

國立交通大學

統計學研究所

碩士論文

設限資料下分量迴歸模型分析之  
文獻回顧

**Quantile Regression Analysis  
based on Censored Data  
A Literature Review**

研究生：林欣穎

指導教授：王維菁 博士

中華民國 一 百 年 五 月

設限資料下分量迴歸模型分析之  
文獻回顧

**Quantile Regression Analysis based on Censored Data**  
**A Literature Review**

研究生：林欣穎

Student: Hsin-Ying Lin

指導教授：王維菁 博士

Advisor: Dr. Wei-Jing Wang

國立交通大學  
統計學研究所  
碩士論文



Submitted to Institute of Statistics  
College of Science

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistics

May 2011

Hsinchu, Taiwan, Republic of China

中華民國 一 百 年 五 月

# 設限資料下分量迴歸模型分析之 文獻回顧

研究生：林欣穎

指導教授：王維菁 博士

國立交通大學理學院

統計學研究所

## 摘要

在此論文中，我們回顧分量迴歸模型的重要文獻。首先，我們介紹在沒有變數之完整資料下，估計分量的推論技巧，藉此了解分量的幾何結構和分析上問題的困難處。接著，我們引進變數的影響和討論不同的估計程序，並提供幾何上的解釋意義。最後，我們加入設限的影響並且討論幾種修正的估計方法。我們提供系統化的架構，讓讀者透過推論方法基本的建構原則，對分量迴歸模型有初步的了解。

**關鍵詞：** 設限, 分量迴歸模型

# Quantile Regression Analysis based on Censored Data A Literature Review

Student: Hsin-Ying Lin

Advisor: Dr. Wei-Jing Wang

Institute of Statistics

National Chiao Tung University

Hsinchu, Taiwan



In the thesis, we review important literature on quantile regression models for survival data. First, we introduce the inference techniques for estimating a quantile based on complete data without covariates. This allows us to see the geometric structure and analytical difficulty of the problem. Then we include the effect of covariates and discuss different estimation procedures. Geometric explanations are also provided. Finally the effect of censoring is incorporated and we discuss several approaches of modification. We aim to provide a systematic framework which allows the readers to understand the quantile regression model from fundamental inference principles.

**Keywords:** Censoring, Quantile regression model

# 致 謝

隨著此篇論文的完成，我的學生生涯即進入尾聲，將要踏出校園，邁向未知的職場生活。在碩士這兩年的求學生涯，不僅學到統計方面的專業知識，更體會到人生中酸甜苦辣的滋味，我所經歷的每件事、所相處的每個人，都充實、豐盛了我的生命。

首先，最想感謝的是我的指導教授-王維菁老師，每次討論完論文後，對於老師的佩服又更加深一層。遵循老師的指導，我體驗到大體架構的重要性，若不了解文章整體結構為何，鑽研枝微末節也只是徒然，不僅點出我一直以來的問題，也訓練我先從大局上著手。另一個重要的是，大膽猜測並用自己的話語表達理解的事物，若沒辦法將所讀加以理解、吸收並用自己的意見表達出來，那不過是抄襲他人的想法罷了。在老師身邊，不只學到做學問的態度，也體悟到做人處事的道理，比起研究，我想這是更重要的，謝謝老師這一年來的教導。

接著，我要感謝的是郭姐和每一位教授熱切的關心和不時的小叮嚀，讓我在所上的日子備感溫馨；還有共同努力的同學們，在沮喪時帶來鼓勵、在開心時一同享樂、在作業堆裡攜手共度難關，謝謝你們。

林欣穎 謹誌

中華民國 一百年夏

于交通大學統計學研究所

# Contents

<b>1. Introduction.....</b>	<b>1</b>
1.1 Different types of regression models.....	1
1.2 Examples of quantile regression models.....	2
1.3 Comparison with other regression models.....	5
<b>2. Inference without Censoring.....</b>	<b>6</b>
2.1 Inference based on homogeneous data.....	6
2.2 Inference based on quantile regression analysis.....	7
2.3 Numerical algorithms.....	9
2.3.1 Minimizing the objective function.....	9
2.3.2 Solving the estimating function by simulated annealing...	12
<b>3. Inference under Right Censoring.....</b>	<b>15</b>
3.1 Data structure.....	15
3.2 Modification based on the objective function.....	15
3.3 Modification based on the estimating function.....	17
3.4 A new approach based on counting process.....	17
3.5 A new approach based on self-consistency algorithm.....	20
<b>4. Conclusion.....</b>	<b>22</b>
<b>References.....</b>	<b>32</b>
<b>Appendix.....</b>	<b>34</b>

# Chapter 1 Introduction

## 1.1 Different types of regression models

Consider the response variable  $Y$  and covariate vector  $Z$ . Regression analysis refers to the situation that one wants to model the behavior of  $Y$  based on  $Z$ . The following linear model is the most popular form:

$$Y = \alpha + Z^T \beta + \varepsilon. \quad (1.1)$$

The major interest is the estimation of  $\beta$  which measures the effect of  $Z$  and  $Y$ . To interpret the meaning of  $\beta$  and also develop valid inference methods, we need to impose additional assumptions on the distribution of  $\varepsilon$ . For example, if we assume that  $\varepsilon$  are identically and independently distributed (*iid*) with mean zero, the  $\alpha$  measures  $E(Y|Z=0)$  and the  $\beta$  measures the change of  $E(Y)$  when the corresponding covariate changes one unit.

However in survival analysis, the mean is often not a useful descriptive measure since it is not robust. Therefore other types of regression model are preferred for analyzing survival data. The Cox proportional hazards (PH) model is the most popular one. Denote  $T > 0$  as the lifetime variable of interest. A natural link for applying model (1.1) in survival analysis is to write  $Y = \log(T)$ . The PH model can be written as

$$\lambda_z(t) = \lambda_0(t) \exp\{Z^T \beta\} \quad (1.2)$$

where  $\lambda_z(t) = \lim_{\Delta \rightarrow 0} \Pr(T \in [t, t + \Delta) | T \geq t, Z) / \Delta$  is the hazard of  $T | Z$  and  $\lambda_0(t)$  is the hazard for the baseline group with  $Z = 0$ . Hence we have the accelerated failure time (AFT) model such that

$$Y = \log(T) = Z^T \beta + \varepsilon. \quad (1.3)$$

Notice that models (1.1) and (1.3) differ in whether an intercept term is included in the right-hand side. The difference comes from the fact that, in survival analysis, the

assumption  $E(\varepsilon) = 0$  is dropped and distribution-free assumption is imposed on  $\varepsilon$ . The resulting inference method will be rank-invariant which makes the intercept parameter to be non-identifiable.

The above discussions imply that the distribution form of  $\varepsilon$  plays a key role in regression analysis based on linear models. In the thesis, we will focus on quantile regression models. Define the quantile of  $Y$  as

$$Q_Y(\tau) = \inf\{y : \Pr(Y \leq y) \geq \tau\},$$

where  $0 < \tau < 1$ . Figure 1.1 shows the location of  $Q_Y(\tau)$  for a continuous random variable  $Y$ . Quantile regression models state that

$$Q_Y(\tau | Z) = \alpha(\tau) + Z^T \beta(\tau). \quad (1.4)$$

Note that the  $\alpha(\tau)$  is the quantile of  $Y$  for the baseline group with  $Z = 0$ . It is important to note that model (1.4) is equivalent to the linear model in (1.1) with the assumption that  $\Pr(\varepsilon \leq 0) = \tau$ . To see this, we have

$$\Pr(Y \leq Q_Y(\tau | Z)) = \Pr(\alpha(\tau) + Z^T \beta(\tau) + \varepsilon \leq Q_Y(\tau | Z)) = \Pr(\varepsilon \leq 0) = \tau.$$

We draw a plot based on the special two-sample case such that  $Q_Y(\tau | Z = 0) = \alpha(\tau)$  and  $Q_Y(\tau | Z = 1) = \alpha(\tau) + \beta(\tau)$ . In Figure 1.2, we see that the two samples differ more in the region at larger quantile. Note that we may imagine that when the two curves cross, the covariate effect may be reserved.

## 1.2 Examples of quantile regression models

The first three examples are originally described in the book of Keonker. In these examples, quantile regression models provide better explanations to the real-world phenomenon.

### *Example 1: Salaries and Experiences in Academia*

The American Statistical Association conducted a salary survey in 1995 on 370 full professors of statistics from 99 departments in U.S. colleges and universities. The



response is the professor's salary and the covariate is the year of experience (years in rank). Simply based on the scatterplot without imposing any model assumption (not presented here), the plot shows that (a) salary increases with experience; (b) the growth rate with the experience increase is different for different quantile group of salary. Specifically the first quantile (25<sup>th</sup> percentile) of the salary distribution has a 7.3% growth rate per year of tenure, whereas the median (50<sup>th</sup> percentile) and the third quantile (75<sup>th</sup> percentile) grow at 14% and 13% respectively. It seems to us that in American Academia, the full professors are treated different. A middle-paid one has the potential of achieving a leader position and hence gets the highest salary rate. A high-paid one who has been quite established has salary increase at about the same speed as the former. A relatively low-paid professor is probably the one who becomes less productive after he/she has been promoted to a full professor and hence gets the lowest rate of salary increase.

*Example 2: Score of Course Evaluation and Size of Class*

A university conducted course evaluation based on 1482 courses over the period 1980-94. The response is the mean course evaluation questionnaire (CEQ) score and the class size is the main covariate. The observations are classified into three categories (high, median and low quantile) based on their CEQ scores. It is found that larger classes tend to get lower CEQ score but the effect of class size is more significant on the lower quantile than on the upper quantile. In other words, high-evaluated courses may contain different sizes of classes but low-evaluated courses are significantly related to large classes.

*Example 3: Infant Birth Weight and Mother's Background*

The sample contains the birth weight and other covariate information for 198,377 infants. Comparing boys and girls, their difference is larger than 100 grams at the upper quantile, whereas smaller than 100 grams at the lower quantile. (Note boys tend

to be heavier.) Comparing married and unmarried mothers, their difference becomes more obvious in lower quantile. (Note the child of a married mother tend to be heavier.) Comparing white and black mothers, their difference is about 330 grams at the lower quantile, while 175 grams at the upper quantile. (Note the child of a white mother tends to be heavier.) The analysis implies that the difference between different race (white/black) or social (married/unmarried) groups is more obvious for babies in the lower quantile range.

*Example 4: Mortality risk for Dialysis Patients with Restless Legs Syndrome*

Peng and Huang (2008) did an analysis on a cohort of 191 renal dialysis patients from 26 dialysis facilities serving the 23-country area surrounding Atlanta, GA. The restless legs syndrome (RLS) is the main covariate and is classified into two levels of symptoms, mild RLS symptoms and severe RLS symptoms. The interest is to see the effect of RLS symptoms on mortality risk. Comparing the mild and severe RLS symptoms groups, their difference is about 1.5 years at the first quantile (25 percentiles of survival time), while there is no obvious difference at the third quantile (75 percentiles of survival time). In other words, there is a strong association between the RLS symptoms and mortality risk for short-term survivors. The phenomenon can not be detected by the ATF model.

*Example 5: Mortality rate for medflies with different Gender*

Koenker and Geling (2001) studied the relationship between mortality rate and gender on medflies, a study originally conducted by Garey, Liedo, Orozco and Vaupel (1992). According to the survival time, medflies are classed into three categories, the lower (before 20 days), middle (20-60 days) and upper (after 60 days) quantile. It found that males have lower mortality rate than females on the lower quantile, while males have higher mortality rate than females on the middle quantile and there is no obvious difference on the upper quantile.

### 1.3 Comparison with other regression models

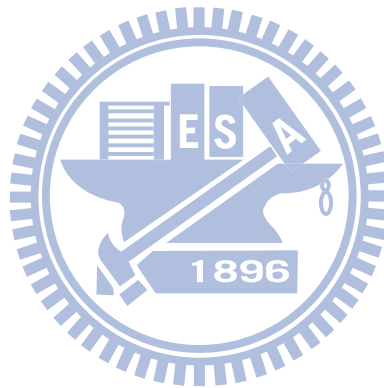
Based on the PH model in (1.2), we obtain

$$Q_T(\tau | Z) = \Lambda_0^{-1}(-\log(1 - \tau) \exp\{-Z^T \beta(\tau)\}),$$

which depends on the form of the baseline  $\Lambda_0(\cdot)$ . Furthermore,  $Q_T(\tau | Z)$  is monotone in  $-\log(1 - \tau)$  for all  $Z$ . This property restricts the application of PH models because it cannot handle the heterogeneity data. Based on the AFT model

$$\log T_i = Z_i^T \beta(\tau) + \varepsilon_i.$$

Usually, it is assumed that  $\varepsilon_i$  ( $i = 1, \dots, n$ ) have the same distribution which is independent of  $Z$ . This property can not well explain the heterogeneity data.



## Chapter 2 Inference without Censoring

### 2.1 Inference based on homogeneous data

We first review the simple case in which we have a random sample  $\{Y_i (i = 1, \dots, n)\}$  from  $Y$  such that the quantile of  $Y$  is of interest. That is, we temporarily ignore the effect of covariates. Recall that the  $\tau$  th quantile of  $Y$  is defined as

$$Q_Y(\tau) = F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\} \quad (0 < \tau < 1),$$

where  $F(y) = P(Y \leq y)$ . Now we discuss possible inference techniques for estimating  $Q_Y(\tau)$ . Intuitively we can use the empirical distribution of  $F(y)$  to find the empirical quantile. That is, we find  $\hat{Q}_Y(\tau)$  which satisfies

$$\sum_{i=1}^n I(Y_i \leq \hat{Q}_Y(\tau)) / n \approx \tau.$$

A more formal approach is to define the following loss function

$$\rho_\tau(u) = u \times (\tau - I(u < 0)) \tag{2.1}$$

which is depicted in Figure 2.1. Notice that although the loss function is continuous and linear, it is not differentiable at 0. For  $u > 0$  the slope is  $\tau$  and for  $u < 0$  the slope is  $\tau - 1$ . Now we aim to minimize

$$\begin{aligned} E[\rho_\tau(Y - \xi)] &= E[(Y - \xi) \cdot \{\tau - I(Y - \xi < 0)\}] \\ &= (\tau - 1) \int_{-\infty}^{\xi} (y - \xi) dF(y) + \tau \int_{\xi}^{\infty} (y - \xi) dF(y). \end{aligned}$$

We can differentiate it with respect to  $\xi$  and solve

$$U(\xi) = (1 - \tau) \int_{-\infty}^{\xi} dF(y) - \tau \int_{\xi}^{\infty} dF(y) = F(\xi) - \tau = 0.$$

This implies that the minimum is achieved when  $\tau = F^{-1}(\tau)$ . Given data  $\{Y_i (i = 1, \dots, n)\}$ , define

$$U_n(\xi) = F_n(\xi) - \tau = \sum_{i=1}^n I(Y_i \leq \xi) / n - \tau. \tag{2.2}$$

Hence the solution to  $U_n(\xi) = 0$  or the minimum of

$$R_n(\xi) = \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - \xi) = \frac{1}{n} \sum_{i=1}^n (Y_i - \xi)(\tau - I(Y_i - \xi < 0)) \quad (2.3)$$

is attained at  $\hat{\xi} = Q_Y(\tau)$ .

Now we plot  $U_n(\xi)$  and  $R_n(\xi)$  for  $Y_i \sim^{iid} N(0,1)$  ( $i=1, \dots, 10$ ). These data points are -2.002, -0.914, -0.526, -0.446, -0.052, 0.120, 0.386, 0.647, 1.379, 1.705. Note that  $\Phi^{-1}(0.3) = -0.524$ ,  $\Phi^{-1}(0.5) = 0$  and  $\Phi^{-1}(0.7) = 0.524$ . In Figure 2.2, we present two plots: (a)  $\xi$  versus  $U_n(\xi)$  and (b)  $\xi$  versus  $R_n(\xi)$  based on  $\rho_\tau(u) = u \times (\tau - I(u < 0))$  with  $\tau = 0.3, 0.5, 0.7$ . Our first purpose is to see the shapes of the two functions and whether the solution to  $U_n(\xi) = 0$  or the minimum of  $R_n(\xi)$  are closed to  $\Phi^{-1}(0.3)$ ,  $\Phi^{-1}(0.5)$  and  $\Phi^{-1}(0.7)$  respectively. In Figure 2.2, we see that is a monotone step function and can be well approximated by a convex function. These nice properties are useful for numerical computations and large-sample analysis.

## 2.2 Inference based on quantile regression analysis

In presence of covariates, the quantile regression model can be written as

$$Q_Y(\tau | Z) = \alpha(\tau) + Z^T \beta(\tau). \quad (2.4)$$

Give data  $\{(Y_i, Z_i) (i = 1, \dots, n)\}$ , the objective function can be written as

$$\begin{aligned} R_n(\alpha(\tau), \beta(\tau)) &= \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - \alpha(\tau) - Z_i^T \beta(\tau)) \\ &= \frac{1}{n} \sum_{i=1}^n (Y_i - \alpha(\tau) - Z_i^T \beta(\tau)) (\tau - I(Y_i - \alpha(\tau) - Z_i^T \beta(\tau) < 0)). \end{aligned} \quad (2.5)$$

The major purpose is to minimize  $R_n$ . For illustration, we may first review the problem based on minimizing the squared loss function such that

$$\tilde{R}_n(\alpha, \beta) = \sum_{i=1}^n (Y_i - \alpha - Z_i^T \beta)^2.$$

By taking the derivative with respect to  $\alpha$  and  $\beta$ , we obtain the following

estimating function

$$\tilde{U}_n(\alpha, \beta) = \begin{pmatrix} \tilde{U}_n^{(1)} \\ \tilde{U}_n^{(2)} \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} 1 \\ Z_i \end{pmatrix} (Y_i - \alpha - Z_i^T \beta). \quad (2.6)$$

Notice that  $\tilde{U}_n^{(k)}(\alpha, \beta)$  ( $k=1,2$ ) are both smooth linear planes. In Figure 2.3, we plot  $(\alpha, \beta, \tilde{R}_n)$  and  $(\alpha, \beta, \tilde{U}_n^{(k)})$  ( $k=1,2$ ) based on 10 simulated data points from a simple regression model.

Now return to the quantile estimation problem. Although  $R_n$  is not differentiable, we can still derive the directional derivative. Specifically the directional derivative of  $R_n$  in direction  $w$  with  $w \in R^p$  and  $\|w\|=1$  can be obtained as follows:

$$\begin{aligned} \nabla R_n(\tilde{\beta}(\tau), w) &\equiv \left. \frac{d}{dt} R_n(\tilde{\beta}(\tau) + tw) \right|_{t=0} \\ &= \left. \frac{d}{dt} \sum_{i=1}^n \rho_\tau(Y_i - \tilde{Z}_i^T(\tilde{\beta}(\tau) + tw)) \right|_{t=0} \\ &= \left. \frac{d}{dt} \sum_{i=1}^n (Y_i - \tilde{Z}_i^T \tilde{\beta}(\tau) - \tilde{Z}_i^T tw) \left\{ \tau - I(Y_i - \tilde{Z}_i^T \tilde{\beta}(\tau) - \tilde{Z}_i^T tw < 0) \right\} \right|_{t=0} \\ &= \sum_{i=1}^n (-\tilde{Z}_i^T w) (\tau - I(Y_i - \tilde{Z}_i^T \tilde{\beta}(\tau) < 0)) \\ &= \sum_{i=1}^n (-\tilde{Z}_i^T w) (\tau - (1 - I(Y_i - \tilde{Z}_i^T \tilde{\beta}(\tau) \geq 0))), \end{aligned}$$

where  $\tilde{Z}_i = (1, Z_i)^T$  and  $\tilde{\beta}(\tau) = (\alpha(\tau), \beta(\tau))^T$ . In general, we get the estimating function

$$U_n(\alpha(\tau), \beta(\tau)) = \begin{pmatrix} U_n^{(1)} \\ U_n^{(2)} \end{pmatrix} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} 1 \\ Z_i \end{pmatrix} (I(Y_i - \alpha(\tau) - Z_i^T \beta(\tau) \geq 0) - (1 - \tau)). \quad (2.7)$$

This implies that  $\alpha(\tau)$  and  $\beta(\tau)$  can be estimated by either minimizing  $R_n$  or solving  $U_n = 0$ .

Now we generate 10 data points from the model

$$Q_{Y_i}(\tau | Z_i) = \alpha(\tau) + \beta(\tau)Z_i \quad (i = 1, \dots, 10)$$

where  $\alpha(\tau)$  is the  $\tau$ -quantile from a standard normal distribution and  $\beta(\tau) = 3.5\tau$ . We will use these points to show several 3-D plots in Figure 2.4 and 2.5. Then we set  $\tau = 0.3$  and plot  $(\alpha(\tau), \beta(\tau), R_n)$  and  $(\alpha(\tau), \beta(\tau), U_n^{(k)})$  where  $k = 1, 2$ ,  $\alpha(\tau)$  is  $\Phi^{-1}(\tau)$  and there are two forms of  $Z$ , namely Binomial( $n, 1/2$ ) and Uniform(0,1).

Comparing Figures 2.4 and 2.5 with Figure 2.3, we see that the estimating functions for the former may not take value at 0. As a result, the numerical algorithms to implement the estimation need special techniques.

### 2.3 Numerical algorithms

To simplify the notations, from now on we let  $Z$  and  $\beta(\tau)$  include the intercept term. Numerical algorithms are needed to find the estimator of  $\beta(\tau)$  which minimizes  $R_n(\beta(\tau))$  or solves  $U_n(\beta(\tau)) \approx 0$ . Both problems are non-trivial since the functions are not differentiable so that the powerful Taylor-expansion technique cannot be directly applied.

#### 2.3.1 Minimizing the objective function

Before introducing the algorithm, it is worthy to discuss the geometric structure behind the minimization problem. Consider the subspace spanned by  $\mathbf{Z}_j^T = (Z_{j1}, \dots, Z_{jn})^T$  for  $j = 1, \dots, p$  and let  $\mathbf{Y}^T = (Y_1, \dots, Y_n)^T$  denote the center. The mechanism of the squared loss function behaves like a ball which inflates until the ball touches the surface of the above subspace. Now consider the problem of minimizing

$$R_n(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - Z_i^T \beta(\tau)) = \frac{1}{n} \sum_{i=1}^n (Y_i - Z_i^T \beta(\tau)) (\tau - I(Y_i - Z_i^T \beta(\tau) < 0)).$$

The mechanism behaves like a polyhedral diamond, also centered at  $\mathbf{Y}$ , expands until it touches the subspace.

Now we compare the solutions based on minimizing  $\tilde{R}_n(\beta)$  and  $R_n(\beta(\tau))$  respectively. Since  $\tilde{R}_n(\beta)$  is differentiable, the solution is obtained by solving  $\partial\tilde{R}_n(\beta)/\partial\beta=0$  equivalent to  $\mathbf{Z}^T(\mathbf{Y}-\mathbf{Z}^T\beta)=0$ . From a geometric viewpoint, it implies that the columns of  $\mathbf{Z}$  are perpendicular to the error vector. Usually the estimator is obtained by solving the linear equations of  $\beta$  analytically. On the other hand, the directional derivatives of  $R_n(\beta(\tau))$  is given by

$$\nabla R_n(\beta(\tau), w) = \sum_{i=1}^n (-Z_i^T w) \left( \tau - (1 - I(Y_i - Z_i^T \beta(\tau) \geq 0)) \right).$$

Geometrically moving away from  $\hat{\beta}(\tau)$  will increase  $R_n(\beta(\tau))$ . A solution which minimizes  $R_n(\beta(\tau))$  subject to  $\nabla R_n(\beta(\tau), w) \geq 0$  will only occur at vertex points. This implies that we only need to examine over all vertex points about which one satisfies the constraint. Analytically, the solution which minimizes  $R_n(\beta(\tau))$  denoted as  $\hat{\beta}(\tau)$  must satisfy  $\nabla R_n(\hat{\beta}(\tau), w) \geq 0$  for all  $w \in R^p$ .

Using the terminology of linear programming, the candidate solutions are called as the “basic solutions”. A nice feature of these basic solutions is that they can be represented by a linear combination of  $p$ -component sub-matrix of  $\mathbf{Z}$  and sub-vector of  $\mathbf{Y}$ . Specifically let  $h \in H$  be the set with each element containing  $p$ -numbers from  $\{1, 2, \dots, n\}$  and hence  $H$  consists of  $\binom{n}{p}$  members. Denote  $\mathbf{Z}(h): p \times p$  as the sub-matrix of  $\mathbf{Z}$  with rows  $(Z_{i_1}, \dots, Z_{i_p})$  for  $i \in h$  and  $\mathbf{Y}(h): p \times 1$  as the sub-vector of  $\mathbf{Y}$  with elements  $Y_i$  for  $i \in h$ . The basic solution denoted as  $\beta_h(\tau)$  satisfies

$$\beta_h(\tau) = \mathbf{Z}(h)^{-1} \mathbf{Y}(h).$$

Hence we only need to check whether



$$\nabla R_n(\beta_h(\tau), w) = \sum_{i=1}^n (-Z_i^T w) \left( \tau - \left( 1 - I(Y_i - Z_i^T \beta_h(\tau) \geq 0) \right) \right) \geq 0$$

for all  $w \in R^p$ . Recall that there are  $\binom{n}{p}$  candidates of  $h$  and how to examine the

above inequality in all dimensions analytically is not straightforward.

The following equivalent expressions are useful for developing the numerical algorithm:

$$\nabla R_n(\beta_h(\tau), v) = -\sum_{i \in h} (\tau - I(-v_i < 0)) v_i - \zeta^T(h) v \quad (2.8)$$

where  $v = \mathbf{Z}(h)w \in R^p$ ,  $\zeta^T(h) = \sum_{i \notin h} \psi_\tau^*(Y_i - Z_i^T \beta_h(\tau), -Z_i^T \mathbf{Z}(h)^{-1} v) Z_i^T \mathbf{Z}(h)^{-1}$  and

$$\psi_\tau^*(u_1, u_2) = \begin{cases} \tau - I(u_1 < 0) & \text{if } u_1 \neq 0 \\ \tau - I(u_2 < 0) & \text{if } u_1 = 0 \end{cases}$$

It suffices to check whether  $\nabla R_n(\beta_h(\tau), v) \geq 0$  for all  $v = \pm e_k$  ( $k = 1, \dots, p$ ) which is equivalent to the following condition

$$-\tau \mathbf{1}_p \leq \zeta(h) \leq (1 - \tau) \mathbf{1}_p.$$

The above analysis simplifies the complicated minimization problem to the analytic objective of finding  $\beta_h(\tau) = \mathbf{Z}(h)^{-1} \mathbf{Y}(h)$  which satisfies  $-\tau \mathbf{1}_p \leq \zeta(h) \leq (1 - \tau) \mathbf{1}_p$ .

This is equivalent to

$$(\tau - 1) \mathbf{1}_p \leq \sum_{i \in h} \left( \frac{1}{2} - \frac{1}{2} \text{sign}(Y_i - Z_i^T \beta_h(\tau)) - \tau \right) Z_i \mathbf{Z}(h)^{-1} \leq \tau \mathbf{1}_p. \quad (2.9)$$

Thus we have  $2p$  inequalities in  $\tau$  of the form which can be viewed as

$$(\tau - 1) \leq a_j + b_j \tau \leq \tau \quad (j = 1, \dots, p) \quad (2.10)$$

where  $(a_j, b_j)$  can be determined given the data and the basic solution. Notice that the above inequalities can be used to (a) find an appropriate  $\beta_h(\tau)$  from all the vertex points given the value of  $\tau$  or (b) given a vertex point, to find the corresponding value (or range) of  $\tau$ .

Now we briefly describe the algorithm. We can start from  $\tau_0 = 1/n$  which corresponds to the basic solution say  $\beta_{h_0}(\tau) = \mathbf{Z}(h_0)^{-1}\mathbf{Y}(h_0)$  by checking the  $2p$  inequalities from all the  $\binom{n}{p}$  candidates of basic solutions. Then we find  $\tau_1$  which corresponds to the basic solution  $\beta_{h_1}(\tau) = \mathbf{Z}(h_1)^{-1}\mathbf{Y}(h_1)$  where  $h_0$  and  $h_1$  differ in one unit. It seems to us that the algorithm is to use the basic solutions  $\beta_h(\tau) = \mathbf{Z}(h)^{-1}\mathbf{Y}(h)$  to find the corresponding value  $\tau$  satisfying

$$\frac{a_j}{(1-b_j)} \leq \tau \leq \frac{1+a_j}{(1-b_j)} \quad (j = 1, \dots, p).$$

This algorithm only needs to perform a thorough search at the first time based on  $\tau_0$ . Then only a small variation in the basic solution is involved which changes the values of  $(a_j, b_j)$  which are used to find the next  $\tau_1$ . The procedure can go on until the targeted largest  $\tau$  is found.

### 2.3.2 Solving the estimating function by simulated annealing

Recall that

$$U_n(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n Z_i \left( I(Y_i - Z_i^T \beta(\tau) \geq 0) - (1 - \tau) \right).$$

Since  $U_n(\beta(\tau))$  is a step function which takes discrete values, it may happens that there exists no solution to  $U_n(\beta(\tau)) = 0$ . Although the solution can be defined as the value where the estimating function changes signs, an algorithm to find its location based on the data is still needed. One way of finding the solution is to transfer the root-solving problem into another minimization problem. Define  $\hat{\beta}(\tau)$  such that  $\|U_n(\hat{\beta}(\tau))\|$  is minimized where  $\|U_n(\beta(\tau))\|$  is the sum of the absolute values of the  $p$ -components of  $U_n(\beta(\tau))$ . Lin and Geyer (1992) suggested the *simulated annealing* algorithm originally developed for a similar numerical problem in semi-parametric AFT regression analysis. Here we briefly describe this approach

under the context of quantile regression models.

Consider there are an old candidate  $\beta^*(\tau)$  and a new one  $\beta^{**}(\tau)$ . The old solution  $\beta^*(\tau)$  will be replaced by  $\beta^{**}(\tau)$  “for sure” if  $D = \|U_n(\beta^{**}(\tau))\| - \|U_n(\beta^*(\tau))\| \leq 0$ . This step is reasonable since obviously  $U_n(\beta^{**}(\tau))$  is closer to zero. If  $D > 0$ , the chance of  $\beta^*(\tau)$  being replaced by  $\beta^{**}(\tau)$  is  $\exp(-D/c)$ . This means that the algorithm allows for an intermediate solution to move “farer away” from zero with the probability  $\exp(-D/c)$ . We guess that the reason is to avoid the algorithm gets stuck in a local minimum point. Now the next question is how to come up with a new candidate  $\beta^{**}(\tau)$ ? It is suggested that  $\beta^{**}(\tau)$  has a (multivariate) normal random distribution with mean  $\beta^*(\tau)$  and a diagonal covariance matrix. This step implies that the position of the new  $\beta^{**}(\tau)$  is related to the old one  $\beta^*(\tau)$ . Now we may ask what’s the principle for setting up the covariance matrix? It is suggested that, when the iterations continue, the variance of each component of  $\beta^{**}(\tau)$  and the value of  $c > 0$  should decrease. The former implies that  $\beta^{**}(\tau)$  gets closer to  $\beta^*(\tau)$ . The latter implies that if  $\beta^{**}(\tau)$  does not reduce  $\|U_n(\beta(\tau))\|$ , the chance that it replaces  $\beta^*(\tau)$  gets smaller in the iterations.

The simulated annealing algorithm can be summarized below.

- a. Give the initial values of  $\beta^{(0)}(\tau)$ ,  $c^{(0)} > 0$  and  $\Sigma^{(0)}$  which is a diagonal matrix;
- b. Generate a new  $\beta^{(1)}(\tau)$  which follows  $N(\beta^{(0)}(\tau), \Sigma^{(0)})$ ;
- c. If  $D = \|U_n(\beta^{(1)}(\tau))\| - \|U_n(\beta^{(0)}(\tau))\| \leq 0$ , then replace  $\beta^{(0)}(\tau)$  by  $\beta^{(1)}(\tau)$  and otherwise replace  $\beta^{(0)}(\tau)$  by  $\beta^{(1)}(\tau)$  with the probability  $\exp(-D/c^{(0)})$ ;
- d. Change the diagonal components of  $\Sigma^{(0)}$  and  $c^{(0)}$  to smaller numbers denoted as  $\Sigma^{(k)}$  and  $c^{(k)}$  (say  $c^{(k)} = \rho c^{(k-1)}$ ) respectively;

e. Repeat Steps (b) ~ (d) for  $k = 1, \dots, m$  where  $m$  is a fixed value.

As for the initial values, Lin and Geyer (1992) made the suggestion: set  $\beta^{(0)}(\tau)$  to zero; set the number of steps  $m = 1000p$ , where  $p$  is the dimension of  $\beta(\tau)$ ; set “the cooling rate”  $\rho$  for  $\rho^{m/2} \approx 0.0005$ ; set initial variance as 0.1 and the cooling rate for the variances such that the diagonal components of  $\Sigma^{(m)}$  is about 0.0005. The estimator is  $\hat{\beta}(\tau) = \beta^{(m)}(\tau)$  which hopefully will get to the minimum of  $\|U_n(\beta(\tau))\|$ .

Now based on the simulated example, we present the second component of  $\|U_n(\beta(\tau))\|$  and the movement of the algorithm. Figure 2.6 represents the cases of different types of  $Z$ .

Now we present the intermediate steps of the algorithm in Figure 2.7. We see that the points still can jump up but finally the points come to the bottom and will not jump up anymore (convergence is reached). From our experience, we found that the choice of  $c^{(0)}$  is influential. When this value is too large, the probability of jumping up becomes bigger while if it is small, there is higher chance that it remains unmoved.

We set  $c^{(0)} = \|U_n(\beta^{(0)}(\tau))\| - 0.001$  and also set the initial  $\Sigma^{(0)} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$ .

## Chapter 3 Inference under Right Censoring

### 3.1 Data structure

Now we discuss quantile regression analysis based on survival data. Here the variable of interest is the time to the occurrence of an event of interest, denoted as  $T$  which is subject to censoring by  $\tilde{C}$ . Let  $Y = \log(T)$  and  $C = \log(\tilde{C})$ . The quantile regression model can be written as

$$Q_Y(\tau | Z) = Z^T \beta(\tau) \quad (3.1a)$$

or, based on the original failure time, such that

$$Q_T(\tau | Z) = \exp\{Z^T \beta(\tau)\}. \quad (3.1b)$$

Note that here  $Z$  and  $\beta(\tau)$  include the intercept term. Observed variables become  $(X, \delta, Z)$  where  $X = Y \wedge C$  and  $\delta = I(T \leq \tilde{C})$ . The main purpose is to estimate  $\beta(\tau)$  based on data  $(X_i, \delta_i, Z_i)$  ( $i=1, \dots, n$ ) which is a random sample from  $(X, \delta, Z)$ .

Several approaches to modifying the methods under censoring have been proposed. Here we summarize their main ideas and comment on their properties including drawbacks.

### 3.2 Modification based on the objective function

The paper by Powell (1986) was motivated by economic examples. The original construction is based on the situation that  $Y_i$  is subject to left censoring by 0, a common phenomenon in economic applications. The idea can be easily modified for the situation that  $Y_i$  is subject to right censoring by a fixed value  $C_i$ . He used the idea that minimizing a loss function based on  $Y_i - Z_i^T \beta(\tau)$  is equivalent to minimizing the same loss function based on  $\{Y_i \wedge C_i\} - \{Z_i^T \beta(\tau) \wedge C_i\}$ . Accordingly the modified objective function can be written as

$$R_n^P(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n \rho_\tau(X_i - \min\{C_i, Z_i^T \beta(\tau)\}). \quad (3.2)$$

The corresponding estimating function becomes

$$U_n^P(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n Z_i \left( \tau - I(X_i - Z_i^T \beta(\tau) < 0) \right) \cdot I(Z_i^T \beta(\tau) < C_i).$$

An obvious drawback of this approach is that  $C_i$  may be a random variable and also subject to censoring. Furthermore  $R_n^P(\beta(\tau))$  in (3.2) is no longer a convex function and needs not be unimodal which creates numerical problems. To find the global minimum, excessive computation time is required.

To handle the censoring problem of  $C_i$ , Honoré, Khan and Powell (2002) proposed to replace the loss function by its conditional expectation given  $(X_i, \delta_i, Z_i)$  in (3.2). That is, the objective function can be written as

$$R_n^{HP}(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n E \left[ \rho_\tau(X_i - \min\{Z_i^T \beta(\tau), C_i\}) \mid (X_i, \delta_i, Z_i) \right].$$

Accordingly

$$E[\rho_\tau(X_i - \min\{Z_i^T \beta(\tau), C_i\}) \mid X_i, \delta_i = 0, Z_i] = (1 - \delta_i) \rho_\tau(X_i - \min\{Z_i^T \beta(\tau), X_i\})$$

and

$$\begin{aligned} & E[\rho_\tau(X_i - \min\{Z_i^T \beta(\tau), C_i\}) \mid X_i, \delta_i = 1, Z_i] \\ &= \frac{\delta_i \int \rho_\tau(X_i - \min\{Z_i^T \beta(\tau), c\}) \cdot I(X_i < c) dG(c)}{G(X_i)} \end{aligned}$$

where  $G(t) = \Pr(C > t)$  can be estimated by the Kaplan-Meier estimator

$$\hat{G}(t) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(X_i = u, \delta = 1_i)}{\sum_{i=1}^n I(X_i \geq u)} \right\}.$$

Notice that the above derivation requires that  $G(t)$  does not depend on  $Z$ . The corresponding estimating function becomes

$$U_n^{HP}(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n Z_i \left( \tau \cdot I(X_i - Z_i^T \beta(\tau) \geq 0) - \frac{\delta_i \cdot (1-\tau) \cdot I(X_i - Z_i^T \beta(\tau) < 0) \hat{G}(Z_i^T \beta(\tau))}{\hat{G}(X_i)} \right).$$

### 3.3 Modification based on the estimating function

Ying et al. (1995) proposed to directly modify the estimating function

$$U_n(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n Z_i \left( I(Y_i - Z_i^T \beta(\tau) \geq 0) - (1-\tau) \right).$$

In presence of censoring,  $I(Y_i - Z_i^T \beta(\tau) \geq 0)$  is not observable. Instead we observe  $I(X_i - Z_i^T \beta(\tau) \geq 0)$ . Under the assumption that  $T_i$  and  $C_i$  are independent, it follows that

$$E \left[ I(X_i - Z_i^T \beta_0(\tau) \geq 0) \right] = (1-\tau) G(Z_i^T \beta_0(\tau))$$

where  $\beta_0(\tau)$  is the true value of  $\beta(\tau)$  and  $G(t) = \Pr(C > t)$  which can be estimated by the Kaplan-Meier estimator. Applying the technique of “inverse-probability-of-censoring-weighting”, they proposed the modified estimating function

$$U_n^{YJW}(\beta(\tau)) = \frac{1}{n} \sum_{i=1}^n Z_i \left( \frac{I(X_i - Z_i^T \beta(\tau) \geq 0)}{\hat{G}(Z_i^T \beta(\tau))} - (1-\tau) \right) \quad (3.2)$$

where  $I(X_i - Z_i^T \beta(\tau) \geq 0) / \hat{G}(Z_i^T \beta(\tau))$  is zero if  $\hat{G}(Z_i^T \beta(\tau))$  is zero. The unconditional independence between  $T_i$  and  $C_i$  may be too restrictive and also the weighting approach becomes invalid if the censoring support is shorter than the support of the failure variable.

### 3.4 A new approach based on counting process

The framework of counting process is useful for analyzing survival data in presence of censoring. Peng and Huang (2008) proposed a novel approach based on counting process to estimate the quantile regression parameter. We try to find a

motivation for their idea. Define  $F_T(t) = \Pr(T \leq t)$  and the corresponding cumulative hazard function can be written as  $\Lambda_T(t) = -\log\{1 - F_T(t)\}$ . Consider the counting process  $N(t) = I(T \leq t)$ . Under the standard setting, one often use the property that

$$M(t) = N(t) - \int_0^t I(T \geq s) d\Lambda_T(s) \quad (3.3)$$

is a mean-zero martingale when  $\Lambda_T(\cdot)$  is the true function. One major advantage of the above decomposition is that  $N(t)$  and  $I(T \geq s)$  can easily be modified under right censoring.

Notice that (3.3) holds when  $t = Q_T(\tau)$ . That is,

$$M(Q_T(\tau)) = N(Q_T(\tau)) - \int_0^{Q_T(\tau)} I(T \geq s) d\Lambda_T(s)$$

which has a mean-zero property. The key result of Peng and Huang (2008) is the re-expression of the cumulative hazard function by substituting the range of integration from the scale of time to the scale of probability. Notice that

$$\Lambda_T(Q_T(\tau)) = \int_0^{Q_T(\tau)} d\Lambda_T(t) = \int_0^{F_T^{-1}(\tau)} d\Lambda_T(t) = \int_0^\tau d\Lambda_T \circ F_T^{-1}(u).$$

Furthermore define

$$H(u) = \Lambda_T \circ F_T^{-1}(u) = -\log(1 - F_T \circ F_T^{-1}(u)) = -\log(1 - u).$$

Thus

$$\Lambda_T(Q_T(\tau)) = \int_0^{Q_T(\tau)} d\Lambda_T(t) = \int_0^{F_T^{-1}(\tau)} d\Lambda_T(t) = \int_0^\tau dH(u).$$

In the paper, the original model assumption based on  $Y_i = \log T_i$  with

$$Q_{Y_i}(\tau | Z_i) = Z_i^T \beta(\tau)$$

which is re-expressed based on  $T_i$  such that

$$Q_{T_i}(\tau | Z_i) = \exp\{Z_i^T \beta(\tau)\} \text{ for } \tau \in (0,1).$$

Another ‘‘trick’’ they used is to re-express  $I(X_i \geq s)$  for  $s \leq \exp\{Z_i^T \beta_0(\tau)\}$  as

$I(X_i \geq \exp\{Z_i^T \beta_0(u)\})$  for  $u \leq \tau$ . Accordingly they recognized that



$$N_i(\exp\{Z_i^T \beta_0(\tau)\}) - \int_0^\tau I(X_i \geq \exp\{Z_i^T \beta_0(u)\}) dH(u), \quad (3.4)$$

where  $\beta_0(\tau)$  is the true value of  $\beta(\tau)$  and  $N_i(t)$  is sample analogs of  $N(t)$  for  $i=1, \dots, n$ . Note that (3.4) has a mean-zero property. Their proposed estimating function can be written as

$$U_n^{PH}(\beta(\tau)) = n^{-1/2} \sum_{i=1}^n Z_i \left[ N_i(\exp\{Z_i^T \beta(\tau)\}) - \int_0^\tau I(X_i \geq \exp\{Z_i^T \beta(u)\}) dH(u) \right]. \quad (3.5)$$

To solve the equation  $U_n^{PH}(\beta(\tau)) = 0$  is not an easy task since to estimate  $\beta(\tau)$  one needs to know the (estimated) value of  $\beta(u)$  for all  $u \leq \tau$ . Consider the set of grid points:  $S_L = \{0 = \tau_0 \leq \tau_1 \leq \dots \leq \tau_L \leq \tau_U < 1\}$ , where  $\tau_U$  is a constant representing the upper bound of the identifiable region due to censoring. The authors suggested a grid-based estimation procedure for  $\beta_0(\tau)$ . Denote  $\hat{\beta}(\tau)$  as an estimator of  $\beta_0(\tau)$ , which is a right-continuous piecewise-constant function that jumps only on the grid  $S_L = \{0 = \tau_0 \leq \tau_1 \leq \dots \leq \tau_L \leq \tau_U < 1\}$ . Note that the attention of this approach is  $\tau \in (0, \tau_U]$ , not  $\tau \in (0, 1)$ . It is suggested that the equation is solved successively for  $j = 1, \dots, L$  such that

$$n^{-1/2} \sum_{i=1}^n Z_i \left[ N_i(\exp\{Z_i^T \beta(\tau_j)\}) - \sum_{k=0}^{j-1} I(X_i \geq \exp\{Z_i^T \hat{\beta}(\tau_k)\}) (H(\tau_{k+1}) - H(\tau_k)) \right] = 0, \quad (3.6)$$

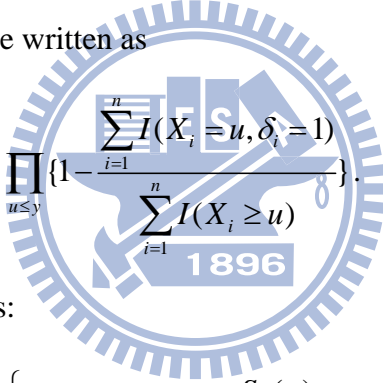
where  $\hat{\beta}(\tau_j)$  is the solution to the equation. Notice that since the equation is not continuous, there may not exist an exact root. The algorithm of simulated annealing can be applied. A nice feature of the above estimating function is that it is a monotone function. This property is useful to develop a more efficient algorithm for numerical computation. Specifically, Peng and Huang (2008) found that solving (3.6) is equivalent to minimize the following objective function

$$l_j(h) = \sum_{i=1}^n \left| \delta_i \log X_i - \delta_i h^T Z_i \right| + \left| R^* - h^T \sum_{l=1}^n -\delta_l Z_l \right| + \left| R^* - h^T \sum_{r=1}^n \left[ 2Z_r \sum_{k=0}^{j-1} I \left[ X_r \geq \exp \{ Z_r^T \hat{\beta}(\tau_k) \} \right] (H(\tau_{k+1}) - H(\tau_k)) \right] \right|$$

where  $R^*$  is a very large number and  $j = 1, \dots, m$ . Peng and Huang (2008) mentioned that one advantage of the latter formulation is that the Barrodale-Roberts algorithm (Barrodale and Roberts 1974) can be directly applied. We do not investigate further in this direction.

### 3.5 A new approach based on self-consistency algorithm

Portnoy (2003) proposed a method which utilizes the idea of self-consistency based on the Kaplan-Meier estimator. Specifically the Kaplan-Meier based on data  $\{(X_i, \delta_i) (i = 1, \dots, n)\}$  can be written as



$$\hat{S}_Y(y) = \prod_{u \leq y} \left\{ 1 - \frac{\sum_{i=1}^n I(X_i = u, \delta_i = 1)}{\sum_{i=1}^n I(X_i \geq u)} \right\}.$$

Alternatively it also satisfies:

$$S_Y(y) = \frac{1}{n} \sum_{i=1}^n \left\{ I(X_i \leq y, \delta_i = 0) \frac{S_Y(y)}{S_Y(X_i)} + I(X_i > y) \right\}.$$

That is, the “weight” contributed to estimate  $S_Y(y)$  for an observation  $X_i$  with  $X_i \leq y$  and  $\delta_i = 0$  is  $\frac{S_Y(y)}{S_Y(X_i)}$ . Now we return to the quantile problem. Setting

$S_Y(y) = 1 - \tau$  and  $S_Y(X_i) = 1 - \tau_i$  where  $\tau_i < \tau$ , we have

$$1 - \tau = \frac{1}{n} \sum_{i=1}^n \left\{ I(X_i \leq Q_Y(\tau), \delta_i = 0) \frac{1 - \tau}{1 - \tau_i} + I(X_i > Q_Y(\tau)) \right\}.$$

This is equivalent to

$$\tau = \frac{1}{n} \sum_{i=1}^n \left\{ I(X_i \leq Q_Y(\tau), \delta_i = 1) + I(X_i \leq Q_Y(\tau), \delta_i = 0) \frac{\tau - \tau_i}{1 - \tau_i} \right\}. \quad (3.7)$$

Figure 3.1 shows the idea of (3.7). For example, there is a single sample with 10 observations at  $X = 1, 2, \dots, 10$  with  $X = 5, 7, 8$  be censored points.

$y$	1	2	3	4	5	6	7	8	9	10
$S_Y(y)$	0.9	0.8	0.7	0.6	0.6	0.48	0.48	0.48	0.24	0
$\tau$	0.1	0.2	0.3	0.4	0.4	0.52	0.52	0.52	0.67	1

In Figure 3.2, we can see that  $S_Y(y)$  and the empirical quantile  $\tau$  are both step functions but the former is decreasing, while the latter is increasing. Notice that the value of empirical quantile  $\tau$  is equivalent to 1 minus the corresponding value of  $S_Y(y)$ . Now we select  $Q_Y(0.52) = 6$ . It is clear that

$$0.52 = \frac{1}{n} \sum_{i=1}^n \left\{ I(X_i \leq 6, \delta_i = 1) + I(X_i \leq 6, \delta_i = 0) \frac{0.52 - \tau_i}{1 - \tau_i} \right\},$$

which results in

$$0.52 = \frac{1}{10} \left( 5 + \frac{0.52 - 0.4}{1 - 0.4} \right).$$

This implies that a censored observation with  $\delta_i = 0$  will contribute the weight  $w_i(\tau) = \frac{\tau - \tau_i}{1 - \tau_i}$ . Note that the proposed algorithm is based on minimizing

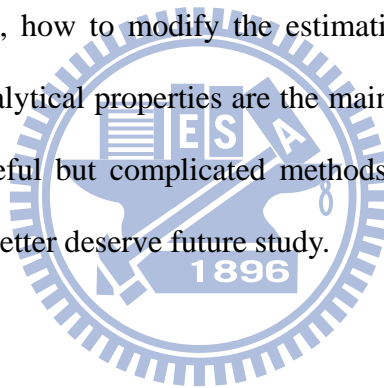
$$\sum_{i=1}^n w_i(\tau) \rho_\tau(X_i - Z_i^T \beta(\tau))$$

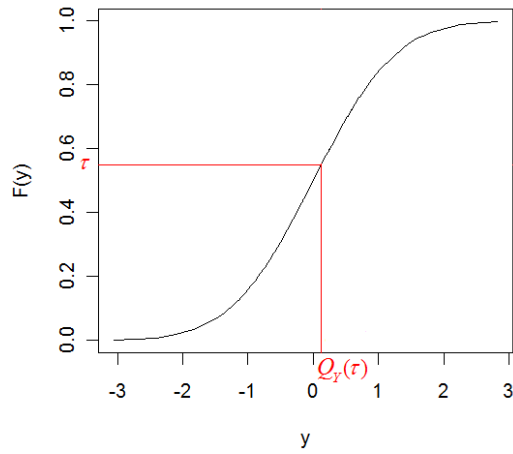
$$\text{where } w_i(\tau) = \begin{cases} 0, & \tau_i > \tau \\ \frac{\tau - \tau_i}{1 - \tau_i}, & \tau_i \leq \tau \text{ and } \delta_i = 0. \\ 1, & \tau_i \leq \tau \text{ and } \delta_i = 1 \end{cases}$$

The idea proposed by Portnoy (2003) is novel, but it is hard to understand the details. Later work by Neocleous, Branden and Portnoy (2006) and Portnoy and Lin (2010) provide more explanations on the self-consistency algorithm.

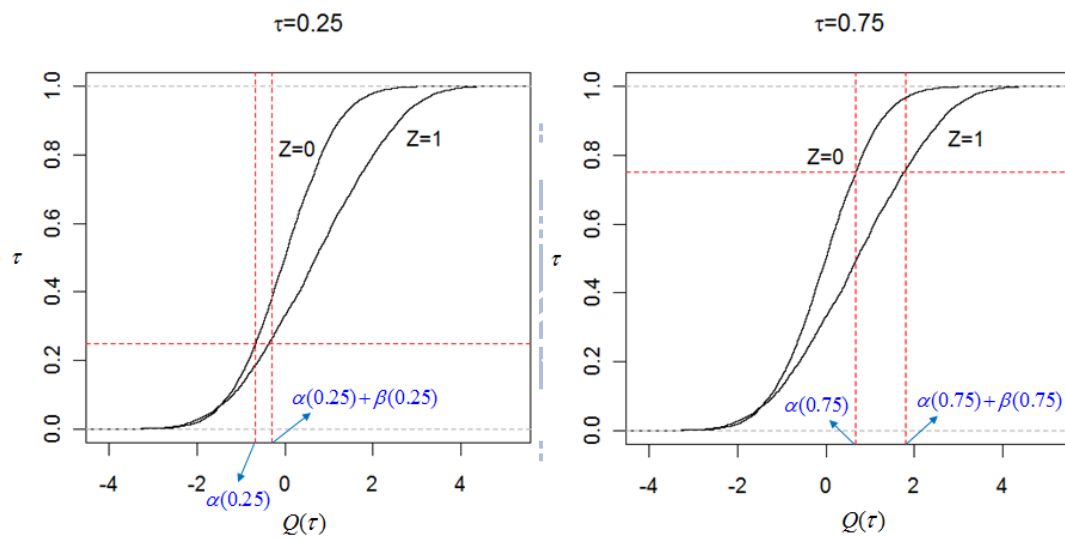
## Chapter 4 Conclusion

In this thesis, we review literature on quantile regression models. Comparing with other regression models, the quantile regression model is more flexible without imposing strong assumption on the error distribution and can well explain the heterogeneity data. Unlike the squared loss function in the least squared regression method, the loss function for estimating the quantile is not differentiable. Accordingly many nice analytic approaches such as the Newton-Raphson method are no longer applicable. Nevertheless, without censoring, some nice analytic properties such as monotonicity and convexity for the estimating and objective functions still exist. They are also helpful for developing numerical algorithms to implement the estimation. When censoring is present, how to modify the estimation procedures and how the modifications affect the analytical properties are the main issues. The thesis is just an initial review for these useful but complicated methods. Thorough investigation to understand these methods better deserve future study.

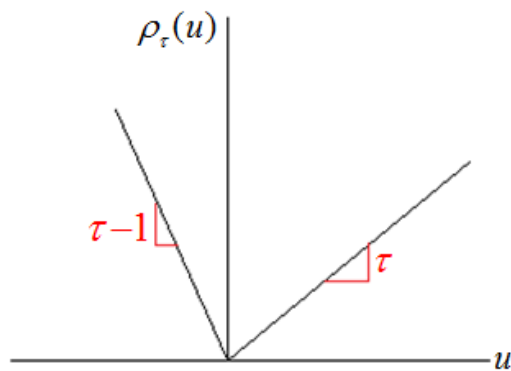




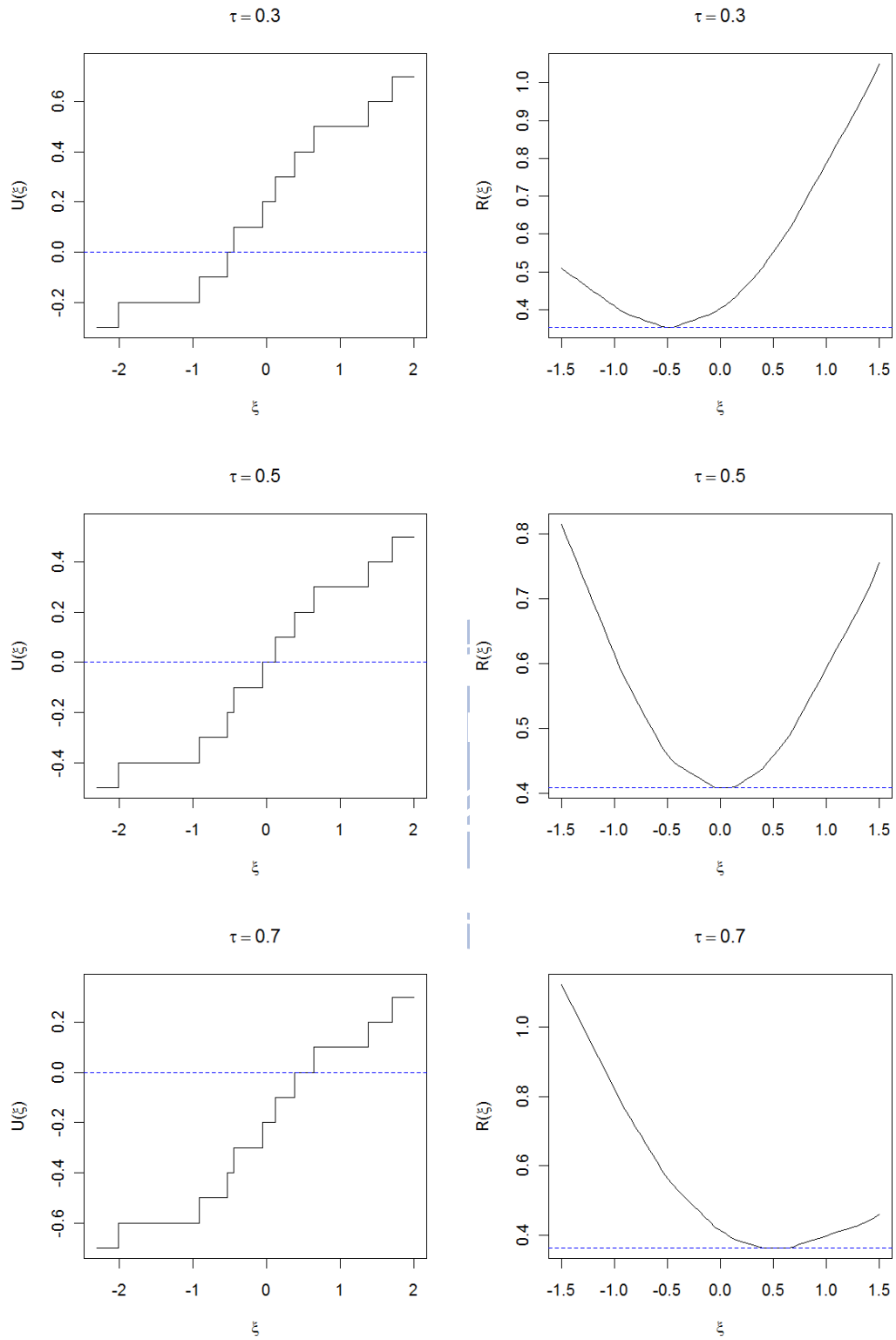
**Figure 1.1: Definition of  $\tau$  th quantile.**



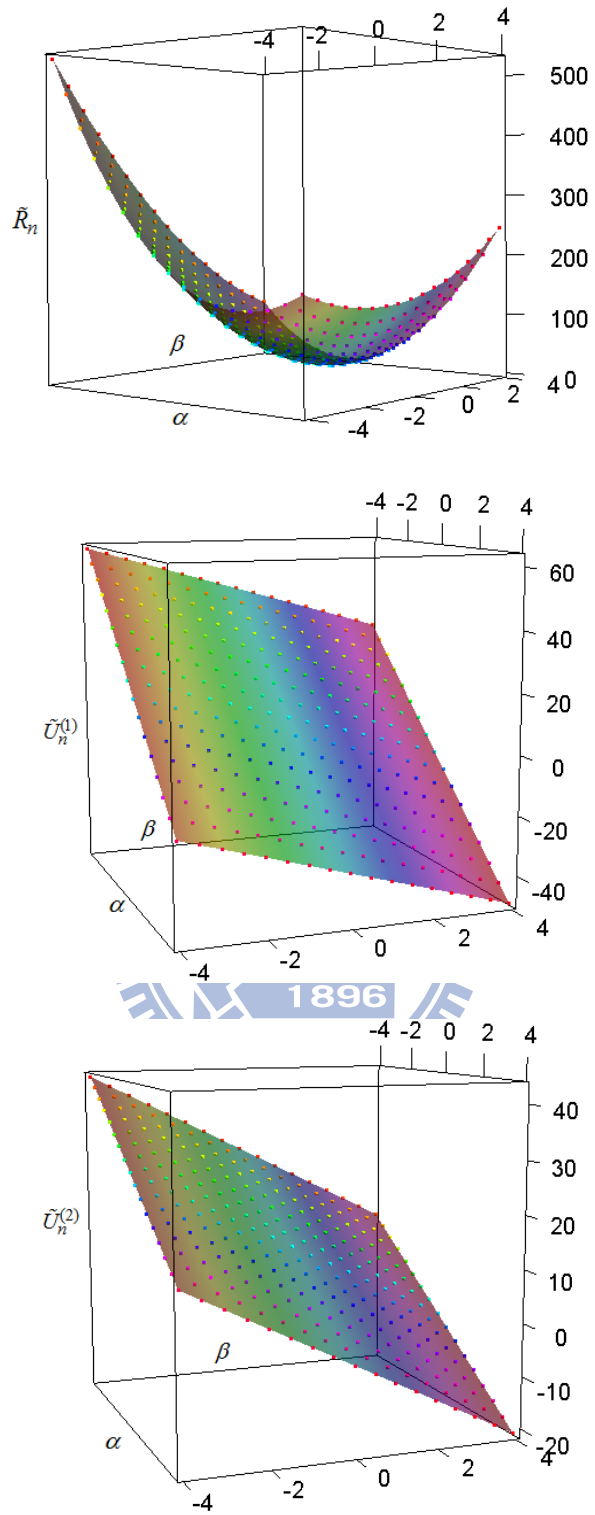
**Figure 1.2: Quantiles for two samples.**



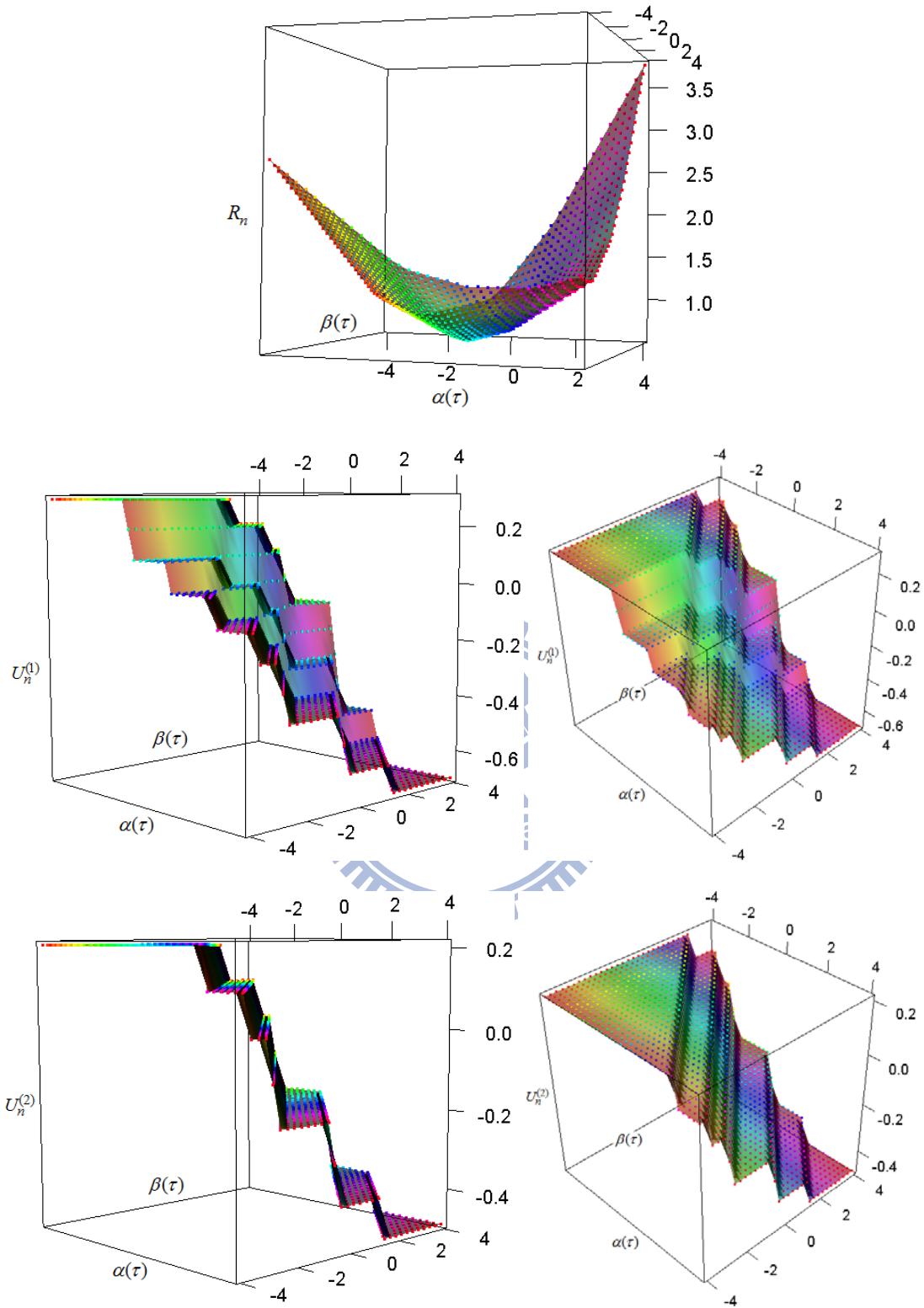
**Figure 2.1: The loss function.**



**Figure 2.2: Plots of  $U_n(\xi)$  and  $R_n(\xi)$ .**

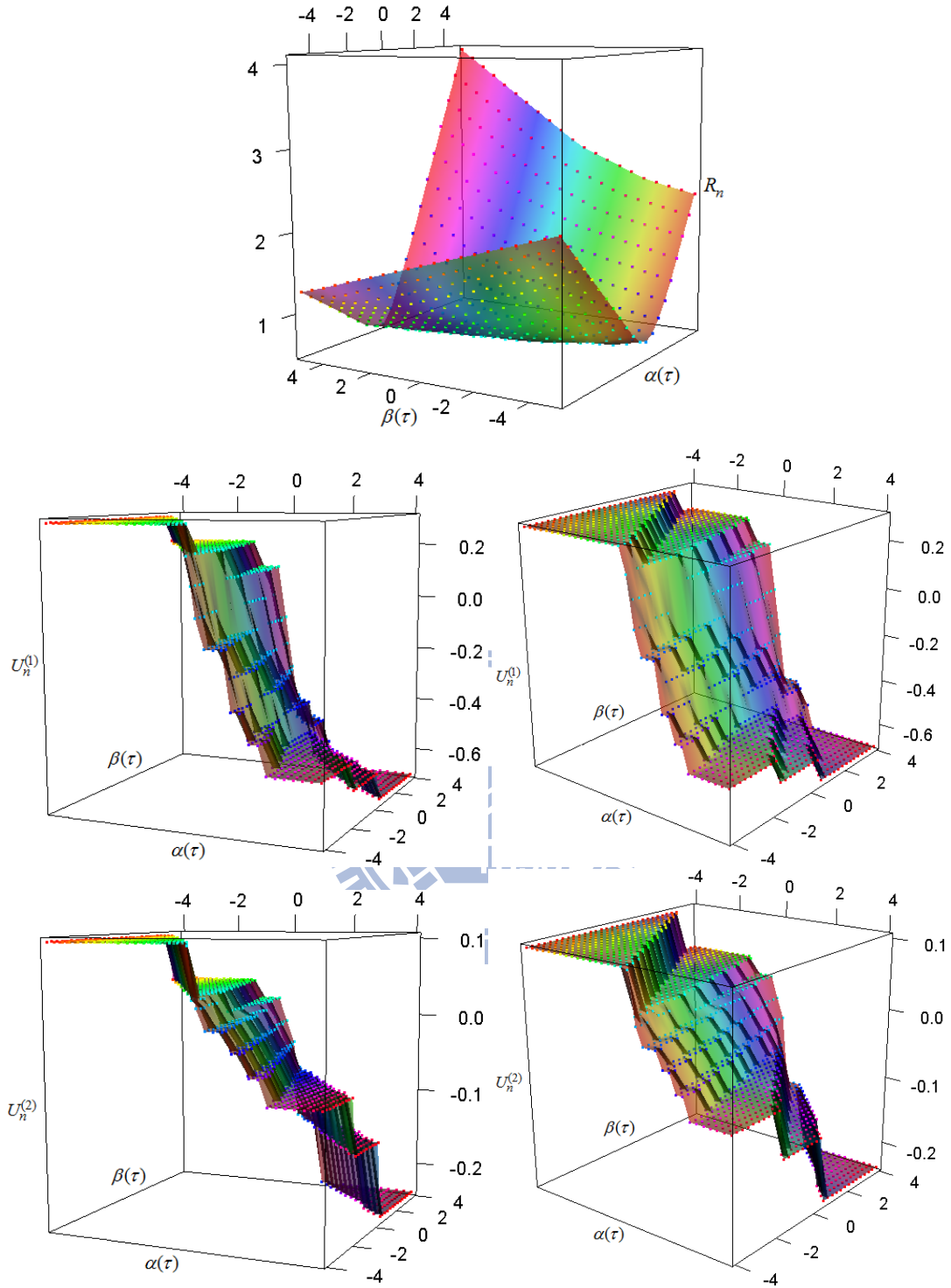


**Figure 2.3: Objective and estimating functions based on LSE with**  
 $Z_i \sim \text{Bernoulli}(1/2)$ .



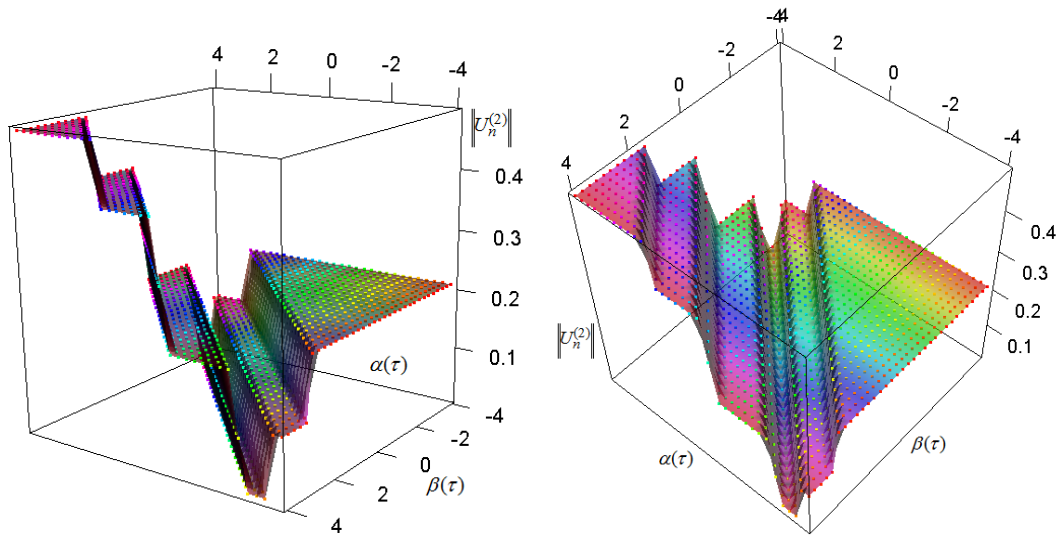
**Figure 2.4: Plots of  $(\alpha(\tau), \beta(\tau), R_n)$  and  $(\alpha(\tau), \beta(\tau), U_n^{(k)})$  ( $k = 1, 2$ ) based on quantile loss function with  $\tau = 0.3$ . Here  $Z_i \sim \text{Bernoulli}(1/2)$ .**



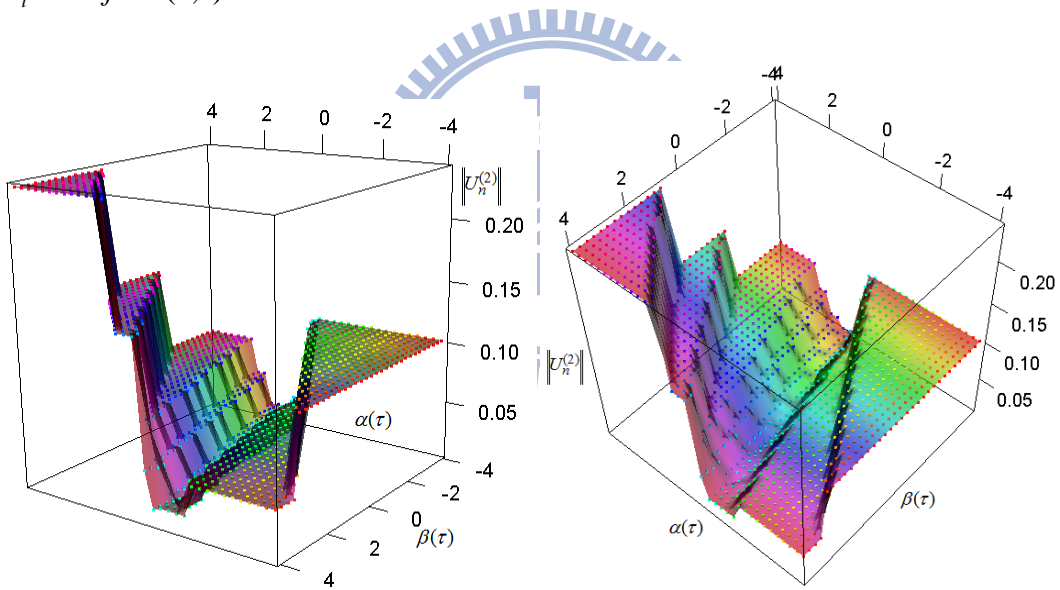


**Figure 2.5: Plots of  $(\alpha(\tau), \beta(\tau), R_n)$  and  $(\alpha(\tau), \beta(\tau), U_n^{(k)})$  ( $k = 1, 2$ ) based on quantile loss function with  $\tau = 0.3$ . Here  $Z_i \sim Uniform(0, 1)$ .**

$Z_i \sim \text{Bernoulli}(1/2)$  :

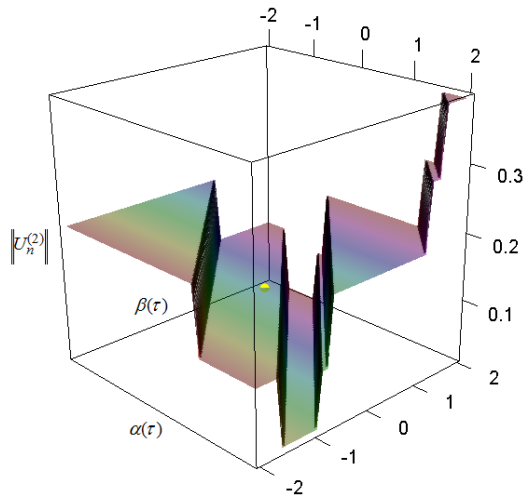


$Z_i \sim \text{Uniform}(0,1)$  :

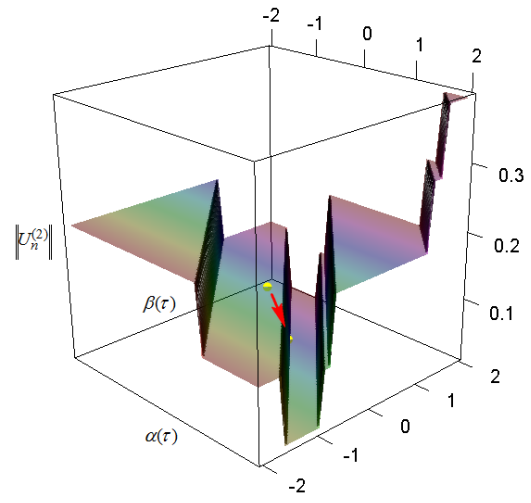


**Figure 2.6: Plots of  $\|U_n^{(2)}(\beta(\tau))\|$  based on  $Z_i \sim \text{Bernoulli}(1/2)$  and  $Z_i \sim \text{Uniform}(0,1)$ .**

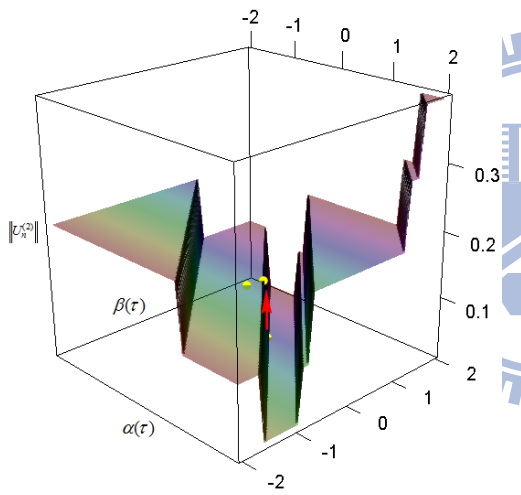
Step 1: initial point



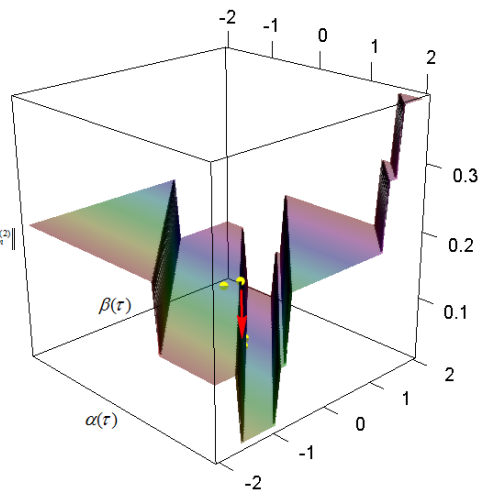
Step 2:



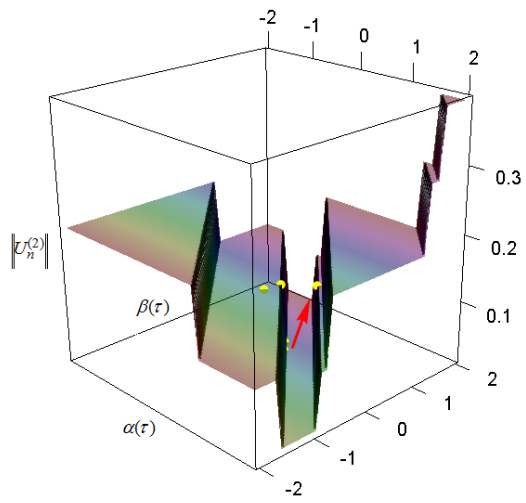
Step 3:



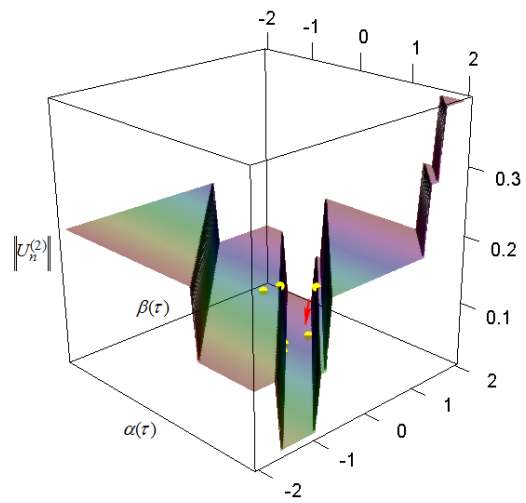
Step 4:



Step 5:



Step 6:



Final step:

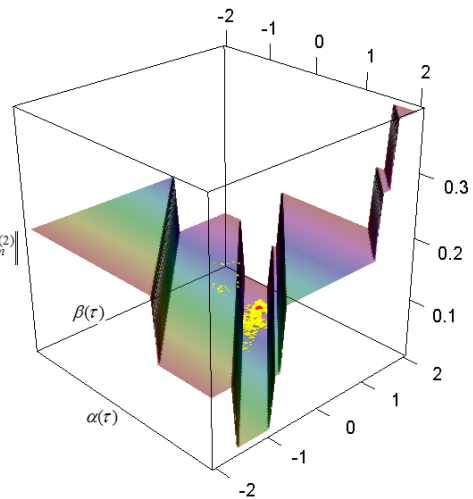
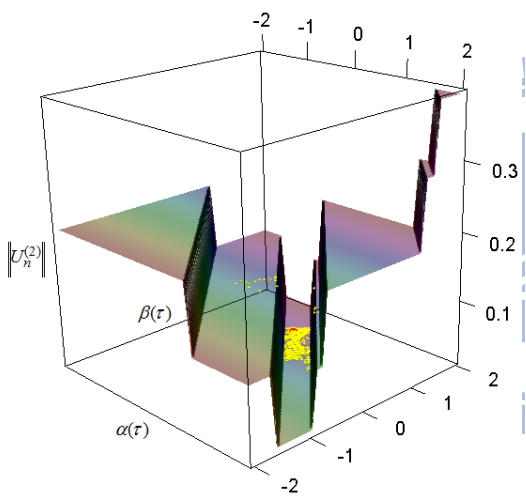


Figure 2.7: Intermediate steps for the simulated annealing algorithm.

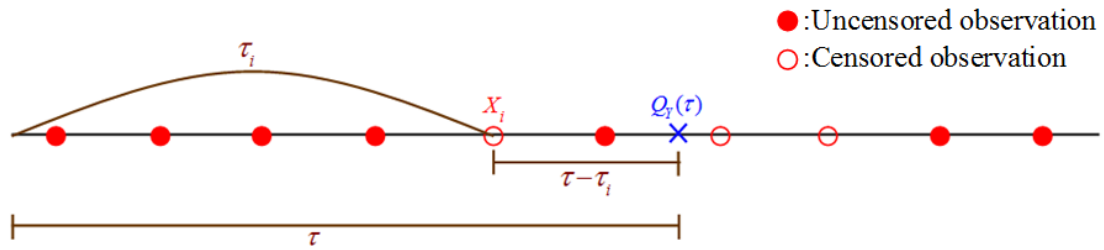


Figure 3.1: Idea of equation (3.7)

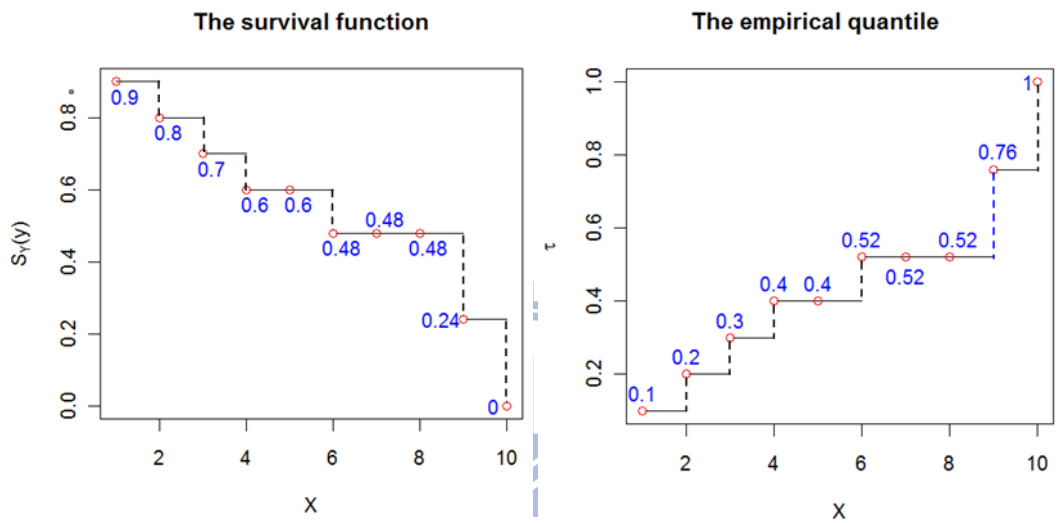


Figure 3.2: The survival function  $S_Y(y)$  and the empirical quantile  $\tau$ .

## References:

- [1] Barrodale, I., and Roberts, F. D. K. (1974), "Solution of an Overdetermined System of Equations in the  $\ell_1$  Norm," *Communications of the ACM*, 17, 319-320.
- [2] Honoré, B., Khan, S., and Powell J. L. (2002), "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 109, 67-105.
- [3] Lin, D.Y., and Geyer, C. J. (1992), "Computational Methods for Semiparametric Linear Regression with Censored Data," *Journal of Computational and graphical Statistics*, 1, 77-90.
- [4] Necleous, T., Branden, K. V., and Portnoy, S. (2006), "Correction to Censored Regression Quantiles by S. Portnoy, 98 (2003), 1001-1012," *Journal of the American Statistical Association*, 101, 869-861.
- [5] Peng, L., and Huang, Y. (2008), "Survival Analysis with Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637-649.
- [6] Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001-1012; Corr, 101, 860-861.
- [7] Portnoy, S., and Lin, G. (2010), "Asymptotics for censored regression quantiles," *Journal of Nonparametric Statistics*, 22, 115-130.
- [8] Powell, J. L. (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-155.
- [9] Ying, Z., Jung, S. H., and Wei, L. J. (1995), "Survival Analysis with Median Regression Models," *Journal of the American Statistical Association*, 90, 178-184.
- [10] Koenker, R., and Geling O. (2001), "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis," *Journal of the American Statistical Association*, 96, 458-468.

- [11] Koenker, R., and V. D’Orey (1987), “Computing Regression Quantiles,” *Journal of the Royal Statistical Society*, 36, 383-393.
- [12] Koenker, R. (2005), *Quantile Regression*, Cambridge: Cambridge University Press.



## Appendix

### A.1 Data generation from a quantile regression model

Usually to generate  $Y$  which has the distributional function  $F_Y(\cdot)$ , we can generate  $U \sim U(0,1)$  and set  $Y = F_Y^{-1}(U)$ . Consider the objective that, given  $Z_i$ , we want to generate  $Y_i$  which follows

$$\Pr(Y_i \leq Z_i^T \beta(\tau)) = \tau \quad (i = 1, \dots, n).$$

Generate  $U_i \stackrel{iid}{\sim} U(0,1)$ . Setting  $Y_i = Z_i^T \beta(U_i)$  and suppose that  $Z^T \beta(t)$  is a monotone increasing function of  $t$ , it follows that

$$\Pr(Y_i \leq Z_i^T \beta(\tau) | Z_i) = \Pr(Z_i^T \beta(U_i) \leq Z_i^T \beta(\tau)) = \Pr(U_i \leq \tau) = \tau.$$

The next question is how to set the distribution of  $Y_i$  as a target one. Consider a simple two-sample case with  $\tilde{Z}_i \sim \text{Bernoulli}(p)$ . Notice that

$$\Pr(Y_i \leq \alpha(\tau) | \tilde{Z}_i = 0) = F_0(\alpha(\tau)) = \tau$$

which means that the intercept  $\alpha(\tau) = F_0^{-1}(\tau)$  is the quantile of the baseline group.

Suppose we let  $Y_i | \tilde{Z}_i = 0$  be the standard normal distribution, we should set

$$\alpha(\tau) = \Phi^{-1}(\tau)$$

which is the quantile of  $N(0,1)$ . It follows that  $Y_i | \tilde{Z}_i = 1 \sim N(\tilde{\beta}(\tau), 1)$ . The algorithm is summarized below:

Step1: Generate  $\tilde{Z}_i \sim \text{Bernoulli}(p)$  and  $U_i \sim \text{Uniform}(0,1)$ .

Step2: Let  $\alpha(U_i) = \Phi^{-1}(U_i)$  and  $\tilde{\beta}(U_i) = c_0 U_i$  where  $\Phi$  is the cumulative distribution function for  $N(0,1)$  and  $c_0$  is a constant.

Step3: Set  $Y_i = \alpha(U_i) + \tilde{Z}_i \tilde{\beta}(U_i)$ .

Step4: Repeat Steps (1) ~ (3) for  $i = 1, \dots, n$ .



Now let  $n = 500$ ,  $p = 1/2$  and  $c_0 = 1.5$ . Then we generate data from a quantile regression model by the following codes of R:

```
n=500; b=1.5; Z=sample(c(0,1), size=n, replace=TRUE);
u=runif(n,0,1); btau=b*u; Y=array(0,n);
Y=qnorm(u)+Z*btau
windows(); plot(Z, Y);
```

Figure A.1 shows that the two groups of generated data  $(\tilde{Z}_i, Y_i)$  ( $i = 1, \dots, n$ ) with the control group ( $\tilde{Z}_i = 0$ ) and the treatment group ( $\tilde{Z}_i = 1$ ). To check the normality for the control group, we can examine Figure A.2 which provides a histogram plot and the Q-Q plot which show that  $Y_i | \tilde{Z}_i = 0$  is exactly generated from a standard normal distribution. Notice that

$$\Pr(Y_i \leq \alpha(\tau) + \tilde{\beta}(\tau) | \tilde{Z}_i = 1) = F_1(\alpha(\tau) + \tilde{\beta}(\tau)) = \tau$$

which means that the slope  $\tilde{\beta}(\tau) = F_1^{-1}(\tau) - F_0^{-1}(\tau)$  is the quantile treatment effect, illustrated in Figure A.3. To check the accuracy of generated data, we plot the empirical quantile functions and the corresponding theory functions of  $F_0$  and  $F_1$  in Figure A.4. Note that the empirical functions of  $F_0$  and  $F_1$  are

$$\hat{F}_0(y) = \frac{\sum_{i=1}^n I(Y_i \leq y, \tilde{Z}_i = 0)}{\sum_{i=1}^n I(\tilde{Z}_i = 0)} \quad \text{and} \quad \hat{F}_1(y) = \frac{\sum_{i=1}^n I(Y_i \leq y, \tilde{Z}_i = 1)}{\sum_{i=1}^n I(\tilde{Z}_i = 1)}.$$

From Figure A.4, we see that both empirical quantile functions match the corresponding theory functions very well. That is, our data generation from a quantile regression model is correct.

Because the objective function is not differentiable, to evaluate the estimation procedures, we make 2-D plots of  $(\alpha(\tau), R_n(\alpha(\tau)))$  and  $(\tilde{\beta}(\tau), R_n(\tilde{\beta}(\tau)))$  for given  $\tau = 0.3, 0.5, 0.7$  by minimizing the objective function

$$\begin{aligned}
R_n(\alpha(\tau), \tilde{\beta}(\tau)) &= \sum_{i=1}^n \rho_\tau(Y_i - \alpha(\tau) - \tilde{Z}_i \tilde{\beta}(\tau)) \\
&= \sum_{i=1}^n (Y_i - \alpha(\tau) - \tilde{Z}_i \tilde{\beta}(\tau)) (\tau - I(Y_i - \alpha(\tau) - \tilde{Z}_i \tilde{\beta}(\tau) < 0)).
\end{aligned}$$

Note that, to get the plots of  $(\alpha(\tau), R_n(\alpha(\tau)))$ , we set  $\tilde{\beta}(\tau) = c_0 \tau$  which is known, whereas  $\alpha(\tau) = \Phi^{-1}(\tau)$  as known for getting the plots of  $(\tilde{\beta}(\tau), R_n(\tilde{\beta}(\tau)))$ . In Figure A.5, we see that there is a minimum for the parabolic curve in each panel and these are the estimators of  $\alpha(\tau)$  and  $\tilde{\beta}(\tau)$  which are marked by blue points. It shows that the estimators of  $\alpha(\tau)$  and  $\tilde{\beta}(\tau)$  are close to the true values  $\alpha_0(\tau)$  and  $\tilde{\beta}_0(\tau)$  which are marked red dash lines. Figure A.5 is made by the following codes of

R:

```

alpha=seq(-5, 5, by=0.1); beta=seq(-5, 8, by=0.1);
for(k in 2:4) {
  tau=0.1*(2*k-1); windows(); par(mfrow=c(1,2));
  Ra=array(0, length(alpha)); Rb=array(0, length(beta))
  amax=-2^31; amin=2^31; bmax=amax; bmin=amin;
  for(i in 1:length(alpha)) {
    for(j in 1:n) {
      if(Y[j]<alpha[i]+Z[j]*b*tau)
        Ra[i]=Ra[i]+(Y[j]-alpha[i]-Z[j]*b*tau)*(tau-1)
      else Ra[i]=Ra[i]+(Y[j]-alpha[i]-Z[j]*b*tau)*tau }
    if(Ra[i]> amax) amax=Ra[i]
    if(Ra[i]< amin) {
      amin=Ra[i]; ax=alpha[i] } }
  plot(alpha, Ra, type='l', xlab=expression(alpha(tau)),
        ylab=expression(paste('R(',alpha(tau),')')))
  title(main=substitute(list(tau)==list(a), list(a=tau)))
  abline(v=qnorm(tau), col='red', lty=2)
  text(qnorm(tau)+0.3, amax-200, substitute(list(alpha[0](tau))==list(a),

```

```

list(a=round(qnorm(tau),3)), col='red')
points(ax, amin, col='blue')
text(ax, amin+350, substitute(list(hat(alpha)(tau))==list(a),
list(a=round(ax,3))), col='blue')
for(i in 1:length(beta)) {
  for(j in 1:n) {
    if(Y[j]<qnorm(tau)+Z[j]*beta[i])
      Rb[i]=Rb[i]+(Y[j]-qnorm(tau)-Z[j]*beta[i])*(tau-1)
    else Rb[i]=Rb[i]+(Y[j]-qnorm(tau)-Z[j]*beta[i])*tau }
    if(Rb[i]> bmax) bmax=Rb[i]
    if(Rb[i]< bmin) {
      bmin=Rb[i]; bx=beta[i]; } }
plot(beta, Rb, type='l', xlab=expression(beta(tau)),
ylab=expression(paste('R(',beta(tau),')')))
title(main=substitute(list(tau)==list(a), list(a=tau)))
abline(v=b*tau, col='red', lty=2)
text(b*tau+0.2, bmax-100, substitute(list(beta[0](tau))==list(a),
list(a=round(b*tau,3))),col='red')
points(bx, bmin,col='blue')
text(bx, bmin+250, substitute(list(hat(beta)(tau))==list(a),
list(a=round(bx,3))), col='blue') }

```

Now we do the data generation from a quantile regression model 100 times. The mean, variance and biased for the estimators of  $\alpha(\tau)$  and  $\tilde{\beta}(\tau)$  are presented in the following:

	Estimator of $\alpha(\tau)$			Estimator of $\tilde{\beta}(\tau)$		
	Mean	Variance	Bias	Mean	Variance	Bias
$\tau = 0.1$	-1.269	0.0076	0.0125	0.164	0.019	0.014
$\tau = 0.3$	-0.524	0.0052	$4 \times 10^{-4}$	0.46	0.01777	0.01
$\tau = 0.5$	$2 \times 10^{-16}$	0.0056	$2 \times 10^{-16}$	0.747	0.01625	-0.003
$\tau = 0.7$	0.511	0.0056	-0.0134	1.038	0.01874	-0.012
$\tau = 0.9$	1.273	0.0066	-0.0085	1.331	0.01953	-0.019

From the table, we see that the values of bias are very small for all  $\tau$ . That is, the average values of the estimators are closed to the true value  $\alpha_0(\tau)$  and  $\tilde{\beta}_0(\tau)$ .

## A.2 How to generate the 3-D plots

To show how to generate the 3-D plots in *R*, we take Figure 2.4, Figure 2.5 and Figure 2.6 as examples. Note that the functions for 3-D plots are in package *scatterplot3d* and *rgl* for *R*. If you are the first time to make a 3-D plot, you must install them and load them into *R*. The code of 3-D plot for *R* is as follow:

```
####Loading these packages
library(scatterplot3d) ; library(rgl);

#### Data generate
n=10; b=3.5; tau=0.3;
Z=sample(c(0,1),size=n,replace=TRUE); u=runif(n,0,1); btau=b*u;
Y=array(0,n); Y=qnorm(u)+Z*btau;

#### Range of alpha and beta
alpha=seq(-4, 4, by=0.5); beta=seq(-4, 4, by=0.5)
a=array(rep(c(alpha), each=length(beta)), length(alpha)*length(beta))
b=array(rep(c(beta), times=length(alpha)), length(alpha)*length(beta))

#### 3-D plot for R(.)
R=array(0, length(alpha)*length(beta)) ; NR=array(0, length(alpha)*length(beta));
for(k in 1: length(alpha)) {
  for(s in 1: length(beta)) {
    for(j in 1: n) {
      if(Y[j]< alpha[k]+beta[s]*Z[j])
        R[s+(k-1)*length(beta)]=
          R[s+(k-1)*length(beta)]+(Y[j]-alpha[k]-beta[s]*Z[j])*(tau-1)/n
      else
        R[s+(k-1)*length(beta)]=
          R[s+(k-1)*length(beta)]+(Y[j]-alpha[k]-beta[s]*Z[j])*tau/n }
    }
  }
}
```

```

NR[k+(s-1)*length(alpha)]=R[s+(k-1)*length(beta)] } }
col=rainbow(length(alpha)*length(beta))
open3d()
plot3d(a, b, R, pch=176, col=col, type='p', xlab="alpha", ylab="beta",
                                             zlab=" R(tau)", smooth=FALSE)
surface3d(alpha, beta, NR, col=col, alpha=0.65)

### 3-D plot for U1(.)
U1=array(0, length(alpha)*length(beta)); NU1=array(0, length(alpha)*length(beta))
for(k in 1: length(alpha)) {
  for(s in 1: length(beta)) {
    for(j in 1: n) {
      if(Y[j]< alpha[k]+beta[s]*Z[j])
        U1[s+(k-1)*length(beta)]=U1[s+(k-1)*length(beta)]+(tau-1)/n
      else
        U1[s+(k-1)*length(beta)]=U1[s+(k-1)*length(beta)]+tau/n }
    NU1[k+(s-1)*length(alpha)]=U1[s+(k-1)*length(beta)] } }
col=rainbow(length(alpha)*length(beta))
open3d()
plot3d(a, b, U1, pch=176, col=col, type='p', xlab="alpha", ylab="beta",
                                             zlab=" U1(tau)")
surface3d(alpha, beta, NU1, col=col, alpha=0.65)

### 3-D plot for U2(.)
U2=array(0, length(alpha)*length(beta)); NU2=array(0, length(alpha)*length(beta))
for(k in 1: length(alpha)) {
  for(s in 1: length(beta)) {
    for(j in 1: n) {
      if(Y[j]< alpha[k]+beta[s]*Z[j])
        U2[s+(k-1)*length(beta)]=U2[s+(k-1)*length(beta)]+Z[j]*(tau-1)/n
      else
        U2[s+(k-1)*length(beta)]=U2[s+(k-1)*length(beta)]+Z[j]*tau/n }
    NU2[k+(s-1)*length(alpha)]=U2[s+(k-1)*length(beta)] } }

```

```

col=rainbow(length(alpha)*length(beta))
open3d()
plot3d(a, b, U2, pch=176, col=col, type= 'p',xlab="alpha", ylab="beta",
                                             zlab=" U2(tau)")
surface3d(alpha, beta, NU2, col=col, alpha=0.65)

### 3-D plot for ||U(.)||
U=array(0, length(alpha)*length(beta)); NU=array(0, length(alpha)*length(beta))
for(k in 1: length(alpha)) {
  for(s in 1: length(beta)) {
    for(j in 1: n) {
      if(Y[j]< alpha[k]+beta[s]*Z[j])
        U[s+(k-1)*length(beta)]=U[s+(k-1)*length(beta)]+Z[j]*(tau-1)/n
      else
        U[s+(k-1)*length(beta)]=U[s+(k-1)*length(beta)]+Z[j]*tau/n }
      U[s+(k-1)*length(beta)]=abs(U[s+(k-1)*length(beta)])
      NU[k+(s-1)*length(alpha)]=U[s+(k-1)*length(beta)] } }
col=rainbow(length(alpha)*length(beta))
open3d()
plot3d(a, b, U, col=col, type= 'p',xlab="alpha", ylab="beta", zlab=" U(tau)")
surface3d(alpha, beta, NU, col=col, alpha=0.65)

```

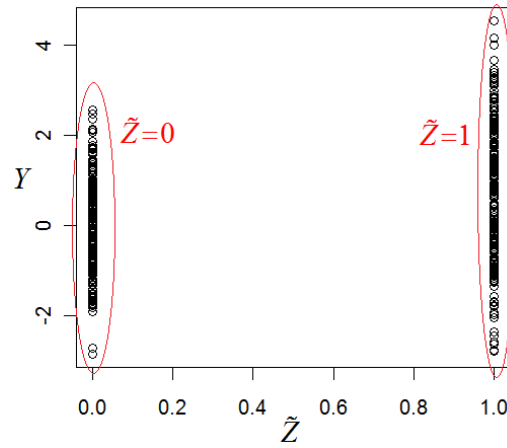


Figure A.1: Scatterplot of generated data from a quantile regression model.

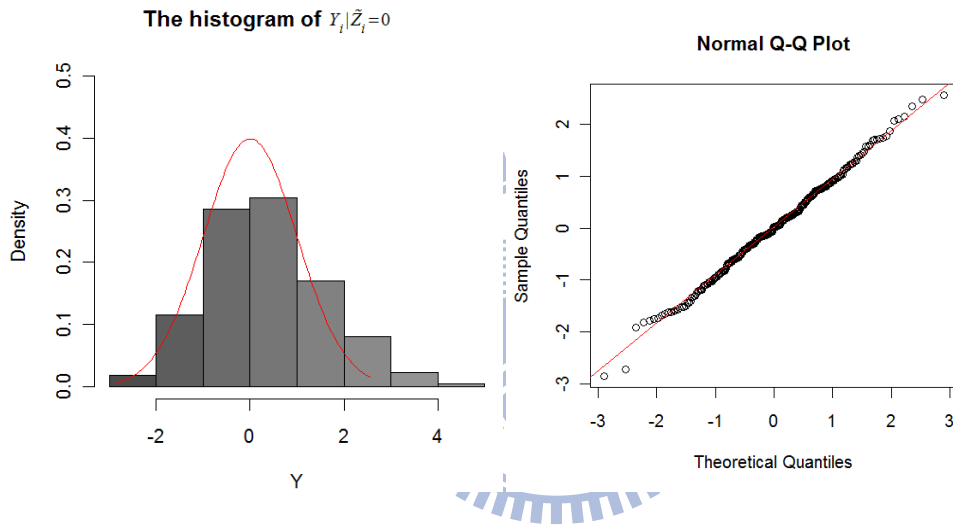


Figure A.2: Histogram plot and Q-Q plot of  $Y_i | \tilde{Z}_i = 0$ .

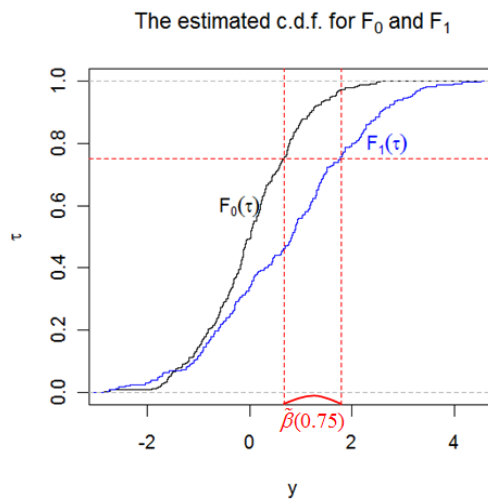
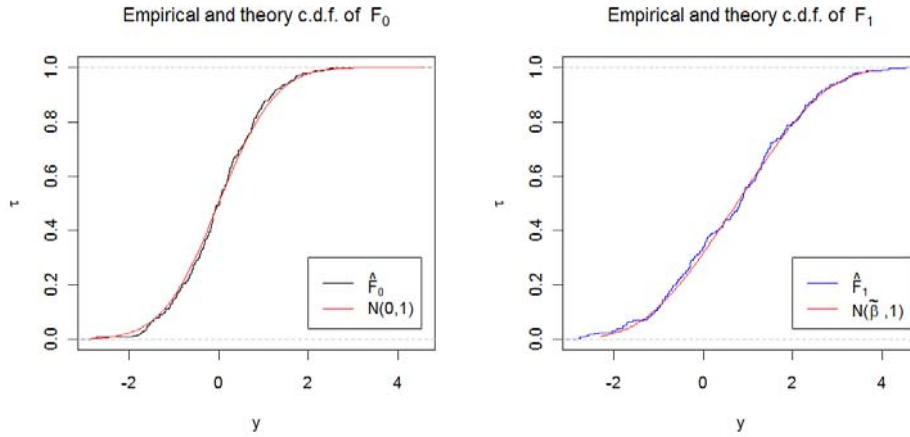
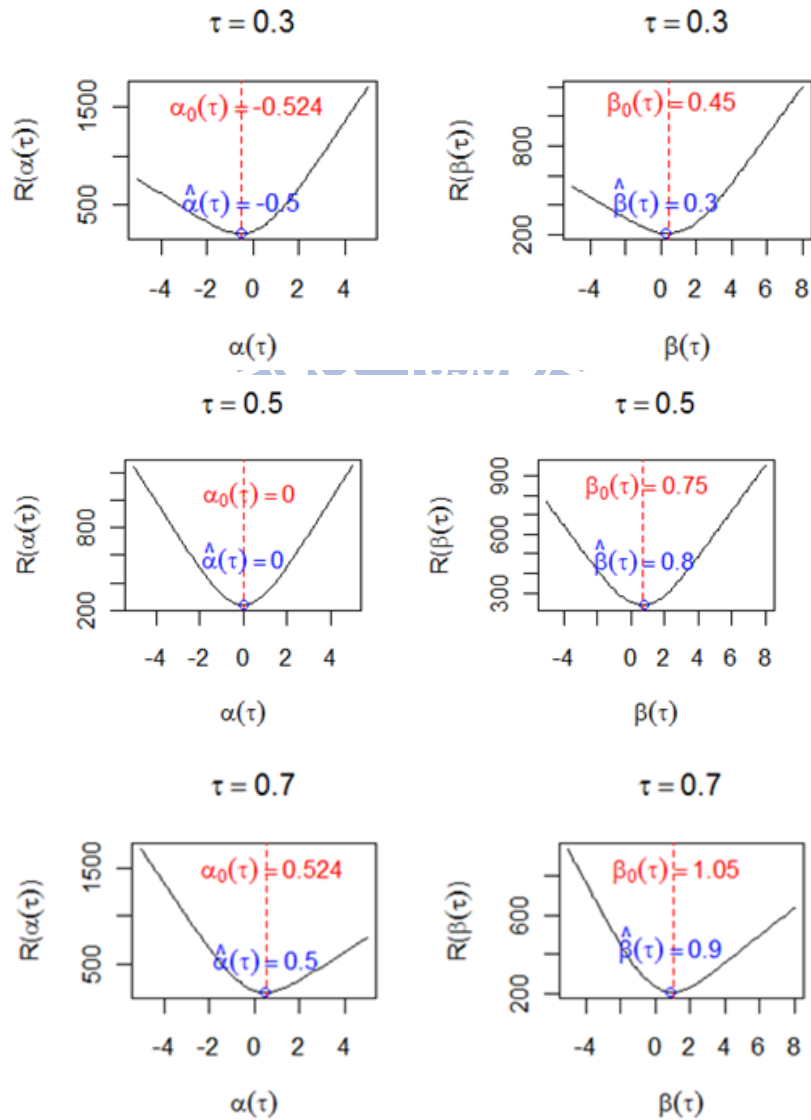


Figure A.3: Plot of the quantile treatment effect  $\tilde{\beta}(\tau)$ .



**Figure A.4:** Plots of the empirical quantile functions and the corresponding theory functions of  $F_0$  and  $F_1$ .



**Figure A.5:** Plots of  $(\alpha(\tau), R_n(\alpha(\tau)))$  and  $(\tilde{\beta}(\tau), R_n(\tilde{\beta}(\tau)))$ .