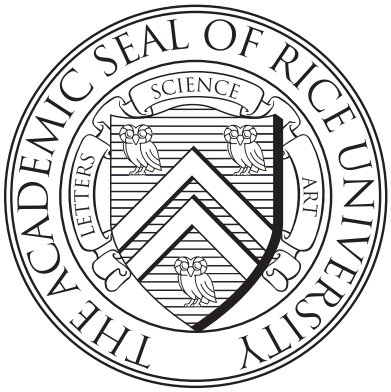
Bayesian Quantile Regression

Tyler Bagwell, Meredith Kruse, Kevin McCoy

Department of Statistics
Rice University
December 8th, 2023



Abstract

The goal of this project is to explore Bayesian quantile regression and its applications. Specifically, we will be discussing ordinary Bayesian quantile regression, longitudinal Bayesian quantile regression, and how to handle missing longitudinal data with Bayesian quantile regression. Within these topics we will present demonstrations with both simulated and real world data. For the ordinary Bayesian quantile regression approach we make comparisons to frequentist quantile regression and assess how both methods perform with different error terms. In the longitudinal quantile regression section we discuss different methods of posterior sampling and compare their performances. We will also briefly touch on the missing data problem to show why understanding longitudinal Bayesian quantile regression is important.

Contents

1 Ordinary Bayesian Quantile Regression

1.1 Background

Bayesian quantile regresion (BQR) was first introduced by Keming Yu and Ranna A Moyeed in [4], but before diving into BQR it is important to understand the fundamentals of standard quantile regression. As opposed to least squares regression which estimates conditional mean functions, quantile regression focuses on estimating families of conditional quantile functions. For quantile regression the standard linear model is given by $y_t = \mu(\boldsymbol{x_t}) + \epsilon_t$ where $\mu(\boldsymbol{x_t}) = \boldsymbol{x_t'}\boldsymbol{\beta}(\tau)$ with $0 < \tau < 1$ being the specified quantile of interest. The τth regression quantile is defined as any solution, $\hat{\beta}(\tau)$, to the quantile regression minimisation problem:

$$\min_{\beta} \sum_{t} \rho_\tau(y_t - \mathbf{x_t'}\beta) \tag{1}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the loss function being minimized over. This minimisation problem is equivalent to the maximization of a likelihood function formed by combining independently distributed asymmetric Laplace densities (ALD). A random variable U has an asymmetric Laplace distribution if its probability density is given by:

$$f_\tau(u; \mu, \sigma) = \frac{\tau(1-\tau)}{\sigma} exp\{-\rho_\tau(\frac{u-\mu}{\sigma})\} \tag{2}$$

where μ is a location parameter and σ is a scale parameter.

Now that the framework for general quantile regression has been explained we will now look specifically at BQR. In the Bayesian framework we are interested in the conditional quantile, $q_\tau(y_i|\boldsymbol{x_i})$, rather than the conditional expectation. Given the observations $\boldsymbol{y} = (y_1, ..., y_n)$, the posterior distribution of $\boldsymbol{\beta}$ is given by:

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}) \tag{3}$$

where $p(\boldsymbol{\beta})$ is the prior distribution of $\boldsymbol{\beta}$ and the likelihood is:

$$L(\boldsymbol{y}|\boldsymbol{\beta}) = p^n(1-p)^n exp\{-\sum_i \rho_\tau(y_i - \mathbf{x_i'}\beta)\} \tag{4}$$

using the location parameter $\mu_i = \mathbf{x_i'}\beta$. Ideally we would be able to choose an informative prior for $\boldsymbol{\beta}$, but in the case where there is no obvious prior choice we are able to use an improper uniform prior. Using this framework we will proceed to show an example of BQR using simulated and real data.

1.2 Simulation

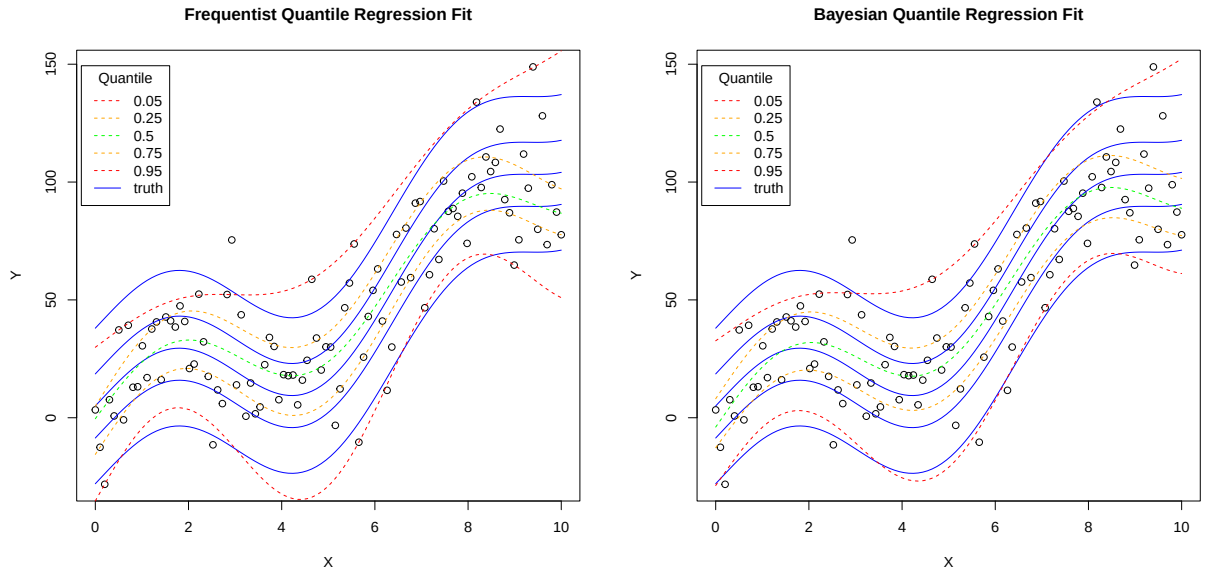
In order to demonstrate the effectiveness of BQR, synthetic data was generated according to the following equation:

$$y_i = 5 + x + x^2 + 20\sin(x) + \epsilon \tag{5}$$

where ϵ is the noise term added. We used two different distributions of noise to test the relative strength of BRQ in each setting, including normal and χ^2 distributed errors.

1.2.1 Normal error

As a baseline, we started with normal noise distributed $\epsilon \sim \mathcal{N}(0, 20^2)$. This yields a a signal to noise ratio (SNR), $SNR \approx 2.42$, where we define SNR as the ratio of variance of the data to the variance of the noise, $SNR = \frac{Var(\mathbf{y})}{\sigma^2}$. Off-the-shelf implementations of both BRQ and frequentist quantile regression (FQR) were used to compare the performance between frequentist and Bayesian methods. Figure 11 below illustrates the output of the two models at quantiles of $\tau = \{0.05, 0.25, 0.5, 0.75, 0.95\}$. Overlaid in blue are the true / theoretical quantiles based on the normal noise implemented. Because the error is normal here, the third blue line represents both the true 50th percentile and the generating equation with zero noise.



(a) Frequentist Quantile Regression (FQR)

(b) Bayesian Quantile Regression (BQR)

Figure 1: The $\tau = \{0.05, 0.25, 0.5, 0.75, 0.95\}$ quantiles are plotted, with the predicted values shown as the dashed lines, and the true values shown as solid blue lines.

One feature of BRQ that is missing in a frequentist implementation is the ability to derive credible intervals from its estimates. Illustrated below in Figure 2, the 0.05 and 0.95 quantiles are shown in dashed red lines, with their 95% credible intervals shown in solid black. The true quantiles are again shown in solid blue.

Next, we conducted a sensitivity analysis. A prior of $\beta(p) \sim \mathcal{N}(\mathbf{1}, \Sigma)$ was used, where $\Sigma = \sigma^2 * I$. Thus, a lower value of σ^2 indicates a higher confidence that the β's are equal to 1. Therefore, the smaller values of σ^2 correspond to smoother regression lines. The output of this sensitivity analysis is shown in Figure 3. Finally, the posterior distributions of the β's are shown below in Figure 4. You can see that higher degrees of confidence correspond to more peaked distributions around 1, our prior mean.

MCMC diagnostics were also performed. The ACF plots are shown below in Appendix A: MCMC Diagnostics, Figure 9. Effective mixing was only achieved in β_0 and β_3.

1.2.2 χ^2 error

BQR was also compared to frequentist methods using an asymmetrical error component. Thus, the previous experiments were repeated with the data generated according to:

$$y_i = 5 + x + x^2 + 20\sin(x) + 10 * \epsilon\epsilon \sim \chi^2_4 \implies SNR \approx 3.06 \tag{6}$$

The results of these experiments can be found in Appendix B: BQR with χ^2 error.

1.3 Real Data

The use of quantiles and quantile regression is flush with opportunities. We decided to investigate the impact of world factors and quality of life indicatorson BMI. The quality of life data was gathered by the Organisation for Economic Cooperation and Development Better Life Index [2]. The BMI data was taken from NCD Risk Factor Collaboration [1]. BQR was employeed with a diffuse prior, yielding the posteriors shown below in Figure 5. Unfortunately, no parameters except the intercept were found to be significant (all 95% credible intervals crossed 0). However, there is still plentiful information that can be gleamed from how

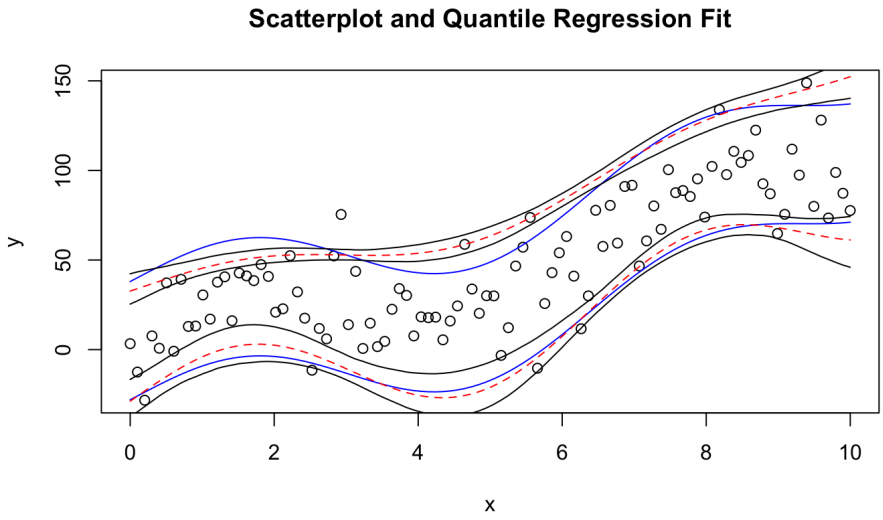


Figure 2: 0.05, 0.95 quantiles and their 95% Credible Interval

the parameters change over various posteriors. For example, higher disposable income is associated with higher 95% BMIs.

2 Longitudinal Bayesian Quantile Regression

2.1 Background

An especially important data type is longitudinal data which are data characterized by repeated measurements on the same subject over time. Instances of longitudinal data are wide and varied, including data collected in clinical trials, epidemiological studies, sociology studies, consumer research, and political polling. This type of data and measurement scheme undoubtedly introduces within-subject dependence and correlation in the observations. To handle these issues with longitudinal data, adding random effects to your model can account for over dispersion caused by unobserved heterogeneity among subjects and can account for correlation in time-dependent data which ultimately helps practitioners to avoid bias in parameter estimation. Due to the importance of the domains where longitudinal data analysis is required, it seems only natural to explore extensions of the ordinary BQR model introduced in Section 1 to longitudinal data by introducing linear mixed-effects BQR models.

In this section, we will introduce two linear mixed-effects BQR models derived by the work of Luo et. al. [3] that ultimately forms a Metropolis-Hastings sampler and a Gibbs sampler. Firstly, we will introduce the the generalized Bayesian hierarchical quantile regression model for linear mixed-effects. Next we will introduce the two sampling algorithms based on this Bayesian hierarchical quantile regression model including all parameters' posterior distributions. Then, we will compare the performances of both these two sampling algorithms along with ordinary BQR applied to synthetic longitudinal data via numerical simulations. Finally, we will write out how to extend this BQR model when there are intermittent missing observational data.

2.2 Two Posterior Sampling Algorithms for Longitudinal Data

We present here two posterior sampling algorithms for longitudinal data models with BQR methods based on the work by Luo et al. Firstly, consider the classical linear random effects model

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\alpha} + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \quad \sum_i n_i = N \tag{7}$$

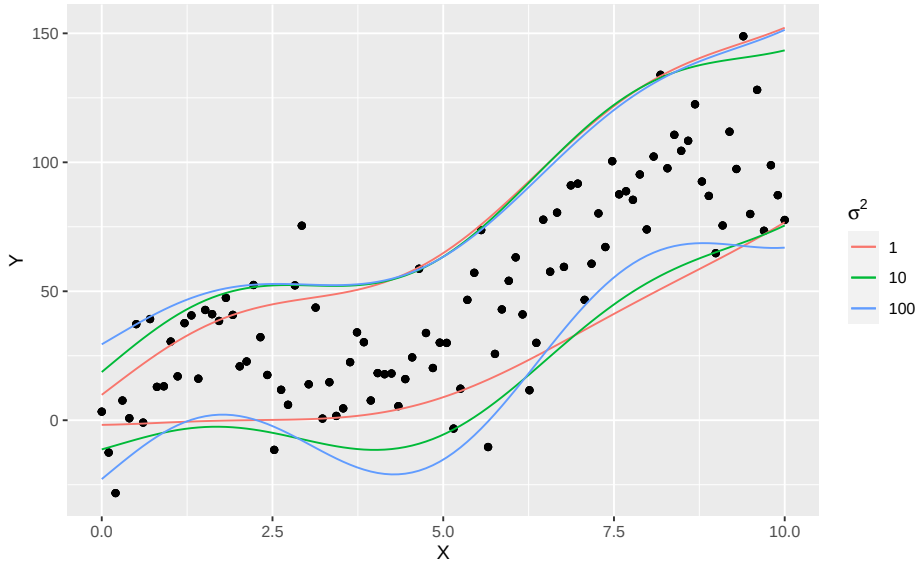


Figure 3: 0.05, 0.95 quantile regression lines as σ^2 is varied. Here, a lower value of σ^2 corresponds to a higher confidence in our prior.

where y_{ij} is the j-th scalar measurement of a continuous random variable on the i-th subject, $\boldsymbol{ij}^T$ are row vectors of a known design matrix of dimension $N \times k$ and $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed regression coefficients, $\boldsymbol{z}_{ij}$ is a $p \times 1$ covariate associated with random effects, and $\boldsymbol{\alpha}_i$ is a $p \times 1$ vector of random effects.

Under this model, it is the case that the quantile regression is now dependent on both the $\boldsymbol{\beta}$'s and $\boldsymbol{\alpha}$'s. Thus, the linear mixed quantile functions of the response y_{ij} are

$$Q_{y_{ij}}(\tau|\boldsymbol{x}_i j, \boldsymbol{\alpha}_i) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\alpha}, \tag{8}$$

where, unlike in the ordinary BQR seen in Section 1, $Q_{y_{ij}}$ is now also a function of the random effects $\boldsymbol{\alpha}_i$. We assume that the sampling distribution of the y_{ij}s is still asymmetric Laplace but now with location parameter $\mu_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\alpha}$. The fully generalized Bayesian hierarchical quantile regression model is

$$y_{ij} \sim \text{ALD}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\alpha}, \sigma, \tau) \tag{9}$$
$$\boldsymbol{\beta} \sim \pi(\boldsymbol{\beta}) \tag{10}$$
$$\boldsymbol{\alpha_i}|\Sigma \sim f(\boldsymbol{\alpha}_i|\Sigma) \tag{11}$$
$$\Sigma \sim \pi(\Sigma) \tag{12}$$
$$\sigma \sim \pi(\sigma), \tag{13}$$

with priors defined on parameters via the $\pi(\cdot)$'s.

Assuming a set of observations for each subject, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, the joint posterior distribution of the unknown parameters takes the form

$$\pi(\boldsymbol{\beta}, \sigma, \Sigma, \boldsymbol{\alpha}|\boldsymbol{y}) \propto f(\boldsymbol{y}, \boldsymbol{\alpha}|\boldsymbol{\beta}, \sigma, \Sigma)\pi(\boldsymbol{\beta})\pi(\sigma)\pi(\Sigma). \tag{14}$$

However, our main concern is with estimating the fixed effects $\boldsymbol{\beta}$, hence, we desire to compute

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) = \int \cdots \int \pi(\boldsymbol{\beta}, \sigma, \Sigma, \boldsymbol{\alpha}|\boldsymbol{y}) d\sigma \; d\Sigma \; d\boldsymbol{\alpha}, \tag{15}$$

which is generally an intractable integral even when assuming well-known distributions as priors for some of the parameters due to the asymmetric Laplace distribution having no known conjugate prior itself.

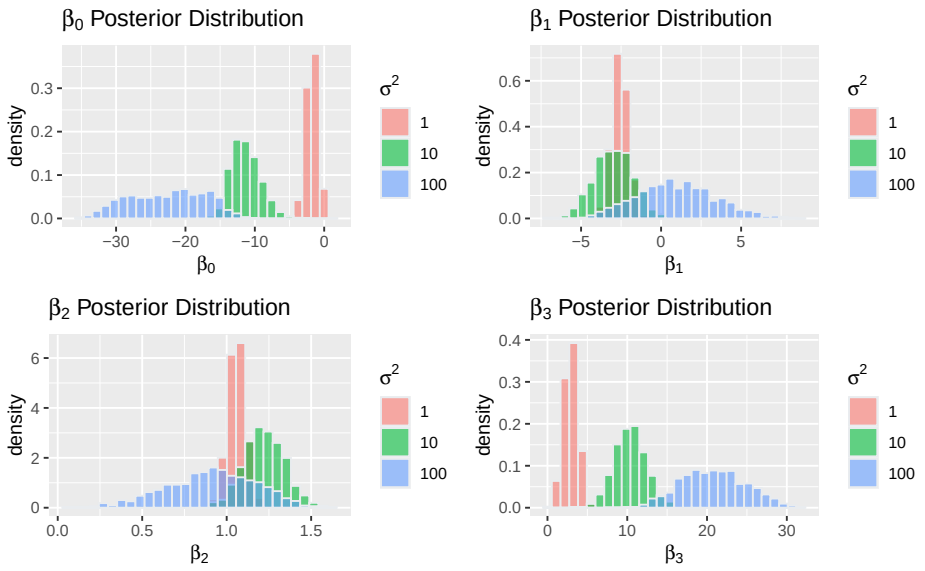


Figure 4: 95% β posteriors as σ^2 is varied. Here, a lower value of σ^2 corresponds to a higher confidence in our prior.

For the following two posterior sampling models discussed, we will assume assume the following priors:

$$\boldsymbol{\beta} \sim \text{MVN}_k(\boldsymbol{b}_0, \boldsymbol{B}_0) \tag{16}$$
$$\sigma \sim \text{IG}(c_0, d_0) \tag{17}$$
$$\boldsymbol{\alpha}_i \overset{\text{i.i.d}}{\sim} \text{MVN}_p(\boldsymbol{0}, \phi^2 \mathbb{I}) \tag{18}$$
$$\phi^2 \sim \text{IG}(k_0, w_0), \tag{19}$$

note that we have assumed no interdependence among the $\boldsymbol{\alpha}_i$'s and have written the covariance matrix Σ simply as identity matrix times a dispersion constant ϕ^2. Fortunately, the joint posterior distribution for $\boldsymbol{\beta}, \sigma, \Sigma, \boldsymbol{\alpha}|\boldsymbol{y}$ can be evaluated via common MCMC methods – explicitly for this study – Metropolis-Hastings sampling and Gibbs sampling.

2.2.1 Bayesian Quantile Regression via a Metropolis-Hastings Sampler

Based on the joint posterior distribution, the full conditional distributions of the parameters can be written as

$$\boldsymbol{\alpha}_i|\boldsymbol{y}, \boldsymbol{\beta}, \sigma, \phi^2 \sim \exp\left[-\frac{1}{2\phi^2}\boldsymbol{\alpha}_i^T\boldsymbol{\alpha}_i - \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - \boldsymbol{z}_{ij}^T\boldsymbol{\alpha})\right], \tag{20}$$

$$\boldsymbol{\beta}_i|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma, \phi^2 \sim \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{b}_0)^T\boldsymbol{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{b}_0) - \frac{1}{\sigma}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - \boldsymbol{z}_{ij}^T\boldsymbol{\alpha})\right], \tag{21}$$

$$\sigma|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi^2 \sim \text{IG}\left(N + c_0, d_0 + \sum_{i=1}^{n}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - \boldsymbol{z}_{ij}^T\boldsymbol{\alpha})\right), \tag{22}$$

$$\phi^2|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma \sim \text{IG}\left(\frac{np}{2} + k_0, w_0 + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^T\boldsymbol{\alpha}_i\right), \tag{23}$$

where $\rho_\tau(\cdot)$ is the check function explained in Section 1.

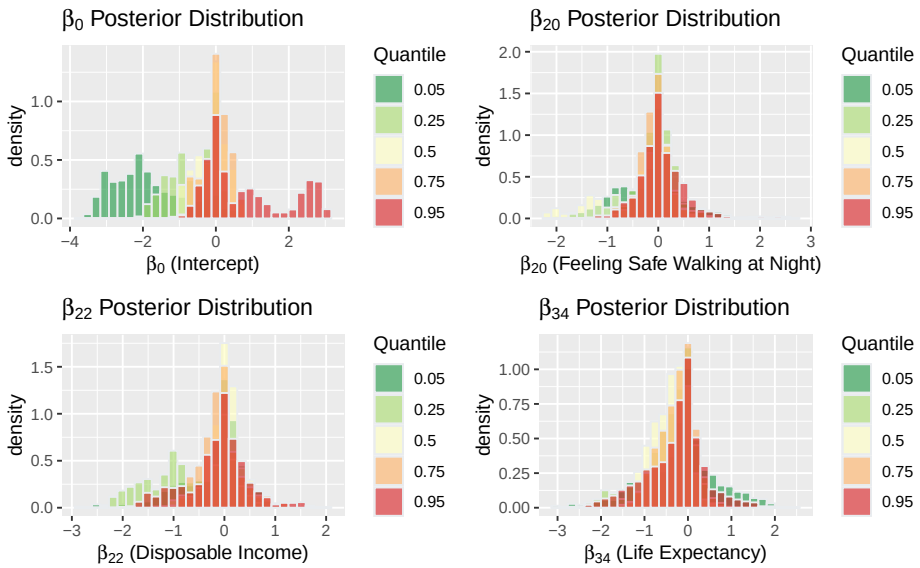


Figure 5: β posteriors over various quantiles linking world factors and quality of life to country and age level BMI. Only the intercept was found to be significant at 95% credible levels.

The two proposal distributions employed for choosing potential updates for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i$ are $\text{MVN}_k(\boldsymbol{\beta}^c, \sigma_\beta^2 \Sigma_\beta)$ and $\text{MVN}_p(\boldsymbol{\alpha}_i^c, \sigma_\alpha^2 \Sigma_\alpha)$, respectively, where $\boldsymbol{\beta}^c$ is the current value of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i^c$ is the current value of $\boldsymbol{\alpha}_i$. The hyperparameters σ_β and σ_α are tuning parameters for the acceptance probabilities of accepting the proposed values over the current values for updating $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i$. As will be shown later, these tuning parameters must be optimized for different values of N, n, and τ to ensure efficient Metropolis-Hastings sampling.

2.2.2 Bayesian Quantile Regression via a Gibbs Sampler

Luo et al. [3] also propose a Gibbs sampling algorithm based on a location-scale mixture representation of the asymmetric Laplace distribution which does not require any approximations. Let $y \sim \text{ALD}(\mu, \sigma, \tau)$, z be a standard normal random variable, and e be an exponential random variable with parameter σ. Then Luo et al. propose the following representation

$$y = \sqrt{\frac{2\sigma e}{\tau(1-\tau)}} z + \frac{1-2\tau}{\tau(1-\tau)} e + \mu. \tag{24}$$

Denoting $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$ and $k_2 = \frac{2}{\tau(1-\tau)}$, the above is rewritten as

$$y = k_1 e + \sqrt{k_2 \sigma e} z + \mu. \tag{25}$$

Using this result, the response y_{ij} can be rewritten as

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{\alpha} + k_1 e_{ij} + \sqrt{k_2 \sigma e_{ij}} z_{ij}, \tag{26}$$

where $\boldsymbol{\alpha}_i \sim \text{MVN}_p(\boldsymbol{0}, \phi^2 \mathbb{I})$, $e_{ij} \sim \exp(1/\sigma)$, and $z_{ij} \sim \text{N}(0,1)$ are mutually independent, therefore the sampling distribution of y_{ij} becomes

$$f(y_{ij}, \boldsymbol{\beta}, \boldsymbol{\alpha}_i, e_{ij}, \sigma) = (2\pi k_2 \sigma e_{ij})^{-1/2} \exp\left[-\frac{1}{2k_2 \sigma e_{ij}} (y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta} - \boldsymbol{z}_{ij}^T \boldsymbol{\alpha} - k_1 e_{ij} + \sqrt{k_2 \sigma e_{ij}})^2 \right]. \tag{27}$$

Based on this representation, the full conditional distributions of the parameters can be derived and are families of well-known distributions. We have for the $\boldsymbol{\beta}$

$$\pi(\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{e}, \sigma) \sim \text{MVN}_k(\boldsymbol{Bb}, \boldsymbol{B}), \tag{28}$$

where

$$\boldsymbol{B}^{-1} = \frac{1}{k_2\sigma}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T}{e_{ij}} + \boldsymbol{B}_0^{-1}, \tag{29}$$

$$\boldsymbol{b} = \frac{1}{k_2\sigma}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}(y_{ij} - \boldsymbol{z}_{ij}^T\boldsymbol{\alpha}_i - k_1 e_{ij})}{e_{ij}} + \boldsymbol{B}_0^{-1}\boldsymbol{b}_0. \tag{30}$$

The full conditional density for $\boldsymbol{\alpha}_i$ is

$$\pi(\boldsymbol{\alpha}_i|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{e}_i,\sigma,\phi^2) \sim \text{MVN}_p(\boldsymbol{A}\boldsymbol{a},\boldsymbol{A}), \tag{31}$$

where

$$\boldsymbol{A}^{-1} = \frac{1}{k_2\sigma}\sum_j \frac{\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T}{e_{ij}} + \frac{1}{\phi^2}\mathbb{I}, \tag{32}$$

$$\boldsymbol{a} = \frac{1}{k_2\sigma}\sum_j \frac{\boldsymbol{z}_{ij}(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - k_1 e_{ij})}{e_{ij}}. \tag{33}$$

The full conditional of the e_{ij}'s is generalized inverse Gaussian (GIG) distribution, thus

$$\pi(e_{ij}|\boldsymbol{y},\boldsymbol{\alpha}_i,\boldsymbol{\beta}) \sim \text{GIG}(\frac{1}{2},\hat{\delta}_{ij},\hat{\gamma}_{ij}), \tag{34}$$

where

$$\hat{\delta}_{ij}^2 = \frac{(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - \boldsymbol{z}_{ij}^T\boldsymbol{\alpha}_i)^2}{k_2\sigma}, \tag{35}$$

$$\hat{\gamma}_{ij}^2 = \frac{k_1^2}{k_2\sigma} + \frac{2}{\sigma}. \tag{36}$$

The full conditional for the remaining parameters σ and ϕ^2 are

$$\pi(\sigma|\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{e},\boldsymbol{\beta},\phi^2) \sim \text{IG}(\nu,\omega), \tag{37}$$

where

$$\nu = \frac{N}{2} + c_0, \tag{38}$$

$$\omega = d_0 + \frac{1}{k_2}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \frac{(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - \boldsymbol{z}_{ij}^T\boldsymbol{\alpha}_i - k_1 e_{ij})^2}{e_{ij}}, \tag{39}$$

and

$$\pi(\phi^2|\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma) \sim \text{IG}(\xi,\bar{\omega}), \tag{40}$$

where

$$\xi = \frac{np}{2} + k_0, \tag{41}$$

$$\bar{\omega} = w_0 + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^T\boldsymbol{\alpha}_i. \tag{42}$$

Fortunately, the full conditional posterior distributions for all the parameters above are well known, therefore, they can be easily sampled from widely available scientific computing packages and Gibbs sampling can be employed. Additionally, a Gibbs sampler does not require any proposal sample tuning unlike that seen in the Metropolis-Hastings algorithm, hence, one can postulate that this Gibbs sampler should generally be more efficient than the aforementioned Metropolis-Hastings algorithm. Our simulation results below give numerical evidence to this proposition.

2.3 Numerical Simulations

2.3.1 Simple Regression Model

For the purposes of model comparison, we will use the following simple linear mixed-effects model to generate data similar to that employed by Luo et al. [3]

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{i0} + \alpha_{i1} x_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, 20, \quad j = 1, \ldots, 5, \tag{43}$$

where $x_{ij} \sim \text{Uniform}(0, 1)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^T = (1, 5)^T$. We assumed the random effects are drawn as $\boldsymbol{\alpha}_i = (\alpha_{i0}, \alpha_{i1})^T \sim \text{MVN}_2(\mathbf{0}, \mathbb{I})$. For this first study we will assume that the errors are Gaussian noise, i.e., $\epsilon_{ij} \sim \text{N}(0, 1)$. Keeping with Luo et al. [3], we will also assume weak prior information and assume an $\text{MVN}_2(\mathbf{0}, 100\mathbb{I})$ prior on $\boldsymbol{\beta}$ and $\text{IG}(0.01, 0.01)$ priors on σ and ϕ^2. We will perform simulations on five different levels of $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. Under standard normal (Gaussian) error, we can compute the theoretical regression quantiles when $\boldsymbol{\beta} = (\beta_0, \beta_1)^T = (1, 5)^T$. They are summarized in the following table:

Table 1: Theoretical regression quantiles for (β_0,β_1) for given τ.

τ	$\beta_0(\tau)$	$\beta_1(\tau)$
0.10	-0.22	5.00
0.25	0.33	5.00
0.50	1.00	5.00
0.75	1.67	5.00
0.90	2.28	5.00

It is obvious from this table that the quantile regression lines for all five values of τ are merely parallel lines with identical slope $\beta_1 = 5$. For clearer exposition, we plotted a single synthetic data set below along with the theoretical quantile regression lines under the assumed Gaussian error.

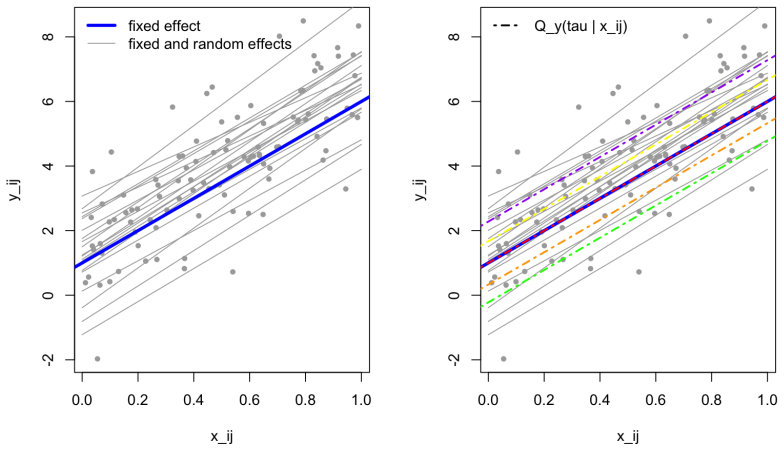


Figure 6: A simulated data set from the simple linear mixed-effects model. The plot on the right superimposes the theoretical quantile regression lines (dashed lines) for the five values of $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ on top of the simulated data.

2.3.2 Sampler Diagnostics

We coded up both the BQR Gibbs sampler (BQRGS) and Metropolis-Hastings sampler (BQRMH) based on the algorithms explained in the previous section. To get a sense of the overall efficiency for each sampling algorithm, we can compare the trace and autocorrelation plots of the drawn posterior samples from both algorithms. These plots of the β_1 posterior draws for both algorithms from a single simulated dataset can be seen in Figure 7 for $\tau = 0.5$. For this single simulated dataset, there were a total of 5000 sample draws with the first 500 thrown away as burn-in.

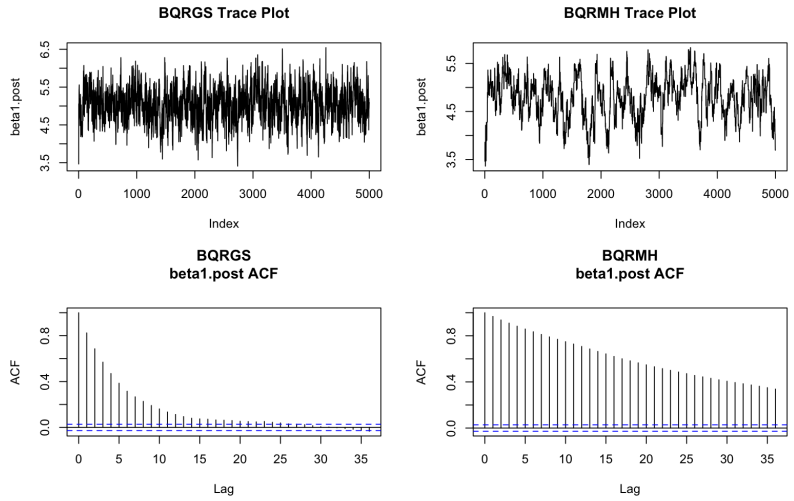


Figure 7: Trace and autocorrelation plots of the posterior samples for β_1 from the Gibbs sampler (BQRGS) and the Metropolis-Hastings sampler (BQRMH) for $\tau = 0.5$. Note that the trace plots show all draws including burn-in, but the ACF plots are computed with burn-in draws removed.

From the plots seen in Figure 7, it is quite obvious that the Gibbs sampler is a much more efficient sampler than the Metropolis-Hastings sampler. This can be seen in both the trace and ACF plots. One important point to note is that for every specific value of τ, the practitioner must find the set $(\sigma_\alpha^2, \sigma_\beta^2)$ that jointly optimizes the acceptances rates of both the $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\beta}$'s, which is generally a nuisance, especially compared to the Gibbs sampler which does not rely on deciding between the current value of a parameter and a proposed one, but directly samples a new value from a draw-able distribution. Ideally, the practitioner tries to obtain acceptance rates in the interval $[0.2, 0.4]$. Even under optimal acceptance rates (as shown in Figure 7) the Metropolis-Hastings sampler still remains much *stickier* than the Gibbs sampler and this is further evidenced by the high values of autocorrelation for large values of lag seen in the ACF plots of the posterior samples of β_1 drawn from the Metropolis-Hastings algorithm compared to those drawn from the Gibbs sampler. This result also seemed to hold for other parameters as well across the two algorithms. Therefore, since both models do no require any approximations in deriving the forms of the parameter posterior distributions, we will generally recommend researchers to employ the Gibbs sampler over the Metropolis-Hastings sampler as a rule of thumb. Despite the relative inefficiency of the Metropolis-Hastings sampler, the two algorithms' parameter estimation performance are generally quite similar as will be show in the following subsection.

2.3.3 Model Performance Comparisons

Finally, we compare the performance of parameter estimation of the two mixed-effects BQR samplers (BQRGS and BQRMH) along with the performance of parameter estimation of ordinary Bayesian quantile regression (OBQR) introduced in Section 1 under data generated from the simple linear mixed-effects model introduced at the beginning of this section. To perform this comparison, we simulated 100 datasets and

drew 5000 samples (minus 500 burn-in) for each parameter for each of the three models for the five levels of $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. These posterior samples were then used to compute the parameter estimation bias and MSE from these three models. The results of which are plotted in Figure 8 (please enjoy this plot, it took my little MacBook around 4 hours to compute these results).

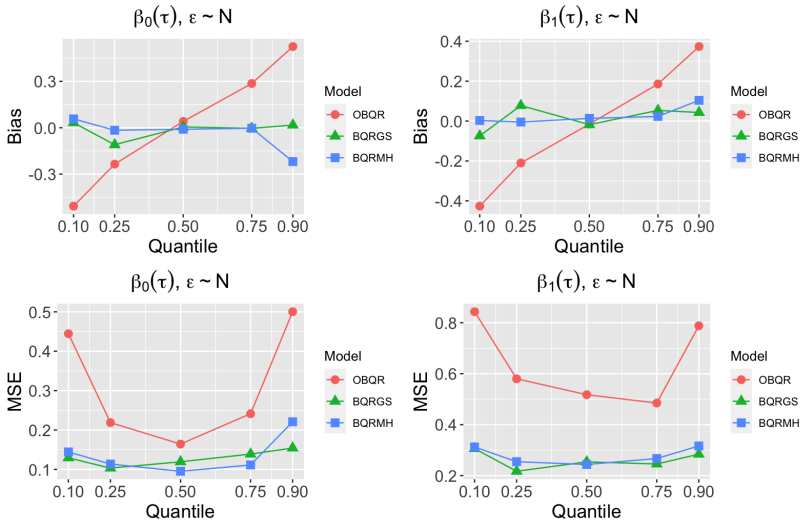


Figure 8: Bias and MSE of estimations of β_0 and β_1 from the three algorithms under Gaussian error: OBQR, BQRQS, and BQRMH.

For the simple linear mixed-effects model, the BQRGS and BQRMH algorithms perform quite similarly and the simulation results show approximately zero bias for both estimations of β_0 and β_1 across the five values of τ. With regards to MSE, both models show a slight uptick in MSE as you move away on either side from the median regression quantile which is to be expected since there are less data to evaluate in these more extreme regions under Gaussian error. Meanwhile, the OBQR model clearly performs much worse than BQRGS and BQRMH exhibiting both relatively large bias and MSE across the values of τ. The only exception is at the median ($\tau = 0.5$), where our simulations show roughly no bias in the estimation of β_0 and β_1 for the OBQR model, however, the MSE lines indicate that, although the median regression quantile estimation is unbiased for $\beta_0(\tau = 0.5)$ and $\beta_1(\tau = 0.5)$, there is still high variance in the computed point estimations using OBQR severely hurting reliable parameter estimation compared to BQRGS and BQRMH. Therefore, both BQRGS and BQRMH are superior algorithms compared to OBQR under linear mixed-effects models.

2.4 Extensions to Missing Data Problems

Lastly, we briefly discuss extensions of the longitudinal BQR model to deal with intermittent missing data in observations based on the work of Yuan and Yin. These two researchers propose the penalized model

$$\sum_{i=1}^{n} \sum_{j \in J_{obs}} \rho_\tau \left(y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta} - \boldsymbol{z}_{ij}^T \boldsymbol{\alpha}_i \right) + \frac{1}{2} \sum_{i=1}^{n} \boldsymbol{\alpha}_i^T \Lambda^{-1} \boldsymbol{\alpha}_i, \tag{44}$$

where Λ is a symmetric matrix. This model can be recast into a mixed-effects model, like those explored in the preceding subsections that extends to missing data issues. The model assumes that one can link the missing data process with the longitudinal outcome process while simultaneously assuming they share the same random effect $\boldsymbol{\alpha}_i$. This compounds the likelihood where instead of just a marginal likelihood for the observations $\boldsymbol{y}$, you have a joint likelihood for both the observed response AND whether or not a certain response is observed $(\boldsymbol{y}, \boldsymbol{s})$. As such, a missing data model would also require the computation of transition probabilities $\pi_{ij}^{\text{Observed}}$ and $\pi_{ij}^{\text{Missing}}$. We plan to implement this model in future work.

3 Conclusion

Throughout this report we explored multiple different types of quantile regressions and methods of doing Bayesian quantile regression. First, we found that Bayesian quantile regression does a better job at estimating quantiles than the frequentist method especially in the case when the error terms are not normally distributed. This was seen through both simulated and real data examples. This makes the Bayesian framework a more flexible option and thus more viable for use in real data examples. Second, we explored longitudinal Bayesian quantile regression and its potential applications. We recommend using the Gibbs sampler when doing longitudinal Bayesian quantile regression as it is more efficient than the Metropolis Hastings method and the two methods give similar results. Lastly, we briefly touched on longitudinal Bayesian quantile regression with missing data. This is a particularly interesting use of our methods, but given the time constraints we felt it could better be explored in future work. Throughout this project we learned the many uses of quantile regression and why implementing it using a Bayesian method is extremely useful in real world problems.

Code Availability

All code is written using `R` and is made available on GitHub at github.com/kmccoy3/bayesian-qr. Please contact Kevin McCoy `<kevin@kmccoy.net>` if you have any issues downloading or running the code.

Author Contributions

OBQR theoretical derivations, M.K.; simulated and real data exploration for OBQR, K.M.; BQRGS and BQRMH theoretical derivations, M.K.; simulated and real data exploration for BQRGS and BQRMH, T.B.; writing, review, and editing, T.B., M.K., K.M.

References

[1] James Bentham, Mariachiara Di Cesare, V BIlano, Lynne M Boddy, et al. Worldwide trends in children's and adolescents' body mass index, underweight and obesity, in comparison with adults, from 1975 to 2016: a pooled analysis of 2,416 population-based measurement studies with 128.9 million participants. *Lancet*, 2017.

[2] Marc Fleurbaey. Beyond gdp: The quest for a measure of social welfare. *Journal of Economic literature*, 47(4):1029–1075, 2009.

[3] Youxi Luo, Heng Lian, and Maozai Tian. Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 2012. URL https://www.tandfonline.com/doi/abs/10.1080/00949655.2011.590488.

[4] Keming Yu and Rana A Moyeed. Bayesian quantile regression. 2001.

[5] Ying Yuan and Guosheng Yin. Bayesian Quantile Regression for Longitudinal Studies with Nonignorable Missing Data. *Biometrics*, 66(1):105–114, March 2010. ISSN 0006-341X, 1541-0420. doi: 10.1111/j.1541-0420.2009.01269.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1541-0420.2009.01269.x.

Appendix A: MCMC Diagnostics

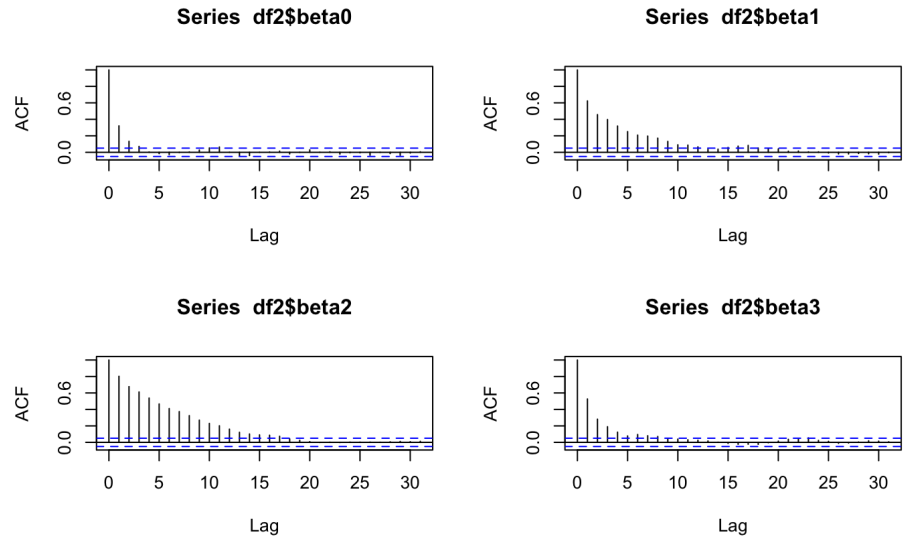


Figure 9: Autocorrelation Plots

Appendix B: BQR with χ^2 error

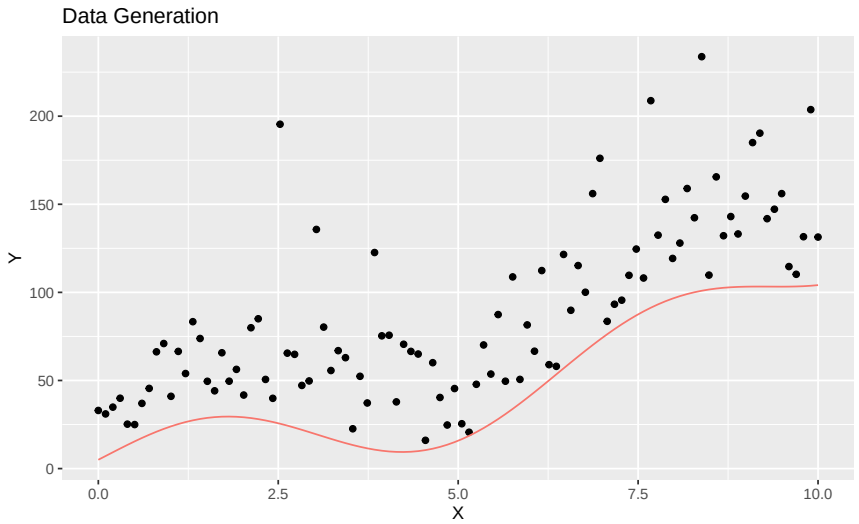
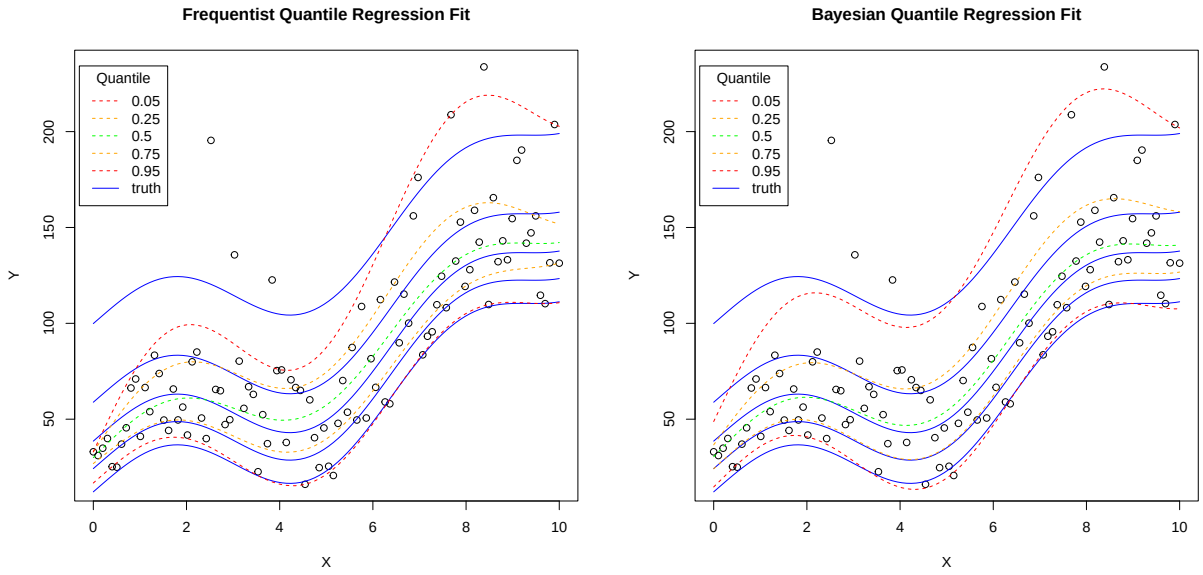


Figure 10: Data generated according to equation 6.



(a) Frequentist Quantile Regression (b) Bayesian Quantile Regression

Figure 11: The $\tau = \{0.05, 0.25, 0.5, 0.75, 0.95\}$ quantiles are plotted, with the predicted values shown as the dashed lines, and the true values shown as solid blue lines.

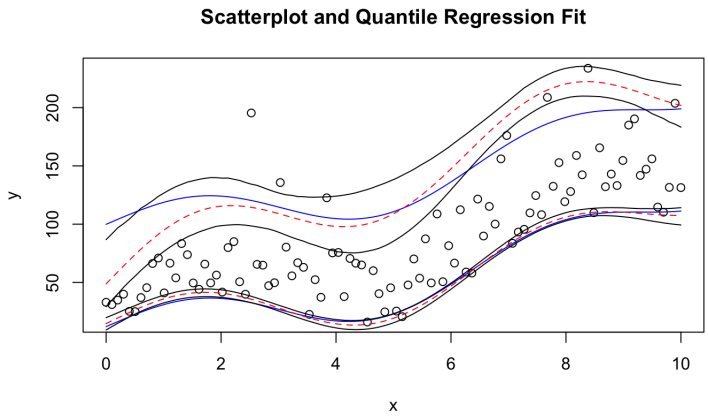


Figure 12: 0.05, 0.95 Quantiles and their 95% Credible Interval