# Flexible Bayesian quantile regression for independent and clustered data

**Brian J. Reich**,[1] Department of Statistics,

North Carolina State University, Campus Box 8203, Raleigh, NC 27695-8203
Email: reich@stat.ncsu.edu

**Howard D. Bondell**, Department of Statistics,

North Carolina State University, Campus Box 8203, Raleigh, NC 27695-8203
Email: bondell@stat.ncsu.edu

**Huixia Wang**, Department of Statistics,

North Carolina State University, Campus Box 8203, Raleigh, NC 27695-8203
Email: wang@stat.ncsu.edu

**Abstract**

Quantile regression has emerged as a useful supplement to ordinary mean regression. Traditional frequentist quantile regression makes very minimal assumptions on the form of the error distribution, and thus is able to accommodate non-normal errors which are common in many applications. However, inference for these models is challenging, particularly for clustered or censored data. A Bayesian approach enables exact inference and is well-suited to incorporate clustered, missing, or censored data. In this paper, we propose a flexible Bayesian quantile regression model. We assume that the error distribution is an infinite mixture of Gaussian densities subject to a stochastic constraint which enables inference on the quantile of interest. This method outperforms the traditional frequentist method under a wide array of simulated data models. We extend the proposed approach to analyze clustered data. Here we differentiate between and develop conditional and marginal models for clustered data. We apply our methods to analyze a multi-patient apnea duration data set.

**Key words:** Bayesian semiparametric modeling; clustered data; quantile regression; stick-breaking prior.

[1]Corresponding author.

# 1 Introduction

Quantile regression has emerged as a useful supplement to ordinary mean regression. As might be expected, the upper or lower quantiles of the response variable may depend on the covariates very differently from the center. Therefore, quantile regression can provide a more complete description of functional changes than focusing solely on the mean. The value of "going beyond the conditional mean model" has been demonstrated in rapidly expanding literatures in econometrics, social sciences, and biomedical studies; see Koenker (2005) for a comprehensive review. In addition, quantile regression makes very minimal assumptions on the form of error distribution, and thus is able to accommodate non-normal errors, which are common in many applications. Although asymptotic theory for quantile regression is well studied, the development of convenient inference procedures has been challenging, as the asymptotic covariance matrix of quantile estimates involves the unknown error density function which can not be estimated reliably.

Another serious challenge in quantile regression lies in the analysis of clustered data. At present time, few options exist for quantile inference for such data. Jung (1996) proposed a quasi-likelihood method for median regression estimation, where the dependency structure is captured by the covariance matrix of the sign of residuals. Employing the idea of Jung's estimator, Lipsitz et al. (1997) and Yin and Cai (2005) discussed quantile regression for correlated data in different contexts, and considered resampling methods for inference by treating each cluster as a sampling unit. The cluster-resampling scheme generally performs well for large sample sizes, but it may lose control of the false positive rate for small number of clusters; see Wang and Fygenson (2008). Koenker (2004) proposed to estimate the cluster-specific fixed effects through penalizing the cluster effects. The results of this penalization

approach depend on the choice of some penalization parameter, and the practical use of inference for fixed effects was not studied. Based on estimates obtained under the working assumption of independence, Wang and He (2007) developed a quantile rank score test for clustered data by incorporating the intra-subject correlation in the test statistic. Even though it was shown to be robust to modest heteroscedasticity, the validity of the score test relies on the common error distribution assumption. Geraci and Bottai (2007) developed a parametric model with an asymmetric Laplace error distribution.

This paper provides an appealing quantile inference approach through Bayesian modeling. The Bayesian framework enables exact inference and is well-suited to incorporate clustered, missing, or censored data. From the frequentist point of view, the quantile regression problem is tackled by minimization of an objective function whose population minimizer is the desired quantile. The estimate is equivalent to the maximum likelihood solution under an asymmetric Laplace error distribution. A Bayesian approach to quantile regression must specify a likelihood, and thus a natural choice for the likelihood is the asymmetric Laplace distribution. The asymmetric Laplace distribution has been used to construct Bayesian quantile regression models for independent data (e.g., Yu and Moyeed, 2001).

Other researchers considered nonparametric approaches to avoid the restrictive parametric assumption. Walker and Mallick (1999), and Hanson and Johnson (2002) proposed median regression models using a diffuse Polya tree and a mixture of continuous Polya trees, respectively. Scaccia and Green (2003) modeled the conditional distribution of the response variable $y$ given a single covariate (time) with a discrete normal mixture with nonparametric time-dependent weights. The centile curves were then evaluated numerically from the MCMC ouput at a grid of time points. Focusing on the univariate data without covariates, Hjort (2003), and Hjort and Petrone (2007) discussed the posterior distributions of quan-

tiles, which were induced by the Dirichlet process prior distribution of the data. Taddy and Kottas (2007) focused on nonlinear regression, and they modeled the joint distribution of the response $y$ and convariates $x$ through a Dircichlet process mixture of multivariate normal distributions. Bayesian inference on the conditional quantiles of $y$ were based on draws obtained through numerical integration of the posterior conditional density of $y$, implied by the DP mixture model. Different from the above methods, in this paper, we model the conditional quantiles of $y$ directly by imposing a quantile-constrained Dirichlet process prior on the residuals. Under such formulation, Bayesian inference could be done automatically on the quantile of interest, and no numerical evaluation is required.

For independent data, Kottas and Gelfand (2001) proposed to model the error distribution by two families of median zero distributions based on mixtures with Dirichlet process priors on the mixture distributions. Recently Kottas and Krnjajic (2008) extended the idea to quantile regression for arbitrary quantiles, and their resulting nonparametric error distributions was able to capture general forms of skewness and tail behavior. However, like the asymmetric Laplace distribution, these densities necessarily have their mode at the quantile of interest, which can be a very restrictive property, particularly when modeling extreme quantiles. In addition, these constructed densities are discontinuous at the mode.

To create a fully Bayesian framework for quantile regression inference that allows for the full span of quantile-restricted error distributions, we introduce a flexible nonparametric density along with a novel sampling method to fit the model with the newly proposed error distribution. The likelihood is taken to be an infinite mixture of Gaussian densities. We make no further assumptions about the shape of the error distribution except for a stochastic constraint that permits inference on the quantile of interest. We show that our prior for the residual density spans the entire space of densities that satisfy the quantile constraint and

that our prior is stochastically-centered on the asymmetric Laplace density.

We also extend this model to analyze clustered data from a recent study of apnea duration, the period of nasal airflow cessation during swallowing. In this study, apnea duration is measured several times for each subject so within-subject correlation must be taken into account. Geraci and Bottai (2007) account for within-subject correlation by adding a random subject effect to the quantile and modeling the residual distribution with an asymmetric Laplace density. However, unlike mean regression, marginalizing over the random effects in quantile regression alters the desired quantile level and the fixed effects can no longer be interpreted in terms of the population quantile which is often the focus of the analysis. To overcome this problem, we develop a marginal quantile regression model and show that this leads to different results than the independent model.

The remainder of the paper proceeds as follows. Section 2 introduces the flexible Bayesian quantile regression model for independent data. This approach is extended to model clustered data in Section 3. Section 4 conducts a simulation study to compare the proposed method with traditional frequentist quantile regression and the Bayesian asymmetric Laplace model. The nonparametric Bayesian model improves estimates of the regression coefficients and maintains the nominal frequentist coverage probability for a wide array of error distributions. Section 5 analyzes the multi-patient study of apnea duration and Section 6 concludes. Some theoretical results and MCMC details are relegated to the Appendix.

# 2 Bayesian quantile regression for independent data

Following He (1997), we assume the heteroskedastic linear regression model

$$y_i = \mathrm{x}_i\boldsymbol{\beta} + \mathrm{x}_i\boldsymbol{\gamma}\varepsilon_i, \tag{2.1}$$

where $\mathrm{x}_i\boldsymbol{\gamma}$ is constrained to be positive for all $\mathrm{x}_i$ and the residuals $\varepsilon_i$ are independent and identically distributed. Under this model, $y_i$'s $\tau^{th}$ quantile is $\mathrm{x}_i[\boldsymbol{\beta} + \Psi_\varepsilon^{-1}(\tau)\boldsymbol{\gamma}]$, where $\Psi_\varepsilon^{-1}(\tau)$ is $\varepsilon_i$'s $\tau^{th}$ quantile, and all of $y_i$'s quantiles are in the span of $\mathrm{x}_i$. Note that this model not only covers the simple linear quantile model, but can also be used to model nonparametric quantile curves by including a set of basis functions for each covariate.

Model (2.1) may be rewritten

$$y_i = \mathrm{x}_i\boldsymbol{\beta}^{(\tau)} + \mathrm{x}_i\boldsymbol{\gamma}^{(\tau)}\varepsilon_i^{(\tau)}, \tag{2.2}$$

where $\varepsilon_i^{(\tau)} = \varepsilon_i - \Psi_\varepsilon^{-1}(\tau)$ has $\tau^{th}$ quantile equal to zero. To analyze $y_i$'s $\tau^{th}$ quantile, $\mathrm{x}_i\boldsymbol{\beta}^{(\tau)}$, we consider only distributions for the residual term with $\tau^{th}$ quantile equal to zero. For simplicity of notation, we omit the superscript $\tau$ for the remainder of the paper, although we note the dependence on the quantile of interest $\tau$. Also, we fix the element of $\boldsymbol{\gamma}$ corresponding to the intercept at one to separate out the scale of the errors from $\boldsymbol{\gamma}$.

Traditional quantile regression techniques make few assumptions about the residual distribution. Bayesian nonparametric methods avoid specifying a particular residual distribution by placing a prior on the residual distribution, e.g., the Dirichlet process mixture prior (Ferguson, 1973; Ferguson, 1974). Kottas and Gelfand (2001) and Kottas and Krnjajic (2008) apply the Dirichlet process mixture prior to quantile regression by placing separate priors

on the areas above and below zero to ensure the correct amount of mass in each region. However, this leads to a discontinuity at the mode and the mode is also forced to be at the quantile of interest, so that the distribution is not as flexible as one would hope.

We propose an alternative approach which avoids these restrictions. Our approach is to build a flexible residual distribution $h$ as an infinite mixture of simple densities $f$ that each satisfy the desired quantile constraint. As shown below, mixing these simple constrained densities yields an arbitrarily flexible residual distribution. We assume $\varepsilon_i$'s distribution is the infinite mixture

$$h(\varepsilon|\boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \sum_{k=1}^{\infty} p_k f(\varepsilon|\boldsymbol{\mu}_k, \sigma_k^2, q_k), \tag{2.3}$$

where the $p_k$ are the mixture proportions with $\sum_{k=1}^{\infty} p_k = 1$. The base density $f(\varepsilon|\boldsymbol{\mu}_k, \sigma_k^2, q_k)$ is the quantile-restricted two-component mixture

$$f(\varepsilon|\boldsymbol{\mu}_k, \sigma_k^2, q_k) = q_k\phi(\mu_{1k}, \sigma_{1k}^2) + (1 - q_k)\phi(\mu_{2k}, \sigma_{2k}^2) \tag{2.4}$$

where $\phi(\mu, \sigma^2)$ is the normal density with mean $\mu$ and variance $\sigma^2$, and $q_k \in (0, 1)$. To ensure $\int_{-\infty}^{0} f(\varepsilon|\boldsymbol{\mu}_k, \sigma_k^2, q_k)d\varepsilon = \tau$ the mixture proportion is set to

$$q_k = \frac{\tau - \Phi(-\mu_{2k}/\sigma_{2k})}{\Phi(-\mu_{1k}/\sigma_{1k}) - \Phi(-\mu_{2k}/\sigma_{2k})} \tag{2.5}$$

where $\Phi$ is the standard normal distribution function. By construction, $\int_{-\infty}^{0} f(\varepsilon|\boldsymbol{\mu}_k, \sigma_k^2, q_k)d\varepsilon = \tau$ and thus $\sum_k p_k \int_{-\infty}^{0} f(\varepsilon|\boldsymbol{\mu}_k, \sigma_k^2, q_k)d\varepsilon = \sum_k p_k\tau = \tau$, the desired quantile constraint.

Although the form of the base distribution $f$ is rather simple, the resulting prior for the residual distribution $h$ is arbitrarily flexible. In particular, draws from this prior may be multi-modal. To see this complete flexibility, note that as $\sigma_{1k}, \sigma_{2k} \to 0$ for all $k$, this

becomes an infinite mixture of point mass distributions. In this limiting case, $q_k = \tau$ and $\mu_{1k} < 0 < \mu_{2k}$ for all $k$.

Denote the mixing distribution by $dP(s,t)$. Then the residual density can be written as

$$h(\varepsilon|\boldsymbol{\mu}, 0) = \int (\tau\delta_{\mu_1}(\varepsilon) + (1-\tau)\delta_{\mu_2}(\varepsilon))\, dP(\mu_1, \mu_2), \tag{2.6}$$

with $\mu_1 < 0 < \mu_2$ and $\delta_\mu(\varepsilon)$ is the point mass distribution for $\varepsilon$ with point mass at $\mu$. Varying the bivariate mixing distribution $dP(s,t)$ in (2.6) over the class of all distributions with support on the product space $(-\infty, 0) \times (0, \infty)$ generates the class of all distributions having $\tau^{th}$ quantile zero (see Hoff, 2003). If a prior for this mixing distribution is chosen so that any distribution on the product space has nonzero prior mass, then the resulting prior will give nonzero mass to any distribution having $\tau^{th}$ quantile at zero. Although we assume $\sigma_{1k}, \sigma_{2k} > 0$, this result implies that any density can be well-approximated by a density in the span of our prior using small $\sigma_{1k}$ and $\sigma_{2k}$.

To specify a prior for this mixing distribution to span the appropriate space, we take $\mu_{1k}$ and $\mu_{2k}$ to be independent draws from the asymmetric Laplace (ASL) distribution with density

$$p(\mu|\lambda, \tau) \propto \lambda^{-1} \exp\left(-\frac{\mu}{\lambda} \cdot (\tau - I[\mu \leq 0])\right). \tag{2.7}$$

The ASL is commonly used for parametric Bayesian quantile regression since its $\tau^{th}$ quantile is zero; the double exponential distribution is a special case of the asymmetric Laplace distribution with $\tau = 0.5$. The standard deviations $\sigma_{1k}, \sigma_{2k} \sim$ Uniform(0,$c_1$) for some large constant $c_1$ (Gelman, 2006). Each mixture component must also satisfy the constraint that

$0 \leq q_k \leq 1$, which leads to the truncated prior

$$\begin{aligned} p(\mu_{1k}, \mu_{2k}, \sigma_{1k}, \sigma_{2k} | \lambda, \tau, c_1) &\propto \exp\left(-\frac{\mu_{1k}}{\lambda} \cdot (\tau - I[\mu_{1k} \leq 0]) - \frac{\mu_{2k}}{\lambda} \cdot (\tau - I[\mu_{2k} \leq 0])\right) \\ &\times I(0 \leq \sigma_{1k} \leq c_1) I(0 \leq \sigma_{2k} \leq c_1) I(0 \leq q_k \leq 1). \end{aligned} \tag{2.8}$$

To specify a prior for $p_k$, we use the stick-breaking representation (see, e.g., Ishwaren and James (2001) and references therein) to model the mixture proportions. Specifically, the proportions are defined through the latent variables $V_k \overset{iid}{\sim} \text{Beta}(1, D)$. The first is $p_1 = V_1$. Successive proportions are given by

$$p_k = V_k(1 - \sum_{j<k} p_j) = V_k \prod_{j<k} (1 - V_j) \tag{2.9}$$

where $1 - \sum_{j<k} p_j = \prod_{j<k}(1 - V_j)$ is the mass not accounted for by the first $k-1$ components and $V_k$ is the proportion of the remaining mass attributed to the $k^{th}$ component. By construction, $\sum_{k=1}^{\infty} p_k = 1$ almost surely.

With $\sigma_{1k} = \sigma_{2k} = 0$ for all $k$, the prior for the residual distribution $h$ is centered on the asymmetric Laplace distribution, while $D$ controls the strength of the prior. To see this, let $F$ be the distribution function corresponding to (2.4). Then Appendix A.1 shows that when $\sigma_{1k} = \sigma_{2k} = 0$,

$$\begin{aligned} E(F(u)) &= F_{ASL}(u|\lambda) \\ V(F(u)) &= F_{ASL}(u|\lambda)\left[1 - F_{ASL}(u|\lambda)\right]/(D+1)^2 \end{aligned} \tag{2.10}$$

where $u \in \mathcal{R}$, expectations are taken with respect to $\{V_k, \mu_{1k}, \mu_{2k}\}$, and $F_{ASL}(u|\lambda)$ is the asymmetric Laplace distribution function. Although the residual density is centered on a

unimodal density with mode at zero, this does not require draws from the model to have a single mode at zero. In practice, we do not take $\sigma_{1k} = \sigma_{2k} = 0$ so the mean and variance in (2.10) are only approximate. Hanson et al. (2005) discuss the discrepancy between the centering distributions of the Dirichlet process versus mixture Dirichlet process approaches, and recommend the choice of centering distribution be based on the limiting discrete Dirichlet process.

For computational and conceptual purposes, this model can also be written as a mixture model. We introduce latent variables $G_i \in \{1, 2, ...\}$ and $H_i \in \{1, 2\}$ to indicate the mixture component from which the $i^{th}$ observation is drawn, and write the model as

$$\begin{aligned} y_i &= \mathrm{x}_i\boldsymbol{\beta} + \mathrm{x}_i\boldsymbol{\gamma}e_i \text{ where } e_i \sim \mathrm{N}(\mu_{H_iG_i}, \sigma^2_{H_iG_i}) \\ G_i &\sim \text{Categorical}(p_1, p_2, ...) \text{ and } H_i \sim \text{Categorical}(q_{G_i}, 1 - q_{G_i}) \end{aligned} \tag{2.11}$$

and $q_k$, $\mathrm{p}(\mu_{1k}, \mu_{1k}, \sigma_{1k}, \sigma_{2k})$, and $p_k$ are given in (2.5), (2.8), and (2.9), respectively. The regression coefficients and scale parameters are given diffuse priors, $\beta_j \sim \mathrm{N}(0,\, c_2)$ for large $c_2$ and $\lambda \sim$ Gamma(0.1,0.1). The regression scale parameters $\gamma_j$ also have vague normal priors, subject to $\mathrm{x}_i\boldsymbol{\gamma} > 0$ for all $\mathrm{x}_i$ and we restrict the first element of $\boldsymbol{\gamma}$ corresponding to the intercept to be one to identify the scale of the residuals. Appendix A.2 describes an algorithm to analyze this infinite mixture model using retrospective MCMC sampling. Alternatively, this model could be truncated to be an $M$ component mixture by fixing $V_M = 1$. It would be straight-forward to fit this finite mixture model in freely available WinBUGS software. R code is available from the first author on request.

# 3 Bayesian quantile regression for clustered data

In this section we describe our quantile regression models for clustered data, such as the swallowing data described in Section 1. We consider two models: Section 3.1 introduces random subject effects to account for within-subject correlation, Section 3.2 integrates over the random subject effects and models the quantile of the marginal distribution.

## 3.1 Conditional model

One approach to quantile regression for clustered data is to add a random subject effect $\alpha_s$ to the model in (2.2), that is,

$$y_{is} = x_{is}\boldsymbol{\beta} + \alpha_s + x_{is}\boldsymbol{\gamma}\varepsilon_{is}, \tag{3.12}$$

where $y_{is}$ is the $i^{th}$ measurement for subject $s$ and $\varepsilon_{is}$ follows the quantile-restricted mixture distribution in (2.3). We refer to this as the conditional model, since conditional on $\alpha_s$, the $\tau^{th}$ quantile of $y_{is}$ is $\mathrm{x}_{is}\boldsymbol{\beta} + \alpha_s$. Conceivably any distribution could be used for the random effects. Regardless of the random effects' distribution, observations from different subjects are independent, while observations from the same subject are dependent with

$$\mathrm{cor}(y_{is}, y_{i's}) = \frac{\sigma_\alpha^2}{\sqrt{\left[(x_{is}\boldsymbol{\gamma})^2\sigma_\varepsilon^2 + \sigma_\alpha^2\right]\left[(x_{i's}\boldsymbol{\gamma})^2\sigma_\varepsilon^2 + \sigma_\alpha^2\right]}}, \tag{3.13}$$

where $\sigma_\alpha^2$=Var($\alpha_s$) and

$$\sigma_\varepsilon^2 = \left[\sum_{k=1}^{\infty} q_k p_k(\sigma_{1k}^2 + \mu_{1k}^2) + (1-q_k)p_k(\sigma_{2k}^2 + \mu_k^2)\right] - \left[\sum_{k=1}^{\infty} q_k p_k \mu_{1k} + (1-q_k)p_k\mu_{2k}\right]^2.$$

## 3.2 Marginal model

After integrating over the random effects, Section 3.1's conditional model may no longer have the desired quantile relationship. For example, assuming the random effects are Gaussian with mean zero and $x_{is}\boldsymbol{\gamma} = 1$ for all $x_{is}$, integrating over $\alpha_s$ gives

$$P(y_{is} < \mathrm{x}_{is}\boldsymbol{\beta}) = \sum_{k=1}^{\infty} p_k \int_{-\infty}^{0} q_k \phi(r|\mu_{1k}, \sigma_\alpha^2 + \sigma_{1k}^2) + (1-q_k)\phi(r|\mu_{2k}, \sigma_\alpha^2 + \sigma_{2k}^2)dr. \tag{3.14}$$

If $\sigma_\alpha^2$ is large, $\int_{-\infty}^{0} \phi(r|\mu_{jk}, \sigma_\alpha^2 + \sigma_{jk}^2)dr \approx 0.5$ for all $j$ and $k$ and $P(y_{is} < \mathrm{x}_{is}\boldsymbol{\beta}) \approx 0.5$, which is inappropriate for an analysis of extreme quantiles. Therefore, the conditional model is appropriate if the focus of the study is to estimate each cluster's quantile. However, unlike mean regression, $\boldsymbol{\beta}$ should not be interpreted in terms of the population's $\tau^{th}$ quantile.

When the focus of the study is to estimate the population's quantile, we must specify a model that both accounts for within-subject correlation as well as guarantees that the $\tau^{th}$ marginal quantile of $y_{is}$ is $\mathrm{x}_{is}\boldsymbol{\beta}$. To do so, consider the random effect model

$$y_{is} = \mathrm{x}_{is}\boldsymbol{\beta} + x_{is}\boldsymbol{\gamma}(\alpha_s + \varepsilon_{is}), \tag{3.15}$$

where the $\tau^{th}$ quantile of $\alpha_s + \varepsilon_{is}$ is zero. For simplicity, we assume a conjugate prior for the random effects, i.e., $\alpha_s \overset{iid}{\sim} N(0, \sigma_\alpha^2)$. The residuals $\varepsilon_{is}$ are modeled as an infinite mixture of normals as before, except that we now must account for the random effect variability in the quantile restriction. We modify (2.5) as

$$q_k = \frac{\tau - \Phi\left(-\frac{\mu_{2k}}{\sqrt{\sigma_{2k}^2+\sigma_\alpha^2}}\right)}{\Phi\left(-\frac{\mu_{1k}}{\sqrt{\sigma_{1k}^2+\sigma_\alpha^2}}\right) - \Phi\left(-\frac{\mu_{2k}}{\sqrt{\sigma_{2k}^2+\sigma_\alpha^2}}\right)}. \tag{3.16}$$

Under this model, observations from the same subject have correlation

$$\text{cor}(y_{is}, y_{i's}) = \frac{\sigma_\alpha^2}{(x_{i's}\boldsymbol{\gamma})(x_{is}\boldsymbol{\gamma})(\sigma_\varepsilon^2 + \sigma_\alpha^2)}, \tag{3.17}$$

where $\sigma_\alpha^2$ and $\sigma_\varepsilon^2$ are defined in (3.13).

To show this model gives the desired quantile relationship, let $r_{is} = (y_{is} - \mathrm{x}_{is}\boldsymbol{\beta})/\mathrm{x}_{is}\boldsymbol{\gamma}$. Using the mixture representation in (2.11) and integrating over $\alpha_s$, we get $r_{is} \sim N(\mu_{H_{is}G_{is}}, \sigma^2_{H_{is}G_{is}} + \sigma_\alpha^2)$, where $H_{is}$ and $G_{is}$ indicate the observation's residual mixture component. Marginalizing over $H_{is}$ and $G_{is}$ gives

$$p(r_{is} \leq 0) = \sum_{k=1}^{\infty} p_k \left[ q_k \Phi(-\frac{\mu_{1k}}{\sigma_{1k} + \sigma_\alpha}) + (1 - q_k)\Phi(-\frac{\mu_{2k}}{\sigma_{2k} + \sigma_\alpha}) \right] = \sum_{k=1}^{\infty} p_k \tau = \tau$$

by the definition of $q_k$.

# 4 Simulation Study

We conduct a simulation study to assess the performance of the proposed semiparametric Bayesian quantile regression approach for independent data for both homosckedastic and heterosckedastic models.

## 4.1 Design

Our simulation design follows Kocherginsky, He, and Mu (2005). We generate data from five model designs with uncorrelated errors:

- **Design 1** : $y_i = 1 + x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_{1i}$
- **Design 2** : $y_i = 1 + x_{1i}\beta_1 + x_{2i}\beta_2 + \pi_i\varepsilon_{1i} + (1 - \pi_i)(\varepsilon_{2i})$

- **Design 3** : $y_i = 1 + x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_{3i}$
- **Design 4** : $y_i = 1 + x_{3i}\beta_1 + (1.1 - x_{3i})\varepsilon_{1i}$
- **Design 5** : $y_i = 1 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{4i}\beta_3 + \varepsilon_{1i}$

where $x_{1i}$, $x_{2i} \overset{iid}{\sim}$ N(0,1), $\varepsilon_{1i} \sim$ N(0,1), $\varepsilon_{2i} \sim$ N(3,3), $\varepsilon_{3i} \overset{iid}{\sim}$ DExp(0,1), $x_{3i} \overset{iid}{\sim}$ Unif(-1,1), $\pi_i \overset{iid}{\sim}$ Bern(0.8), and $x_{4i} \overset{iid}{\sim} |t_2|$. All covariates and error terms are mutually independent. We set all slopes to one, i.e., $\beta_k = 1$, $k = 1, 2, 3$. Models 1, 2, and 3 are simple location shift models with different error distributions. Model 4 has heteroskedastic errors. In Model 5, the predictor $x_4$ has heavy tails which is troublesome for Frequentist asymptotics. For each model we generate 200 data sets assuming the sample size is $n = 100$. We analyze both $\tau = 0.5$ and $\tau = 0.9$.

Each simulated data set is analyzed using three methods. We use our flexible Bayesian quantile regression model ("FBQR") proposed in Section 2. For our FBQR model we choose $D = 1$ in the stick-breaking prior so that the prior on the weights is uniform. We also consider the parametric Bayesian model assuming the errors follow an asymmetric Laplace ("ASL") distribution with $\boldsymbol{\beta} \sim N(0, 100 \cdot I_p)$ and $\sigma \sim U(0, 10)$. We compare these methods with the standard frequentist quantile regression approach ("QReg") using the "quantreg" package in R (R Development Core Team, 2006) using the default "rank" method to obtain confidence intervals.

Methods are evaluated based on mean squared error

$$MSE = \frac{1}{p}\sum_{j=1}^{p}\left(\beta_j - \hat{\beta}_j\right)^2, \tag{4.18}$$

where $p$ is the number of covariates (excluding the intercept), $\beta_j$ is the true value, and $\hat{\beta}_j$ is the estimate (we use the posterior median for Bayesian methods). MSE is computed for

each data set and each method, and we report the mean (standard error) of the 200 MSEs for each method. We also report the coverage probabilities for the 90% intervals for each procedure.

## 4.2 Results

Table 1 gives the results of the simulation study. In all cases other than Design 3 with $\tau = 0.50$, our FBQR procedure gives smaller $MSE$ than its competitors. In these cases the parametric form of the error distribution for the Bayesian ASL model is incorrect which permeates into the regression coefficients' estimates. The difference in $MSE$ is particularly dramatic in the analyses of extreme quantiles. The data provide less information about extreme quantiles so the parametric form of the residual distribution has a greater effect.

As expected, the ASL procedure has the lowest $MSE$ for Design 3 with $\tau = 0.50$. In this case the ASL model has the smallest $MSE$ because the data are generated from this model. The QReg procedure also performs well in this case as the estimated coefficient vector is actually the posterior mode from the Bayesian ASL model. Interestingly, the posterior median from the ASL model gives considerably smaller $MSE$ than the QReg procedure for this (and all other) design(s). In contrast, the FBQR procedure has the smallest $MSE$ of all the procedures for Design 3 with $\tau = 0.90$. In this case, the mode of the double exponential error distribution is in the center, but the ASL model assumes the mode is in the right tail. This causes the ASL's MSE to increase and the coverage probability of the 90% intervals to dip to 0.63.

The two procedures (QReg and our proposed FBQR) that do not assume a parametric form for the residual distribution have coverage probabilities near the nominal 90% level for all simulations. The parameteric Bayesian ASL procedure's coverage probability sinks to as

low as 56%. While both the FBQR and QReg procedures have correct coverage probabilities, the QReg confidence intervals are generally much wider than our FBQR intervals. For example, the average interval width for the QReg procedure for the Design 2 with $\tau = 0.9$ is 1.97, compared to an average width of 1.02 for the FBQR procedure.

We also use the simulated data to study sensitivity to the hyperparameter, $D$, that controls the strength of stick-breaking prior. Figure 1 plots the posterior mean density for 4 simulated data sets. For these data, $D = 0.1$, $D = 1$, and $D \sim$ Gamma(1,1) give similar density estimates. Especially for small sample sizes, $D = 10$ gives slightly different density estimates. However, Table 2 shows that the posterior mean and standard deviation of the regression coefficients $\beta$, which are the primary focus, are robust to changes in $D$.

# 5 Analysis of the swallowing data

In this section we use quantile regression methods to analyze the swallowing data previously analyzed in Perlman et al. (2005) and Zhou and He (2008). The response, apnea duration, is the period of nasal airflow cessation during swallowing. Longer apnea durations are related to longer cycle-time breathing patterns, and often indicate age-related function changes for seniors. There are 23 elderly women in the study. The apnea duration of each subject was measured multiple times while they were swallowing either water or pudding with different volumes and under different feeding conditions, giving 1286 total observations. The purpose of the study is to determine how apnea duration is affected by three covariates: feeding condition (self-fed or examiner-fed), viscosity (liquid or pudding), and volume (5ml or 10ml). The covariates vary within subject and we assume a two-way interaction model.

Apnea duration is highly right-skewed so we perform the analysis on the log scale. The

data are shown graphically in Figure 2. The heteroscedastic FBQR model seems to be appropriate here, as the data are non-Gaussian even after the log transformation and the covariate effects appear to be different in the right tail than in the center, especially for viscosity.

We compare three models: the frequentist quantile regression model, our FBQR model assuming the data are independent, and the FBQR model assuming marginal subject random effects. We begin our analysis of the swallowing data by temporarily ignoring the within-subject correlation and using Section 2's model for independent data, referred to as "IID". At $\tau = 0.50$, the main effects for feeding condition and volume and the interaction between viscosity and volume are moderately strong predictors of swallowing time (Table 3). The results are quite different for the analysis of the upper tail. At $\tau = 0.90$, the main effect of feeding condition is far more pronounced than with $\tau = 0.5$ and the interaction between feeding condition and viscosity emerges as a strong predictor of swallowing time. These differential effects are also apparent in Figure 2; the distributions have the heavier right tails when the viscosity is liquid and the examiner controls the feeding.

The Bayesian 95% intervals are narrower and closer to zero than the frequentist intervals (Table 3), especially for $\tau = 0.9$. For example, at $\tau = 0.9$, the frequentist interval for the viscosity main effect is (-0.56, -0.06) compared to the Bayesian interval (-0.27, -0.01). The Bayesian density estimates in Figure 3 are also smoothed. Although the density estimates fit the data well, the estimated curves have slightly less mass in the right tail than the residual histograms; this is the effect of Bayesian smoothing. Section 4's simulation study, especially Design 2's right-skewed data, shows that this Bayesian smoothing results in more stable estimates of the regression coefficients (Table 1).

Figure 4 presents a sensitivity analysis. Our model with $D = 1$, $c_1 = 1$ (roughly twice

the sample standard deviation of the response), and $c_2 = 100$ was refit five times, each time modifying one of the hyperparameters (given in the legend of Figure 4). With this large sample size of more than 1,000 observations, both the density estimate and posterior of the regression coefficient are robust to the choice of hyperparameters.

Ignoring the within-subject correlation appears to be inappropriate for these data. The 95% interval for the within-subject correlation under Section 3.2's marginal model is (0.06, 0.26) with $\tau = 0.5$ and (0.06, 0.28) with $\tau = 0.9$. Adding subject random effects influences several of the regression coefficients. The marginal model identifies an additional significant predictor, viscosity for $\tau = 0.5$. Also, the 95% interval for the viscosity by feeding condition interaction not longer excludes zero under the marginal model.

For clustered data, within-subject correlation can reduce power for testing between-subject effects (e.g., the intercept, or the treatment effect in an experiment where a subject is either assigned to a treatment or a control but not both) because the correlation reduces the effective sample size for each subject. However, intra-subject correlation can improve power for testing within-subject effects (e.g., feeding condition, volume, and viscosity in this example) by reducing within-subject variability and isolating the covariate effects. This is evident in the width of the intervals for the swallowing data; the intervals for the intercept are larger for the marginal model than the independence model, but the majority of the intervals for the other effects are smaller for the marginal model than the independence model. See Wang and Fygenson (2008) for a similar discussion in a different setup.

Finally, we conduct a posterior predictive model check to demonstrate that the marginal random effects model adequately fits the data. 10% of the observations are randomly selected to be removed from the data set. For each withheld observation we compute the posterior predictive median and 95% interval. These predictive quantiles are plotted against the

withheld data in Figure 5. The shape of the predictive densities match the right-skewness of the withheld data and coverage probabilities of the 95% intervals is 97%. It appears our model is well-calibrated.

# 6 Discussion

This paper presents a new Bayesian quantile regression model. The residual distribution is modeled using a stick-breaking construction equipped with stochastic constraints to ensure that the desired quantile relationship is satisfied almost surely. The simulation study demonstrates that this flexible approach improves estimation compared to the usual frequentist procedure when the true residual distribution is non-Laplacian. The Bayesian quantile regression model is extended to model clustered data. We differentiate between and develop conditional and marginal models and illustrate that accounting for within-subject correlation in the swallowing data affects the posterior of the regression coefficients.

Section 3's models for correlated data assume that the random effects follow independent normal distributions. Although exploratory analysis suggests that this is adequate for the swallowing data, it is possible to replace the normal prior with any parametric distribution or even a nonparametric model with for example a Dirichlet process prior for the random effects' distribution. Also, it would be straight-forward to accommodate spatial or temporal correlation structures. For example, spatial correlation could be introduced by modeling the vector of random effects with a Gaussian spatial prior or a nonparameteric spatial prior (Gelfand, Kottas, and MacEachern, 2005; Griffin and Steel, 2006; Reich and Fuentes, 2007).

# References


Ferguson TS (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.

Ferguson TS (1974). Prior distribution on spaces of probability measures. *The Annals of Statistics*, **2**, 615–629.

Gelfand AE, Kottas A, MacEachern SN (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, **100**, 1021–1035.

Gelman A (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, **1**, 515-534.

Geraci M, Bottai M (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, 140-154.

Griffin JE, Steel MFJ (2006) Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, **101**, 179–194.

Hanson T, Johnson WO (2002). Modeling Regression Error With a Mixture of Polya Trees. *Journal of the American Statistical Association*, **97**, 1020–1033.

Hanson T, Sethuraman J, Xu L (2005). On choosing the centering distribution in Dirichlet process mixture models. *Statistics and Probability Letters*, **72**, 153–162.

He X (1997). Quantile curves without crossing. *The American Statistician*, **51**, 186–192.

Hjort NL (2003). Topics in non-parametric Bayesian statistics. In *Highly structured Stochastic Systems*, edited by Green, Hjort and Richardson.

Hjort NL, Petrone S (2007). Nonparametric quantile inference using Dirichlet processes. In *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A. Doksum*, Edited by V Nair.

Ishwaran H, James LF (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, **96**, 161–173.

Jung S (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, **91**, 251–257.

Kocherginsky M, He X, Mu Y (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, **14**, 41–55.

Koenker R (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74-89.

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

Kottas A, Gelfand AE (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, **96**, 1458–1468.

Kottas A, Krnjajić M (2008). Bayesian nonparametric modeling in quantile regression. To appear, *Scandinavian Journal of Statistics*.

Lipsitz SR, Fitzmaurice GM, Molenberghs G, Zhao LP (1997). Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society, Series C*, **46**, 463-76.

Papaspiliopoulos O, Roberts G (2008). Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika*, to appear.

Perlman AL, He X, Barkmeier J, Van Leer E (2005). Bolus location associated with videofluoroscopic and respirodeglutometric events. *Journal of Speech, Language, and Hearing Research*, **48**, 21–33.

R Development Core Team (2006). R: A Language and Environment for Statistical Computing. `http://www.R-project.org`.

Reich BJ, Fuentes M (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, **1**, 249–264.

Scaccia L, Green PJ (2003). Bayesian Growth Curves Using Normal Mixtures With Nonparametric Weights. *Journal of Computational & Graphical Statistics*, **12** 208–331.

Taddy M, Kottas A (2007). A Nonparametric Model-based Approach to Inference for Quantile Regression. *Technical report ams2007-21*, UCSC Department of Applied Math and Statistics.

Walker S, Mallick BK (1999). A Bayesian Semiparametric Accelerated Failure Time Model. *Biometrics*, **55**, 477–483.

Wang H, Fygenson M (2008). Inference for censored quantille regression models in longitudinal studies. *Annals of Statistics*, to appear.

Wang H, He X (2007). Detecting differential expressions in GeneChip microarray studies: a quantile approach. *Journal of American Statistical Association*, **102**, 104–112.

Yin GS, Cai J (2005). Quantile regression models with multivariate failure time data. *Biometrics*, **61**, 151–161.

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

Zhou T, He X (2008). Three-step estimation in linear mixed models with skew-$t$ distributions. *Journal of Statistical Planning and Inference*, **138**, 1542–1555.

# Appendix A.1 – Proof of the centering distribution

If $\sigma_{1k} = \sigma_{2k} = 0$, the constraint that $0 \leq q_k \leq 1$ implies that $\mu_{1k}$ and $\mu_{2k}$ have different signs. Without loss of generality, assume $\mu_{1k} < 0 < \mu_{2k}$. Then (2.5) implies $q_k = \tau$. Let $F_1(u) = P(\mu_{1k} < u | \mu_{1k} < 0)$ and $F_2(u) = P(\mu_{2k} < u | \mu_{2k} > 0)$. Then

$$\begin{aligned} E(F(u)) &= E_v E_\mu \left( \sum_{k=1}^{\infty} p_k (\tau I(\mu_1 < u) + (1-\tau) I(\mu_2 < u)) \right) \\ &= E_v \left( \sum_{k=1}^{\infty} p_k (\tau F_1(u) + (1-\tau) F_2(u)) \right) \end{aligned}$$

$$
\begin{aligned}
&= (\tau F_1(u) + (1-\tau)F_2(u)) \left( E_v \sum_{k=1}^{\infty} p_k \right) \\
&= \tau F_1(u) + (1-\tau)F_2(u)
\end{aligned}
$$

where $E_v$ and $E_\mu$ are expectations with respect to $\{V_k\}$ and $\{\mu_{1k}, \mu_{2k}\}$, respectively. Therefore, $E(F(u)) = F_{ASL}(u|\lambda)$. To see this, let $\mu \sim \text{ASL}(\lambda, \tau)$. Then

$$
\begin{aligned}
P(\mu < u) &= P(\mu < 0)P(\mu < u|\mu < 0) + P(\mu > 0)P(\mu < u|\mu > 0) \\
&= \tau F_1(u) + (1-\tau)F_2(u)
\end{aligned}
$$

The calculation of $V(F(u))$ proceeds similarly.

# Appendix A.2 – MCMC details

MCMC sampling is carried out in R (R Development Core Team, 2006). We generate 25,000 samples and discarded the first 5,000 as burn-in. We describe below the algorithm for the independent data model in Section 2. The regression parameters $\boldsymbol{\beta}$, $D$ and $\lambda$ have conjugate priors and are updated using Gibbs sampling. The full conditional for $\boldsymbol{\beta}$ is multivariate normal with mean $(X'WX)^{-1}X'D\text{r}$ and covariance $(X'WX)^{-1}$ where $W$ is diagonal with diagonal elements $1/(\text{x}_i\boldsymbol{\gamma}\sigma_{G_i})^2$ and $r_i = y_i - \text{x}_i\boldsymbol{\gamma}\mu_{G_i}$. The remaining parameters are updated using Metropolis-Hastings sampling. Given $N = \max\{G_1, ..., G_n\}$, we only need to update $(\mu_{1k}, \mu_{2k}, V_k, \sigma_{1k}, \sigma_{2k})$ for $k = 1, ..., N$. The remaining terms do not enter the posterior except through their priors. $\mu_{1k}$, $\mu_{2k}$, $V_k$, $\sigma_{1k}$, and $\sigma_{2k}$ are updated individually using Gaussian candidate distributions. Candidates with zero probability are simply rejected. The standard deviation parameters are also updated with Gaussian candidates. Candidates with $\text{x}_i\boldsymbol{\gamma} < 0$

for any $i$ are rejected.

The group indicators $G_i$ are also updated using Metropolis-Hastings sampling. Candidate $G_i$ are generated from the prior $G_i \sim \text{Categorical}(p_1, p_2, ...)$. Following Papaspiliopoulos and Roberts (2008), we generate the candidate by first drawing $w \sim \text{Uniform}(0,1)$. If $w < \sum_{l=1}^{N} p_l$, we take $\min\{G | w < \sum_{l=1}^{G} p_l\}$ as the candidate. If $w \geq \sum_{k=1}^{N} p_k$, we increase $N$, drawing the corresponding $\mu_{1k}$,$\mu_{2k}$, $V_N$, $\sigma_N$ from their priors, until $w < \sum_{l=1}^{N} p_l$ and use the new $N$ as the candidate for $G_i$.

Section 3's models for clustered data add subject random effects, which are updated using Gibbs sampling. We find convergence is dramatically improved by updating the fixed effects and random effects jointly. Given the random effects, the MCMC algorithm above can be used for the remaining parameters.

Table 1: Mean squared error and coverage probabilities of 90% intervals for the simulation study for independent data. Mean squared error is reported as 100 × average (100 × standard error) over the 200 simulated datasets for each simulation setting.

| | | Mean squared error | | | Coverage Probability | | |
|---|---|---|---|---|---|---|---|
| Design | $\tau$ | FBQR | ASL | QReg | FBQR | ASL | QReg |
| 1 | | 1.11 (0.09) | 1.31 (0.08) | 1.58 (0.10) | 0.90 | 0.85 | 0.90 |
| 2 | | 2.10 (0.17) | 2.29 (0.15) | 2.68 (0.17) | 0.88 | 0.89 | 0.88 |
| 3 | 0.5 | 1.37 (0.11) | 1.27 (0.10) | 1.36 (0.12) | 0.92 | 0.91 | 0.90 |
| 4 | | 2.34 (0.39) | 4.18 (0.51) | 4.42 (0.53) | 0.92 | 0.84 | 0.84 |
| 5 | | 1.07 (0.07) | 1.20 (0.07) | 1.33 (0.07) | 0.88 | 0.83 | 0.88 |
| 1 | | 2.38 (0.14) | 2.46 (0.16) | 2.88 (0.19) | 0.92 | 0.73 | 0.89 |
| 2 | | 13.47 (1.07) | 41.35 (2.40) | 48.10 (2.72) | 0.89 | 0.56 | 0.87 |
| 3 | 0.9 | 4.51 (0.34) | 7.27 (0.43) | 8.76 (0.51) | 0.91 | 0.63 | 0.88 |
| 4 | | 3.74 (0.50) | 7.15 (0.70) | 8.49 (0.88) | 0.93 | 0.68 | 0.83 |
| 5 | | 1.86 (0.16) | 1.92 (0.10) | 2.38 (0.13) | 0.91 | 0.73 | 0.89 |

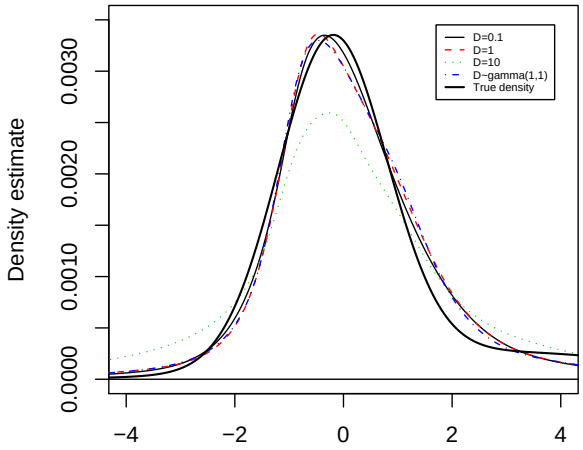
Table 2: Posterior mean (sd) of the regression coefficients for four simulated data sets under different priors/values for $D$.

| | | Simulation Design 2 | | Simulation Design 3 | |
|---|---|---|---|---|---|
| Prior | $n$ | $\beta_1$ | $\beta_2$ | $\beta_1$ | $\beta_2$ |
| $D = 0.1$ | 100 | 1.14 (0.16) | 1.00 (0.14) | 1.02 (0.11) | 0.99 (0.14) |
| $D = 1$ | 100 | 1.13 (0.17) | 1.01 (0.14) | 1.01 (0.10) | 0.99 (0.14) |
| $D = 10$ | 100 | 1.11 (0.17) | 1.00 (0.15) | 1.02 (0.10) | 1.01 (0.13) |
| $D \sim$ G(1,1) | 100 | 1.12 (0.17) | 0.99 (0.14) | 1.03 (0.11) | 1.01 (0.14) |
| $D = 0.1$ | 1000 | 0.94 (0.04) | 0.98 (0.04) | 0.99 (0.04) | 0.99 (0.04) |
| $D = 1$ | 1000 | 0.94 (0.04) | 0.99 (0.04) | 1.00 (0.03) | 1.00 (0.04) |
| $D = 10$ | 1000 | 0.94 (0.04) | 0.99 (0.04) | 1.00 (0.03) | 0.98 (0.04) |
| $D \sim$ G(1,1) | 1000 | 0.94 (0.04) | 0.99 (0.04) | 1.00 (0.03) | 0.99 (0.04) |

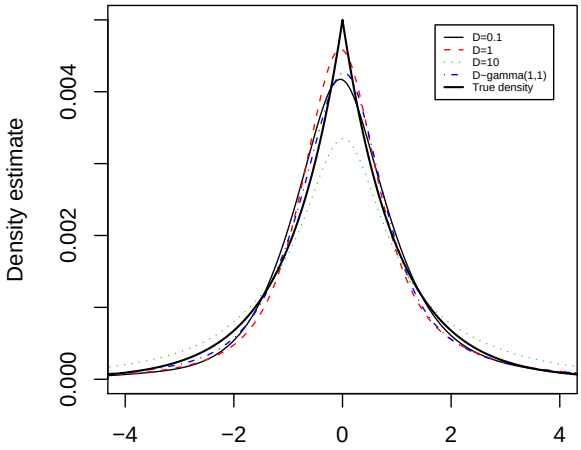
Table 3: 95% intervals for the swallowing data. The frequentist quantile regression procedure assuming independence ("QReg") is compared with three semiparametric Bayesian models: the model assuming independent errors ("IID"), and the marginal random effects model ("Marginal RE").

| | $\tau$ | QReg | IID | Marginal RE |
|---|---|---|---|---|
| Intercept | 0.5 | (-0.14,-0.01) | (-0.11,-0.03) | (-0.11, 0.04) |
| Viscosity (Visc) | 0.5 | (-0.11, 0.05) | (-0.09, 0.01) | (-0.12,-0.03) |
| Feed Cond (FC) | 0.5 | (-0.01, 0.15) | ( 0.01, 0.12) | ( 0.00, 0.09) |
| Volume (Vol) | 0.5 | ( 0.03, 0.21) | ( 0.00, 0.10) | ( 0.02, 0.10) |
| Visc × FC | 0.5 | (-0.14, 0.04) | (-0.11, 0.01) | (-0.07, 0.03) |
| Visc × Vol | 0.5 | (-0.23,-0.05) | (-0.14,-0.02) | (-0.10,-0.01) |
| FC × Vol | 0.5 | (-0.12, 0.06) | (-0.05, 0.07) | (-0.06, 0.03) |
| Intercept | 0.9 | ( 0.61, 0.97) | ( 0.72, 0.92) | ( 0.65, 0.98) |
| Viscosity | 0.9 | (-0.56,-0.06) | (-0.27,-0.01) | (-0.35,-0.11) |
| Feed Cond | 0.9 | ( 0.37, 0.79) | ( 0.10, 0.35) | ( 0.10, 0.37) |
| Volume | 0.9 | (-0.05, 0.35) | (-0.08, 0.19) | (-0.08, 0.28) |
| Visc × FC | 0.9 | (-0.83,-0.34) | (-0.35, -0.03) | (-0.28, 0.05) |
| Visc × Vol | 0.9 | (-0.25, 0.23) | (-0.23, 0.08) | (-0.25, 0.10) |
| FC × Vol | 0.9 | (-0.29, 0.19) | (-0.14, 0.17) | (-0.24, 0.07) |

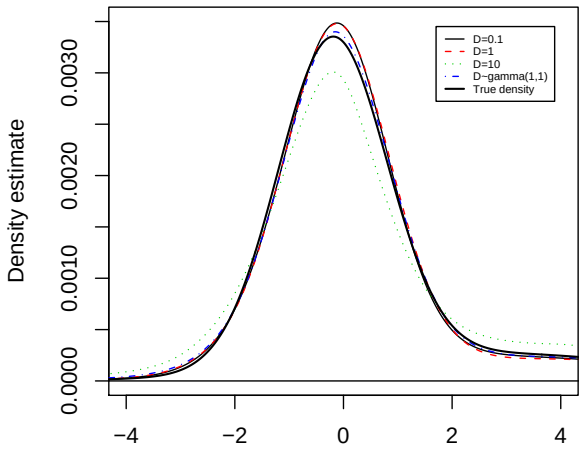
Figure 1: Posterior means residual density (solid lines) for four simulated data sets.



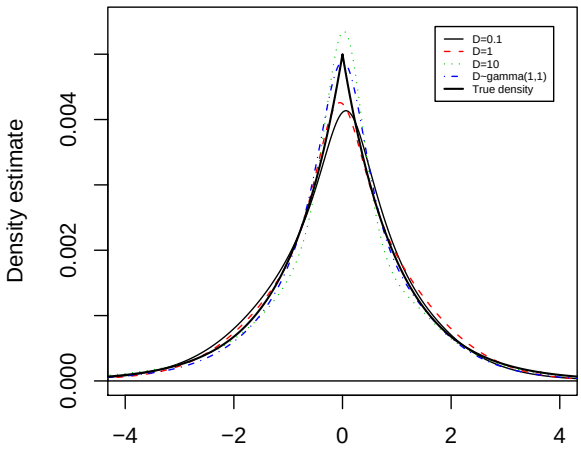

(a) Simulation 2, n=100




(b) Simulation 3, n=100




(c) Simulation 2, n=1000




(d) Simulation 3, n=1000

Figure 2: Boxplots of the log duration (seconds) for each group. Wide boxes represent subjects being fed by the examiners, narrow boxes represent self-feeding. Shaded boxes are pudding, white boxes are liquid. The horizontal lines give the sample quantiles for $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$.

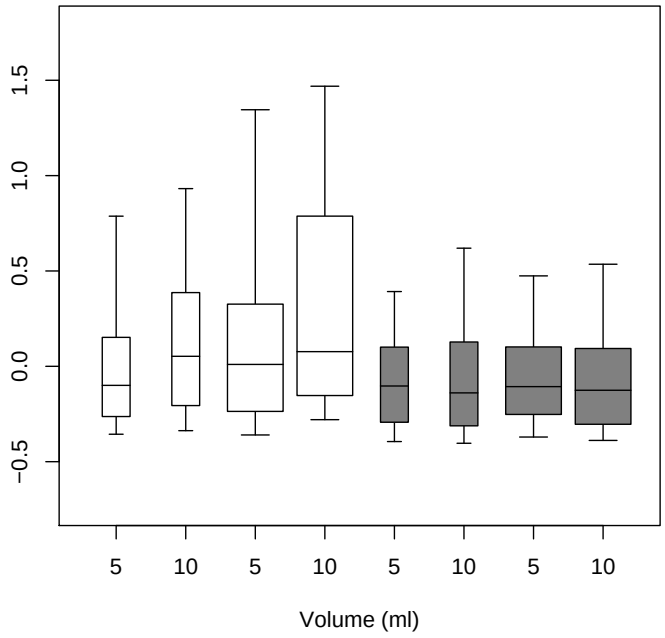
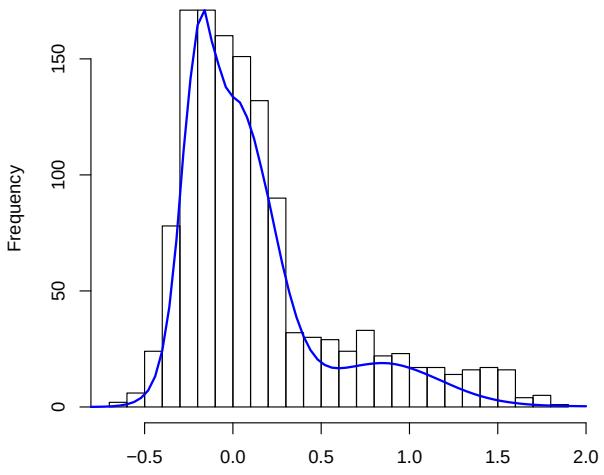
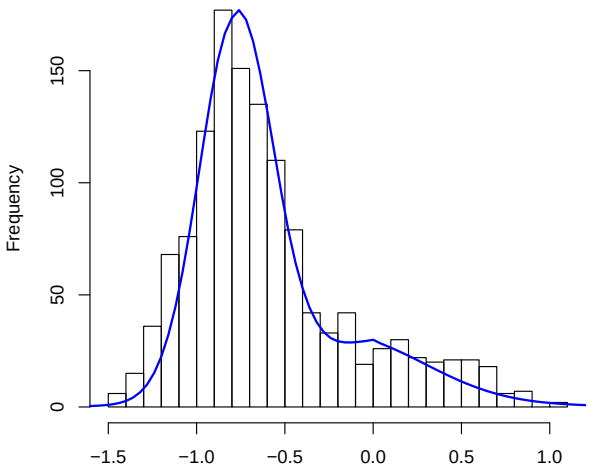



Figure 3: Standardized residuals (bars; i.e., $r_{is} = (y_{is} - x_{is}\hat{\boldsymbol{\beta}})/x_{is}\hat{\boldsymbol{\gamma}}$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ are posterior means) and posterior means of the residual density (solid lines) for the swallowing data assuming independent observations.

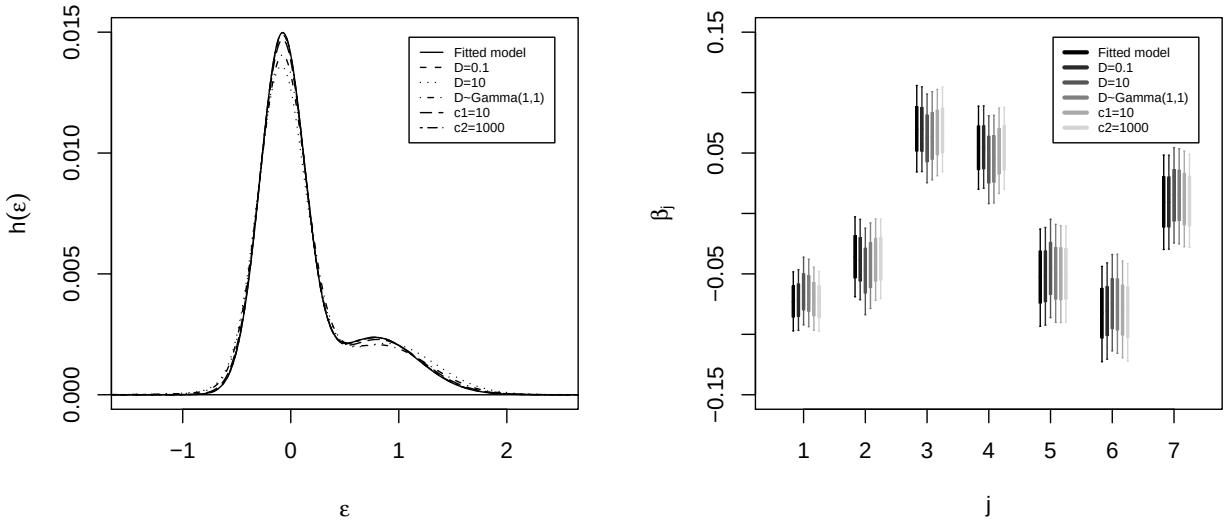



(a) $\tau$=0.5




(b) $\tau$=0.9

Figure 4: Posterior mean of the residual density $h(\varepsilon)$ and the posterior distribution of the regression coefficients $\beta_j$ for different hyperprior combinations.




(a) Posterior mean of $h(\varepsilon)$

(b) Posterior of the $\beta_j$

Figure 5: Posterior predictive model check of the marginal random effects model for the swallowing data with $\tau = 0.9$. The points are the withheld observations and the three lines are the median and 95% intervals of the predictive densities.