Hierarchical Quantile Regression

by

Imran Hassan

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Biostatistics

Department of Mathematical and Statistical Sciences

University of Alberta

Abstract

Quantile regression supplements the ordinary least squares regression and provides a complete view of a relationship between a response variable and a set of covariates. The quantile regression model does not assume any particular error distribution. It is estimated by minimizing an asymmetric absolute error loss function. Bayesian inference of quantile regression is based on the likelihood function formed by independent asymmetric Laplace densities. The asymmetric Laplace distribution is a natural choice for the error distribution of the quantile regression model. However, the model based on the asymmetric Laplace distribution solely focuses on estimation and does not describe the underlying true model. Moreover, it assumes different models for estimating parameters for different quantile levels.

In this project, we introduce a hierarchical quantile regression model that removes ambiguities of the quantile regression model based on the asymmetric Laplace distribution. The proposed hierarchical model treats the intercept and the slope of the linear quantile regression model as random effects. The model is estimated by the data cloning method which works in the Bayesian framework exploiting the computational advantage of the Markov Chain Monte Carlo (MCMC) algorithm, but gives the maximum likelihood estimates with standard errors.

A simulation study with 50 repetitions has been performed to assess the parameter estimates. We have compared our results to the regular quantile

regression estimates for different quantile levels.

Our proposed hierarchical model gives a greater insight into the overall quantile regression picture. The model is easily extendable to accommodate more complex situations and provides room for further research.

Acknowledgements

All praise is due to Allah, the most merciful and the most forgiving. First of all, I am thankful to my creator for enabling me to conduct this project.

I would like to express my sincere gratitude to my supervisor, Dr. Subhash Lele. Without his guidance, constant encouragement and support, this thesis could never be possible. It has been a rewarding opportunity for me to work under his supervision. He has a great impact on my thought process. I am also greatly benefited from the courses taught by him.

I am grateful to Dr. Narasimha Prasad. I have taken four graduate courses with him. He has a great contribution to my understanding and knowledge of statistics. I have learned a lot from him. I specially thank him for being a member of my thesis committee.

I would like to thank Dr. Linglong Kong and Dr. Ivor Cribben for being a part of my examination committee.

I would like to take this opportunity to extend my gratitude to my professors, specially Dr. Wasimul Bari, Mr. MA Jalil and Dr. Jafar Ahmed Khan, at the University of Dhaka, Bangladesh. They made me interested in statistics through their outstanding teaching and research. My sincere thanks goes also to all my friends and colleagues in Bangladesh and here in Canada.

I am heartily grateful to my parents and sister for their constant encouragement, support and sacrifice throughout my life. They are instrumental in every success of my life. Finally, I thank my beloved wife, Adri Anan Brishty. Without her endless support and encouragement, I could never complete this work in a timely manner.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Regression (a term coined by Sir Francis Galton in 1885) analysis is one of the most popular and useful tools of statistics. It models a variable (called dependent or response variable) as a function of other variables (called independent variables, explanatory variables or covariates) on which it depends. In statistical words, regression analysis studies how the means of the conditional distributions of a dependent variable change with covariates. It is often called mean regression or regression of the conditional mean. However, the mean is not only quantity one is interested in to describe a distribution. One may also be interested in spread, skewness and kurtosis to get more insight into a distribution. Moreover, the mean itself has its own limitations as a measure of central tendency. As the mean does not provide a complete picture of a distribution, the conditional mean regression does not give a complete picture of a relationship between a dependent variable and a set of covariates. It may happen that a relationship may not exist through the central part of the conditional distributions, but it may exist through the upper or the lower tail, or it may be different through the different parts of the conditional distributions.

However, now, we may overcome such limitations of the usual conditional mean regression with the advent of quantile regression.

Quantile regression, introduced by Koenker and Bassett in 1978, allows one to study the relationship through different parts of the conditional distributions by modeling different quantiles. In recent years, quantile regression has gained much popularity in applied research and has started to replace the conditional mean regression in many applications. Despite its wide use in many fields, the quantile regression model is still difficult to understand. Though the estimation method is quite established in the semi-parametric (functional form of the model is specified but the error distribution is left unspecified) case, there is no sound parametric model available. In this project, we attempt to make quantile regression clear and understandable. We address the issue that the available parametric quantile regression model does not describe the underlying process at all, rather concentrates on estimation. We introduce a hierarchical quantile regression model that describes the underlying process of quantile regression well.

We discuss the motivation to quantile regression, the quantile regression model and its estimation in chapter 2. The philosophy of estimating the quantile regression model is closely tied with the least squares method. Estimation of quantile regression is based on the fact that quantiles can be estimated through a simple minimization problem. The quantile regression model is estimated by minimizing an asymmetric absolute error loss function. This is similar to the fact that the conventional mean regression is estimated by the least squares method through minimizing the squared error loss function. In section 2.1, 2.2 and 2.3, we follow chapter 1 of Koenker (2005). In the same chapter, we give a brief description of Bayesian quantile regression (Yu and Moyeed, 2001) and the asymmetric Laplace distribution. Bayesian quantile regression is based on the likelihood function of the asymmetric Laplace distribution. Minimizing the asymmetric absolute error loss function turns out to be the same as maximizing a likelihood function formed by combining independent asymmetric Laplace densities.

In chapter 3, we discuss the issue that available parametric quantile regression model based on the asymmetric Laplace distribution does not provide the true model of the data. In this chapter, we introduce a hierarchical quantile regression model as an attempt to give a clear picture of the true model of the data. Later on we discuss the estimation process of the proposed model followed by simulation studies, results, discussion and limitations.

Finally, we talk about the possibilities of future research in this area and conclude the report in chapter 4.

Chapter 2

Quantile Regression

2.1 Motivation to Quantile Regression

Consider a simple linear regression model,

$$y_i|x_i = \beta_0 + \beta_1 x_i + \epsilon_i, \tag{2.1}$$

where y_i is the response of the ith individual, x_i is the corresponding covariate, and ϵ_i is independent and identically distributed (iid) error with mean zero. Thus, the conditional mean of y given x is

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i.$$

The least squares regression captures how the conditional mean of y given x changes with x.

Figure 2.1 illustrates artificial data (size 100) generated from model (2.1) with normal-distributed errors (where $\beta_0 = 2$, $\beta_1 = 0.5$). The line through the

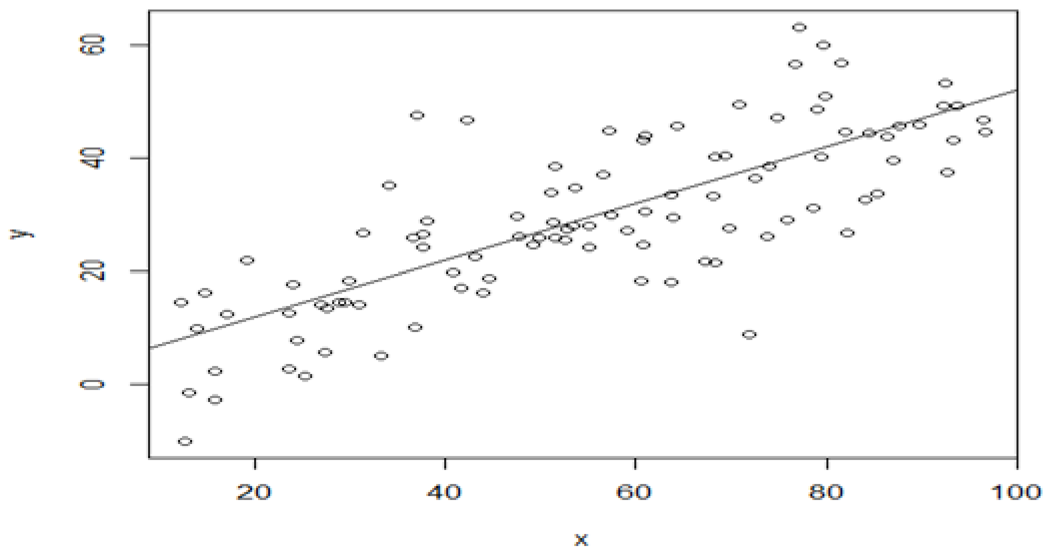

Figure 2.1: Conditional mean function

data points is the true conditional mean line.

Now, from model (2.1), the conditional quantile functions of y are

$$Q_{y_i}(\tau|x_i) = \beta_0 + \beta_1 x_i + F_\epsilon^{-1}(\tau) = \{\beta_0 + F_\epsilon^{-1}(\tau)\} + \beta_1 x_i,$$

which can be written as

$$Q_{y_i}(\tau|x_i) = \beta_0(\tau) + \beta_1 x_i, \tag{2.2}$$

where, $Q_{y_i}(\tau|x_i)$ is the τ^{th} conditional quantile of y given x, F_ϵ denotes cumulative distribution function (cdf) of the error ϵ, hence, $F_\epsilon^{-1}(\tau)$ is the τ^{th} quantile of the error, and $\beta_0(\tau) = \beta_0 + F_\epsilon^{-1}(\tau)$.

In the conditional quantile functions of (2.2), intercepts of the conditional quantile lines change with the quantile levels, however, slopes remain the same for all quantiles.

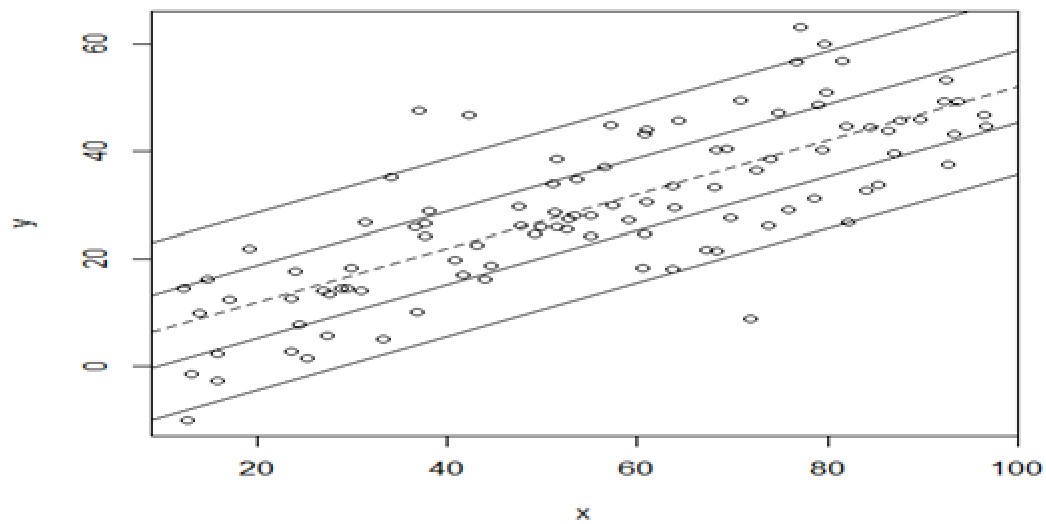

Figure 2.2: Conditional quantile functions (homoscedastic case)

Figure 2.2 shows several true conditional quantile functions for the simulated data from model (2.1). The lines, from top to bottom, are $\{0.05^{th}$, 0.25^{th}, 0.5^{th}, 0.75^{th}, $0.95^{th}\}$ conditional quantile lines respectively. The conditional median function (0.5^{th} quantile) is the same as the conditional mean function depicted by dashed line in the figure, because the mean and the median is the same for symmetric distributions. All lines share the same slope, with only the intercepts varying for different quantiles. Thus, the marginal change of the conditional quantiles for the marginal change in x is the same for all quantiles. In other words, the relationship between y and x is the same in all parts of the conditional distribution of y given x. Thus, we do not need to study the relationship between y and x for the entire conditional distribution. Hence, there is a little need of quantile regression in the iid linear model case. The study of the conditional mean by the least squares regression is sufficient in this case.

Now, we consider a heteroscedastic situation where the model takes the following form.

$$y_i|x_i = \beta_0 + \beta_1 x_i + \sigma(x_i)\epsilon_i. \tag{2.3}$$

We assume, $\sigma(x_i) = \gamma_0 + \gamma_1 x_i$ and ϵ_i are again iid with cdf F_ϵ. Thus,

$$y_i|x_i = \beta_0 + \beta_1 x_i + (\gamma_0 + \gamma_1 x_i)\epsilon_i. \tag{2.4}$$

The conditional quantile functions of y are

$$Q_{y_i}(\tau|x_i) = \beta_0 + \beta_1 x_i + (\gamma_0 + \gamma_1 x_i)F_\epsilon^{-1}(\tau),$$

which can be rewritten as

$$Q_{y_i}(\tau|x_i) = \{\beta_0 + \gamma_0 F_\epsilon^{-1}(\tau)\} + \{\beta_1 + \gamma_1 F_\epsilon^{-1}(\tau)\}x_i,$$

that can be expressed as

$$Q_{y_i}(\tau|x_i) = \beta_0(\tau) + \beta_1(\tau)x_i,$$

where $\beta_0(\tau) = \{\beta_0 + \gamma_0 F_\epsilon^{-1}(\tau)\}$, and $\beta_1(\tau) = \{\beta_1 + \gamma_1 F_\epsilon^{-1}(\tau)\}$.
That is, in the heteroscedastic situation, both the intercepts and slopes of the conditional quantile lines depend on the quantile levels.

In Figure 2.3, we exhibit an artificial sample of size 100 generated from

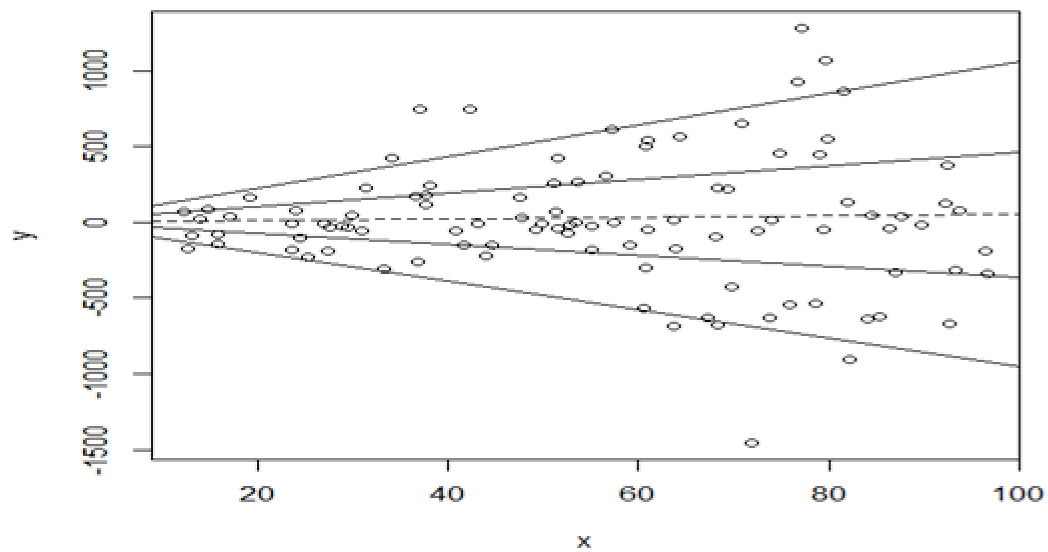

Figure 2.3: Conditional quantile functions (heteroscedastic case)

model (2.4) with normal-distributed error (where $\beta_0 = 2$, $\beta_1 = 0.5$, $\gamma_0 = 1$, $\gamma_1 = 0.7$) and quantile functions for different quantile levels. The lines exhibited in Figure 2.3, from top to bottom, are $\{0.05^{th}$, 0.25^{th}, 0.5^{th}, 0.75^{th}, $0.95^{th}\}$ conditional quantile functions respectively. As shown in Figure 2.3, the relationship between y and x is different on the different parts of the conditional distribution of y. For instance, the relationship is positive on the upper tail and negative on the lower tail of the conditional distribution. In the heteroscedastic case described by model (2.3), the least squares regression fails to capture the overall relationship between the response variable and the covariate. The dashed line in Figure 2.3 is the conditional mean line (which is the same for the median (0.5^{th} quantile) in this case) which shows a weak relationship between the two variables. However, the relationship is very strong (and opposite to each other in sign) on either tail of the conditional distribution depicted by the top (0.95^{th} quantile) and the bottom (0.05^{th} quantile) lines in the figure. So, we need to study the relationship between y and x for the entire conditional distribution in heteroscedastic case. We can do so by quantile regression. It enables us to estimate conditional quantile functions for different quantile levels, and thus gives a complete view of the relationship.

2.2 Quantile Regression Model

We first consider the least squares regression model,

$$y_i|x_i = E(y_i|x_i) + \epsilon_i,$$

where y_i is the response of the ith individual, x_i is the corresponding covariate, ϵ_i is assumed to be independently and identically distributed (iid) errors with mean zero and constant variance.

For the simple linear regression model,

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i.$$

Similarly, the quantile regression model for τ^{th} quantile is constructed as

$$y_i|x_i, \tau = Q_{y_i}(\tau|x_i) + u_i(\tau),$$

where $u_i(\tau)$'s are independently and identically distributed (iid) errors and its τ^{th} quantile is zero.

For the linear quantile regression,

$$Q_{y_i}(\tau|x_i) = \beta_0(\tau) + \beta_1(\tau)x_i.$$

Thus, the linear quantile model for the τ^{th} quantile is

$$y_i|x_i, \tau = \beta_0(\tau) + \beta_1(\tau)x_i + u_i(\tau), \qquad (2.5)$$

where parameters are specified for the τ^{th} quantile and they differ for different quantile levels.

A special case of model (2.5) is model (2.4), where $\beta_0(\tau) = \{\beta_0 + \gamma_0 F_\epsilon^{-1}(\tau)\}$ and $\beta_1(\tau) = \{\beta_1 + \gamma_1 F_\epsilon^{-1}(\tau)\}$.

From model 2.4 and 2.5, we get

$$u_i(\tau) = \epsilon_i - F_\epsilon^{-1}(\tau).$$

Hence, the τ^{th} quantile of $u_i(\tau)$ is zero.

2.3 Estimation

The least squares estimation technique of the conditional mean regression provides a framework for the estimation of quantile regression. Like the sample mean, sample quantiles can be obtained through a simple minimization problem. This optimization problem replaces the problem of sorting observations to find quantiles. This leads to a estimation technique similar to the least squares method.

We know the mean can be obtained by minimizing a squared error loss function. Similarly, quantiles can be obtained by minimizing an asymmetric version of the absolute error loss function. For the τ^{th} quantile, the loss function is defined as

$$\rho_\tau(u) = u(\tau - I_{\{u<0\}}), \tag{2.6}$$

where $I_{\{u<0\}}$ is an indicator function which takes value 1 when $u < 0$ and value 0 otherwise. The loss function in (2.6) is a piecewise linear function. This function is also known as check function. For $\tau = 0.5$, that is median, the check function becomes the symmetric absolute error loss function. The check function is illustrated in Figure 2.4 for $\tau = 0.5$ and $\tau = 0.75$.

A motivation to use this asymmetric loss function to obtain quantiles is as follows. Let X be a random variable with distribution function F(x). In order to find $\hat{x}$ to minimize the expected loss, we need to minimize

$$E\rho_\tau(X - \hat{x}) = (\tau - 1)\int_{-\infty}^{\hat{x}} (x - \hat{x})\mathrm{d}F(x) + \tau \int_{\hat{x}}^{\infty} (x - \hat{x})\mathrm{d}F(x).$$

Taking derivative with respect to $\hat{x}$ and setting it to 0, we have

$$(1 - \tau)\int_{-\infty}^{\hat{x}} \mathrm{d}F(x) - \tau \int_{\hat{x}}^{\infty} \mathrm{d}F(x) = 0,$$

$$\Rightarrow \int_{-\infty}^{\hat{x}} \mathrm{d}F(x) - \tau\{\int_{-\infty}^{\hat{x}} \mathrm{d}F(x) + \int_{\hat{x}}^{\infty} \mathrm{d}F(x)\} = 0,$$

$$\Rightarrow F(\hat{x}) - \tau = 0.$$

When the solution is unique, $\hat{x} = F^{-1}(\tau)$,
where $F^{-1}(\tau) = inf\{x : F(x) \geq \tau\}$.

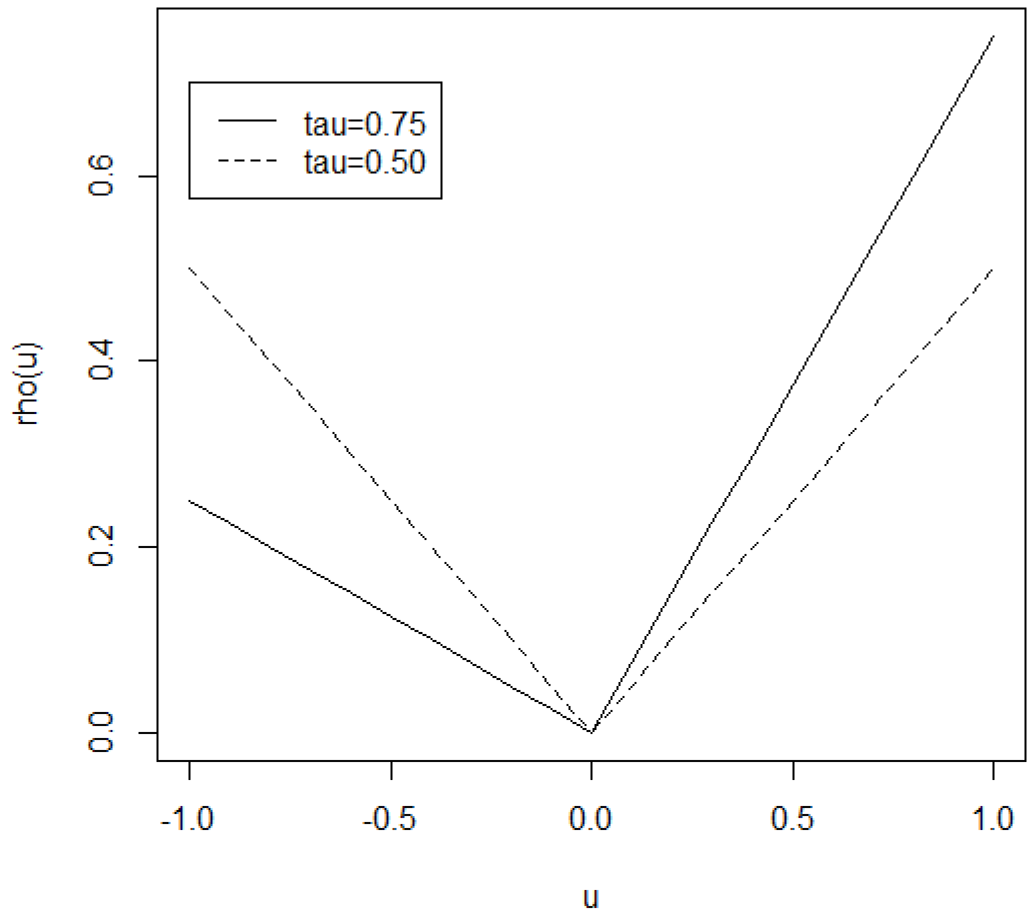

Figure 2.4: Check function

Now, we start looking through the least squares regression to get an insight into the estimation technique of the quantile regression.

As we know, the sample mean is obtained by solving

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} (y_i - \mu)^2.$$

If $x^T\beta$ is the conditional mean of y given x, then the above knowledge leads to estimate β by solving,

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2.$$

As median is obtained by minimizing the absolute error loss function,

$$\rho(u) = |u|,$$

sample median may be obtained by solving

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} |y_i - \mu|.$$

As quantiles are estimated by minimizing the loss function,

$$\rho_\tau(u) = u(\tau - I_{\{u<0\}}),$$

the τ^{th} sample quantile may be obtained by solving

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \mu).$$

Now, if the τ^{th} conditional quantile of y given x, $Q_y(\tau|x)$ is expressed as

$$Q_y(\tau|x) = x^T \beta(\tau),$$

then, $\beta(\tau)$ may be estimated by solving the problem

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \beta(\tau)).$$

The above problem can be transformed into the linear programming problem,

$$\min_{\beta(\tau),u,v} \{\tau 1_n^T u + (1-\tau) 1_n^T v | X\beta(\tau) + u - v = y\},$$

where X and y are $n \times p$ and $n \times 1$ matrices respectively, 1_n denotes an n-vector of 1, u and v correspond to the positive and negative parts of the residual vector $y - X\beta(\tau)$ respectively. The estimate of $\beta(\tau)$ can be obtained through solving the linear programming problem. We refer the reader to Koenker (2005) for details.

2.4 Bayesian Quantile Regression

Unlike the conventional approach, Bayesian inference gives the entire distribution of the parameter of interest. One can use Markov Chain Monte Carlo (MCMC) algorithm to obtain samples from a posterior distribution. The MCMC algorithm makes Bayesian inference easier and attractive. When posterior distribution contains high dimensional integral, one can avoid evaluating such integral by using MCMC. Bayesian quantile regression employs a likelihood function based on the asymmetric Laplace distribution. This idea is proposed by Yu and Moyeed (2001). The asymmetric Laplace distribution is a natural choice here as it is closely related to quantile regression.

2.4.1 Asymmetric Laplace Distribution

The asymmetric Laplace distribution is closely related to the quantile regression in a sense that the minimization of the check function defined by (2.6) is exactly the same as the maximization of a likelihood function formed by combining independent asymmetric Laplace densities. The probability density function of the Asymmetric Laplace distribution with location parameter $\mu \in (-\infty, \infty)$, scale parameter $\sigma \in (0, \infty)$ and skewness parameter $\tau \in (0, 1)$ is defined as

$$f(x; \mu, \sigma, \tau) = \frac{1}{\sigma}\tau(1-\tau)\exp\{-\frac{1}{\sigma}\rho_\tau(x-\mu)\},$$

where $\rho_\tau(u)$ is defined as

$$\rho_\tau(u) = u(\tau - I_{\{u<0\}}),$$

where $I_{\{u<0\}}$ is an indicator function which takes value 1 when $u < 0$ and value 0 otherwise. $\rho_\tau(u)$ defined above is the same as asymmetric loss function defined in (2.6). If a random variable X has an asymmetric Laplace distribution with location parameter μ, scale parameter σ and skewness parameter τ, we write $X \sim ALD(\mu, \sigma, \tau)$. For $\tau = 0.5$, the asymmetric Laplace distribution reduces to the symmetric Laplace distribution which is also known as the double exponential distribution. For $\tau > 0.5$, it is left skewed, and for $\tau < 0.5$, it is skewed to the right. Moreover, the τ^{th} quantile of X is μ, which is also the mode of the distribution. Figure 2.5 illustrates the density function of the symmetric and the asymmetric Laplace distribution for $\tau = 0.5$ and $\tau = 0.75$ respectively.

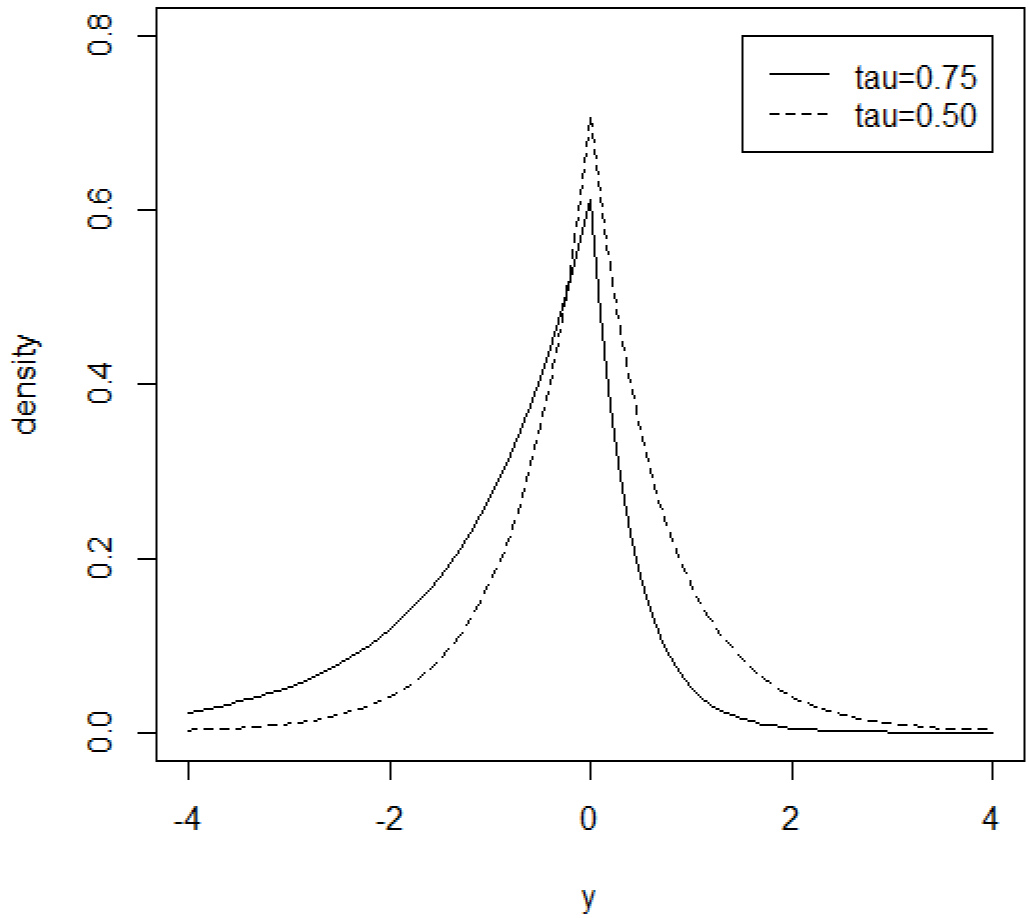

Figure 2.5: Asymmetric Laplace distribution

2.4.2 Asymmetric Laplace as a Mixture of Normal Distributions

The asymmetric Laplace distribution may be represented as a mixture of normal distributions (C. Reed and K. Yu, 2009) by the lemma given below. This representation makes data generation from an asymmetric Laplace distribution easier. We use this representation to simulate data from the asymmetric Laplace distribution in this project.

Lemma:

Let z_i be a standard normal variable, w_i be an exponential variable with mean σ and $x_i \sim ALD(\mu, \sigma, \tau)$. Then one can represent x_i as a scale mixture of normal given by

$$x_i \stackrel{d}{=} \sqrt{\frac{2w_i\sigma}{\tau(1-\tau)}}z_i + \frac{1-2\tau}{\tau(1-\tau)}w_i + \mu,$$

where, $\stackrel{d}{=}$ denotes equality in distribution.

Using the above lemma we may simulate data from the asymmetric Laplace distribution by using the following two steps.

1. Generate a random variable w_i from $Exponential(\sigma)$.
2. Generate data from $Normal(\frac{1-2\tau}{\tau(1-\tau)}w_i + x_i^T\beta, \frac{2w_i\sigma}{\tau(1-\tau)})$.

2.4.3 Bayesian Quantile Regression Using Hierarchical Representation of the Asymmetric Laplace Distribution

By using the lemma given in section 2.4.2, one may express Bayesian qunatile regression model as a hierarchical model.

Let, w_i's are latent weight independently and identically distributed as

$$w_i|\beta, \sigma, \tau \sim Exponential(\sigma).$$

Then,

$$y_i|w_i, \beta, \sigma, \tau \sim Normal(\frac{1-2\tau}{\tau(1-\tau)}w_i + x_i^T\beta, \frac{2w_i\sigma}{\tau(1-\tau)}).$$

It can be seen that the marginal distribution of y_i marginalized over the latent weight is $ALD(\mu, \sigma, \tau)$.

In the Bayesian quantile regression, posterior distribution is obtained by the MCMC method. This method allows one to choose any prior.

Chapter 3

Hierarchical Quantile Regression

3.1 Likelihood Based Inference of Quantile Regression

We recall from chapter 2, the linear quantile regression model is

$$y_i|x_i, \tau_i = \beta_0(\tau) + \beta_1(\tau)x_i + u_i(\tau),$$

where parameters are specified for τ^{th} quantile and they differ for different quantile levels. We may assume, $u_i(\tau) \sim ALD(0, \sigma, \tau), i = 1, 2, ..., n$ are independent. Thus, $u_i(\tau)$ has τ^{th} quantile zero. Hence, y_i conditional on x_i, for $i = 1, 2, ..., n$, are independently distributed as $ALD(\beta_0(\tau) + \beta_1(\tau)x_i, \sigma, \tau)$. The τ^{th} quantile of the conditional distribution of y_i is $\beta_0(\tau) + \beta_1(\tau)x_i$, which depends on τ. Thus, we actually assume different model for y_i for different τ.

Now, we consider the estimation of quantile regression based on the asymmetric Laplace distribution. We usually seek to estimate parameters for different quantile levels. For estimating parameters for different quantile levels, we assume different models for y_i. For instance, for $\tau = 0.5$, the distribution of y_i is symmetric and for $\tau = 0.95$, it is very left skewed. In reality, we must not have different model for y_i.

Yu and Moyeed (2001) considered a likelihood based on the asymmetric Laplace distribution to estimate quantile regression. They ignored the original distribution of y_i. Using Bayesian inference, they suggested that use of the asymmetric Laplace distribution irrespective of the original distribution of the data was quite satisfactory.

However, Yu and Moyeed and other papers on quantile regression do not explicitly explain the data generation mechanism underneath quantile regression. The goal of this thesis is to clarify the data generation mechanism of quantile regression through a hierarchical quantile regression model. We describe the hierarchical quantile regression model and its estimation procedure in the next section.

3.2 Hierarchical Quantile Regression Model

As discussed in the last section, in the likelihood based inference, one assumes different models for y_i for different quantile levels. We introduce a hierarchical quantile regression model that gives one a greater insight into quantile regression. The linear hierarchical quantile regression model is described by using the following steps.

Step 1:

$$y_i|x_i, \tau_i, \sigma, \beta_{0i}, \beta_{1i} = \beta_{0i} + \beta_{1i}x_i + u_i; u_i \sim ALD(0, \sigma, \tau_i)$$

The error u_i follows the asymmetric Laplace distribution with location 0, scale σ and skewness parameter τ_i. The intercept, β_{0i} and the slope, β_{1i} are random.

Step 2:

$$\beta_{0i}|\tau_i, a_1, b_1, \eta_1 \sim Normal(a_1 + b_1\tau_i, \eta_1^2)$$

$$\beta_{1i}|\tau_i, a_2, b_2, \eta_2 \sim Normal(a_2 + b_2\tau_i, \eta_2^2)$$

We assume β_{0i} and β_{1i} are linearly related to quantiles with normally distributed errors.

Step 3:

$$\tau_i \sim Uniform(0, 1)$$

Finally, we assume the quantiles are uniformly distributed between 0 and 1. One may assume other distributions that range from 0 to 1 for τ_i.

The above hierarchical model allows one to generate data from this model. Whereas, the model assumed by Yu and Moyeed (2001) does not allow data generation. It only provides estimation without much understanding about the underlying model. In the next section, we discuss the estimation procedure of the proposed hierarchical model.

3.3 Estimation of the Hierarchical Quantile Regression Model

The hierarchical quantile regression model defined in the previous section is

$$
\begin{gathered}
y_i|x_i, \tau_i, \sigma, \beta_{0i}, \beta_{1i} = \beta_{0i} + \beta_{1i}x_i + u_i; u_i \sim ALD(0, \sigma, \tau_i), \\
\beta_{0i}|\tau_i, a_1, b_1, \eta_1 \sim Normal(a_1 + b_1\tau_i, \eta_1^2), \\
\beta_{1i}|\tau_i, a_2, b_2, \eta_2 \sim Normal(a_2 + b_2\tau_i, \eta_2^2), \\
\tau_i \sim Uniform(0, 1).
\end{gathered}
$$

(3.1)

We estimate the model by using data cloning. The data cloning method introduced by Lele et al. (2007) is a very useful method for obtaining maximum likelihood (ML) estimates and their standard errors for complex hierarchical models. Using the Bayesian framework, the method utilizes Markov Chain Monte Carlo (MCMC) algorithm to produce frequentist inferences. Data cloning method removes the subjectivity of the Bayesian priors and it is computationally simple because of the use of the MCMC algorithm. One needs to construct a full Bayesian model with proper priors for the unknown parameters. Instead of using the likelihood function for the observed data, the data cloning method uses the likelihood function of the k copies (clones) of the data, where k is large. The method assumes that the copies are independent of each other. The posterior distribution is then obtained by MCMC. The mean of the posterior distribution equals the ML estimate of the unknown parameters and the k times variance of the posterior distribution equals the asymptotic variance of the ML estimate. We refer the reader to Lele et al. (2007, 2010) for details on the data cloning algorithm.

Like the Bayesian framework of the quantile regression model discussed in section 2.4.3, we use the representation of the asymmetric Laplace distribution as scale mixtures of normal distributions. By using the lemma (section 2.4.2), we can write the hierarchical quantile regression model (3.1) in the following fashion:

$$\begin{gathered} y_i|x_i, \tau_i, \sigma, \beta_{0i}, \beta_{1i} \sim Normal(\tfrac{1-2\tau}{\tau(1-\tau)}w_i + \beta_{0i} + \beta_{1i}x_i, \tfrac{2w_i\sigma}{\tau(1-\tau)}), \\ w_i|\tau_i, \sigma, \beta_{0i}, \beta_{1i} \sim Exponential(\sigma), \\ \beta_{0i}|\tau_i, a_1, b_1, \eta_1 \sim Normal(a_1 + b_1\tau_i, \eta_1^2), \\ \beta_{1i}|\tau_i, a_2, b_2, \eta_2 \sim Normal(a_2 + b_2\tau_i, \eta_2^2), \\ \tau_i \sim Uniform(0, 1). \end{gathered}$$

(3.2)

In order to estimate the model by data cloning, we need to assume priors for unknown parameters a_1, b_1, a_2, b_2, η_1, η_2 and σ. As the data cloning method does not depend on the choice of priors, we can make the priors moderately informative to get faster convergence of the data cloning algorithm.

3.4 Simulation Study and Results

We generated samples of size, $n = 100$ from the hierarchical quantile regression model (3.2) with $a_1 = 1$, $a_2 = 2$, $b_1 = 3$, $b_2 = 5$, $\eta_1 = 0.5$, $\eta_2 = 0.4$, $\sigma = 1$ and $x \sim Uniform(0, 80)$. A scatter plot of y vs x from a simulated data set is given in Figure 3.1.

The data cloning method is implemented in the statistical computing environment R using the package dclone. The following priors are used in the estimation of the model by data cloning:

$$a_1 \sim Normal(0, 10),$$

$$a_2 \sim Normal(0, 10),$$

$$b_1 \sim Normal(0, 10),$$

$$b_2 \sim Normal(0, 10),$$

$$log\eta_1 \sim Normal(0, 0.5),$$

$$log\eta_2 \sim Normal(0, 0.5),$$

$$log(1/\sigma) \sim Normal(0, 0.2).$$

As the parameter estimates obtained by the data cloning method are independent of priors, priors are made informative to make MCMC chains converge. We put prior on inverse-scale $(1/\sigma)$ because the convergence for inverse-scale $(1/\sigma)$ is better than that of scale (σ). For most of the repetitions, we used 100,000 iterations and we discarded the first 11,000 observations. However, for some repetitions, we used a few more iterations for the MCMC chains to

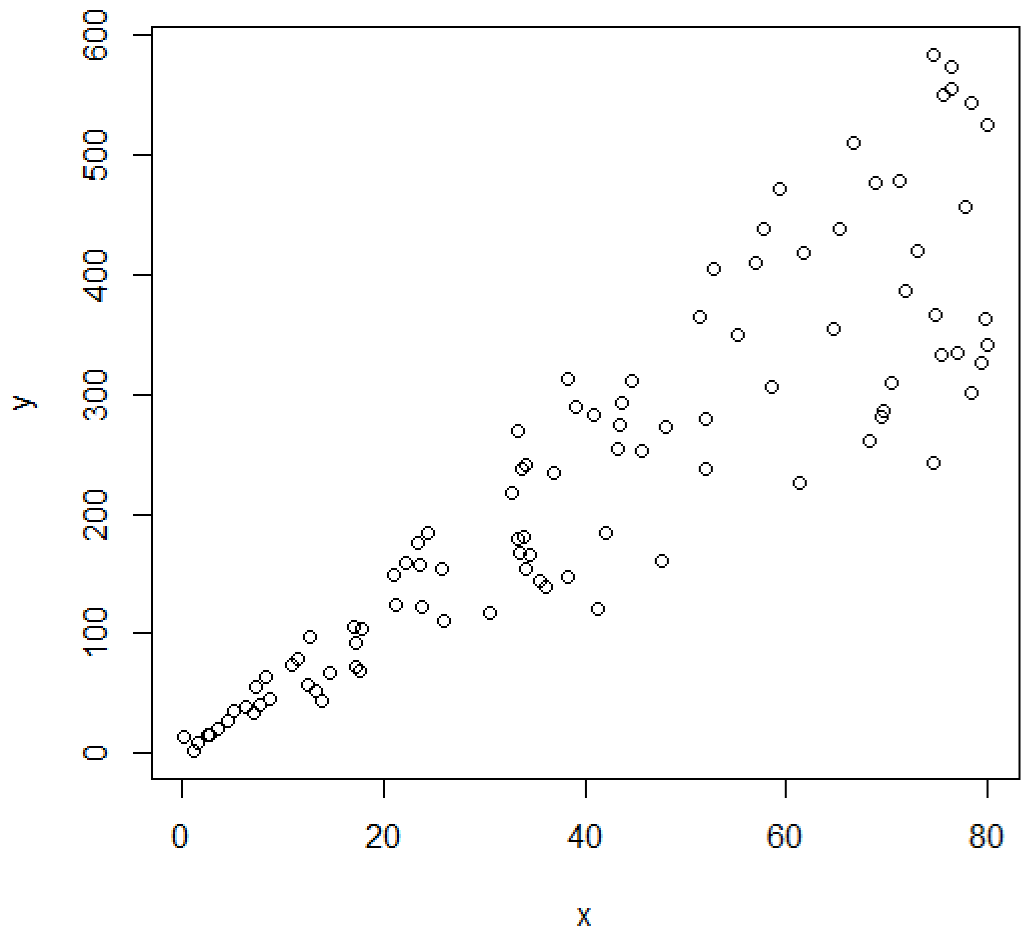

Figure 3.1: A scatter plot from a simulated data set from the hierarchical model

converge. We started with 2 clones. For the samples (repetitions) for which MCMC chains did not converge, we gradually increased the number of clones up to 10 to make the MCMC chains converge. Convergence of the MCMC chains was assessed by Gelman and Rubin diagnostics. $\hat{R}$ values for parameters a_1, a_2, b_1, b_2 and η_2 were less than 1.2 for all repetitions. However, $\hat{R}$ value was more than 1.2 (1.6) for η_1 for a repetition and $\hat{R}$ values were between 1.2 and 1.6 for $1/\sigma$ for 11 repetitions.

Table 3.1 compares the average of the parameter estimates to the true values. It shows lower and upper limits of 95% confidence intervals averaged over 50 repetitions. The coverage of the confidence intervals was 100% for a_1, b_1 and $1/\sigma$, 98% for b_2, η_1, and η_2, and 96% for a_2.

Apart from the parameter estimates, one is interested to assess the effect of a covariate on the conditional quantiles of y. In the hierarchical quantile regression model, $E(\beta_{1i}|\tau_i)$ serves this purpose. Table 3.2 represents the estimates of the expected random slopes, $E(\beta_{1i}|\tau_i) = a_2 + b_2\tau_i$, for different quantile levels, τ_i, with their 95% confidence intervals. $E(\beta_{1i}|\tau_i)$ is estimated by $\widehat{E(\beta_{1i}|\tau_i)} = \hat{a_2} + \hat{b_2}\tau$, where $\hat{a_2}$ and $\hat{b_2}$ are the Maximum Likelihood estimates (MLE) of a_2 and b_2 respectively obtained by the data cloning method. The standard error of $\widehat{E(\beta_{1i}|\tau_i)}$ is obtained by using the covariance matrix, which is provided by the data cloning method, of the model parameters. The entries of the third, fourth and fifth columns of the Table 3.2 are the average values over 50 samples (repetitions). The coverage of the 95% confidence intervals was 100% for all the quantile levels.

Table 3.3 exhibits the estimates of the slopes of the regular linear quantile regression model for different quantiles with their 95% confidence intervals. We compare the slope estimates of the regular linear quantile regression model to the expected value of the random slope, $E(\beta_{1i}|\tau_i)$, of the hierarchical linear quantile regression model for different quantiles. The entries of the tables are the average values over 50 samples (repetitions). The coverage of the 95% confidence intervals was approximately 60% in the upper (0.8 and 0.9) and lower (0.1 and 0.2) quantiles. It is approximately 80% between quantiles 0.3 and 0.7.

Table 3.1: Parameter Estimates by the Data Cloning Method (sample size, n=100)

parameter	true value	mean(estimate)	sd(estimate)	lower limit	upper limit
a1	1	1.37	1.85	−3.70	6.44
b1	2	0.30	1.90	−7.86	8.47
a2	3	3.17	0.55	2.28	4.06
b2	5	4.73	1.03	3.07	6.38
$\eta 1$	0.5	1.3	1.04	−1.28	3.90
$\eta 2$	0.4	0.46	0.15	0.09	0.83
$1/\sigma$	1	1.17	0.54	−0.03	2.38

Table 3.2: Estimate of $E(\beta_{1i}|\tau_i)$ at Different Quantiles by the Data Cloning Method (sample size, n=100)

quantiles	$E(\beta_{1i}\|\tau_i)$	estimate of $E(\beta_{1i}\|\tau_i)$	lower limit	upper limit
0.1	3.5	3.64	2.91	4.38
0.2	4.0	4.12	3.53	4.70
0.3	4.5	4.59	4.14	5.04
0.4	5.0	5.06	4.73	5.39
0.5	5.5	5.54	5.27	5.80
0.6	6.0	6.01	5.67	6.34
0.7	6.5	6.48	6.03	6.93
0.8	7.0	6.95	6.36	7.54
0.9	7.5	7.43	6.69	8.16

It is important to have similar results with different priors in the data cloning method. So, we repeated the study by using different priors. We used $Uniform(-5, 5)$ prior for a_1, b_1, a_2 and b_2, and $Uniform(0, 1)$ prior for η_1, η_2 and $1/\sigma$ and got the similar results.

Table 3.3: Regular Quantile Regression Estimates (sample size, n=100)

quantiles	expected random slopes of the hierarchical model, $E(\beta_{1i}\|\tau_i)$	estimate of the slope of regular quantile regression model	lower limit	upper limit
0.1	3.5	3.76	3.46	4.24
0.2	4.0	4.14	3.86	4.54
0.3	4.5	4.62	4.22	5.03
0.4	5.0	5.06	4.62	5.48
0.5	5.5	5.50	5.09	5.93
0.6	6.0	5.92	5.53	6.28
0.7	6.5	6.44	6.03	6.80
0.8	7.0	6.83	6.44	7.13
0.9	7.5	7.27	6.71	7.55

We also tried to estimate model (3.1) by the Bayesian approach. We used non-informative uniform priors. The MCMC chains does not converge well for the non-informative priors. We found that the chains converged only for 30% of the repetitions for most of the parameters. The convergence rate is worse for non-informative normal and log-normal priors.

3.5 Discussion

Table 3.1 shows that the estimates of the parameters associated with the random slope (a_2, b_2 and η_2) are close to the corresponding true values. Their confidence intervals are narrower compared to the intervals of the other parameters and coverage is very good. Convergence of the MCMC chains for these three parameters are also good. However, the estimates of the parameters associated with the random intercept (a_1, b_1 and η_1) are not quite close to their true values. Their confidence intervals are wider and contain zero. The point estimate of the inverse-scale ($1/\sigma$) is good, but its interval estimate is poor. There was convergence problem in some repetitions for $1/\sigma$. Often intercept of the linear model is not of scientific interest and may be treated as a nuisance parameter. The scale parameter of the asymmetric Laplace distribution in the hierarchical model may also be treated as nuisance. One focuses on estimating the coefficients of covariates. The point and interval estimates of all the parameters associated with the covariate by the data cloning method are quite good. So, our proposed hierarchical quantile regression model works well. However, the coverage of the confidence intervals are greater than the expectation. This may be due to the use of small number of clones than required in the model fitting using the data cloning algorithm. Data cloning diagnostics may be performed to determine the data cloning convergence and the appropriate number of clones.

From Table 3.2, we see that the average values of the estimated $E(\beta_{1i}|\tau_i)$ over repetitions are quite close to the true $E(\beta_{1i}|\tau_i)$ for different quantile levels. However, the coverage of their 95% confidence intervals is 100% that indicates large standard errors of the estimates.

We also compare the regular quantile regression estimates of the slope of the usual linear quantile regression model to the true $E(\beta_{1i}|\tau_i)$ for different quantiles. Table 3.3 shows that the regular quantile regression estimates are quite close to the corresponding values assumed by the hierarchical model on an average. Hence, the regular semi parametric quantile regression estimates the expected random slope of the hierarchical model. However, the confidence intervals based on the regular quantile regression does not cover the corresponding hierarchical model parameters well.

In conclusion, we are able to obtain the Maximum Likelihood estimates of all the important parameters of the hierarchical model by the data cloning method. Hence, our proposed hierarchical quantile regression model is a successful model. This model is a parametric quantile regression model that describes the data well and is estimable, whereas the available parametric quantile regression model focuses on the estimation ignoring the distribution of the data totally. The proposed model clearly shows the data generation mechanism in quantile regression.

3.6 Limitation

The major limitation of the proposed hierarchical model is that each additional covariate adds three more parameters in the model. So, the model may become very complex with the large number of covariates. One should balance between the complexity of the model and the information content in the data.

One may have problem with the convergence of the MCMC chains and may need to play with the number of clones, number of iterations and burn-in periods to make the chains converge. Data cloning diagnostics may be performed to assess the data cloning convergence.

Chapter 4

Future Research and Conclusion

4.1 Future Research

In this project, we make an effort to understand the quantile regression process. Our model produces a different picture of quantile regression that requires more attention and research. We propose a very simple hierarchical quantile regression model that assumes a linear relationship of the random intercept and slope with quantiles. One may modify the model assuming a non-linear relationship between the random effects and quantiles. Although the data cloning method does an excellent job in estimating the model, there is also a room to use different estimation techniques to improve results. This model may provide an excellent aid in measurement error problem in quantile regression. Censoring may also be easily incorporated in the model.

4.2 Conclusion

Quantile regression has emerged as a widely used method in many applications. However, the quantile regression model is not easy to understand. Assumption of the asymmetric Laplace distribution for the error term provides the likelihood based inference of quantile regression. Bayesian inference of quantile regression is based on this assumption. This approach focuses on estimation and completely fails to describe the underlying process of the quantile regression. However, a model should make an attempt to describe the true process. Our proposed hierarchical quantile regression model provides a greater insight into the quantile regression process. It also provides a tool for data generation to test the performance of Bayesian quantile regression as one cannot generate data under the Bayesian model.

Bibliography

Reed, C., & Yu, K. (2009). A Partially collapsed Gibbs sampler for Bayesian quantile regression. *Tech. Rep., Department of Mathematical Sciences, Brunel University, UK.*

Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1, 412-420.

Galton, F. (1885). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute*, 15, 246-263.

Koenker, R.(2005). *Quantile regression.* New York: Cambridge University Press.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.

Lele, S. R., Dennis, B., & Lutscher, F. (2007). Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10, 551-563.

Lele, S. R., Nadeem, K., & Schmuland, B. (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105, 1617-1625.

Plummer, M. (2013, August 30). JAGS version 3.4.0 user manual. Retrieved from http://sourceforge.net/projects/mcmc-jags/files/Manuals/3.x/

Sólymos, P. (2010). dclone: Data Cloning in R. *The R Journal*, 2, 29-37.

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

Yu, K., & Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics-Theory and Methods*, 34, 1867-1879.