epsilon_{ij}}{\sigma_{2i}} & \text{if } 0 \le \epsilon_{ij} < \sigma_{2i} \\ 1 & \text{if } \epsilon_{ij} \ge \sigma_{2i} \end{cases}$$

We can use the inverse CDF algorithm to generate ϵ_{ij} from the mixture distribution, which is a truncated distribution over interval $(y_{ij}^L - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}, y_{ij}^R - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}]$. For a distribution $F_\tau(\epsilon_{ij};\sigma_{1i},\sigma_{2i})$, we generate uniform random variate U on the interval $(F_\tau(y_{ij}^L - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i};\sigma_{1i},\sigma_{2i}), F_\tau(y_{ij}^R - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i};\sigma_{1i},\sigma_{2i})]$ and then apply the inverse CDF,

$$F_\tau^{-1}(U;\sigma_{1i},\sigma_{2i}) = \frac{(U-\tau)\sigma_{1i}}{\tau} 1_{(0,\tau)}(U) + \frac{(U-\tau)\sigma_{2i}}{1-\tau} 1_{[\tau,1)}(U)$$

the resulting value ϵ_{ij} follow the $F_\tau(\epsilon_{ij};\sigma_{1i},\sigma_{2i})$ distribution truncated to $(y_{ij}^L - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}, y_{ij}^R - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}]$, which is the same as y_{ij} truncated on $(y_{ij}^L, y_{ij}^U]$. With latent mixing parameters σ_{1i} and σ_{2i} for each response $\boldsymbol{y}_i$, the full Bayesian model has the hierarchical structure

$$\begin{aligned}
y_{ij}|\boldsymbol{\beta},\boldsymbol{b_i},\sigma_{1i},\sigma_{2i} &\overset{ind}{\sim} k_\tau^\tau(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i};\sigma_{1i},\sigma_{2i}), \text{truncated on } (y_{ij}^L, y_{ij}^U] \\
&\quad i = 1,\ldots,n, j = 1,\ldots,n_i \\
\sigma_{ri}|G_r &\overset{iid}{\sim} G_r, r = 1,2, i = 1,\ldots,n \\
G_r|\alpha_r, d_r &\sim DP(\alpha_r G_{r0}), r = 1,2, \\
\boldsymbol{\beta} &\sim N_p(\mu_0,\Sigma_0) \\
\boldsymbol{b_i}|G_b &\sim G_b, i = 1,\ldots,n \\
G_b|\alpha_b, G_{0b} &\sim DP(\alpha_b G_{0b})
\end{aligned}$$

4.3 Posterior Computation

Here, we present the MCMC method for posterior inference under the model developed in Section 4.2. Detailed expressions for the algorithm follow in Section 2.3. Posterior computing details for QAFT can be found in Appendix.B

As presented in the Bayesian hierarchical modeling in section 4.2, DP priors are given to both scale parameters, σ_{1i} and σ_{2i}, and the random effects vector $\boldsymbol{b}_i$. The appeal of DP priors in this method owes much to the Pólya urn formulation for the posterior predictive distribution as a mixture of point masses on the observed data and the underlying prior base measure αG_0. $\alpha_1, \alpha_2, \alpha_b$ correspond to precision parameters for σ_{1i}, σ_{2i}, and $\boldsymbol{b}_i$ respectively. The optimal values of α_r, $r = 1,2,b$, are decided by the grid search. G_0 is a probability distribution that is the mean of the process. We adopt a uniform distribution for G_{r0} [Gelman, 2006]. A normal base measure with zero mean is specified to G_{0b}. An assumed DP prior with zero mean may still imply a non-zero mean for random effect distribution, to ensure that

$E(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$, the additional constraint $E(\boldsymbol{b}_i = \mathbf{0})$ is needed. As noted before, the error $y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}$ must be within $(-\sigma_{1i}, \sigma_{2i})$ for all i. $\boldsymbol{X}_i$ is an $n_i \times p$ matrix, $Z_{ij} = X_{ij}$. $\boldsymbol{\beta}$ is a $1 \times p$ regression coefficient vector. Due to the nature of the error distribution, $y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}$ is restricted to be within $[-\sigma_{1i}, \sigma_{2i}]$ when updating each parameter.

Independent $p-$dimensions multivariate normal distribution are given to fixed effect $\boldsymbol{\beta}$, $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0})$. The posterior distribution of $\boldsymbol{\beta}$ can be written as

$$\begin{aligned}
p(\boldsymbol{\beta}|Y) &\propto \text{prior}(\boldsymbol{\beta}) \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}) \\
&\propto N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma_0}) \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i}) \\
&\propto \frac{1}{(2\pi)^{p/2} \mid \boldsymbol{\Sigma_0} \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu_0})^T {\boldsymbol{\Sigma_0}}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu_0}) \right) \\
&\times \prod_{i=1}^{n}\prod_{j=1}^{n_i} \frac{\tau}{\sigma_{1i}} 1_{(-\sigma_{1i},0)}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) + \frac{1-\tau}{\sigma_{2i}} 1_{[0,\sigma_{2i})}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i})
\end{aligned}$$

We use M-H algorithm to sample posteriors for $\boldsymbol{\beta}$. we suppose the proposal(jump) distribution for $\boldsymbol{\beta}$ is multivariate normal with covariance matrix T_0,

$$J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1}) = \frac{1}{(2\pi)^{p/2} \mid T_0 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{t-1})^T T_0^{-1} (\boldsymbol{\beta}^* - \boldsymbol{\beta}^{t-1}) \right)$$

$$J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*) = \frac{1}{(2\pi)^{p/2} \mid T_0 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}^{t-1} - \boldsymbol{\beta}^*)^T T_0^{-1} (\boldsymbol{\beta}^{t-1} - \boldsymbol{\beta}^*) \right)$$

Since the jump distribution is symmetric, $J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1}) = J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)$, we compute acceptance ratio

$$\begin{aligned}
r &= \frac{p(\boldsymbol{\beta}^*|Y)/J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1})}{p(\boldsymbol{\beta}^{t-1}|Y)/J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)} = \frac{p(\boldsymbol{\beta}^*|Y)}{p(\boldsymbol{\beta}^{t-1}|Y)} = \frac{\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma_0}|^{1/2}}}{\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma_0}|^{1/2}}} \\
&\times \frac{\exp\left(-\frac{1}{2}(\boldsymbol{\beta}^* - \boldsymbol{\mu_0})^T {\boldsymbol{\Sigma_0}}^{-1} (\boldsymbol{\beta}^* - \boldsymbol{\mu_0}) \right) \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}^*, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i})}{\exp\left(-\frac{1}{2}(\boldsymbol{\beta}^{t-1} - \boldsymbol{\mu_0})^T {\boldsymbol{\Sigma_0}}^{-1} (\boldsymbol{\beta}^{t-1} - \boldsymbol{\mu_0}) \right) \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}^{t-1}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i})}
\end{aligned}$$

We accept a given proposal with the acceptance probability which is the outcome of the acceptance function described above. Operationally, we draw a random number uniformly between 0 and 1, and if log$r$ is larger than the random number, we accept the proposal; otherwise we reject it.

$$\begin{aligned}
\log r &= \log[p(\boldsymbol{\beta}^*|Y)] - \log[p(\boldsymbol{\beta}^{t-1}|Y)] \\
&= \log\Big(\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i})\Big) - \frac{1}{2}(\boldsymbol{\beta}^* - \boldsymbol{\mu_0})^T {\boldsymbol{\Sigma_0}}^{-1}(\boldsymbol{\beta}^* - \boldsymbol{\mu_0}) \\
&\quad -\Big(\log\big(\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}^{t-1}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i})\big) - \frac{1}{2}(\boldsymbol{\beta}^{t-1} - \boldsymbol{\mu_0})^T {\boldsymbol{\Sigma_0}}^{-1}(\boldsymbol{\beta}^{t-1} - \boldsymbol{\mu_0})\Big)
\end{aligned}$$

Prior specifications and posterior computation for b_i

We write the prior distribution for the random effect vector $\boldsymbol{b}_i$ as follows, to indicate that a DP prior is used for the random distribution G_b. It involves hyper priors α_b and G_{0b}.

$$\boldsymbol{b_i}|G_b \sim G_b, i = 1, \ldots, n$$

$$G_b|\alpha_b, G_{0b} \sim DP(\alpha_b G_{0b})$$

Since random effects do not have to be positive and are usually considered to be normal, we assume the base G_{0b} to be a multivariate normal distribution $N_q(\mathbf{0}, \boldsymbol{D})$. Instead of placing a gamma prior on the DP precision parameter, we use fixed values by grid selection from several possible values.

We can draw values for each $\boldsymbol{b_i}$ from its conditional distribution given both the data and the $\boldsymbol{b_i}$ for $g \neq i$ (which is written as $\boldsymbol{b_{-i}}$).

Conditional prior has this form,

$$\boldsymbol{b_i}|\boldsymbol{b_{-i}} \sim \frac{\alpha_b}{\alpha_b + n - 1} G_{0b}(\boldsymbol{b_i}) + \frac{1}{\alpha_b + n - 1}\sum_{g \neq i}\delta(\boldsymbol{b}_g)$$

Sampling from posterior distributions for each $\boldsymbol{b}_i$, suppose $\boldsymbol{D}$ is known:

$$\begin{aligned}
p(\boldsymbol{b_i}|\boldsymbol{y_i}) &\propto \text{prior}(\boldsymbol{b_i}|\boldsymbol{b_{-i}}) \times \prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i}) \\
&\propto \left(\frac{\alpha_b}{\alpha_b + n - 1} G_{0b}(\boldsymbol{b_i}) + \frac{1}{\alpha_b + n - 1} \sum_{g \neq i} \delta(\boldsymbol{b}_g) \right) \prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i}) \\
&\propto \frac{\alpha_b}{\alpha_b + n - 1} N_q(\boldsymbol{0}, \boldsymbol{D}) \times \left\{ \prod_{j=1}^{n_i} \frac{\tau}{\sigma_{1i}} 1_{(-\sigma_{1i}, 0)}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) + \right. \\
&\quad \left. \frac{1-\tau}{\sigma_{2i}} 1_{[0, \sigma_{2i})}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) \right\} + \\
&\quad \frac{1}{\alpha_b + n - 1} \sum_{g \neq i} \delta(\boldsymbol{b_g}) \times \left\{ \prod_{j=1}^{n_i} \frac{\tau}{\sigma_{1i}} 1_{(-\sigma_{1i}, 0)}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) + \right. \\
&\quad \left. \frac{1-\tau}{\sigma_{2i}} 1_{[0, \sigma_{2i})}(y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}) \right\}
\end{aligned}$$

which can be written as $r_i H_i + \sum_{g \neq i} q_{i,g} \delta(\boldsymbol{b}_g)$, H_i is the posterior distribution for $\boldsymbol{b}_i$ based on the G_{0b} and cluster i, with likelihood $F(\boldsymbol{y_i}, \boldsymbol{b_i}) = \prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i})$. The values of the $q_{i,g}$ and of r_i are written as follows, where b is such that $\sum_{g \neq i} q_{i,g} + r_i = 1$.

$$\begin{aligned}
r_i H_i &\propto \frac{\alpha_b}{\alpha_b + n - 1} N_q(\boldsymbol{0}, \boldsymbol{D}) F(\boldsymbol{y_i}, \boldsymbol{b_i}) \\
r_i &= b\alpha_{bi} \int F(\boldsymbol{y_i}, \boldsymbol{b_i}) dG_{0b}(\boldsymbol{b_i}) \\
H_i &= G_{0b}(\boldsymbol{b_i}) F(\boldsymbol{y_i}, \boldsymbol{b_i}) \\
&= \frac{1}{(2\pi)^{n/2} \mid \boldsymbol{D} \mid^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{b_i} - \boldsymbol{0})^T \boldsymbol{D}^{-1} (\boldsymbol{b_i} - \boldsymbol{0}) \right) F(\boldsymbol{y_i}, \boldsymbol{b_i}) \\
q_{i,g} &= bF(\boldsymbol{y_i}, \boldsymbol{b_g}), g = 1, \ldots, i-1, i+1, \ldots, n
\end{aligned}$$

Update Φ_c using M-H sampler for all $c \in c_1, \ldots, c_n$: draw a new value from $\Phi_c|y_i$ such that $c_i = c$: we suppose the proposal(jump) distribution for $\boldsymbol{b_i}$ is multivariate normal with covariance matrix T_1, $\boldsymbol{b_i}^*$ is the proposed candidate, $\boldsymbol{b_i}^{t-1}$ is the current vector value.

$$J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1}) = \frac{1}{(2\pi)^{p/2} \mid T_1 \mid^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{b_i}^* - \boldsymbol{b_i}^{t-1})^T T_1^{-1} (\boldsymbol{b_i}^* - \boldsymbol{b_i}^{t-1}) \right)$$

$$J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*) = \frac{1}{(2\pi)^{p/2} \mid T_1 \mid^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{b_i}^{t-1} - \boldsymbol{b_i}^*)^T T_1^{-1} (\boldsymbol{b_i}^{t-1} - \boldsymbol{b_i}^*)\Big)$$

Since the jump distribution is symmetric, $J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1}) = J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*)$, we compute acceptance ratio

$$\begin{aligned}
r &= \frac{p(\boldsymbol{b_i}^*|\boldsymbol{y_i})/J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1})}{p(\boldsymbol{b_i}^{t-1}|\boldsymbol{y_i})/J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*)} = \frac{p(\boldsymbol{b_i}^*|\boldsymbol{y_i})}{p(\boldsymbol{b_i}^{t-1}|\boldsymbol{y_i})} \\
&= \frac{\frac{1}{(2\pi)^{p/2}|\boldsymbol{D}|^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{b_i}^* - \boldsymbol{0})^T \boldsymbol{D}^{-1} (\boldsymbol{b_i}^* - \boldsymbol{0})\Big) F_{c_i=c}(\boldsymbol{y_i}, \boldsymbol{b_i}^*)}{\frac{1}{(2\pi)^{p/2}|\boldsymbol{D}|^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{b_i}^{t-1} - \boldsymbol{0})^T \boldsymbol{D}^{-1} (\boldsymbol{b_i}^{t-1} - \boldsymbol{0})\Big) F_{c_i=c}(\boldsymbol{y_i}, \boldsymbol{b_i}^{t-1})} \\
\log r &= \log[p(\boldsymbol{b_i}^*|\boldsymbol{y_i})] - \log[p(\boldsymbol{b_i}^{t-1}|\boldsymbol{y_i})] \\
&= \log F_{c_i=c}(\boldsymbol{y_i}, \boldsymbol{b_i}^*) - \frac{1}{2}(\boldsymbol{b_i}^* - \boldsymbol{0})^T \boldsymbol{D}^{-1} (\boldsymbol{b_i}^* - \boldsymbol{0}) \\
&\quad - \Big(\log F_{c_i=c}(\boldsymbol{y_i}, \boldsymbol{b_i}^{t-1}) - \frac{1}{2}(\boldsymbol{b_i}^{t-1} - \boldsymbol{0})^T \boldsymbol{D}^{-1} (\boldsymbol{b_i}^{t-1} - \boldsymbol{0})\Big)
\end{aligned}$$

Then, we can directly sample $\boldsymbol{D}$ using a Wishart conjugate prior, $\boldsymbol{D}^{-1}$ is usually used and $\boldsymbol{D}^{-1} \sim \text{Wishart}_d(\nu_0, \Phi_0)$

$$p(\boldsymbol{D}^{-1}|\nu_0, \Phi_0) \propto \mid \boldsymbol{D}^{-1} \mid^{(\nu_0 - d - 1)/2} \exp\Big(-\frac{1}{2}\text{tr}(\Phi_0^{-1} \boldsymbol{D}^{-1})\Big)$$

The posterior can be written as

$$\begin{aligned}
p(\boldsymbol{D}^{-1}|\boldsymbol{b_i}) &\propto \prod_{i=1}^{n} p(\boldsymbol{b_i}|\boldsymbol{D}) p(\boldsymbol{D}^{-1}|\nu_0, \Phi_0) \\
&\propto \frac{1}{(2\pi)^{n/2} \mid \boldsymbol{D} \mid^{n/2}} \exp\Big(-\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{b_i} - \boldsymbol{0})^T \boldsymbol{D}^{-1} (\boldsymbol{b_i} - \boldsymbol{0})\Big) \\
&\times \mid \boldsymbol{D}^{-1} \mid^{(\nu_0 - d - 1)/2} \exp\Big(-\frac{1}{2}\text{tr}(\Phi_0^{-1} \boldsymbol{D}^{-1})\Big) \\
&\propto \mid \boldsymbol{D}^{-1} \mid^{(n + \nu_0 - d - 1)/2} \exp\Big(-\frac{1}{2}\Big(\text{tr}(\Phi_0^{-1} \boldsymbol{D}^{-1}) - \frac{1}{2}\text{tr}(\boldsymbol{D}^{-1} \sum_{i=1}^{n} \boldsymbol{b_i} \boldsymbol{b_i}^T)\Big)\Big) \\
&\propto \mid \boldsymbol{D}^{-1} \mid^{(n + \nu_0 - d - 1)/2} \exp\Big(-\frac{1}{2}\Big(\text{tr}(\Phi_0^{-1} \boldsymbol{D}^{-1} + \sum_{i=1}^{n} \boldsymbol{b_i} \boldsymbol{b_i}^T \boldsymbol{D}^{-1})\Big)\Big)
\end{aligned}$$

Thus, we can directly sample from the conjugate posterior

$$\boldsymbol{D}^{-1}|\boldsymbol{b_i} \sim \text{Wishart}_d\Big(n + \nu_0, \Big(\Phi_0^{-1} + \sum_{i=1}^{n} \boldsymbol{b_i} \boldsymbol{b_i}^T\Big)^{-1}\Big)$$

Prior specifications and Posterior computation for σ_{ri}

Because σ_{ri} depends on y_{ij} through $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$, and $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ is supposed to be known. In other words, $l(y_{ij}|\sigma_{ri})p(\sigma_{ri})$ is equivalent to $l(y_{ij})p(\boldsymbol{\beta}|\sigma_{ri})p(\boldsymbol{b_i}|\sigma_{ri})p(\sigma_{ri})$. The likelihood of i-th observation is $F(\boldsymbol{y_i}, \sigma_{ri}) = \prod_{j=1}^{n_i} l(y_{ij}|\sigma_{ri})$

$$\begin{aligned}
p(\sigma_{ri}|y_{ij}, G_{r0}) &\propto F(\boldsymbol{y_i}, \sigma_{ri})\text{prior}(\sigma_{ri}) \\
&\propto F(\boldsymbol{y_i}, \sigma_{ri})\text{prior}(\sigma_{ri}) \\
&\propto \frac{\alpha_r}{\alpha_r + n - 1} G_{r0}(\sigma_{ri}) F(\boldsymbol{y_i}, \sigma_{ri}) + \frac{1}{\alpha_r + n - 1} \sum_{g \neq i} \delta(\sigma_{rg}) F(\boldsymbol{y_i}, \sigma_{ri}) \\
&\propto \frac{\alpha_r}{\alpha_r + n - 1} \text{Unif}(d_r) F(\boldsymbol{y_i}, \sigma_{ri}) + \frac{1}{\alpha_r + n - 1} \sum_{g \neq i} \delta(\sigma_{rg}) F(\boldsymbol{y_i}, \sigma_{ri})
\end{aligned}$$

which can be written as $r_i H_i + \sum_{g \neq i} q_{i,g} \delta(\theta_g)$, H_i is the posterior distribution for σ_r based on the prior G_{r0} and the single observation i, with likelihood $F(\boldsymbol{y_i}, \sigma_{ri})$. The values of the $q_{i,g}$ and of r_i are written as follows, where b is such that $\sum_{g \neq i} q_{i,g} + r_i = 1$.

$$\begin{aligned}
r_i H_i &\propto \frac{\alpha_r}{\alpha_r + n - 1} \text{Unif}(d_r) F(\boldsymbol{y_i}, \sigma_{ri}) \\
r_i &= b\alpha_r \int \frac{1}{d_r} F(\boldsymbol{y_i}, \sigma_{ri}) d\sigma_{ri} \\
H_i &= \frac{1}{d_r} F(\boldsymbol{y_i}, \sigma_{ri}) \\
q_{i,g} &= bF(\boldsymbol{y_i}, \sigma_{rg}), g = 1, \ldots, i-1, i+1, \ldots, n
\end{aligned}$$

M-H to all $c \in c_1, \ldots, c_n$, update $\Phi_c|y_i$ such that $c_i = c$: we suppose the proposal density is lognormal, so $\sigma^*_{rc} \propto \text{logNormal}$ and $X = \log\sigma^*_{rc} \propto \text{Normal}$

$$f(X) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-(x-\mu)^2/(2\tau^2)}$$

$$f(\sigma^*_{rc}) = \frac{1}{\sqrt{2\pi\tau^2}} \frac{e^{-(\sigma^*_{rc}-\mu)^2/(2\tau^2)}}{\sigma^*_{rc}}$$

The ratio

$$r = \frac{p(\sigma^*_{rc}|y)/J(\sigma^*_{rc}|\sigma^{t-1}_{rc})}{p(\sigma^{t-1}_{rc}|y)/J(\sigma^{t-1}_{rc}|\sigma^*_{rc})}$$

$$J(\sigma^*_{rc}|\sigma^{t-1}_{rc}) = \frac{1}{\sqrt{2\pi\tau^2}\sigma^*_{rc}} e^{-(\log\sigma^*_{rc} - \log\sigma^{t-1}_{rc})^2/(2\tau^2)}$$

$$J(\sigma_{rc}^{t-1}|\sigma_{rc}^{*}) = \frac{1}{\sqrt{2\pi\tau^2}\sigma_{rc}^{t-1}} e^{-(\log\sigma_{rc}^{t-1} - \log\sigma_{rc}^{*})^2/(2\tau^2)}$$

$$\log r = \log[p(\sigma_{rc}^{*}|y)] - \log[p(\sigma_{rc}^{t-1}|y)] + \log\sigma_{rc}^{*} - \log\sigma_{rc}^{t-1}$$

where $p(\sigma_{rc}^{*}|y) = \prod_{i\in c} H_i$

4.4 Simulation Study

In order to evaluate the proposed model, we analyze simulated data with comparison to a parametric model, QAFT, under two scenarios: errors from ALD and a bimodal normal mixture distribution. A sample size of 50 subjects with 10 repeated measures for each is applied for each scenario. The logarithm of the uncensored times $\log(T_{ij}) = y_{ij}$ are generated using a linear mixed model $\log(T_{ij}) = y_{ij} = \boldsymbol{\beta}^T X_{ij} + \boldsymbol{b}_i^T Z_{ij} + \epsilon_{ij}, i = 1, \ldots, 50, j = 1, \ldots, 10$. The fixed covariate vector is X_{ij}, $X_{ij} = (1, X_{1ij}, X_{2ij})'$. We generated continuous covariate X_{1ij} from $Unif(-2, 2)$, binary covariate X_{2ij} from Bernoulli(0.5). $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ is the regression coefficient vector for fixed effects. $\beta_1 = \beta_2 = 1$, β_0 are set to be 1, 2 and 3 at quantile 0.3, 0.5 and 0.75. $\boldsymbol{b_i} = (b_{i,1}, \ldots, b_{i,q})^T$ are i.i.d. random effects vectors. We set a normal random effect distribution, $b_{0i} \sim$ Normal$(0, 0.1)$ and a bimodal normal mixture distribution $b_{1i} \sim 1/2N(-0.1, 0.01)+1/2N(0.1, 0.01)$, and b_{2i} as 0. Z_{ij} is the covariate vector for random effects, and we assumed $Z_{ij} = X_{ij}$.

The random error ϵ_{ij} is with a density $k_\tau(\epsilon_{ij})$ as in equation (2.4). Two types of error densities are considered, first ϵ_{ij} is distributed as an ALD with three parameters, location, scale and skewness. we can write it as $\epsilon_{ij} \sim$ ALD$(0, 0.2, \tau)$; second, we assume ϵ_{ij} is distributed as a mixture of two normal densities, and write it as $\epsilon_{ij} \sim 3/5N(-0.3, 0.02) + 2/5N(0.2, 0.03)$. For model QAFT, we gave DP prior for scale parameter and random effects, thus there are two precision parameters unknown. For model BNQAFT, there are three precision parameters unknown since there are

two scale parameters. We do grid search to find the optimal combination of three precision parameters and optimal values are chosen with the smallest DIC.

For the jth subject in the ith cluster, the exact event time of interest, T_{ij}, cannot be observed and is only known to lie in the interval $(y_{ij}^L, y_{ij}^R]$. Thus, we used a Poisson distribution with mean 25 to randomly generate intervals within ranges sampled from $U(0, 20)$. Censoring indicators were created based on the cummulative time intervals for each observation. The Possion distribution parameters are revised a little bit for each quantile level such that data censoring types consisted of about $(10-15\%)$ right censoring, $(10-15\%)$ left censoring, and $(70-80\%)$ interval-censoring observations.

Here we show the traceplots and Gewke' diagnostic for regression coefficients $\beta_0, \beta_1, \beta_2$, from the first simulated dataset. We used tests in the 'coda?package in R to test if the chains appear to be converged. The Geweke' diagnostic produces a Z-score for a test of equality of means between the first 10% and last 50% parts of the chain. Based on the Z-score, a corresponding p-value can be calculated. Table 4.1 and Table 4.2 present the Z-scores and P-values for the Geweke' tests for MCMC chains of the QAFT model parameters under ALD errors and normal mixture errors. Geweke' convergence diagnostic for MCMC chains of the proposed BNQAFT model parameters are shown in Table 4.3 and Table 4.5. All the p-values are greater than 0.05. Based on both graphic and non-graphic diagnostics, it seems that the MCMC chains mix well and converge.

Each simulation is repeated 500 times at three quantile levels, 0.3, 0.5 and 0.75. Table 4.6 summarized the 500 simulated datasets results for both QAFT and BN-QAFT models under ALD errors, the bias of posterior means, sample standard deviations (SSD), estimated standard deviations (ESD) and the corresponding 95% coverage probabilities (CPs) for the fixed effect coefficients are presented. The parameters were estimated using the proposed BNQAFT model fairly accurately, while estimation in QAFT model have larger bias under all three quantile levels though CPs

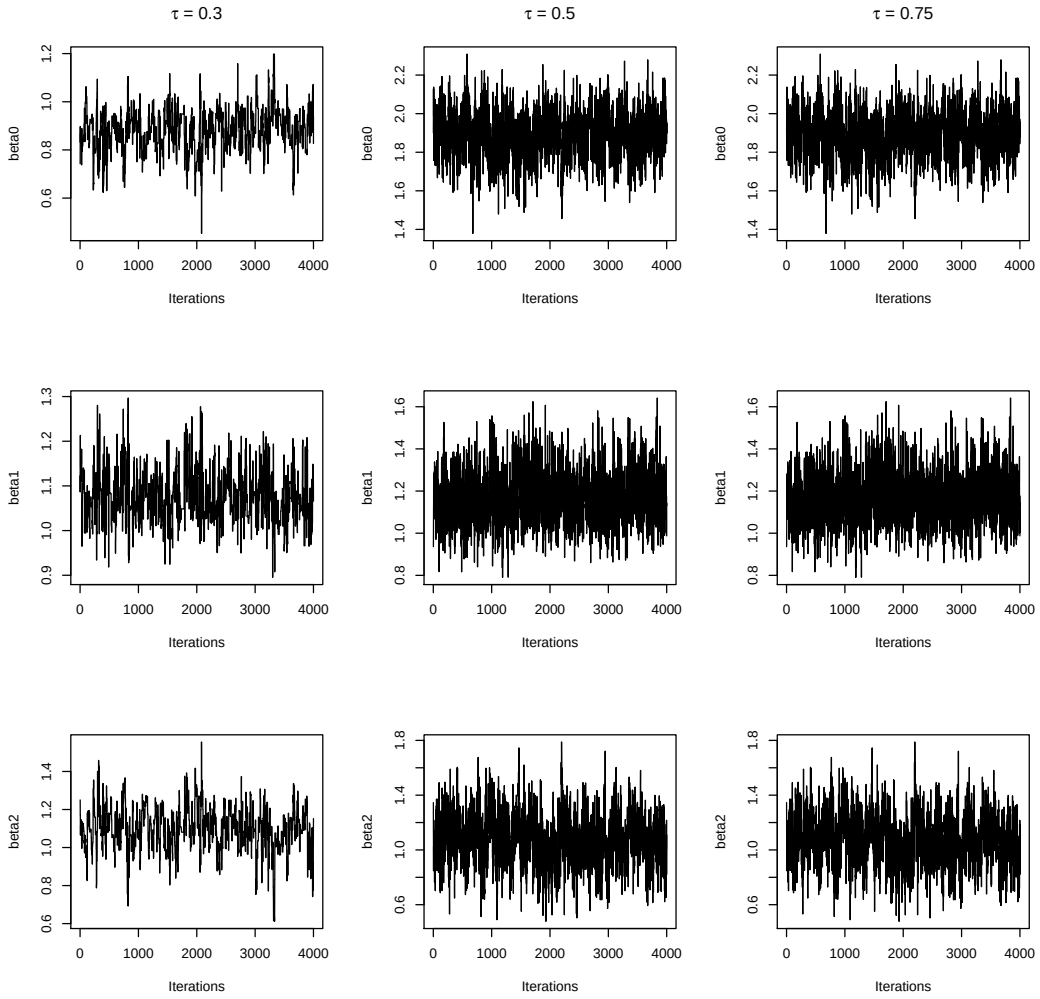


Figure 4.1 Traceplots of posteriors for MCMC chains of the model parameters with ASL errors

are quite high. It can be noted that the ESD is much larger than SSD in the QAFT and ESD in BNQAFT models. The ESD is the mean of the standard deviations of the parameter estimates, which indicates QAFT model's parameter estimates have large standard deviations such that the variations in MCMC chains are larger than the variations in BNQAFT model. In BNQAFT model, SSD and ESD are comparable with each other. For the 500 simulated datasets using Normal Mixture errors, the bias of posterior means, sample standard deviations (SSD), estimated standard deviations (ESD) and the corresponding 95% coverage probabilities (CPs) for the

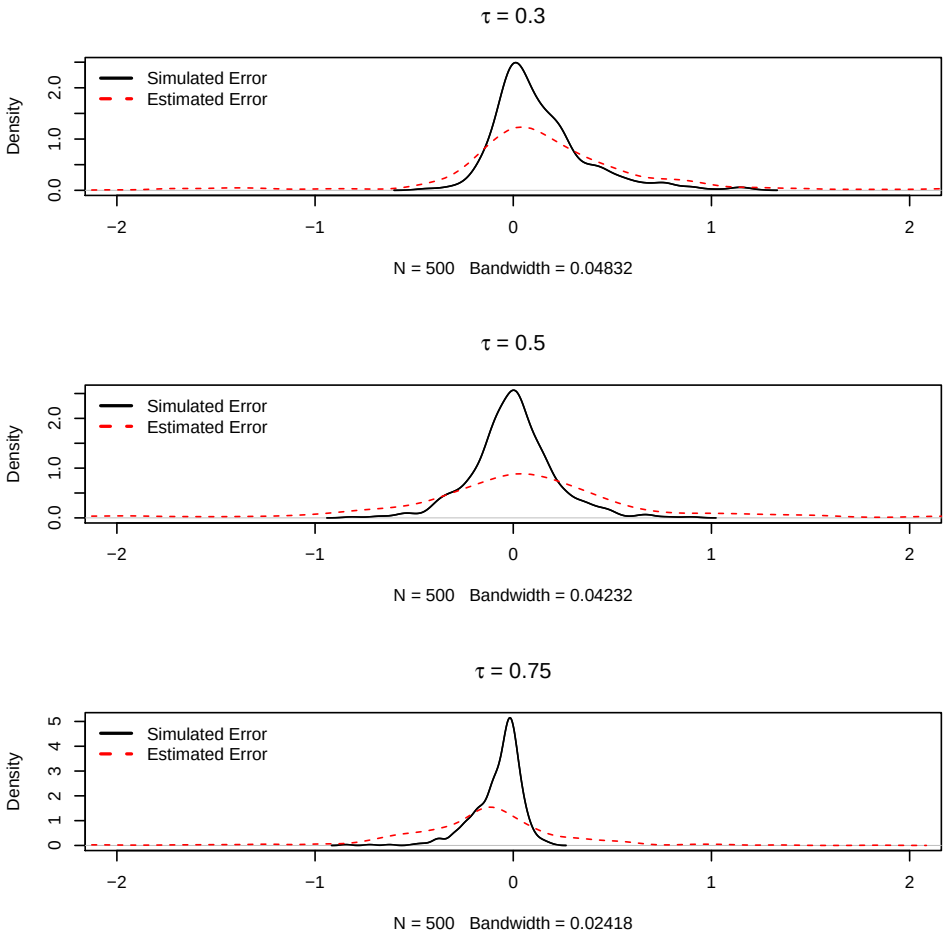


Figure 4.2 Error comparison for QAFT model with ALD errors

fixed effect coefficients using the proposed model and the QAFT are summarized in Table 4.7. The parameters were estimated using the proposed model fairly accurately with high coverage probabilities including the true values. The parameters were estimated using the proposed BNQAFT model fairly accurately, while estimation in QAFT model have larger bias under 0.3 and 0.5 levels. Again, the ESD in QAFT is very large which indicates some problem in estimation in QAFT models. We can say BNQAFT model outperforms QAFT model in these two simulation scenarios.

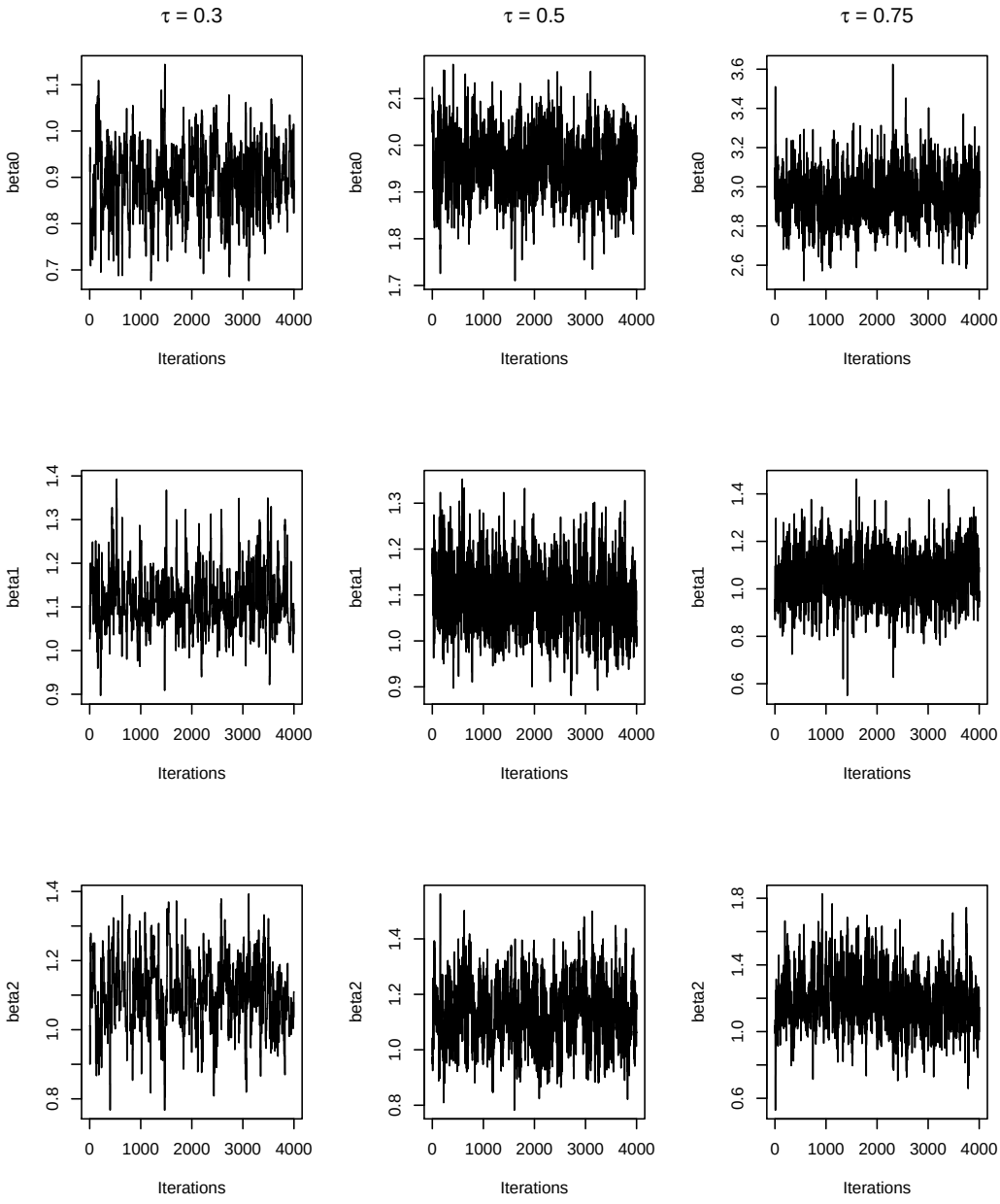


Figure 4.3 Traceplots of posteriors for MCMC chains of the model parameters with Normal mixture errors

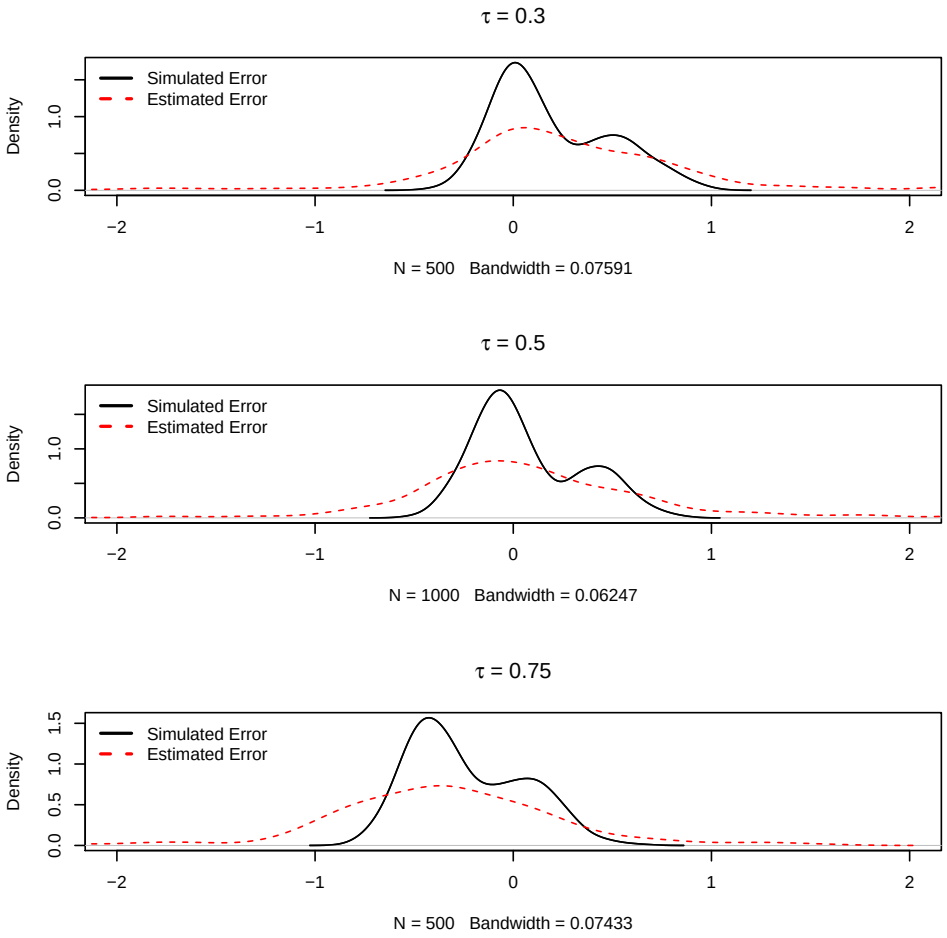


Figure 4.4 Error comparison for QAFT model with Normal mixture errors

4.5 Analysis of Lymphatic Filariasis Data

The proposed method is applied to the lymphatic filariasis study conducted in Recife, Brazil [Dreyer et al., 2006]. The goal is to compare the effectiveness of two treatments, co-administration of diethylcarbamazine and albendazole (DEC/ALB) (new treatment) versus DEC alone (standard treatment) for the treatment of lymphatic filariasis (adult Wuchereria bancrofti infection). The study consists of 47 subjects aging from 16 to 66. Among the 47 study subjects, 22 received co-administration of DEC and ALB while the remaining 25 were given DEC alone. Ultrasound examinations were done to detect worm deaths at $7, 14, 30, 45, 60, 90, 180, 270$ and 360

Table 4.1 Geweke's convergence diagnostic for MCMC chains of the QAFT model parameters with ALD errors

	$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.75$	
Parameter	Z score	P-value	Z score	P-value	Z score	P-value
β_0	-0.0436	0.9652	0.9273	0.3538	-0.7706	0.4409
β_1	0.2848	0.7758	0.3864	0.6992	0.5179	0.6046
β_2	0.3633	0.7164	0.1369	0.8911	0.3153	0.7525

Table 4.2 Geweke's convergence diagnostic for MCMC chains of the QAFT with Normal mixture errors

	$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.75$	
Parameter	Z score	P-value	Z score	P-value	Z score	P-value
β_0	-1.1069	0.2683	-0.5479	0.5838	-0.5479	0.5838
β_1	0.7577	0.4487	0.9359	0.3493	0.9359	0.3493
β_2	0.6158	0.5380	1.1677	0.2429	1.1677	0.2429

Table 4.3 Geweke's convergence diagnostic for MCMC chains of the BNQAFT with ALD errors

	$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.75$	
Parameter	Z score	P-value	Z score	P-value	Z score	P-value
β_0	0.2852	0.7755	-0.2312	0.8171	1.3499	0.1770
β_1	-0.8427	0.3994	-0.5075	0.6117	0.6830	0.4946
β_2	-1.0528	0.2924	-1.2749	0.2023	-0.5075	0.6118

days. The response variable of interest is the extinction time of nests. In total, 78 adult worm nests were detected by ultrasound; there was a range of $1 \sim 5$ nests per patient. In this study, a patient represents the cluster, and cluster size is number of adult worms nests.

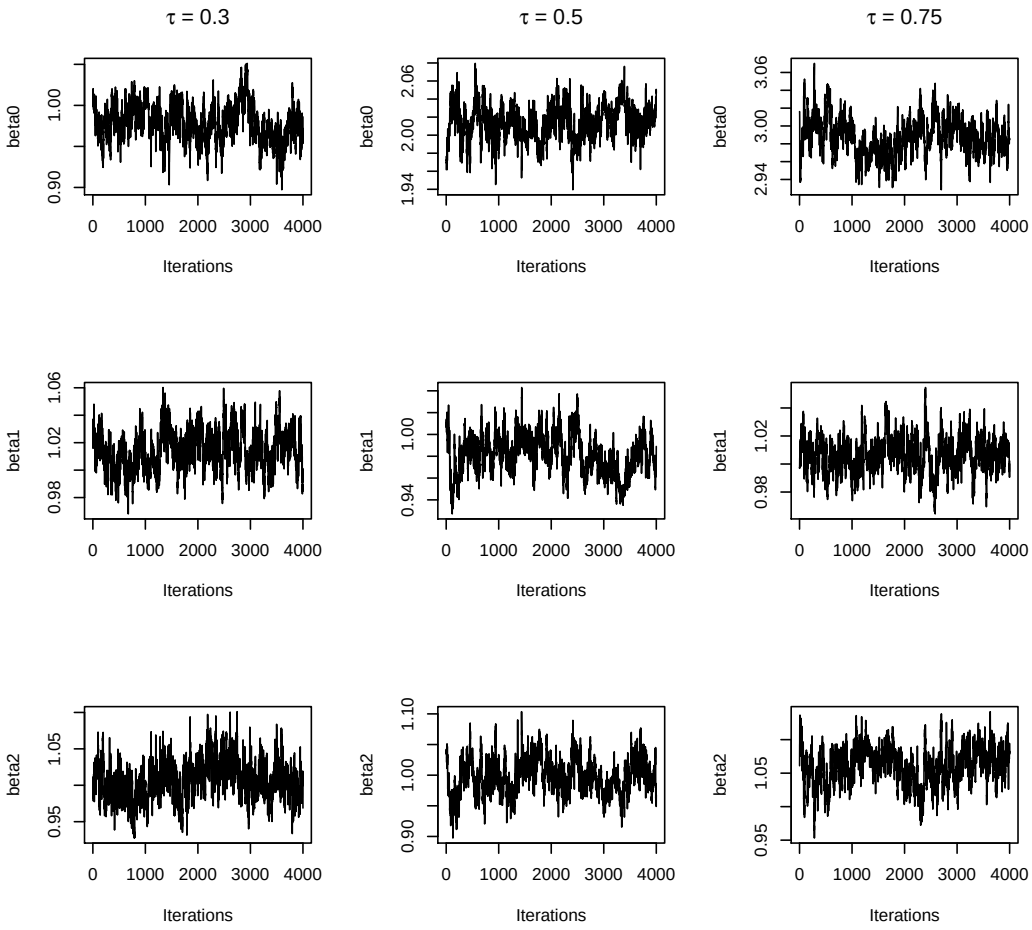


Figure 4.5 Traceplots of posteriors for proposed method MCMC chains of the model parameters with ALD errors

Table 4.4 Geweke's convergence diagnostic for MCMC chains of the BNQAFT with Normal Mixture errors

	$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.75$	
Parameter	Z score	P-value	Z score	P-value	Z score	P-value
β_0	0.2194	0.8263	-1.6914	0.0908	1.5080	0.1315
β_1	-1.0033	0.3157	0.7030	0.4821	-0.7420	0.4581
β_2	0.2901	0.7717	1.6399	0.1010	-1.6327	0.1025

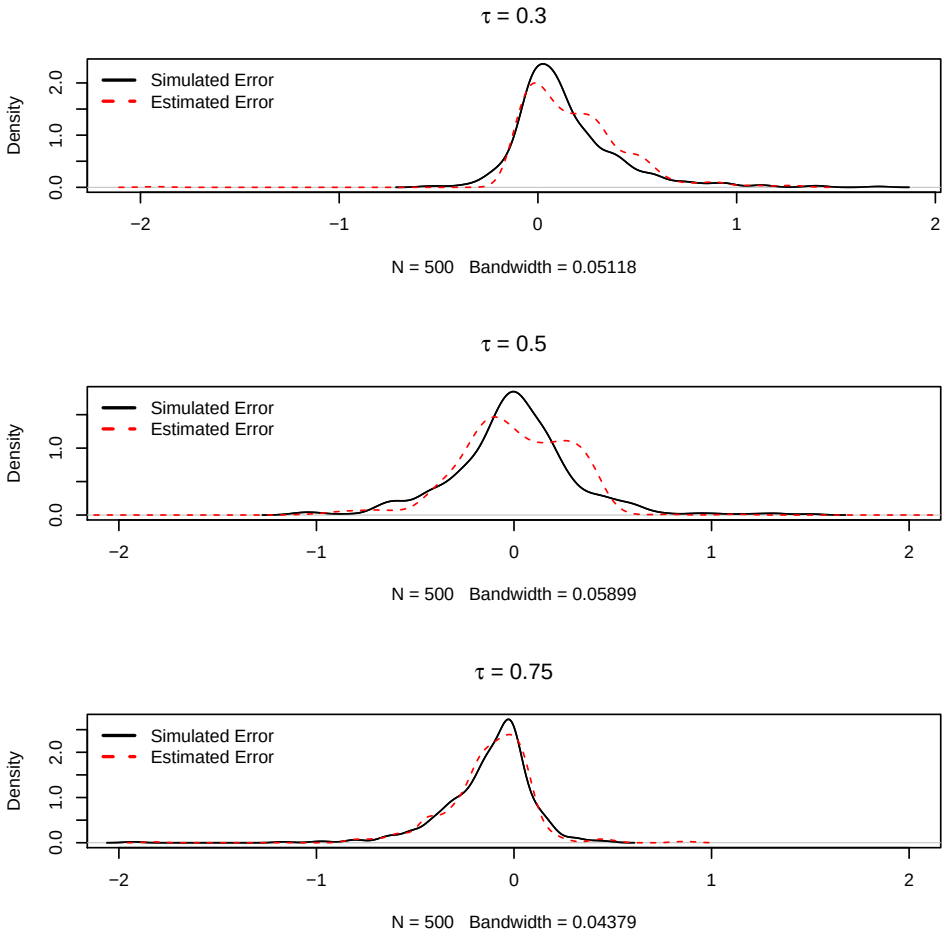


Figure 4.6 Error plot for proposed method each quantile level with ALD errors

Table 4.5 Geweke's convergence diagnostic for MCMC chains of the BNQAFT with Normal mixture errors

	$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.75$	
Parameter	Z score	P-value	Z score	P-value	Z score	P-value
β_0	0.2852	0.7755	-0.2312	0.8171	1.3499	0.1770
β_1	-0.8427	0.3994	-0.5075	0.6117	0.6830	0.4946
β_2	-1.0528	0.2924	-1.2749	0.2023	-0.5075	0.6118

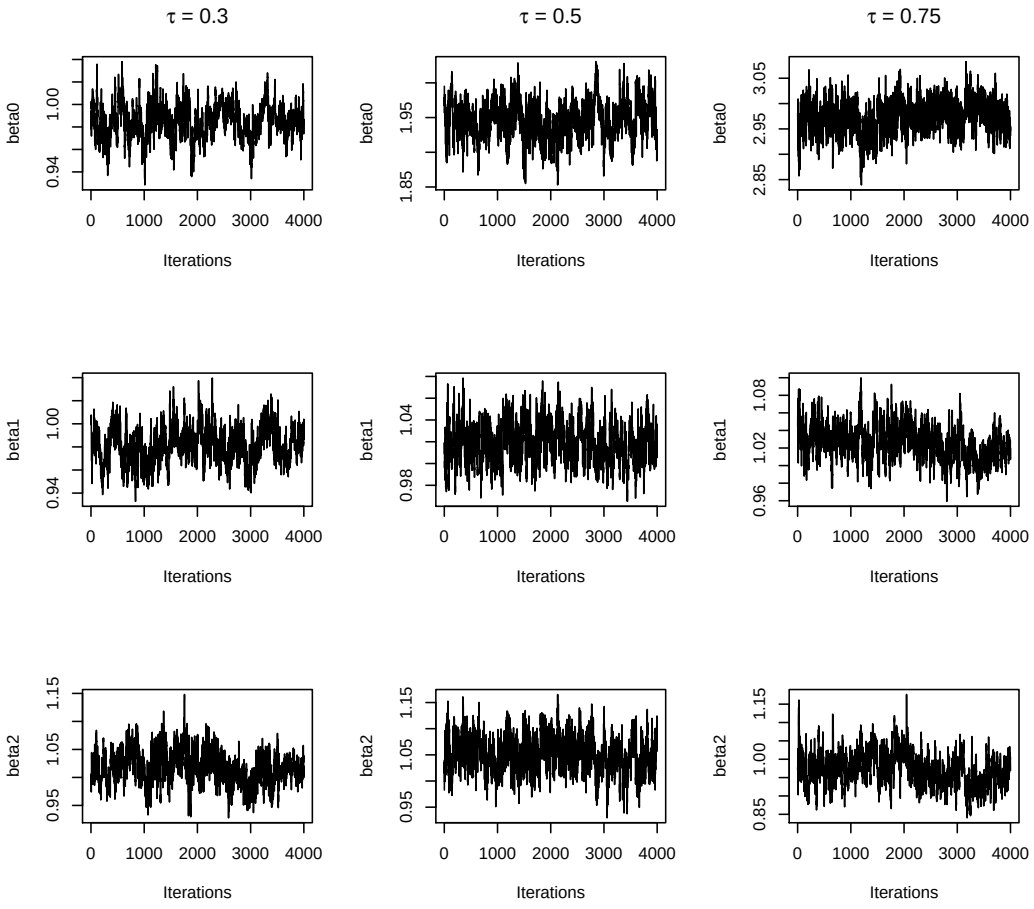


Figure 4.7 Traceplots of posteriors for proposed method MCMC chains of the model parameters with mix errors

This data set has been analyzed by Williamson et al. [2008], Zhang and Sun [2010], Kim [2010], Pan et al. [2015] using different modeling methods and under different assumptions. Thus their analysis suggest that the time-to-clearance of a nest may depend on cluster size in the corresponding patient. Williamson et al. [2008] investigate the estimation of the marginal distribution for multivariate survival data with informative cluster size using cluster-weighted Weibull and Cox proportional hazard model. They found that, as cluster size (i.e. the number of nests) increased, the proportion of nest clearance decreased. Figure 4.9 shows the estimated survival functions according to treatment groups based on the nonparametric maximum likelihood

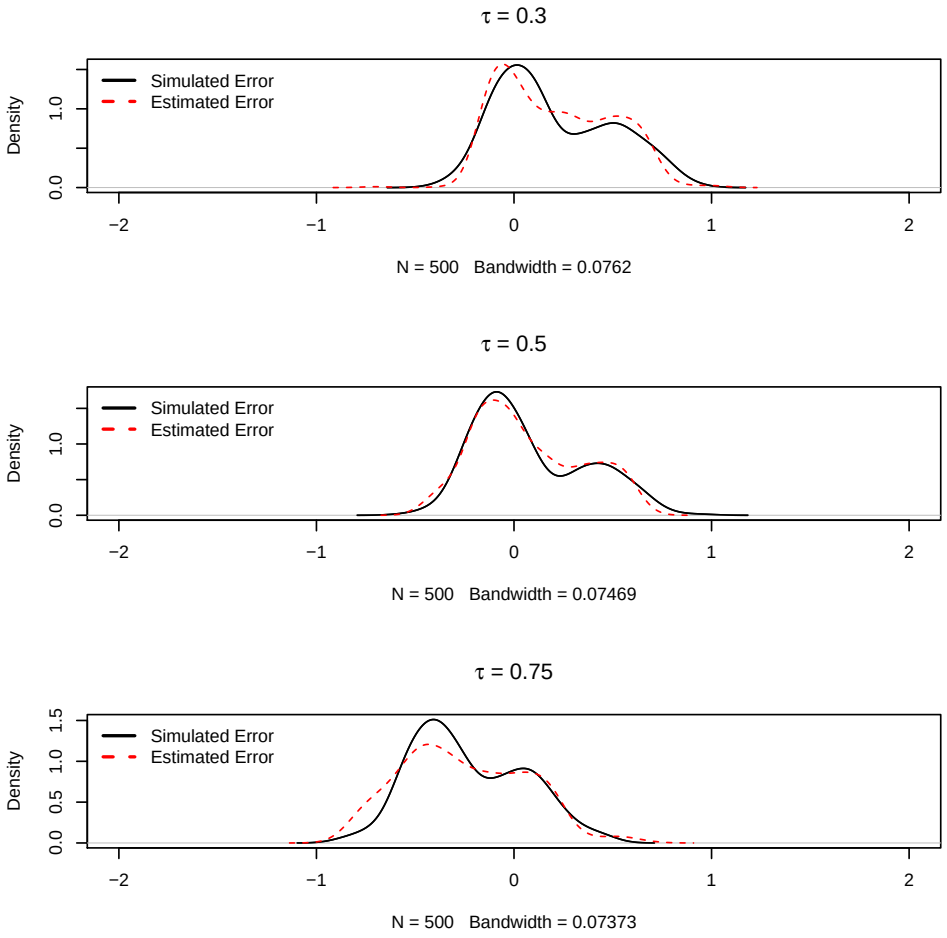


Figure 4.8 Error plot for proposed method each quantile level with mix errors

estimates (NPMLEs) using the Turnbull algorithm [Turnbull, 1976]. This plot does not take account of cluster size and the survival curves indicate that there is insufficient evidence to conclude that there is treatment difference in time to nest clearance. The survival curve for the DEC/ALB treatment is not separated from that for the DEC treatment, which suggests adding ALB does not improve the extinction of worm nests.

The main purpose is to compare the effectiveness of two treatments, DEC/ALB and DEC alone, on the eradication of adult worm nests. We fit our proposed model to the survival data with two covariates: treatment group (DEC/ALB=0, DEC=1) and age in years. The outcome variable is the nest-specific time when all the adult

Table 4.6 Simulation Results for two models with ALD errors

τ	Method		Bias	SSD	ESE	CPs	MSE
0.3	QAFT	β_0	0.11	0.026	0.078	0.85	0.014
		β_1	0.07	0.026	0.057	0.89	0.006
		β_2	0.09	0.032	0.099	0.99	0.009
0.3	BNQAFT	β_0	0.02	0.028	0.022	0.81	0.001
		β_1	0.01	0.016	0.014	0.88	0.000
		β_2	0.00	0.035	0.027	0.89	0.001
0.5	QAFT	β_0	0.10	0.038	0.084	0.98	0.011
		β_1	0.12	0.048	0.084	0.77	0.017
		β_2	0.12	0.036	0.130	0.99	0.016
0.5	BNQAFT	β_0	0.02	0.033	0.021	0.78	0.001
		β_1	0.01	0.021	0.013	0.80	0.001
		β_2	0.01	0.037	0.027	0.84	0.001
0.75	QAFT	β_0	0.08	0.020	0.089	0.99	0.007
		β_1	0.04	0.022	0.077	0.99	0.002
		β_2	0.06	0.027	0.121	0.99	0.004
0.75	BNQAFT	β_0	0.00	0.025	0.022	0.94	0.001
		β_1	0.01	0.017	0.014	0.93	0.000
		β_2	0.02	0.034	0.026	0.79	0.002

worms are cleared in the particular nest. In previous analysis, Williamson et al. [2008] and Kim [2010] reported the influence of age is insignificant on the failure times, but Zhang and Sun [2010]'s results suggest that it took a longer time for old patients to have their worms cleared than for young patients.

To make sure the MCMC chains converge to the stationary distribution, we show the traceplots (Figure 4.10) and Geweke diagnostic in the *CODA* package in R for regression coefficients $\beta_0, \beta_1, \beta_2$ for three quantile levels, 0.25, 0.5 and 0.75. We ran 10000 posterior samples and treated the first 5000 samples as burn-in. Table 4.8

Table 4.7 Simulation Results for two models with Normal Mixture errors

τ	Method		Bias	SSD	ESE	CPs	MSE
0.3	QAFT	β_0	0.08	0.031	0.081	0.98	0.007
		β_1	0.07	0.029	0.060	0.93	0.005
		β_2	0.08	0.042	0.106	0.99	0.008
0.3	BNQAFT	β_0	0.02	0.034	0.030	0.79	0.002
		β_1	0.01	0.019	0.018	0.94	0.000
		β_2	0.01	0.042	0.035	0.92	0.002
0.5	QAFT	β_0	0.02	0.021	0.059	0.99	0.001
		β_1	0.09	0.021	0.062	0.88	0.009
		β_2	0.11	0.033	0.103	0.99	0.012
0.5	BNQAFT	β_0	0.02	0.039	0.032	0.82	0.002
		β_1	0.01	0.023	0.019	0.88	0.001
		β_2	0.01	0.094	0.041	0.89	0.009
0.75	QAFT	β_0	0.01	0.035	0.106	0.99	0.001
		β_1	0.04	0.029	0.085	0.99	0.003
		β_2	0.07	0.052	0.141	0.99	0.008
0.75	BNQAFT	β_0	0.03	0.048	0.041	0.78	0.003
		β_1	0.02	0.030	0.031	0.92	0.001
		β_2	0.03	0.069	0.065	0.93	0.006

presents the z-scores for tests of equality of means and corresponding P-values for the Geweke tests. All the P-values are greater than 0.05. Based on both graphic and non-graphic diagnostics, it seems that the MCMC chains mix well and converge.

In Table 4.9, we have presented the fixed effects of treatment and age from our proposed model. We ran 5000 iterations for each quantile level. After 4000 burn-ins, 1000 iterations were used to calculate estimated parameters. The estimates of treatment group in three quantile levels are 1.240, 1.568 and 1.587, respectively. That is to say, the estimated acceleration factors are $\exp(1.240) = 3.46, \exp(1.568) =$

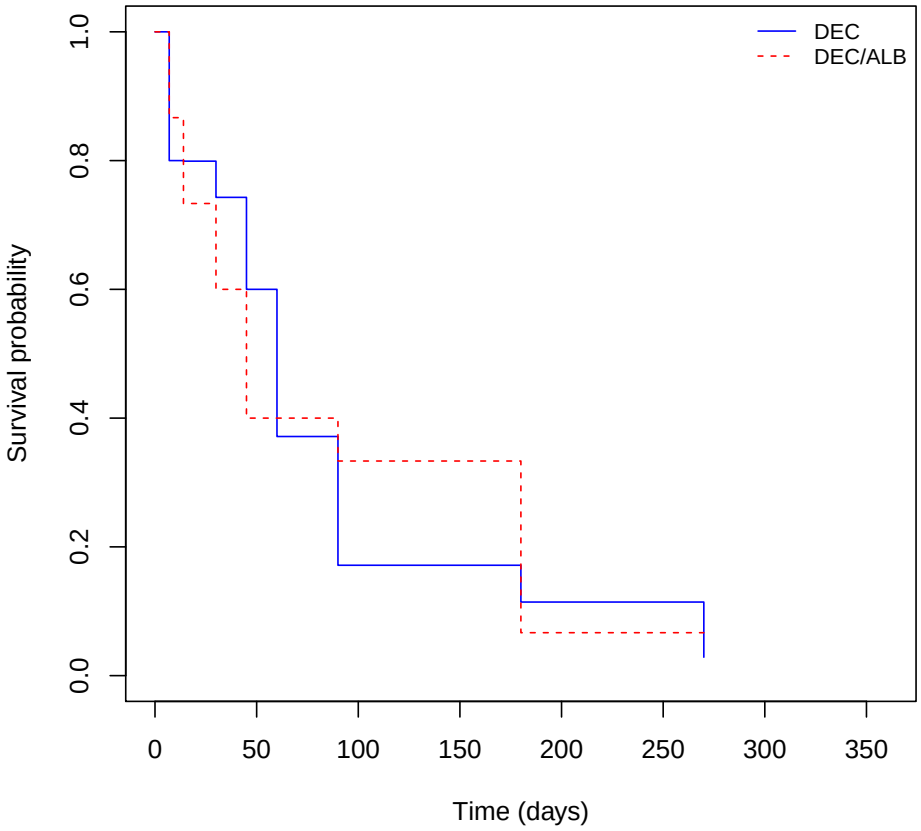


Figure 4.9 Survival curves by treatment group

$4.80, \exp(1.587) = 4.89$, which suggest that the DEC alone is effective for delaying nest extinction by stretching survival time by a factor of 4.80 for median level.

Overall, our estimation results suggest that DEC/ALB combination accelerates the extinction of nests than DEC alone. The treatment effect is not consistent with previous studies. For example, in Zhang and Sun [2010]'s analysis with a weighted estimating procedure, there is essentially no difference on adding ALB to DEC for the LF treatment. According to Kim [2010], they found that nests with treatment DEC alone have faster failures using a proportional hazard frailty model. The difference between results based on previous studies and those given here is due to the method of modeling and the assumption made. In previous approaches, the cluster size was

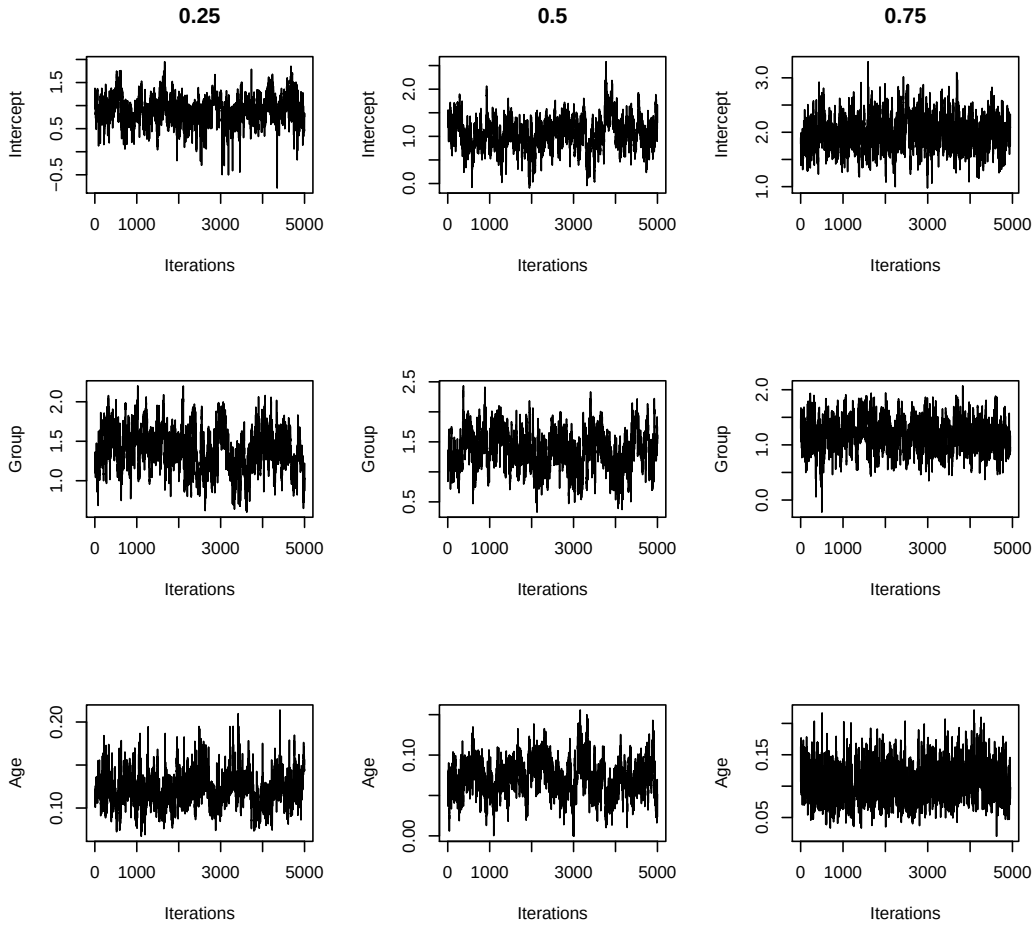


Figure 4.10 Traceplots of posteriors for MCMC chains of the model parameters in the lymphatic filariasis data

regarded as weights and weighted estimating procedures were used to estimate parameters. Besides, they have strong proportional hazard assumption. The relationship between the number of nests and clustered interval-censoring nest-extinction time is also an interesting topic but not explored in this study. Given that the true distributions of the random effects and error density are unknown, our model provide the capability of estimating the survival and random effect distributions nonparametrically.

Table 4.8 Geweke's convergence diagnostic for MCMC chains of the model parameters in the lymphatic filariasis data

	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
Parameter	Z score	P-value	Z score	P-value	Z score	P-value
Intercept	-0.9937	0.8398	0.2832	0.6115	-2.0185	0.9782
Treatment	1.5457	0.9389	0.3878	0.6509	0.3212	0.6260
Age	-0.6846	0.7532	-0.6473	0.7413	-0.4615	0.6778

Table 4.9 Estimates of parameters in the lymphatic filariasis data

τ	Parameter	Estimate	Standard Deviation	95 % CI
0.25	Intercept	1.338	0.292	(0.766, 1.910)
	Treatment	1.240	0.176	(0.890, 1.584)
	Age	0.102	0.015	(0.073, 0.131)
0.5	Intercept	1.933	0.336	(1.474, 2.392)
	Treatment	1.568	0.321	(1.139, 1.997)
	Age	0.051	0.021	(0.014, 0.088)
0.75	Intercept	2.245	0.222	(1.810, 2.680)
	treatment	1.587	0.253	(1.091, 2.083)
	Age	0.042	0.020	(0.002, 0.081)

Chapter 5

Concluding Remarks

In this dissertation, we build a model-based, fully inferential framework for semiparametric quantile regression for clustered data. In particular, we place emphasis on quantile regression models that allow the error density to change non-parametrically with the covariates. A mixture of two uniform densities is used for the error term in quantile regression for longitudinal data and clustered interval-censoring data. The performance of the proposed method was accessed under different error scenarios, such as normal, heavy-tailed, skewed or bimodal mixture distributions. Although capturing more general forms of skewness and tail behaviors, these nonparametric approaches also restrict the error densities to necessarily have their modes at the quantile of interest, particularly when modeling extreme quantiles. In practice, direct quantile estimates at the tails are often unstable due to data sparseness. Based on the extreme value theory, Wang et al. [2012] develop extrapolation methods for estimating the high conditional quantiles associated with heavy-tail distributions. Another drawback of our method is, if more than one quantile regressions are needed, the particular model need to be fitted separately at each quantile level. Through joint-quantile modeling, Jang and Wang [2015]'s method can yield the joint posterior distribution of quantile coefficients at multiple quantiles and meanwhile avoid the quantile crossing issue. They approximate the central density by linearly interpolating the conditional quantile functions of the response at multiple quantiles and estimate the tail densities by adopting extreme value theory.

The classical assumption for random effects is a Normal distribution [Molenberghs,

2000, Ruppert et al., 2003]. While this choice is mathematically convenient, it is often questionable in practice for several reasons. The normal distribution is symmetric, unimodal and has light tails. Since the distributional assumption is made on unobserved quantities, it is typically hard to validate. Possible skewness and multimodality (i.e. unconsidered groups in the data) may be masked when checking the normal distribution in terms of estimated random effects. A finite mixture of normal distributions as a random effects distribution suggested by Verbeke and Lesaffre [1996], Molenberghs [2000] is much more flexible. A data driven choice of the number of mixture components is desirable and could be achieved by a penalization of the mixture weights. Dirichlet process mixture for the random effects distribution is naturally considered. The main advantage of Dirichlet processes is its clustering property that we obtain a clustering of components automatically. Therefore, subgroup identification is possible. Dirichlet process priors for random effects has been mainly used in the Bayesian inference for density estimation and random effects models after being proposed by Kleinman and Ibrahim [1998a] and with the development of MCMC methods for sampling from the posterior distribution of Dirichlet process mixture models, which has enabled the application of nonparametric Bayesian methods to a variety of practical data analysis problems.

The proposed Bayesian quantile approach for inferences on clustered interval-censored failure time data has been proven to have good performance through our simulation study. This approach is especially useful in multi-center clinical trials for cancer or infectious diseases, since the detection of cancer progression or infection is normally made through periodic lab examinations. Since current literature for clustered interval-censored data are mainly involving shared-frailty PH models [Goethals et al., 2009, Zhang and Sun, 2010, Kim, 2010, Li et al., 2012, Pan et al., 2015], our developed model provides a new way to analyze this type of data.

Bibliography

Jason Abrevaya and Christian M Dahl. The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics*, 26(4):379–397, 2008.

Charles E Antoniak. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, pages 1152–1174, 1974.

Rebecca A Betensky, Daniel Rabinowitz, and Anastasios A Tsiatis. Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, 88(3):703–711, 2001.

Rebecca A Betensky, Jane C Lindsey, Louise M Ryan, and MP Wand. A local likelihood proportional hazards model for interval censored data. *Statistics in Medicine*, 21(2):263–275, 2002.

David Blackwell and James B MacQueen. Ferguson distributions via pólya urn schemes. *The annals of statistics*, pages 353–355, 1973.

David M. Blei and Michael I. Jordan. Variational inference for dirichlet process mixtures. *Bayesian Analysis*, 1:121–144, 2005.

Deborah Burr and Hani Doss. A bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association*, 2005.

Christopher A Bush and Steven N MacEachern. A semiparametric bayesian model for randomised block designs. *Biometrika*, 83(2):275–285, 1996.

Tianxi Cai and Rebecca A Betensky. Hazard regression for interval-censored data with penalized spline. *Biometrics*, 59(3):570–579, 2003.

Ivan A Canay. A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386, 2011.

Aysun Çetinyürek Yavuz and Philippe Lambert. Smooth estimation of survival functions and hazard ratios from interval-censored data using bayesian penalized b-splines. *Statistics in medicine*, 30(1):75–90, 2011.

Robert Crouchley and RB Davies. A comparison of population average and random-effect models for the analysis of longitudinal count data with base-line information. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 331–347, 1999.

Gregg E Dinse and SW Lagakos. Regression analysis of tumour prevalence data. *Applied statistics*, pages 236–248, 1983.

Robert M Dorazio, Bhramar Mukherjee, Li Zhang, Malay Ghosh, Howard L Jelks, and Frank Jordan. Modeling unobserved sources of heterogeneity in animal abundance using a dirichlet process prior. *Biometrics*, 64(2):635–644, 2008.

Gerusa Dreyer, David Addiss, John Williamson, and Joaquim NorÃţes. Efficacy of co-administered diethylcarbamazine and albendazole against adult wuchereria bancrofti. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(12):1118 – 1125, 2006. ISSN 0035-9203.

Michael D Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588, 1995.

Michael D Escobar and Mike West. *Computing nonparametric hierarchical models*. Springer, 1998.

Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

Dianne M Finkelstein. A proportional hazards model for interval-censored failure time data. *Biometrics*, pages 845–854, 1986.

Dianne M Finkelstein and Robert A Wolfe. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, pages 933–945, 1985.

Antonio F Galvao. Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164(1):142–157, 2011.

Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–534, 2006.

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis, 2014.

Marco Geraci. Linear quantile mixed models: the lqmm package for laplace quantile regression. *Journal of Statistical Software*, 57:1–29, 2014.

Marco Geraci and Matteo Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154, 2007.

Marco Geraci and Matteo Bottai. Linear quantile mixed models. *Statistics and Computing*, 24(3):461–479, 2014.

John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, 4:169–193, 1992.

Charles Geyer. Introduction to markov chain monte carlo. *Handbook of Markov Chain Monte Carlo*, pages 3–48, 2011.

K. Goethals, B. Ampe, D. Berkvens, H. Laevens, Paul Janssen, and Luc Duchateau. Modeling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(1):1–14, 2009. ISSN 1085-7117. doi: 10.1198/jabes.2009.0001. URL `http://dx.doi.org/10.1198/jabes.2009.0001`.

William B Goggins, Dianne M Finkelstein, David A Schoenfeld, and Alan M Zaslavsky. A markov chain monte carlo em algorithm for analyzing interval-censored data under the cox proportional hazards model. *Biometrics*, pages 1498–1507, 1998.

Philip Hougaard. Fundamentals of survival data. *Biometrics*, 55(1):13–22, 1999.

Philip Hougaard. *Analysis of multivariate survival data*. Springer Science & Business Media, 2012.

Peter J Huber. Robust statistics., 1981.

Woosung Jang and Huixia Judy Wang. A semiparametric bayesian approach for joint-quantile regression with clustered data. *Computational Statistics & Data Analysis*, 84:99–115, 2015.

Yang-Jin Kim. Regression analysis of clustered interval-censored data with informative cluster size. *Statistics in medicine*, 29(28):2956–2962, 2010.

Ken P Kleinman and Joseph G Ibrahim. A semi-parametric bayesian approach to generalized linear mixed models. *Statistics in Medicine*, 17(22):2579–2596, 1998a.

Ken P Kleinman and Joseph G Ibrahim. A semiparametric bayesian approach to the random effects model. *Biometrics*, pages 921–938, 1998b.

Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004.

Roger Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

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.

Arnošt Komárek and Emmanuel Lesaffre. Bayesian accelerated failure time model for correlated censored data with a normal mixture as an error distribution. *Statistica Sinica*, 17:549–569, 2007.

Athanasios Kottas and MILOVAN Krnjanić. Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, 36(2):297–319, 2009.

Minjung Kyung, Jeff Gill, and George Casella. Estimation in dirichlet random effects models. *The Annals of Statistics*, 38(2):979–1009, 2010.

Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

Carlos Lamarche. Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, 157(2):396–408, 2010.

Youngjo Lee, John A Nelder, et al. Conditional and marginal models: another view. *Statistical Science*, 19(2):219–238, 2004.

Junlong Li, Chunjie Wang, and Jianguo Sun. Regression analysis of clustered interval-censored failure time data with the additive hazards model. *Journal of nonparametric statistics*, 24(4):1041–1050, 2012.

Linxiong Li and Zongwei Pu. Rank estimation of log-linear regression with interval-censored data. *Lifetime data analysis*, 9(1):57–70, 2003.

Xiaoyan Lin, Bo Cai, Lianming Wang, and Zhigang Zhang. A bayesian proportional hazards model for general interval-censored data. *Lifetime data analysis*, pages 1–21, 2014.

James K Lindsey and Philippe Lambert. On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in medicine*, 17(4):447–469, 1998.

Jun S Liu. Nonparametric hierarchical bayes via sequential imputations. *The Annals of Statistics*, 24(3):911–930, 1996.

Steven N MacEachern. Estimating normal means with a conjugate style dirichlet process prior. *Communications in Statistics-Simulation and Computation*, 23(3): 727–741, 1994.

Steven N MacEachern and Peter Müller. Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, 7(2):223–238, 1998.

Maria Francesca Marino and Alessio Farcomeni. Linear quantile regression models for longitudinal experiments: an overview. *Metron*, 73(2):229–247, 2015.

Charles E McCulloch and John M Neuhaus. Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*, pages 388–402, 2011.

Geert Molenberghs. *Linear Mixed Models for Longitudinal Data. Springer Series in Statistics*. Springer, 2000.

Peter Müller and Fernando A Quintana. Nonparametric bayesian data analysis. *Statistical science*, pages 95–110, 2004.

Radford M Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265, 2000.

Chun Pan, Bo Cai, and Lianming Wang. Multiple frailty model for clustered interval-censored data with frailty selection. *Statistical methods in medical research*, page 0962280215576987, 2015.

Wei Pan. Extending the iterative convex minorant algorithm to the cox model for interval-censored data. *Journal of Computational and Graphical Statistics*, 8(1): 109–120, 1999.

Zongwei Pu and Linxiong Li. Regression models with arbitrarily interval-censored observations. *Communications in Statistics-Theory and Methods*, 28(7):1547–1563, 1999.

Daniel Rabinowitz, Anastasios Tsiatis, and Jorge Aragon. Regression with interval-censored data. *Biometrika*, 82(3):501–513, 1995.

Brian J Reich, Howard D Bondell, and Huixia J Wang. Flexible bayesian quantile regression for independent and clustered data. *Biostatistics*, 11(2):337–352, 2010.

David Ruppert, Matt P Wand, and Raymond J Carroll. *Semiparametric regression*. Number 12. Cambridge university press, 2003.

Glen A Satten. Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*, 83(2):355–370, 1996.

Jayaram Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

Simon Sheather. *A modern Approach to Regression with R*, volume 58. Springer Science & Business Media, 2009.

Debajyoti Sinha, Ming-Hui Chen, and Sujit K Ghosh. Bayesian analysis and model selection for interval-censored survival data. *Biometrics*, pages 585–590, 1999.

Luke B Smith, Montserrat Fuentes, Penny Gordon-Larsen, Brian J Reich, et al. Quantile regression for mixed models with an application to examine blood pressure trends in china. *The Annals of Applied Statistics*, 9(3):1226–1246, 2015.

Jianguo Sun. Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to aids studies. *Biometrics*, pages 1096–1104, 1995.

Jianguo Sun. *The statistical analysis of interval-censored failure time data*. Springer Science & Business Media, 2007.

Bruce W Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 290–295, 1976.

Florin Vaida and Suzette Blanchard. Conditional akaike information for mixed-effects models. *Biometrika*, 92(2):351–370, 2005.

James W Vaupel. Early death: an american tragedy. *Law and Contemporary Problems*, pages 73–121, 1976.

Geert Verbeke and Emmanuel Lesaffre. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91(433):217–221, 1996.

Huixia Judy Wang, Deyuan Li, and Xuming He. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464, 2012.

Ying Wei, Anneli Pere, Roger Koenker, and Xuming He. Quantile regression methods for reference growth charts. *Statistics in medicine*, 25(8):1369–1382, 2006.

John M Williamson, Hae-Young Kim, Amita Manatunga, and David G Addiss. Modeling survival data with informative cluster size. *Statistics in medicine*, 27(4): 543–555, 2008.

Hongqi Xue, KF Lam, Benjamin J Cowling, and Frank de Wolf. Semi-parametric accelerated failure time regression analysis with application to interval-censored hiv/aids data. *Statistics in medicine*, 25(22):3850–3863, 2006.

Keming Yu and Rana A Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.

Keming Yu and Jin Zhang. A three-parameter asymmetric laplace distribution and its extension. *Communications in StatisticsâĂŤTheory and Methods*, 34(9-10): 1867–1879, 2005.

Keming Yu, Zudi Lu, and Julian Stander. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):331–350, 2003.

Ying Yuan and Guosheng Yin. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, 66(1):105–114, 2010.

Ying Zhang, Lei Hua, and Jian Huang. A spline-based semiparametric maximum likelihood estimation method for the cox model with interval- censored data. *Scandinavian Journal of Statistics*, 37(2):338–354, 2010.

Zhigang Zhang and Jianguo Sun. Interval censoring. *Stat Methods Med Res*, pages 53–70, 2010.

Appendix A

Posterior Sampling under Nonparametric error distributions

The posterior distributions for each parameter are derived one by one. Models with non-conjugate priors can be handled by applying Gibbs sampling with auxilliary parameters. We use this technique to update c_i for a Dirichlet process mixture model without haveing to integrate with respect G_0.

A.1 Sampling from posterior distribution for σ_{ri}

Because σ_{ri} depends on y_{ij} through $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$, and $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ is supposed to be known. In other words, $l(y_{ij}|\sigma_{ri})p(\sigma_{ri})$ is equivalent to $l(y_{ij})p(\boldsymbol{\beta}|\sigma_{ri})p(\boldsymbol{b_i}|\sigma_{ri})p(\sigma_{ri})$. The likelihood of i-th observation is $F(\boldsymbol{y_i}, \sigma_{ri}) = \prod_{j=1}^{n_i} l(y_{ij}|\sigma_{ri})$

$$\begin{aligned}
p(\sigma_{ri}|y_{ij}, G_{r0}) &\propto F(\boldsymbol{y_i}, \sigma_{ri})\text{prior}(\sigma_{ri}) \\
&\propto F(\boldsymbol{y_i}, \sigma_{ri})\text{prior}(\sigma_{ri}) \\
&\propto \frac{\alpha_r}{\alpha_r + n - 1} G_{r0}(\sigma_{ri}) F(\boldsymbol{y_i}, \sigma_{ri}) + \frac{1}{\alpha_r + n - 1} \sum_{g \neq i} \delta(\sigma_{rg}) F(\boldsymbol{y_i}, \sigma_{ri}) \\
&\propto \frac{\alpha_r}{\alpha_r + n - 1} \text{Unif}(d_r) F(\boldsymbol{y_i}, \sigma_{ri}) + \frac{1}{\alpha_r + n - 1} \sum_{g \neq i} \delta(\sigma_{rg}) F(\boldsymbol{y_i}, \sigma_{ri})
\end{aligned}$$

which can be written as $r_i H_i + \sum_{g \neq i} q_{i,g} \delta(\theta_g)$, H_i is the posterior distribution for σ_r based on the prior G_{r0} and the single observation i, with likelihood $F(\boldsymbol{y_i}, \sigma_{ri})$. The

values of the $q_{i,g}$ and of r_i are written as follows, where b is such that $\sum_{g \neq i} q_{i,g} + r_i = 1$.

$$\begin{aligned} r_i H_i &\propto \frac{\alpha_r}{\alpha_r + n - 1} \text{Unif}(d_r) F(\boldsymbol{y_i}, \sigma_{ri}) \\ r_i &= b\alpha_r \int \frac{1}{d_r} F(\boldsymbol{y_i}, \sigma_{ri}) d\sigma_{ri} \\ H_i &= \frac{1}{d_r} F(\boldsymbol{y_i}, \sigma_{ri}) \\ q_{i,g} &= bF(\boldsymbol{y_i}, \sigma_{rg}), g = 1, \ldots, i-1, i+1, \ldots, n \end{aligned}$$

M-H to all $c \in c_1, \ldots, c_n$, update $\Phi_c | y_i$ such that $c_i = c$: we suppose the proposal density is lognormal, so $\sigma^*_{rc} \propto$ logNormal and $X = \log \sigma^*_{rc} \propto$ Normal

$$f(X) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-(x-\mu)^2/(2\tau^2)}$$

$$f(\sigma^*_{rc}) = \frac{1}{\sqrt{2\pi\tau^2}} \frac{e^{-(\sigma^*_{rc}-\mu)^2/(2\tau^2)}}{\sigma^*_{rc}}$$

The ratio

$$r = \frac{p(\sigma^*_{rc}|y)/J(\sigma^*_{rc}|\sigma^{t-1}_{rc})}{p(\sigma^{t-1}_{rc}|y)/J(\sigma^{t-1}_{rc}|\sigma^*_{rc})}$$

$$J(\sigma^*_{rc}|\sigma^{t-1}_{rc}) = \frac{1}{\sqrt{2\pi\tau^2}\sigma^*_{rc}} e^{-(\log\sigma^*_{rc} - \log\sigma^{t-1}_{rc})^2/(2\tau^2)}$$

$$J(\sigma^{t-1}_{rc}|\sigma^*_{rc}) = \frac{1}{\sqrt{2\pi\tau^2}\sigma^{t-1}_{rc}} e^{-(\log\sigma^{t-1}_{rc} - \log\sigma^*_{rc})^2/(2\tau^2)}$$

$$\log r = \log[p(\sigma^*_{rc}|y)] - \log[p(\sigma^{t-1}_{rc}|y)] + \log\sigma^*_{rc} - \log\sigma^{t-1}_{rc}$$

where $p(\sigma^*_{rc}|y) = \prod_{i \in c} H_i$

A.2 Sampling from posterior distributions for $\boldsymbol{\beta}$

$$\begin{aligned} p(\boldsymbol{\beta}|\boldsymbol{y_i}) &\propto \text{prior}(\boldsymbol{\beta}) \prod_{i=1}^{n} L(\boldsymbol{y_i}|\boldsymbol{\beta}) \\ &\propto N_p(\mathbf{0}, \Sigma_0) \prod_{i=1}^{n} p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i}) \\ &\propto \frac{1}{(2\pi)^{1/2} \mid \Sigma_0 \mid^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{\beta} - \mathbf{0})^T \Sigma_0^{-1} (\boldsymbol{\beta} - \mathbf{0}) \right) \end{aligned}$$

$$\times \prod_{i=1}^{n}\prod_{j=1}^{n_i}\frac{p}{\sigma_{1i}}1_{(-\sigma_{1i},0)}(y_{ij}-X_{ij}\boldsymbol{\beta}-Z_{ij}\boldsymbol{b_i})+\frac{1-p}{\sigma_{2i}}1_{[0,\sigma_{2i})}(y_{ij}-X_{ij}\boldsymbol{\beta}-Z_{ij}\boldsymbol{b_i})$$

M-H: we suppose the proposal(jump) distribution for $\boldsymbol{\beta}$ is multivariate normal with covariance matrix T_0,

$$J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1})=\frac{1}{(2\pi)^{1/2}\mid T_0\mid^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{t-1})^T T_0^{-1}(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{t-1})\Big)$$

$$J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)=\frac{1}{(2\pi)^{1/2}\mid T_0\mid^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{\beta}^{t-1}-\boldsymbol{\beta}^*)^T T_0^{-1}(\boldsymbol{\beta}^{t-1}-\boldsymbol{\beta}^*)\Big)$$

Since the jump distribution is symmetric, $J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1})=J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)$, we compute acceptance ratio

$$r=\frac{p(\boldsymbol{\beta}^*|\boldsymbol{y_i})/J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1})}{p(\boldsymbol{\beta}^{t-1}|\boldsymbol{y_i})/J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)}$$

$$\log r=\log[p(\boldsymbol{\beta}^*|\boldsymbol{y_i})]-\log[p(\boldsymbol{\beta}^{t-1}|\boldsymbol{y_i})]$$

A.3 Sampling from posterior distributions for each $\boldsymbol{b_i}$, suppose D is known

$$\begin{aligned}
p(\boldsymbol{b_i}|\boldsymbol{y_i}) &\propto \text{prior}(\boldsymbol{b_i}|D)L(\boldsymbol{y_i}|\boldsymbol{b_i})\\
&\propto N_p(\mathbf{0},D)p(\boldsymbol{y_i}|\boldsymbol{\beta},\boldsymbol{b_i},\sigma_{1i},\sigma_{2i})\\
&\propto \frac{1}{(2\pi)^{1/2}\mid D\mid^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{b_i}-\mathbf{0})^T D^{-1}(\boldsymbol{b_i}-\mathbf{0})\Big)\\
&\times \prod_{j=1}^{n_i}\frac{p}{\sigma_{1i}}1_{(-\sigma_{1i},0)}(y_{ij}-X_{ij}\boldsymbol{\beta}-Z_{ij}\boldsymbol{b_i})+\frac{1-p}{\sigma_{2i}}1_{[0,\sigma_{2i})}(y_{ij}-X_{ij}\boldsymbol{\beta}-Z_{ij}\boldsymbol{b_i})
\end{aligned}$$

M-H: we suppose the proposal(jump) distribution for $\boldsymbol{b_i}$ is multivariate normal with covariance matrix T_1,

$$J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1})=\frac{1}{(2\pi)^{1/2}\mid T_1\mid^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{b_i}^*-\boldsymbol{b_i}^{t-1})^T T_1^{-1}(\boldsymbol{b_i}^*-\boldsymbol{b_i}^{t-1})\Big)$$

$$J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*)=\frac{1}{(2\pi)^{1/2}\mid T_1\mid^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{b_i}^{t-1}-\boldsymbol{b_i}^*)^T T_1^{-1}(\boldsymbol{b_i}^{t-1}-\boldsymbol{b_i}^*)\Big)$$

Since the jump distribution is symmetric, $J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1}) = J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*)$, we compute acceptance ratio

$$r = \frac{p(\boldsymbol{b_i}^*|\boldsymbol{y_i})/J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1})}{p(\boldsymbol{b_i}^{t-1}|\boldsymbol{y_i})/J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*)}$$

$$\log r = \log[p(\boldsymbol{b_i}^*|\boldsymbol{y_i})] - \log[p(\boldsymbol{b_i}^{t-1}|\boldsymbol{y_i})]$$

Then, we can directly sample D using conjugate prior, $D^{-1} \sim \text{Wishart}_p(\nu_0, \Phi_0)$

$$\begin{aligned}
p(D^{-1}|\nu_0, \Phi_0) &\propto |\ D^{-1}\ |^{(\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\text{tr}(\Phi_0^{-1}D^{-1})\Big) \\
p(D^{-1}|\boldsymbol{b_i}) &\propto \prod_{i=1}^{n} p(\boldsymbol{b_i}|D)p(D^{-1}|\nu_0, \Phi_0) \\
&\propto \frac{1}{(2\pi)^{n/2}\ |\ D\ |^{n/2}} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{b_i}-\boldsymbol{0})^T D^{-1}(\boldsymbol{b_i}-\boldsymbol{0})\Big) \\
&\times |\ D^{-1}\ |^{(\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\text{tr}(\Phi_0^{-1}D^{-1})\Big) \\
&\propto |\ D^{-1}\ |^{(n+\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\Big(\text{tr}(\Phi_0^{-1}D^{-1}) - \frac{1}{2}\text{tr}(D^{-1}\sum_{i=1}^{n}\boldsymbol{b_i}\boldsymbol{b_i}^T)\Big)\Big) \\
&\propto |\ D^{-1}\ |^{(n+\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\Big(\text{tr}(\Phi_0^{-1}D^{-1} + \sum_{i=1}^{n}\boldsymbol{b_i}\boldsymbol{b_i}^T D^{-1})\Big)\Big) \\
D^{-1}|\boldsymbol{b_i} &\sim \text{Wishart}_p(n+\nu_0, \Big(\Phi_0^{-1} + \sum_{i=1}^{n}\boldsymbol{b_i}\boldsymbol{b_i}^T\Big)^{-1})
\end{aligned}$$

A.4 Posterior Sampling of QAFT for clustered-interval data

The posterior distributions for each parameter are derived one by one. Models with non-conjugate priors can be handled by applying Gibbs sampling with auxilliary parameters. We use this technique to update c_i for a Dirichlet process mixture model without haveing to integrate with respect G_0.

Table A.1 Determination of precision parameters by DIC at quantile level 0.1

α_1	α_2	α_b	DIC	α_1	α_2	α_b	DIC
0.1	0.1	0.1	1416.809	5	0.1	0.1	1339.026
0.1	0.1	1	1214.19	5	0.1	1	1277.33
0.1	0.1	5	1136.512	5	0.1	5	970.2703
0.1	0.1	20	1125.866	5	0.1	20	936.0066
0.1	1	0.1	1438.607	5	1	0.1	1343.849
0.1	1	1	1161.708	5	1	1	1209.962
0.1	1	5	1059.156	5	1	5	1182.65
0.1	1	20	932.935	5	1	20	1023.581
0.1	5	0.1	1380.349	5	5	0.1	1217.745
0.1	5	1	1194.234	5	5	1	1148.787
0.1	5	5	869.5041	5	5	5	957.5466
0.1	5	20	872.2571	5	5	20	1162.559
0.1	20	0.1	1409.015	5	20	0.1	1312.906
0.1	20	1	905.8409	5	20	1	1137.531
0.1	20	5	974.3051	5	20	5	1129.593
0.1	20	20	1139.726	5	20	20	848.4808

Sampling from posterior distribution for σ_i

Because σ_i depends on y_{ij} through $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$, and $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ is supposed to be known. In other words, $l(y_{ij}|\sigma_i)p(\sigma_i)$ is equivalent to $l(y_{ij})p(\boldsymbol{\beta}|\sigma_i)p(\boldsymbol{b_i}|\sigma_i)p(\sigma_i)$. The likelihood of i-th observation is $F(\boldsymbol{y_i}, \sigma_i) = \prod_{j=1}^{n_i} l(y_{ij}|\sigma_i)$

$$
\begin{aligned}
p(\sigma_i|y_{ij}, G_0) &\propto F(\boldsymbol{y_i}, \sigma_i)\text{prior}(\sigma_i) \\
&\propto F(\boldsymbol{y_i}, \sigma_i)\text{prior}(\sigma_i) \\
&\propto \frac{\alpha}{\alpha+n-1}G_0(\sigma_i)F(\boldsymbol{y_i}, \sigma_i) + \frac{1}{\alpha+n-1}\sum_{g\neq i}\delta(\sigma_g)F(\boldsymbol{y_i}, \sigma_i) \\
&\propto \frac{\alpha}{\alpha+n-1}\text{Unif}(d)F(\boldsymbol{y_i}, \sigma_i) + \frac{1}{\alpha+n-1}\sum_{g\neq i}\delta(\sigma_g)F(\boldsymbol{y_i}, \sigma_i)
\end{aligned}
$$

Table A.2 Determination of precision parameters by DIC at quantile level 0.1 (continued)

α_1	α_2	α_b	DIC	α_1	α_2	α_b	DIC
1	0.1	0.1	1404.649	20	0.1	0.1	1423.355
1	0.1	1	1015.794	20	0.1	1	1194.149
1	**0.1**	**5**	**776.7351**	20	0.1	5	1125.351
1	0.1	20	1031.909	20	0.1	20	1256.755
1	1	0.1	1310.01	20	1	0.1	1322.58
1	1	1	1241.487	20	1	1	1194.209
1	1	5	916.3833	20	1	5	1160.246
1	1	20	805.5571	20	1	20	1239.188
1	5	0.1	1348.073	20	5	0.1	1364.133
1	5	1	1260.412	20	5	1	1138.306
1	5	5	989.2601	20	5	5	1239.804
1	5	20	1047.878	20	5	20	1057.974
1	20	0.1	1444.808	20	20	0.1	1226.757
1	20	1	1173.077	20	20	1	1234.787
1	20	5	1005.932	20	20	5	1225.193
1	20	20	1066.801	20	20	20	1096.868

which can be written as $r_i H_i + \sum_{g \neq i} q_{i,g} \delta(\theta_g)$, H_i is the posterior distribution for σ based on the prior G_0 and the single observation i, with likelihood $F(\boldsymbol{y_i}, \sigma_i)$. The values of the $q_{i,g}$ and of r_i are written as follows, where b is such that $\sum_{g \neq i} q_{i,g} + r_i = 1$.

$$\begin{aligned} r_i H_i &\propto \frac{\alpha}{\alpha + n - 1} \text{Unif}(d) F(\boldsymbol{y_i}, \sigma_i) \\ r_i &= b\alpha \int \frac{1}{d_r} F(\boldsymbol{y_i}, \sigma_i) d\sigma_i \\ H_i &= \frac{1}{d} F(\boldsymbol{y_i}, \sigma_i) \\ q_{i,g} &= bF(\boldsymbol{y_i}, \sigma_g), g = 1, \ldots, i-1, i+1, \ldots, n \end{aligned}$$

M-H to all $c \in c_1, \ldots, c_n$, update $\Phi_c | y_i$ such that $c_i = c$: we suppose the proposal

density is lognormal, so $\sigma_{rc}^* \propto$ logNormal and $X = \log\sigma_c^* \propto$ Normal

$$f(X) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-(x-\mu)^2/(2\tau^2)}$$

$$f(\sigma_c^*) = \frac{1}{\sqrt{2\pi\tau^2}} \frac{e^{-(\sigma_c^*-\mu)^2/(2\tau^2)}}{\sigma_c^*}$$

The ratio

$$r = \frac{p(\sigma_c^*|y)/J(\sigma_c^*|\sigma_c^{t-1})}{p(\sigma_c^{t-1}|y)/J(\sigma_c^{t-1}|\sigma_c^*)}$$

$$J(\sigma_c^*|\sigma_c^{t-1}) = \frac{1}{\sqrt{2\pi\tau^2}\sigma_c^*} e^{-(\log\sigma_c^*-\log\sigma_c^{t-1})^2/(2\tau^2)}$$

$$J(\sigma_c^{t-1}|\sigma_c^*) = \frac{1}{\sqrt{2\pi\tau^2}\sigma_c^{t-1}} e^{-(\log\sigma_c^{t-1}-\log\sigma_c^*)^2/(2\tau^2)}$$

$$\log r = \log[p(\sigma_c^*|y)] - \log[p(\sigma_c^{t-1}|y)] + \log\sigma_c^* - \log\sigma_c^{t-1}$$

where $p(\sigma_c^*|y) = \prod_{i\in c} H_i$

Sampling from posterior distributions for $\boldsymbol{\beta}$

$$\begin{aligned} p(\boldsymbol{\beta}|\boldsymbol{y_i}) &\propto \text{prior}(\boldsymbol{\beta}) \prod_{i=1}^{n} L(\boldsymbol{y_i}|\boldsymbol{\beta}) \\ &\propto N_p(\mathbf{0}, \Sigma_0) \prod_{i=1}^{n} p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_i) \\ &\propto \frac{1}{(2\pi)^{1/2} \mid \Sigma_0 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{0})^T \Sigma_0^{-1} (\boldsymbol{\beta} - \mathbf{0}) \right) \end{aligned}$$

$$\times \prod_{i=1}^{n} \frac{1}{\sigma_i^{n_i}} \exp\left\{ -\sum_{j=1}^{n_i} \left(\frac{y_{ij} - X_{ij}^T\boldsymbol{\beta} - Z_{ij}\boldsymbol{b}_i}{\sigma_i} \right) \left(\tau - I\left(\frac{y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b}_i}{\sigma_i} \leq 0 \right) \right) \right\}$$

M-H: we suppose the proposal(jump) distribution for $\boldsymbol{\beta}$ is multivariate normal with covariance matrix T_0,

$$J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1}) = \frac{1}{(2\pi)^{1/2} \mid T_0 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{t-1})^T T_0^{-1} (\boldsymbol{\beta}^* - \boldsymbol{\beta}^{t-1}) \right)$$

$$J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*) = \frac{1}{(2\pi)^{1/2} \mid T_0 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}^{t-1} - \boldsymbol{\beta}^*)^T T_0^{-1} (\boldsymbol{\beta}^{t-1} - \boldsymbol{\beta}^*) \right)$$

Since the jump distribution is symmetric, $J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1}) = J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)$, we compute acceptance ratio

$$r = \frac{p(\boldsymbol{\beta}^*|\boldsymbol{y_i})/J(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1})}{p(\boldsymbol{\beta}^{t-1}|\boldsymbol{y_i})/J(\boldsymbol{\beta}^{t-1}|\boldsymbol{\beta}^*)}$$

$$\log r = \log[p(\boldsymbol{\beta}^*|\boldsymbol{y_i})] - \log[p(\boldsymbol{\beta}^{t-1}|\boldsymbol{y_i})]$$

Sampling from posterior distributions for each $\boldsymbol{b_i}$, suppose D is known

$$\begin{aligned}
p(\boldsymbol{b_i}|\boldsymbol{y_i}) &\propto \text{prior}(\boldsymbol{b_i}|D)L(\boldsymbol{y_i}|\boldsymbol{b_i}) \\
&\propto N_p(\mathbf{0}, D)p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i}, \sigma_{1i}, \sigma_{2i}) \\
&\propto \frac{1}{(2\pi)^{1/2} \mid D \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{b_i} - \mathbf{0})^T D^{-1}(\boldsymbol{b_i} - \mathbf{0}) \right) \\
&\times \frac{1}{\sigma_i^{n_i}} \exp\left\{ -\sum_{j=1}^{n_i} \left(\frac{y_{ij} - X_{ij}^T\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}}{\sigma_i} \right) \left(\tau - I(\frac{y_{ij} - X_{ij}\boldsymbol{\beta} - Z_{ij}\boldsymbol{b_i}}{\sigma_i} \leq 0) \right) \right\}
\end{aligned}$$

M-H: we suppose the proposal(jump) distribution for $\boldsymbol{b_i}$ is multivariate normal with covariance matrix T_1,

$$J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1}) = \frac{1}{(2\pi)^{1/2} \mid T_1 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{b_i}^* - \boldsymbol{b_i}^{t-1})^T T_1^{-1}(\boldsymbol{b_i}^* - \boldsymbol{b_i}^{t-1}) \right)$$

$$J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*) = \frac{1}{(2\pi)^{1/2} \mid T_1 \mid^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{b_i}^{t-1} - \boldsymbol{b_i}^*)^T T_1^{-1}(\boldsymbol{b_i}^{t-1} - \boldsymbol{b_i}^*) \right)$$

Since the jump distribution is symmetric, $J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1}) = J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*)$, we compute acceptance ratio

$$r = \frac{p(\boldsymbol{b_i}^*|\boldsymbol{y_i})/J(\boldsymbol{b_i}^*|\boldsymbol{b_i}^{t-1})}{p(\boldsymbol{b_i}^{t-1}|\boldsymbol{y_i})/J(\boldsymbol{b_i}^{t-1}|\boldsymbol{b_i}^*)}$$

$$\log r = \log[p(\boldsymbol{b_i}^*|\boldsymbol{y_i})] - \log[p(\boldsymbol{b_i}^{t-1}|\boldsymbol{y_i})]$$

Lastly, we can directly sample D using a conjugate prior, $D^{-1} \sim \text{Wishart}_p(\nu_0, \Phi_0)$

$$
\begin{aligned}
p(D^{-1}|\nu_0, \Phi_0) &\propto |\ D^{-1}\ |^{(\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\text{tr}(\Phi_0^{-1}D^{-1})\Big) \\
p(D^{-1}|\boldsymbol{b_i}) &\propto \prod_{i=1}^{n} p(\boldsymbol{b_i}|D)p(D^{-1}|\nu_0, \Phi_0) \\
&\propto \frac{1}{(2\pi)^{n/2}\ |\ D\ |^{n/2}} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{b_i}-\boldsymbol{0})^T D^{-1}(\boldsymbol{b_i}-\boldsymbol{0})\Big) \\
&\times |\ D^{-1}\ |^{(\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\text{tr}(\Phi_0^{-1}D^{-1})\Big) \\
&\propto |\ D^{-1}\ |^{(n+\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\Big(\text{tr}(\Phi_0^{-1}D^{-1}) - \frac{1}{2}\text{tr}(D^{-1}\sum_{i=1}^{n}\boldsymbol{b_i}\boldsymbol{b_i}^T)\Big)\Big) \\
&\propto |\ D^{-1}\ |^{(n+\nu_0-p-1)/2} \exp\Big(-\frac{1}{2}\Big(\text{tr}(\Phi_0^{-1}D^{-1} + \sum_{i=1}^{n}\boldsymbol{b_i}\boldsymbol{b_i}^T D^{-1})\Big)\Big) \\
D^{-1}|\boldsymbol{b_i} &\sim \text{Wishart}_p(n+\nu_0, \Big(\Phi_0^{-1} + \sum_{i=1}^{n}\boldsymbol{b_i}\boldsymbol{b_i}^T\Big)^{-1})
\end{aligned}
$$