國立交通大學

統計學研究所

碩 士 論 文

線性模型於不同誤差假設下之統計推論

Statistical Inference for Linear Models with Flexible Error Distributions

研 究 生：王文廷

指導教授：王維菁 博士

中 華 民 國 九十七 年 六 月

線性模型於不同誤差假設下之統計推論

Statistical Inference for Linear Models with Flexible Error Distributions

研 究 生：王文廷　　　　Student：Wen-Ting Wang

指導教授：王維菁 博士　　　　Advisor：Dr.Weijing Wang

國立交通大學理學院

統計學研究所

碩 士 論 文

A Thesis
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
June 2008
Hsinchu, Taiwan, Republic of China

中 華 民 國 九十七 年 六 月

線性模型於不同誤差假設下
之統計推論

學生：王文廷

指導教授：王維菁 博士

國立交通大學統計學研究所

摘　　要

傳統線性迴歸模型常假設誤差之分配為常態或呈對稱分配，估計方法也以最小平方法為主軸。然而資料分析時若見到比例不低的極端值，或出現估計誤差有偏斜的狀況，此時最小平方法是否適用，成為值得探討的問題。

在論文中我們回顧了數個常見的推論方法，並討論其在不同誤差分配假設下的適用性。主要的評估標準為方法的穩健性與估計的效能(efficiency)。我們並討論把現有方法延展到設限資料的常用技巧。我們並藉由模擬實驗比較方法的優劣。

關鍵詞：設限，線性模型，穩健性估計量

Statistical Inference for Linear Models with Flexible Error Distributions

Student: Wen-Ting Wang
Advisor: Dr. Weijing Wang

Institute of Statistics
Nation Chiao Tung University
Hsinchu, Taiwan

Abstract

In the thesis, we consider statistical inference for a general class of linear regression models. The assumption on the error distribution plays a crucial role for the development of an appropriate inference method. Here we examine several estimation approaches under four types of error distributions including symmetric (with light/heavy tails) and asymmetric distributions. In particular, we focus on the issues of robustness and efficiency. We also discuss how the existing methods are extended to the situation of censoring. Monte Carlo simulations are performed to evaluate the finite-sample performances of different methods.

Keywords: Censoring, Linear model, Robust estimation

誌謝

時間，奏起了研究所兩年的序曲。

一個懵懂無知的自己，在進入研究所兩年旅程之初，身上僅有探索未知事物的衝勁，但缺乏了許多基本的專業知識。慶幸地，路途上並非全然是孤軍奮戰且安於現狀的畫面。憶起碩一上，六人一起出書賣書的那段期間，從中不僅將基礎知識重新複習，感染到每個人對於完成一件作品不同的觀點以及相同的態度，更重要的是學會了分享；除此之外，統諮分組報告的那段期間，大家各司其職，最後得到的不只是表面上成績。更重要的是，了解付出的多寡不是只為了放在成績天平上而已，是大家合力從這世界挖掘了想要的答案，最後的成就感是無價的。學習是沒有終點，學會分享是幫助我完成這份論文最大原動力以及快樂之所在。

完成這篇論文的重要意義不在於一篇文章，而是在過程中的鍛練。首先感謝我的指導老師王維菁老師，經常將我紊亂的思緒條理分明。還有老師曾點名我有著跳躍式思考，至此我開始將自己腳步放慢，重覆地將各種已讀過的方法再深入研讀。雖然最後成果可能與自己有段差距，但所得到的哲理卻更有意義－不論是處理事或做研究，若只是盲目追隨一個傳統思想進而想要改進現有的問題，很有可能讓自己侷限在一個空間內無法跳脫。也感謝陳志榮老師時常願意解決我課業上所面臨的問題，受益良多。

什麼是真理？或許在每個人心中有不同的詮釋，或許每個人界定的範圍不一樣，可以確定的是這兩年與同學們的相處讓我悟出不一樣的真理，不論是關心我還是批評我都感謝兩年生活在交大的每一位同學在處理事情態度是積極或是消極，是謙虛或是自大、是無我或是自私，從你們身上學會的不只是簡單的思考邏輯以及如何無私分享這兩件事，也學會了如何用真理去迎對未來社會上帶著各種不同面具的人群。特別感謝這些平常多些包容我的同學們：做事都堅守自己原則以及條理的泰佐和翁賢；處理事情會一針見血地剖析的重耕和仲竹、做事有主見的昱緯、以及學習能力很快的瑜達。當然，還有一群曾經共事的同伴以及給予我

機會一同解決問題的同學們。

感謝全所成員一同營造這個環境，讓我像吉普賽人找到了迦南地，自由自在的吸收知識、找尋自己的未來。最後要感謝的是家人、最愛的朋友以及以前的同學們。感謝你們支持我，讓我在這兩年有別於過去，了解與其待在原地抱怨還不如找到對事物的熱沉。

時間，即將畫上休止符，帶著最深的感謝以及祝福每一個我所認識的人。

王文廷 謹誌於

國立交通大學統計研究所

中華民國九十七年六月

Table of Contents

Tables

Chapter 1 Introduction

1.1 Background

In the thesis, we consider regression analysis for analyzing failure time data. Let $\tilde{T}$ be the time to the event of interest and $\mathbf{Z}$ be the vector of covariates. Studying the effect of $\mathbf{Z}$ on $\tilde{T}$ is a practical and important problems in practical applications. Consider the following linear regression model,

$$T = h(\tilde{T}) = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon, \tag{1.1}$$

where $h(.)$ is a monotone function and ε is the error distribution satisfying some criteria. The parameters $\boldsymbol{\beta}$ describe the effect of covariates on T or $h(\tilde{T})$ and hence is of major interest. Statistical inference of $\boldsymbol{\beta}$ can be classified according to whether the form of $h(\cdot)$ is specified and whether the distribution of ε is given.

For the first type of analysis, the form of $h(.)$ is known but the distribution of ε is not specified. For example, if $h(t) = t$, model (1.1) becomes the location-shift model. If $h(t) = \log(t)$, model (1.1) is the accelerated failure time model. The second type of analysis assumes that ε has a known distribution function but leaves $h(.)$ to be unspecified. Such models are often called as transformation models.The Cox proportional hazard model is the most well-known special case with ε being the extreme value distribution. Another useful example is the proportional odds model in which the error has the standard logistic distribution. In recent years, some authors consider the most general structure with both $h(\cdot)$ and the distribution of ε being unspecified

In this thesis, we will focus on the first type of model in which the form of $h(.)$ is given but the distribution of ε is not specified. We would like to study inference methods for estimating $\boldsymbol{\beta}$. In particular, we are interested in studying how the inference methods adjust for the error distribution and how censoring is handled. It is hoped that our review can improve our understanding about different inference ideas.

1.2 Outline

In Chapter 2, we review basic concepts in survival analysis. Chapter 3 considers estimation of $\boldsymbol{\beta}$ in absence of censoring. Statistical inference based on censored data is discussed in chapter 4. Simulation results are presented in Chapter 5. In Chapter 6, we give concluding remarks.

Chapter 2 Preliminary

In this chapter, we review some basic concepts in survival analysis which will be useful for the discussion in Chapter 4. Here T is the failure time of interest.

2.1 Descriptive Measures

The survival function of T is defined as $S(t) = \mathrm{P}(T > t)$ which measures the chance that the failure event has not occurred up to time T. The mean of T can be written as

$$E(T) = -\int_0^{\infty} t dS(t) = \int_0^{\infty} S(t)dt, \tag{2.1}$$

where the second identity is obtained by performing "integration by parts". Another commonly used measure is the mean residual life function defined as

$$mrl(t) = E(T - t | T > t) = \frac{\int_t^{\infty} S(u)du}{S(t)}. \tag{2.2}$$

The above two expressions imply that $E(T)$ and $mrl(t)$ are both related to the survival function $S(t)$. Estimation of the survival function nonparametrically is am important problem in survival analysis. The following product-limit expression, proposed by Kaplan and Meier (1958), is very useful for handling censored data.

$$S(t) = \prod_{u \le t} \left(1 - \frac{\Pr(T \in [u, u + du])}{\Pr(T \ge u)} \right) \tag{2.3}$$

2.2 Nonparametric Inference under Right Censoring

In practice, subjects may drop out from the study or do not developed the event of interest during the study period. Let C be the censoring variable. Under right censoring, we only observe (X, δ) where $X = T \wedge C$ and $\delta = I(T \le C)$. Observe data become $\{(X_i, \delta_i),\ i = 1, \ldots, n\}$, where

$$X_i = T_i \wedge C_i \text{, and } \delta_i = \begin{cases} 1, \text{if } \mathrm{T_i} \le C_i \\ 0, \text{if } \mathrm{T_i} > C_i \end{cases}.$$

A crucial assumption is that T and C are independent. The well-known Kaplan-Meier estimator of $S(t)$ based on (2.3) is given by

$$\hat{S}(t) = \prod_{u \le t} \left(1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1)}{\sum_{i=1}^{n} I(X_i \ge u)} \right). \qquad (2.4)$$

Note that the censoring effect gets cancelled out in the ratio calculation.

The Kaplan-Meier estimator has many nice properties. Here we only present the ideas related our thesis. The survival function can be viewed as the first moment of the event $I(T > t)$, namely

$$S(t) = \Pr(T > t) = E[I(T > t)].$$

When the data are complete, the empirical estimator of $S(t) = E[I(T > t)]$ is given by $\bar{S}(t) = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} [I(T_{\mathrm{i}} > t)]$, which utilizes the method of moment. Under censoring, the value of $I(T_{\mathrm{i}} > t)$ may not be exactly known.

The idea of imputation is to replace $I(T_{\mathrm{i}} > t)$ by an estimate of its conditional expectation given the data. It follows that

$$\hat{E}[I(T_{\mathrm{i}} > t) \mid X_i, \delta_i] = I(X_{\mathrm{i}} > t) + I(X_i < t, \delta_i = 0)\hat{E}[I(T_{\mathrm{i}} > t) \mid X_i, T_i > X_i]$$

$$= I(X_{\mathrm{i}} > t) + I(X_i \le t, \delta_i = 0)\frac{\hat{S}(t)}{\hat{S}(X_i)}.$$

Therefore the following self-consistent equation can be constructed:

$$\hat{S}(t) = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[I(T_{\mathrm{i}} > t) \mid X_i, \delta_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[I(X_{\mathrm{i}} > t) + I(X_i \le t, \delta_i = 0)\frac{\hat{S}(t)}{\hat{S}(X_i)} \right]. \qquad (2.5)$$

The Kaplan-Meier is the only solution to equation (2.5).

Weighting is another way of handling missing data. We can view $I(X \geq t)$ as a proxy of $I(T \geq t)$. To correct the bias of $I(X \geq t)$, we find that $E[\frac{I(X > t)}{G(t)}] = E[I(T > t)]$. The Kaplan estimator $\hat{S}(t)$ can be written as $\hat{S}(t) = \sum_{i=1}^{n} \frac{I(X_i > t)}{\hat{G}(t)}$, where $\hat{G}(t)$ is the Kaplan-Meier estimator of $G(t)$ such that

$$\hat{G}(t) = \prod_{u \leq t} \{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 0)}{\sum_{i=1}^{n} I(X_i \geq u)}\}. \tag{2.6}$$

The ideas of imputation and weighting provide useful skills for handling missing data. We will see that they are also used in the regression framework considered in thesis when data are subject to censoring.

With the estimator $\hat{S}(t)$, it seems that $E(T)$ and $mrl(t)$ can be estimated nonparametrically by $\hat{\mu} = \int_0^{\infty} \hat{S}(t)dt$ and $\hat{mrl}(t) = \int_t^{\infty} \hat{S}(u)du / \hat{S}(t)$. However since $\hat{S}(t)$ does not provide a valid estimator for t locates beyond the data support,

$$\tau^* = \sup_t \{t : \Pr(X > t) > 0\}.$$

Consequently $\hat{\mu}$ and $\hat{mrl}(t)$ will be underestimated due to the tail problem.

Chapter 3 Inference without Censoring

Consider the linear regression model

$$\tilde{\mathbf{T}} = h(\mathbf{T}) = \mathbf{Z}^T \boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where the form of $h(.)$ is given but the distribution of ε is unknown. In this chapter, we consider observed data of form, as $\{(T_i, \mathbf{Z}_i), i = 1, \ldots, n\}$, where $\mathbf{Z}_i^T = (1, Z_1 \ldots, Z_{ip-1})$ denotes the vector of covariates. The main goal is to estimate $\boldsymbol{\beta}^T = (\beta_0, \ldots, \beta_{p-1})$.

Statistical methods for estimating $\boldsymbol{\beta}$ requires making additional assumption on the error distribution. By reviewing existing methods, we can understand how the error distribution affects subsequent inference procedures.

3.1 Least Squares (LS) Estimation

Assume that ε_i are identically and independent distributed. with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. The most well- known result is the least-squares estimator which minimizes

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(T_i - \mathbf{Z}_i^T \boldsymbol{\beta}\right)^2 = \sum_{i=1}^{n} \{e_i(\boldsymbol{\beta})\}^2, \tag{3.1}$$

where $e_i = T_i - \mathbf{Z}_i^T \boldsymbol{\beta}$. The resulting estimator can be obtained by solving $\frac{\partial}{\partial \beta} S(\boldsymbol{\beta}) = 0$. The solution can be written as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{Z}^T \mathbf{Z}\right)^{-1} \mathbf{Z}^T \mathbf{T}, \tag{3.2}$$

where $\mathbf{T} = (T_1, T_2, \ldots, T_n)^T$ and $\mathbf{Z}_i^T = (1, Z_1 \ldots, Z_{ip-1})$. When the error term are i.i.d. with mean-zero, we have $Var(\varepsilon_i) = E(\varepsilon_i^2)$. Under this situation, $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (B.L.U.E.). If the normality assumption is further imposed, efficiency of $\hat{\boldsymbol{\beta}}$ can be established.

The least squares method can be adjusted for unequal variances. For example, if the covariance matrix of $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ can be written as $\sigma^2 \mathbf{V}$, where $\mathbf{V}$ is a non-singular and diagonal matrix, then we can estimate $\boldsymbol{\beta}$ by the weighted least squares estimator $\hat{\boldsymbol{\beta}}^W = \left(\mathbf{Z}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{T}$.

Despite its nice property under the desirable condition, the least squares estimator is known to be sensitive to outliers. We can view that the least squares method minimizes the objective function $\rho(\varepsilon) = \varepsilon^2$. Accordingly, the influence function of the least squares is given by

$$IF(\varepsilon) \propto \frac{\partial \varepsilon^2}{\partial \varepsilon} \propto \varepsilon. \tag{3.3}$$

This implies that the influence of ε is proportional to its size. Therefore when ε has extreme values, its impact on $\rho(\varepsilon)$ is also large. This explains why the least squares method is not robust.

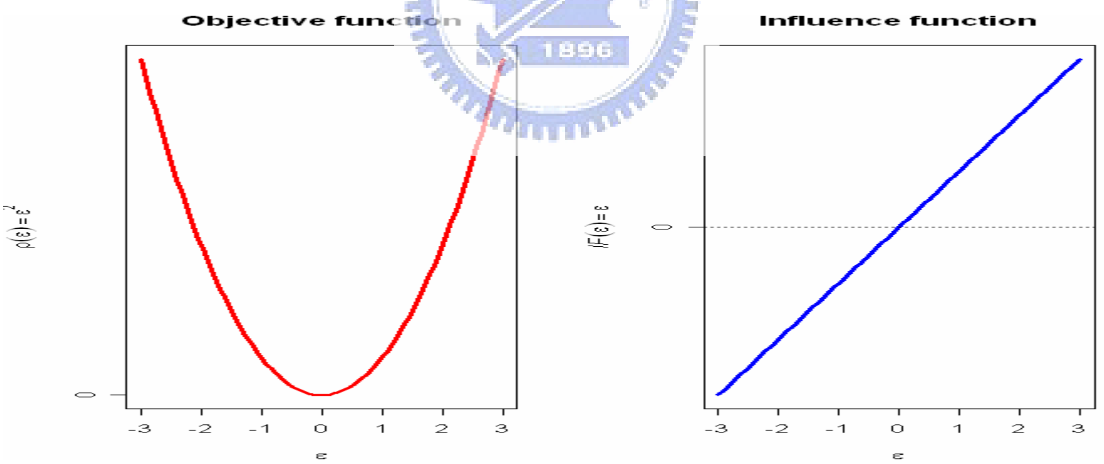

Figure 3.1: The objective function and influence function of least squares method

3.2 Least Absolute Deviations (LAD) Estimation

Assume the distribution of ε is symmetric around zero. The absolute deviance is defined as

$$\sum_{i=1}^{n}\left|T_i - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}\right| = \sum_{i=1}^{n}\left|e_i(\boldsymbol{\beta})\right|. \tag{3.4}$$

It turns out that the median of $(e_1(\boldsymbol{\beta}),\ldots,e_n(\boldsymbol{\beta}))$ minimizes the least absolute deviations objective function in (3.4).

We can write $\rho(\varepsilon)=|\varepsilon|$ .The influence function of the LAD estimator is

$$IF(\varepsilon) \propto \operatorname{sgn}(\varepsilon) = \begin{cases} 1, & \varepsilon > 0 \\ 0, & \varepsilon = 0 \\ -1, & \varepsilon < 0 \end{cases}, \tag{3.5}$$

where we define $\operatorname{sgn}(\varepsilon)$ for $\varepsilon = 0$ despite that it is not differentiable at zero.

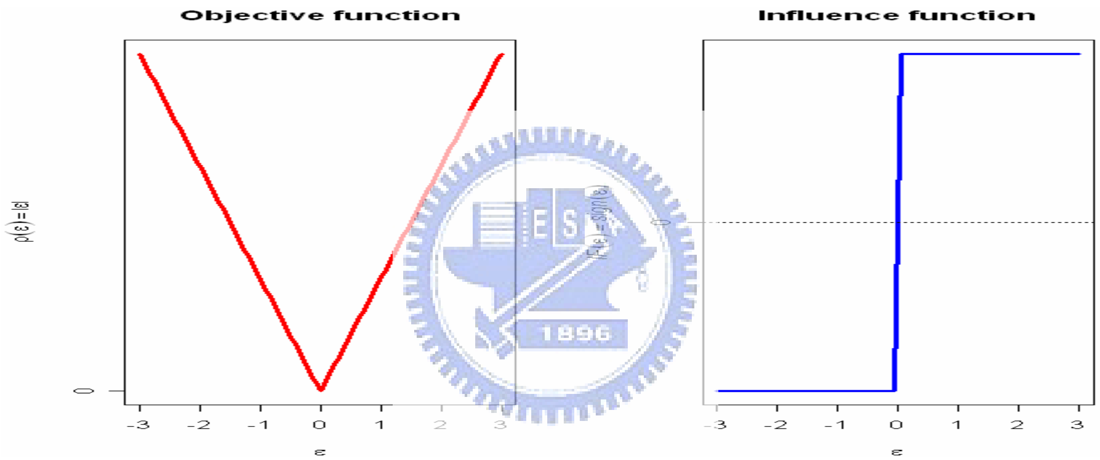

Figure 3.2: The objective function and Influence function of LAD

The above figure shows that the influence of ε on $\rho(\varepsilon)$ only depends on its sign regardless of its size. Since $IF(\varepsilon)$ is bounded, the resulting estimator, namely the median, is resistant or robust to extreme observations. However, $IF(\varepsilon)$ is not related to the size of ε at all. This means large error has the same affect on estimation as the small error.

3.3 M-estimator

The objective function can be written as a flexible form, $\rho(\varepsilon)$. The so-called M-estimator is defined as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(T_i - \mathbf{Z}_i^{T}\boldsymbol{\beta}\right). \quad (3.6)$$

If $\rho(\cdot)$ is differentiable with the derivative ψ, then $\hat{\boldsymbol{\beta}}$ can solve the equation

$$\sum_{i=1}^{n} \psi\left(T_i - \mathbf{Z}_i^{T}\boldsymbol{\beta}\right) = 0 \quad (3.7)$$

Now the next question is how we can determine the form of $\rho(\varepsilon)$.

If the error distribution is symmetric around 0, we can choose the $\rho(\cdot)$ that satisfies $\rho(0) = 0$ and $\rho(\varepsilon) = -\varepsilon$. An appropriate form of $\rho(\varepsilon)$ can adapt to the error distribution to find a general principle for choosing the appropriate form of $\rho(\varepsilon)$, we can view $\rho(\varepsilon)$ as the negative log-likelihood of ε. For instance, if the method of least squares is considered with $\rho(\varepsilon) = \varepsilon^2 \ (\varepsilon \in R)$, it corresponds to a normal distribution with the density proportional to

$$\exp\left(\frac{-\rho(\varepsilon)}{2\sigma^2}\right) = \exp\left(\frac{-\varepsilon^2}{2\sigma^2}\right).$$

For the method of absolute deviations with $\rho(z) = |z|, z \in R$, it corresponds to the situation that ε has the double exponential distribution with the density related to $\exp(-\rho_2(\varepsilon)) = \exp(|\varepsilon|)$.

Accordingly, Huber suggested the following objective function

$$\rho(\varepsilon) = \begin{cases} \frac{1}{2}\varepsilon^2, & |\varepsilon| \le k \\ k|\varepsilon| - \frac{1}{2}k^2, & |\varepsilon| > k \end{cases}, \quad (3.8)$$

where k is given and called tuning constant. The objective function (3.8) can be interpreted that the behavior at the center of the distribution is like the least square; at extremes, like the least absolute deviations.

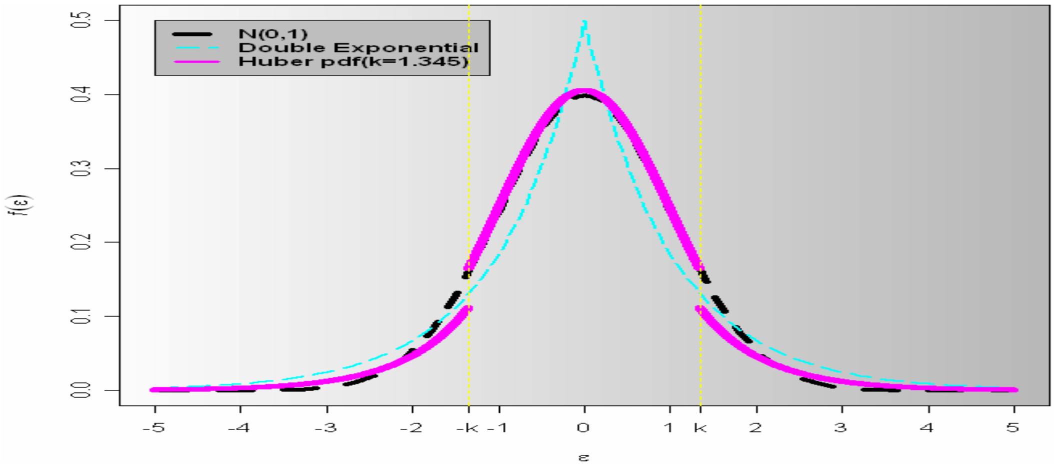

Figure 3.3 The density functions of three error distributions

By changing the value of k, we can manipulate the effect of extreme observation in a flexible way. Huber's objective function can also be related to the negative log-likelihood of a new random variable. In Figure 3.3, we plot the density function of three random variables, namely, the standard normal, the double exponential and Huber's random variable.

The corresponding influence function is given by

$$IF(\varepsilon) \propto \begin{cases} \varepsilon, & |\varepsilon| \le k \\ k \cdot \operatorname{sgn}(\varepsilon), & |\varepsilon| > k \end{cases}. \tag{3.9}$$

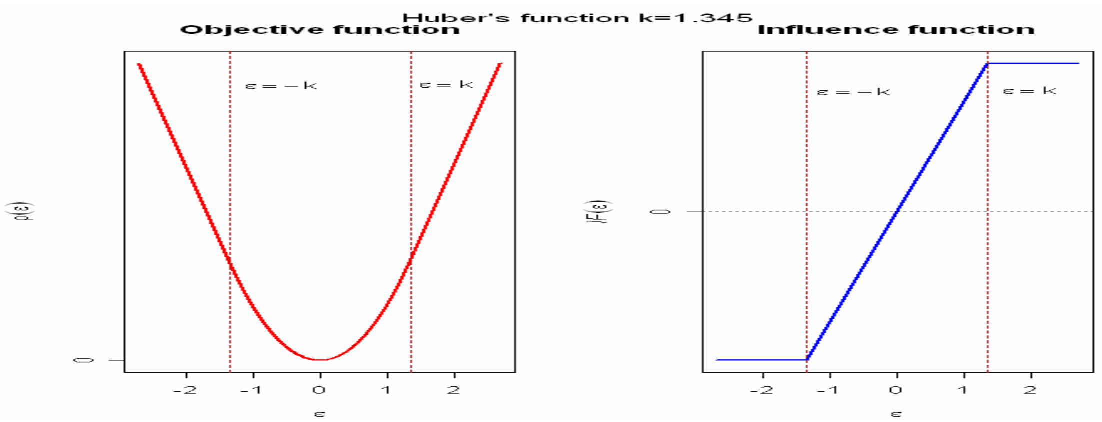

Figure 3.5: The objective function and Influence function of Huber's proposal

The above figure shows that within the range, the information of data is maintained, but, outside the range, the influence of ε is bounded. Therefore, Huber's M estimator possesses the advantages of the LS and LAD estimators but avoid their drawbacks. It still require that ε is symmetric around zero.

3.4 Quantile estimator

When the distribution is not symmetric, the resulting inference procedure has to be adjusted. The quantile regression was first introduced by Koenker and Bassett (1978). Consider the objective function:

$$\rho_\tau(\varepsilon) = (\tau - I(\varepsilon < 0)) \cdot \varepsilon \tag{3.10}$$

where $\tau = P(\varepsilon \le 0)$. We can rewrite (3.10) as

$$\rho_\tau(\varepsilon) = \begin{cases} (1-\tau) \cdot |\varepsilon|, & \text{if } \varepsilon \ge 0 \\ \tau \cdot |\varepsilon|, & \text{if } \varepsilon < 0 \end{cases} \tag{3.11}$$

The corresponding estimator of β is defined as

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho(T_i - \mathbf{Z}_i^T \boldsymbol{\beta}) \tag{3.12}$$

It is more intuitively explainable that if the distribution of error is right-skewed, put more weight on $\varepsilon > 0$, and if the distribution of error is left-skewed, put more weight on $\varepsilon \le 0$. Of course, If we set $\tau = \frac{1}{2}$, then $\rho_{\frac{1}{2}}(\varepsilon) = \frac{1}{2}|\varepsilon|$, and its estimator is equivalent to the least absolute deviations estimator. Therefore, we can view (3.10) as the generalization of the LAD method.

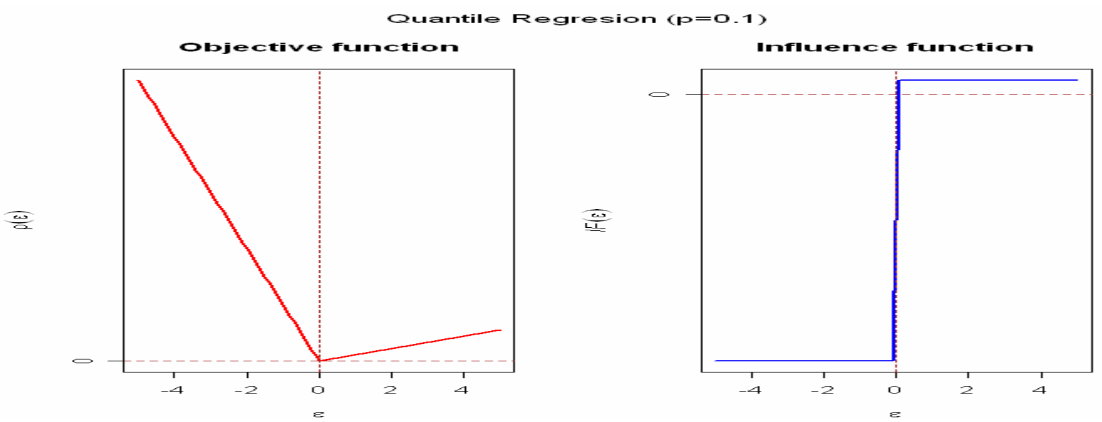

Figure 3.6: The objective function and influence function of quantile estimation

The influence function is

$$IF(\varepsilon) \propto (\tau - I(\varepsilon < 0)). \qquad (3.13)$$

Figure 3.6 shows that the influence is bounded and similar to Figure 3.2 that positive and negative errors contribute to $\rho(\varepsilon)$ with different magnitude. Owing to the bounded influence function, the quantile estimator is not sensitive to outliers of response variable.

3.5 R-estimator

The LAD estimator which minimizes $\sum_{i=1}^{n} |\varepsilon_i|$ is valid if ε_i is symmetric around zero. This assumption is somewhat restricted. If we take two independent replications of ε, denoted as ε_i and ε_j, the distribution of $\varepsilon_i - \varepsilon_j$ is symmetric around zero no matter whether ε is symmetric or not. Accordingly one can consider the objective function

$$\frac{1}{2} \sum \sum_{i<j} |\varepsilon_i - \varepsilon_j|, \qquad (3.14)$$

which is the sum of pairwise absolute deviance. It has been shown that (3.14) can also be written in the form of

$$\sum_{i=1}^{n} (R_i - \overline{R}) \varepsilon_i, \qquad (3.15)$$

where R_i is the rank of ε_i and $\overline{R} = \frac{n+1}{2}$. Equation (3.15) can be compared with the objective function for the LS method $\sum_{i=1}^{n} \varepsilon_i^2$. It is easy to see that, in (3.15), the impact of ε_i is confined to its rank rather than its original size. The rank information preserves some information about the magnitude of ε but is more robust to extreme observations.

Define $\tilde{e}_i = T_i - \sum_{j=1}^{p-1} \beta_j Z_j$ which equal $e_i + \beta_0$. Since the rank relationship is not affected by adding the same constant to each observation, this procedure can only estimate the slope parameters, $(\beta_1, \ldots, \beta_{p-1})$, but not the intercept parameter β_0. The resulting estimator can be expressed as

$$\arg\min \sum_{i=1}^{n} \left(Rank\left(\tilde{e}_i\left(\beta_1, \ldots, \beta_{p-1}\right)\right) - \frac{n+1}{2} \right) \cdot \tilde{e}_i\left(\beta_1, \ldots, \beta_{p-1}\right). \tag{3.16}$$

The idea of using rank information is related to Wilcoxon or Wilcoxon-Mann-Whitney statistics. Jaeckel (1972) pointed out that this estimator is asymptotically equivalent to the following linear rank statistics:

$$U_j(\tilde{\beta}) = \sum_{i=1}^{n} \left(Z_{ij} - \bar{Z}_j \right) \cdot \left(R(\tilde{e}_i) \right) \quad (j = 1, \ldots, p-1) \tag{3.17}$$

where $\bar{Z}_j = \sum_{i=1}^{n} \frac{Z_{ij}}{n}$. The estimator of $(\beta_1, \ldots, \beta_{p-1})$ can be obtained by solving $U_j(\tilde{\beta}) = 0$ for $j = 1, \ldots, p-1$. It is suggested that the intercept term can be estimated as

$$\hat{\beta}_0 = med\left(T_i - \left(Z_1 \hat{\beta}_1 + \ldots + Z_{p-1} \hat{\beta}_{p-1}\right)\right) . \tag{3.18}$$

In summary, the R estimator is valid for symmetric and skewed distributions. It is for estimating the slope parameter but not the intercept. The rank-based procedure is robust to outliers but preserves more data information compared with the LAD method.

Chapter 4 Inference under Censoring

Suppose that the failure time $\tilde{T}$ is subject to censoring by $\tilde{C}$ and hence $h(\tilde{T})$ is subject to censoring by $h(\tilde{C})$. Observed data can be written as

$$\{(X_i, \delta_i, \mathbf{Z}_i),\ i = 1, \ldots, n\},$$

where $X_i = h(\tilde{T}_i) \wedge h(\tilde{C}_i) = T_i \wedge C_i$ and $\delta_i = I(T_i \le C_i)$. We assume that C_i is independent of $(T_i, \mathbf{Z}_i)$. In present of censoring, the error form $e_i(\boldsymbol{\beta}) = T_i - \mathbf{Z}_i^{\mathbf{T}}\boldsymbol{\beta}$ is subject to censoring. Let $e_i^*(\boldsymbol{\beta}) = X_i - \mathbf{Z}_i^{\mathbf{T}}\boldsymbol{\beta}$. Notice that when $\delta_i = 1$, $e_i^*(\boldsymbol{\beta}) = e_i(\boldsymbol{\beta})$; when $\delta_i = 0$, $e_i^*(\beta) < e_i(\beta)$. The Kaplan-Meier estimator for $\Pr(e(\boldsymbol{\beta}) > t)$ is given by

$$\hat{S}_\varepsilon(t|\boldsymbol{\beta}) = 1 - \hat{F}_\varepsilon(t|\boldsymbol{\beta}) = \prod_{i:u \le t}\left(1 - \frac{\sum_{i=1}^{n} I(e_i^*(\boldsymbol{\beta}) = u, \delta_i = 1)}{\sum_{i=1}^{n} I(e_i^*(\boldsymbol{\beta}) \ge u)}\right). \tag{4.1}$$

When $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, the true value of $\boldsymbol{\beta}$, $\hat{F}_\varepsilon(t|\tilde{\boldsymbol{\beta}})$ converges to $\Pr(\varepsilon \le t)$.

In this chapter, we extend the previous discussions to account for the presence of censoring. The challenge is to handle incomplete observation of T_i or $e_i(\boldsymbol{\beta})$. One proposal is to directly modify the objective function, which however was not successful. The other approach is to modify the corresponding estimating function or estimator under censoring. Two useful techniques for handling missing data, namely imputation and weighting have been adopted by many statisticians. We will discuss how these ideas are applied under the least squares estimation in details. For other types of estimation, similar principles can be applied.

4.1 Least Squares Estimation

Several modifications of the least squares method have been proposed to analyze censored data. Miller (1976) proposed to estimate the objective function in (3.2) under censoring using the Kaplan-Meier estimator of ε in (4.1) and then minimize the modified quantity. Buckley-James (1979) estimator and Koul-Susarla-Van Ryzin (1981) estimator both are constructed by directly modifying the least squares solution in (3.2).

4.1.a Modification of the Objective Function

The objective function for the least squares method can be view an estimator of $E(\varepsilon^2)$. In presence of censoring, Miller (1976) suggested to estimate $n \cdot E(\varepsilon^2)$ by

$$n\int_{-\infty}^{\infty} \varepsilon^2 \, d\hat{F}_\varepsilon(\varepsilon \mid \boldsymbol{\beta}),$$

where $\hat{F}_\varepsilon(t \mid \boldsymbol{\beta})$ is given in (4.1). Accordingly, the objective function becomes

$$n\int_{-\infty}^{\infty} \left(e_i^*(\boldsymbol{\beta})\right)^2 d\hat{F}_\beta\left(e_i^*\right) = \sum_{i=1}^{n} \left(e_i^*(\boldsymbol{\beta})\right)^2 d\hat{F}_\beta\left(e_i^*\right) \quad , \tag{4.2}$$

Unfortunately equation (4.2) is a complicated function of $\boldsymbol{\beta}$ and hence is difficult to implement the minimization. Furthermore, the resulting solution does not have reasonable properties, such as consistency.

4.1.b Modification of the LS Solution by Imputation

Buckley and James (1979) suggested to impute T_i by $\hat{E}(T_i | X_i, \delta_i, \mathbf{Z_i})$. This idea is illustrated in Chapter 2. It follows that

$$\mathrm{E}\left(T_i \middle| \delta_i, X_i\right) = \delta_i X_i + (1-\delta_i) E\left(T_i \middle| T_i > C_i, X_i = C_i, \mathbf{Z}_i\right) \; .$$

Under the regression model $E(T_i) = \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}$, it follows that

$$\begin{aligned} & E\left(T_i \middle| T_i > C_i, X_i = C_i, \mathbf{Z}_\mathbf{i}\right) \\ & = E\left(\mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} + \varepsilon_i \middle| \varepsilon_i > X_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}, X_i, \mathbf{Z}_i\right) \\ & = \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} + \frac{\int_{X_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}}^{\infty} u \cdot f_\varepsilon(u) du}{S_\varepsilon\left(X_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}\right)} . \end{aligned} \tag{4.3}$$

Accordingly the above quantity can be estimated by

$$\hat{E}\left(T_i \middle| T_i > X_i, \mathbf{Z}_\mathrm{i}\right) = \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} + \frac{\sum_{k=1}^{n} I\left(e_i^*(\boldsymbol{\beta}) < e_k^*(\boldsymbol{\beta})\right) \cdot d\hat{F}_\varepsilon\left(e_k^* \middle| \boldsymbol{\beta}\right) \cdot e_k^*(\boldsymbol{\beta})}{1 - \hat{F}_\varepsilon\left(e_k^*(\boldsymbol{\beta})\right)} . \tag{4.4}$$

4.1.c Modification of the LS Solution by Weighting

Alternatively, Koul, Susarla, and Van Ryzin (1981) suggested to correct the bias of X_i by weighting. In Chapter 2, we have seen that the Kaplan-Meier estimator can be expressed as a weighted average. It follows that

$$E\left(\delta_i X_i\right) = E\left(I(T_i < C_i) \cdot T_i\right). \tag{4.5}$$

Hence,

$$E\left(\delta_i X_i \middle| T_i\right) = T_i \;\; \Pr(T_i < C_i). \tag{4.6}$$

It follows that if $G(X_i) > 0$, $E\left(\frac{\delta_i X_i}{G(X_i)}\right) = E(T_i)$. They suggested to replace T_i by

$\frac{\delta_i X_i}{\hat{G}(X_i)}$, where

$$\hat{G}(t) = \prod_{u \le t}\left\{1 - \frac{\sum_{i=1}^{n} I\left(e_i^*(\boldsymbol{\beta}) = u, \delta_i = 0\right)}{\sum_{i=1}^{n} I\left(e_i^*(\boldsymbol{\beta}) \ge u\right)}\right\}. \tag{4.7}$$

4.2 Least Absolute Deviation

Minimizing the objective function $\sum_{i=1}^{n}\left|T_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}\right|$ is equivalent to solving the estimating function:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} Z_i\left(I\left(T_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} \geq 0\right) - \frac{1}{2}\right) = 0. \tag{4.8}$$

Since $U(\boldsymbol{\beta})$ is not a continuous function of $\boldsymbol{\beta}$, the solution $\hat{\boldsymbol{\beta}}$ satisfies

$U\left(\hat{\boldsymbol{\beta}}^+\right)U\left(\hat{\boldsymbol{\beta}}^-\right) < 0$.

In presence of censoring censored data, it follows that

$$\begin{aligned} E\left(I\left(X_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} \geq 0\right)\right) &= \Pr\left(T_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} > 0\right)\cdot \Pr\left(C_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} > 0\right) \\ &= \frac{1}{2}G\left(\mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}\right) \end{aligned}. \tag{4.9}$$

Ying, Jung and Wei (1995) proposed the estimating equation resembling (4.8)

$$\sum_{i=1}^{n} Z_i\left(\frac{I\left(X_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} \geq 0\right)}{\hat{G}\left(\mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}\right)} - \frac{1}{2}\right) = 0. \tag{4.10}$$

Notice that the above modification utilizes the weighting technique to correct the censoring bias.

4.3 M-estimator

The M-estimator which minimizes $\sum_{i=1}^{n}\rho\left(T_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}\right)$ can be written as the solution to the following estimating function

$$\sum_{i=1}^{n}\left(\mathbf{Z}_\mathbf{i} - \overline{\mathbf{Z}}\right)\psi\left(T_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}\right) = \mathbf{0}, \tag{4.11}$$

where $\psi(t) = \partial\rho(t)/\partial t$. Ritov (1990) suggested to impute $T_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}$ by an estimator of its conditional expected value given the data. Recall that

$e_i^*(\boldsymbol{\beta}) = X_i - \mathbf{Z}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}$. It follows that

$$E[e_i(\boldsymbol{\beta}) \mid X_i, \delta_i, \mathbf{Z}_i] = \delta_i e_i^*(\boldsymbol{\beta}) + (1-\delta_i)\frac{\int I(\varepsilon > e_i^*(\boldsymbol{\beta}))\varepsilon \cdot dF_{\boldsymbol{\beta}}(\varepsilon)}{S_\beta(e_i^*(\boldsymbol{\beta}))}.$$

The resulting estimating function becomes

$$\sum_{i=1}^{n}(\mathbf{Z_i} - \overline{\mathbf{Z}})\left[\delta_i e_i^*(\boldsymbol{\beta}) + (1-\delta_i)\frac{\int I(\varepsilon > e_i^*(\boldsymbol{\beta}))\varepsilon \cdot d\hat{F}_{\boldsymbol{\beta}}(\varepsilon)}{\hat{S}_{\boldsymbol{\beta}}(e_i^*(\boldsymbol{\beta}))}\right] = \mathbf{0}, \qquad (4.12)$$

where $\hat{S}_{\boldsymbol{\beta}}(t)$ is given in (4.1)

4.4 Quantile Regression

Quantile regression assumes that $\Pr(\varepsilon < 0) = \Pr(T - \mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta} < 0) = \tau$. In presence of censoring, $X = T \wedge C$ is observed instead of T. Now we discuss how this change affects the quantile calculation. It is easy to see that

$$T - \mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta} < 0 \Leftrightarrow T < \mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta} \Leftrightarrow T \wedge \mathrm{C} < \{\mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta} \wedge \mathrm{C}\}.$$

Writing $\mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta} \wedge \mathrm{C} = \min\{\mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta}, \mathrm{C}\}$, we have

$$\Pr(T - \mathbf{Z}^{\mathbf{T}}\beta < 0) = \tau = \Pr(X - \min\{\mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta}, \mathrm{C}\} < 0).$$

This implies that as long as we replace the original $e(\boldsymbol{\beta}) = T - \mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta}$ by $X - \min\{\mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta}, \mathrm{C}\}$, the quantile criteria is the same. That is, the objective function can be written as

$$\rho_\tau(T - \mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta}) = \rho_\tau(\mathrm{X} - \min\{\mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta} \wedge C\}), \qquad (4.13)$$

where $\rho_\tau(\varepsilon) = (\tau - I(\varepsilon < 0)) \cdot \varepsilon$.

However the above expression is still not directly applicable since C is subject to censoring by T. The imputation principle is also applies to replace the unknown C by an estimator of its conditional expected value. Notice that when $\delta = 0$, C is observed but when $\delta = 1$, C is censored. In summary to minimize $(\rho_\tau(X - \min(\mathbf{Z}^{\mathbf{T}}\boldsymbol{\beta}, \mathrm{C})))$, Hornoré, Khan and Powell (1992) proposed to use

$$\frac{1}{n}\sum_{i=1}^{n}\Big((1-\delta_i)\cdot\rho_\tau\left(X_i-\min\{\mathbf{Z_i^T\beta},X_i\}\right) + \delta_i\frac{\int\rho_\tau(X_i-\min\{\mathbf{Z_i^T\beta},c\})\cdot I(X_i<c)d\hat{G}(c)}{1-\hat{G}(X_i)}\Big),$$

where $d\hat{G}(c)=\hat{G}(c-)-\hat{G}(c)$ and $\hat{G}(c)$ is the Kaplan-Meier estimator of $\Pr(C>c)$. Note that the same idea of imputation has been used by Hornoré, Khan and Powell (2002) in M-estimation.

4.5 Rank-based estimators

If we do not consider the previous way to change the structure of data, we can attempt to use the primary information of data – rank. We have discuss that we can explain R-estimator in different meanings but the equivalent result. The following methods are based on modify the solving equation (3.18).

In presence of censoring, direct ranking is impossible. Therefore the alternative expression in terms of pairwise comparison in (3.17) becomes useful. Fygenson and Ritov (1994) suggested to select "comparable pair" $I\left(e_j^*(\boldsymbol{\beta})>e_i^*(\boldsymbol{\beta}),\delta_i=1\right)$. Note that as long as $\delta_i=1$, we know the value $I\left(e_j(\boldsymbol{\beta})>e_i(\boldsymbol{\beta})\right)$ despite of censoring. They proposed the following estimating function

$$n^{\frac{-3}{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{Z_i}-\mathbf{Z_i})\cdot\delta_i\cdot I\left(e_j^*(\boldsymbol{\beta})>e_i^*(\boldsymbol{\beta})\right), \quad (4.14)$$

which has a nice monotonic property that guarantees unit-root. Note that the resulting estimator is a U-statistic useful for large-sample analysis.

Tsiatis (1990) proposed a log-rank type estimator of the form

$$\sum_{i=1}^{n}w_i(\boldsymbol{\beta})\cdot\delta_i\cdot\left[\mathbf{Z_i}-\bar{Z}_i(\boldsymbol{\beta})\right]=\mathbf{0}, \quad (4.15)$$

where $w_i(.)$ is some nonnegative weight function and

$$\overline{Z}_i(\boldsymbol{\beta}) = \frac{\sum_{j=1}^{n} I\left(e_j^*(\boldsymbol{\beta}) \geq e_i^*(\boldsymbol{\beta})\right) \cdot \mathbf{Z_j}}{\sum_{j=1}^{n} I\left(e_j^*(\boldsymbol{\beta}) \geq e_i^*(\boldsymbol{\beta})\right)} \tag{4.16}$$

can be interpreted as the average value of covariate for subjects in the risk set:

$$R(t \mid \boldsymbol{\beta}) = \{ j : e_j(\boldsymbol{\beta}) \geq t \},$$

for $t = e_i(\boldsymbol{\beta})$ with $\delta_i = 1$. When $w_i(\boldsymbol{\beta}) = 1$, the estimator is the log-rank estimator.

Chapter 5 Numerical Analysis

In this chapter, w evaluate finite-sample performances of several estimators by Monte-Carlo simulations. We consider the following model with $h(t) = \log(t)$ such that

$$\log(T_i) = \beta_0 + \beta_1 Z_i + \varepsilon_i,$$

where $Z_i \sim N(0,1)$ and ε_i follows different types of distributions. Under censoring, the observable variable becomes

$$X = (\beta_0 + \beta_1 Z + \varepsilon) \wedge C,$$

where the censoring variable C is generated from a uniform variable distributed in the interval [-1.5, 1.5]. The overall censoring rates vary between 25% and 30%. Sample size set to be n= 100 and 200 with 1000 replications.

5.1 Error with the Standard Normal Distribution

Consider that ε has the standard normal distribution with the density depicted in Figure 5.1:

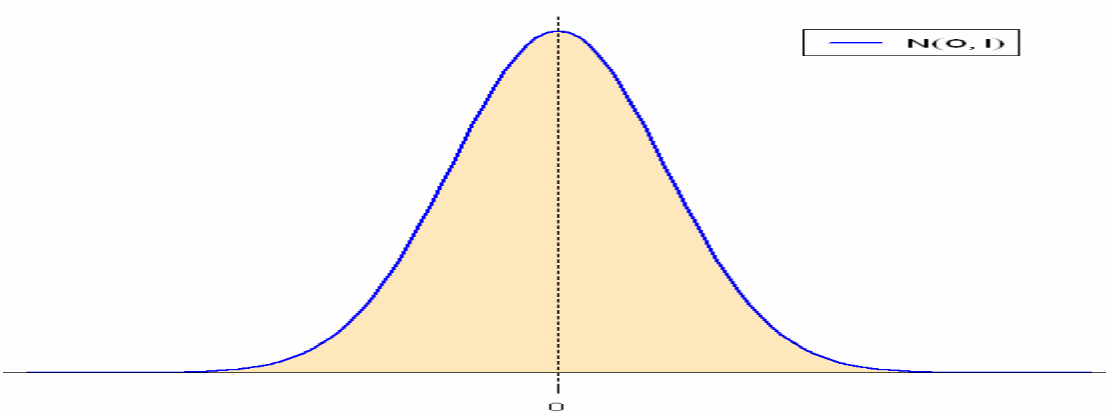

Figure 5.1: $\varepsilon \sim N(0,1)$

The results are given in Table 5.1. Under this case, LAD and the quantile estimator $(\pi = 0.5)$ yield the same result. The two rank-based methods also have nice results. Since ε is symmetric around zero and the chance of observing extreme observations is pretty low, we can expect that the LS has the best performance and all the methods should be valid.

5.2 Error with the Student's T Distribution

Consider that ε has the student's t distribution $T_{(v)}$ with the density

$$f(\varepsilon)=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{\varepsilon^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}\quad(\varepsilon\in R),$$

where v is the degree of freedom. The density for $T_{(2)}$ is depicted in Figure 5.2.

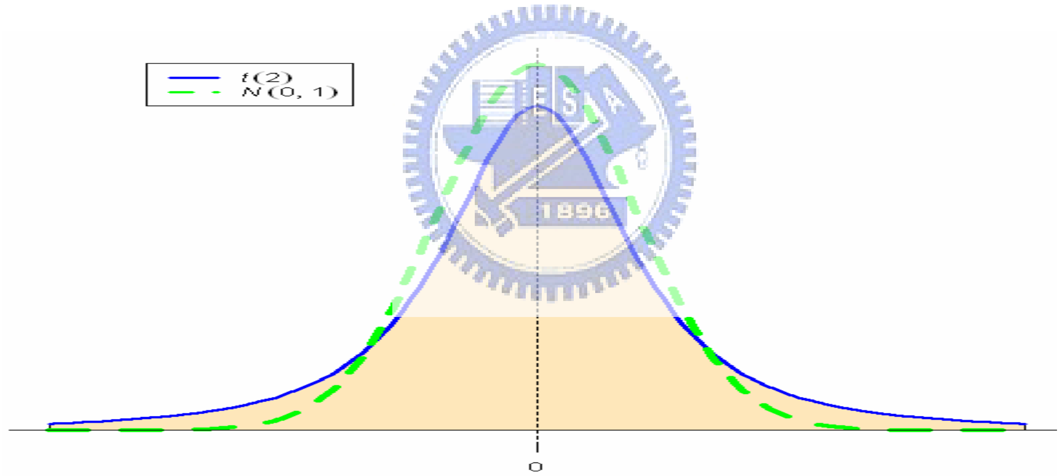

Figure 5.2: $\varepsilon \sim T_{(2)}\ vs.\ \varepsilon \sim N(0,1)$

For comparison, we also plot the density of the standard normal distribution. We see that $T_{(2)}$ tends to produce more extreme observations. Under the adaptive choice of k, Huber's estimator performs the best. The LAD and quantile methods with $\pi = 0.5$ are also the same since ε is symmetric around zero. They become superior to the LS method, which is vulnerable to outliers, under the T distribution with heavy tails. The two log-rank methods still have nice results without being affected by extreme observations.

5.3 Error with the Gumbel Distribution (right-skewed)

Consider that ε has the Gumbel distribution with the density

$$f(\varepsilon)=\frac{1}{\gamma}e^{-\frac{(\varepsilon-\alpha)}{\gamma}}e^{-e^{-\frac{(\varepsilon-\alpha)}{\gamma}}} \quad (\varepsilon\in R),$$

where α is the location parameter and γ is the scale parameter. The density with $(\alpha,\gamma)=(0,5)$ is depicted in Figure 5.3:

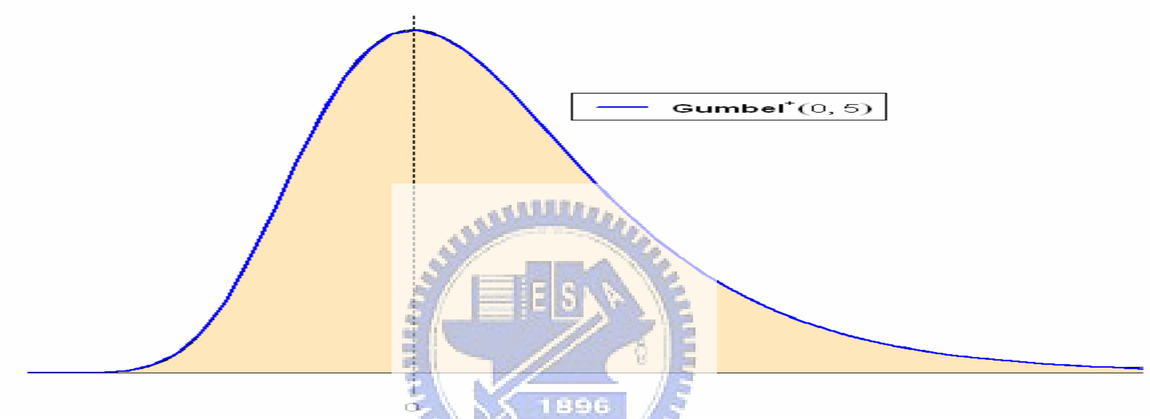

Figure 5.3: $\varepsilon \sim Gumbel^{+}(0,5)$

The distribution is asymmetric such that it yields positive extreme values but with low frequency. Based on the results in Table 5.3, we see that most of the methods cannot accurately estimate β_0, the intercept term, except for the quantile method. Notice that for $Z=0$, $\log(T)=\beta_0+\varepsilon$. It is easy to see that rank-based procedure cannot detect β_0. The first three methods fail too since ε is asymmetric around zero. The quantile can flexibly adjust for this situation. For the slope estimation, all the methods are valid.

5.4 Error with the Gumbel Distribution (left-skewed)

Consider that ε has the Gumbel distribution (left-skewed) with the density

$$f(\varepsilon)=\frac{1}{\gamma}e^{\frac{(\varepsilon-\alpha)}{\gamma}}e^{-e^{\frac{(\varepsilon-\alpha)}{\gamma}}}, \quad \varepsilon\in R,$$

where α is the location parameter and γ is the scale parameter. The density with $(\alpha,\gamma)=(0,5)$ is depicted in Figure 5.4:

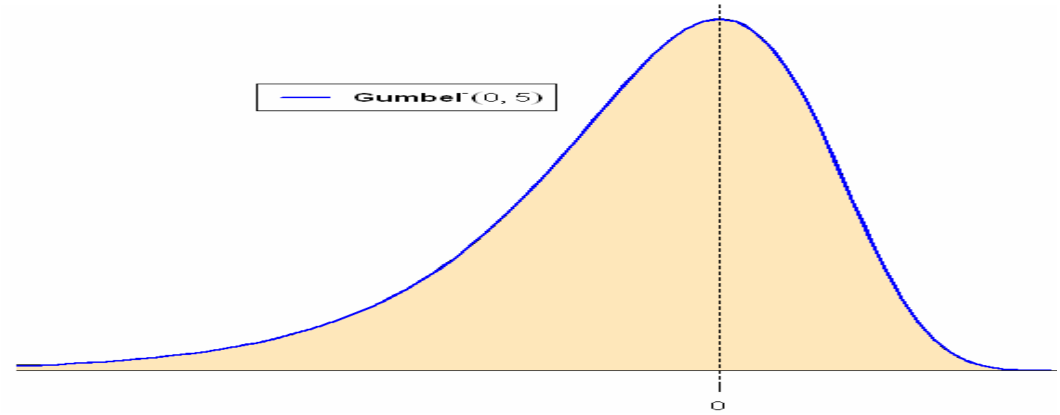

Figure 5.4: $\varepsilon \sim Gumbel^{-}(0,5)$,

The distribution is asymmetric such that yields negative extreme values. Table 5.4 indicates that most of the methods significantly underestimate β_0, except for the quantile method. Similarly the rank-based methods cannot detect β_0 either. The asymmetry of ε also violates the assumption of the first three methods still. Only the quantile estimator can handle this problem. For the slope estimation, all the methods are valid.

5.5 Performances under Censoring

Here we focus on the estimation of β_1, the slope parameter under censoring. From Table 5.5 to Table 5.8, we see the same pattern again that the LS method has better performance under the normal error distribution. Here the LAD and qunatile method with $\tau=0.5$ are no longer the same since they use different methods to adjust for censoring. The imputation approach discussed in §4.4 seems to perform better than the weighting method. Besides, the quantile method still can handle asymmetric error data as well. Most methods, which use the Kaplan-Meier estimator in estimation is much affected by the censoring rate. The linear rank estimator without using Kaplan-Meier estimator is more robust to the censoring condition.

Chapter 6 Conclusion

In this thesis, we consider a class of general linear model and the major objective is to estimate the regression parameter. The underlying error distribution is unknown but usually practitioners may have some prior knowledge about its shape. Such information is useful for choosing an appropriate inference method for parameter estimation. For example, the least squares estimator is a suitable choice if the error is symmetric around zero with light tails like the normal distribution. On the other hand, if the chance of extreme observations is not low, the LAD estimator or rank-based procedures are more appropriate. If the error is skewed, most of the methods can not capture the intercept information except the quantile regression estimator. Generally speaking, the rank-based estimators have nice performances in most situations.

There are two different ways of handling missing data. One is imputation and the other is weighting. We have seen that these two approaches are also adopted for analyzing censored data as well. In the previous discussions, the LS estimator is more sensitive to the tail behavior. It is also more vulnerable to censoring since the Kaplan-Meier estimator does not estimate the tail well. The of quantile estimator which uses the imputation approach to handle censoring is valid in most situations. The rank-based estimators perform pretty well even under censoring. In fact, they are frequently adopted in related problems.

As mentioned at the beginning, another area of research considers the situation that the distribution of ε is known but the form of transformation is unspecified. Such a model is called the transformation model. It seems that most research in this area has not addressed on the relationship between ε and the objective function. We think that this may deserve further investigation.

References

[1] Buckley, J.; James, I. (1979). Linear Regression with Censored Data. *Biometrika,* **66**, 429-436.

[2] Cai, T.; Wei, L. J.; Wilcox, M. (2000) Semiparametric Regression Analysis for Cluster Failure Time Data. *Biometrika*, **87**, 867-878

[3] Cheng, S. C.; Wei, L. J.; Ying, Z. (1995). Analysis of Transformation Models with Censored Data. *Biometrika*, **82**, 835-845.

[4] Fine, J. P.; Ying, Z.;Wei, L. J.(1998). On the linear transformation model for censored data. *Biometrika*, **85**, 980-986.

[5] Fygenson, M.; Ritov, Y. (1994). Monotone Estimating Equations for Censored Data. *The Annals of Statistics*, **22**, 732-746.

[6] Harrington, D. P.; Fleming, T. R. (1982). A Class of Rank Test Procedures for Censored Survival Data. Biometrika, **69**, 553-566.

[7] Honoré, B. E.; Khan, S.; Powell, J. L. (2002). Quantile Regression under Random Censoring. *Journal of Econometrics*, **109**, 67-106.

[8] Huang, J.; Ma, S.; Xie, H. (2005). Least Absolute Deviations Estimation for the Accelerated Failure Time Model. The University of Iowa, Department of Statistics and Actuarial Science, Technical Report No. 350.

[9] Huber, Peter. J. (1981). *Robust Statistics*. Wiley.

[10] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. Ann. *Math. Statist*., **43**, 1449-1458.

[11] Khan, S.; Tamer, E. (2007). Partial Rank Estimation of Duration Models with General Forms of Censoring. *Journal of Econometrics*, **136**, 251–280.

[12] Koenker, R.; Bassett, G.S.(1978). Regression quantiles. *Econometrica*, **46**, 33–50.

[13] Koul, H.; Susarla, V.; Ryzin, J. V. (1981). Regression Analysis with Randomly Right-Censored Data. *The Annals of Statistics,* **9**, 1276-1288.

[14] McKean, J., W. (2004). Robust analysis of linear models. *Statistical Science,* **19**, 562–570.

[15] Miller, R. G. (1976). Least Squares Regression with Censored Data. *Biomttrika*, **63**, 449-64.

[16] Portoy, S. (2003). Censored Regression Quantiles. *Journal of the American Statistical Association*, **98**, 1001-1012.

[17] Powell, J. L. (1984). Least Absolute Deviations Estimation for the Censored Regression Model. *Journal of Econometrics*, **25**, 303-325.

[18] Ritov, Y. (1990). Estimation in a Linear Regression Model with Censored Data. *The Annals of Statistics*, **18**, 303-328.

[19] Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J.* Amer. *Statist. Assoc*. **63** 1379-1389.

[20] Song, X.; Huang, J.; Zhou, X. H. (2006). A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics,* **0**, 1–15.

[21] Tsiatis, A. A. (1990). Estimating Regression Parameters Using Linear Rank Tests for Censored Data. *The Annals of Statistics*, **18**, 354-372.

[22] Ying , Z.; Jung, S. H.; Wei, L. J. (1995). Survival Analysis with Median Regression Models. *Journal of the American Statistical Association*, **90**, 178-184.

[23] Yang, S. (1999). Censored Median Regression Using Weighted Empirical Survival and Hazard Functions. *Journal of the American Statistical Association*, **94**, 137-145.

Appendix

Table 5.1. A: Performances of Different Estimators

censoring rate $= 0$, $\varepsilon \sim N(0,1)$, ***n = 100***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.5$	rank	
$\beta_0 = -2$							
Bias	-0.0013	-0.0012	0.0016	-0.0002	0.0012	0.0014	0.0017
(SD)	(0.1010)	(0.1262)	(0.1049)	(0.0716)	(0.1262)	(0.1012)	(0.1010)
$\beta_1 = 3$							
Bias	0.0024	-0.0025	0.0021	-0.0023	-0.0025	-0.0014	-0.0020
(SD)	(0.0999)	(0.1275)	(0.1044)	(0.1010)	(0.1275)	(0.1100)	(0.1031)

Table 5.1. B: Performances of Different Estimators

censoring rate $= 0$, $\varepsilon \sim N(0,1)$, ***n = 200***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.5$	rank	
$\beta_0 = -2$							
Bias	-0.0003	-0.0001	0.0000	-0.0002	-0.0001	-0.0012	-0.0013
(SD)	(0.0710)	(0.0880)	(0.0735)	(0.0716)	(0.0880)	(0.0701)	(0.0703)
$\beta_1 = 3$							
Bias	-0.0000	-0.0011	0.0007	-0.0003	-0.0011	-0.0006	0.0010
(SD)	(0.0720)	(0.0901)	(0.0750)	(0.0729)	(0.0900)	(0.0719)	(0.0772)

Table 5.2. A: Performances of Different Estimators

censoring rate $= 0$, $\varepsilon \sim T(2)$, ***n = 100***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.5$	rank	
$\beta_0 =$ -2							
Bias	0.0019	0.0018	0.0009	0.0008	0.0018	0.0028	0.0032
(SD)	(0.3515)	(0.1438)	(0.1366)	(0.1445)	(0.1438)	(0.3474)	(0.3475)
$\beta_1 = 3$							
Bias	0.0144	-0.0058	-0.0065	-0.0071	-0.0058	-0.0064	-0.0066
(SD)	(0.3628)	(0.1491)	(0.1400)	(0.1463)	(0.1491)	(0.1823)	(0.1443)

Table 5.2. B: Performances of Different Estimators

censoring rate $= 0$, $\varepsilon \sim T(2)$, ***n= 200***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.5$	rank	
$\beta_0 =$ -2							
Bias	-0.0016	-0.0011	-0.0009	-0.0005	-0.0011	-0.0016	-0.0015
(SD)	(0.2682)	(0.1423)	(0.0936)	(0.0987)	(0.1423)	(0.2674)	(0.2673)
$\beta_1 = 3$							
Bias	-0.0044	0.0040	0.0020	0.0020	0.0040	0.0025	0.0023
(SD)	(0.2665)	(0.1009)	(0.0963)	(0.1009)	(0.1009)	(0.1284)	(0.0997)

Table 5.3. A: Performances of Different Estimators

censoring rate $= 0$, $\varepsilon \sim$ ***Gumbel***$^{+}$***(0, 5), n= 100***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.33$	rank	
$\beta_0 = -2$							
Bias	2.9002	1.8733	1.8856	1.9030	0.0410	2.9025	2.9003
(SD)	(0.6455)	(0.7255)	(0.6942)	(0.6781)	(0.6557)	(0.6455)	(0.6467)
$\beta_1 = 3$							
Bias	-0.0018	0.0038	0.0026	0.0011	-0.0024	-0.0006	-0.0031
(SD)	(0.6497)	(0.7335)	(0.6946)	(0.6791)	(0.6623)	(0.5916)	(0.7776)

Table 5.3. B: Performances of Different Estimators

censoring rate $= 0$, $\varepsilon \sim$ ***Gumbel***$^{+}$***(0, 5), n= 200***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.33$	rank	
$\beta_0 = -2$							
Bias	2.8751	1.8348	1.8512	1.8680	0.0049	2.8756	2.8751
(SD)	(0.4487)	(0.5020)	(0.4741)	(0.4642)	(0.4531)	(0.4484)	(0.4531)
$\beta_1 = 3$							
Bias	0.0176	0.0185	0.0172	0.0169	0.0156	0.0157	0.0196
(SD)	(0.4457)	(0.5086)	(0.4850)	(0.4726)	(0.4627)	(0.4405)	(0.5354)

Table 5.4. A: Performances of Different Estimators

censoring rate = 0, $\varepsilon \sim$ ***Gumbel⁻(0, 5), n = 100***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.67$	rank	
β_0 = -2							
Bias	-2.9044	-1.8955	-1.9058	-1.9220	-0.0700	-2.9022	-2.9042
(SD)	(0.6420)	(0.7200)	(0.6867)	(0.6706)	(0.6633)	(0.6402)	(0.6400)
$\beta_1 = 3$							
Bias	0.0182	0.0167	0.0166	0.0176	0.0263	0.0212	-0.0137
(SD)	(0.6534)	(0.7358)	(0.6867)	(0.6839)	(0.6540)	(0.5984)	(0.5265)

Table 5.4. B: Performances of Different Estimators

censoring rate = 0, $\varepsilon \sim$ ***Gumbel⁻(0, 5), n = 200***

	LS	LAD	M-Huber		Quantile	Linear	Log-rank
			k=1.2	k=1.8	$\tau = 0.67$	rank	
β_0 = -2							
Bias	-2.8918	-1.8479	-1.8634	-1.8800	-0.0327	-2.8914	-2.8921
(SD)	(0.4577)	(0.5148)	(0.4886)	(0.4786)	(0.4722)	(0.4573)	(0.4570)
$\beta_1 = 3$							
Bias	0.0116	0.0020	0.0031	0.0022	0.0012	0.0029	-0.0052
(SD)	(0.4571)	(0.5063)	(0.4867)	(0.4752)	(0.4593)	(0.4109)	(0.3607)

Table 5.5. A: Performances of Different Estimators

censoring rate = 28.1%, $\varepsilon \sim N(0,1)$, ***n = 100***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	$\tau = 0.5$		
$\beta_0 = -2$ Bias (SD)	-0.1861 (0.1321)	-0.3278 (0.3092)	-0.1284 (0.2215)	-0.0062 (0.4535)	0.4136 (0.5294)	-0.6435 (0.2560)	0.7760 (0.2568)
$\beta_1 = 3$ Bias (SD)	-0.1285 (0.2497)	0.0221 (0.7729)	-0.0959 (0.2646)	-0.0551 (0.6660)	0.1766 (0.0650)	0.0004 (0.1448)	-0.2669 (0.1712)

Table 5.5. B: Performances of Different Estimators

censoring rate = 27.6, $\varepsilon \sim N(0,1)$, ***n = 200***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	$\tau = 0.5$		
$\beta_0 = -2$ Bias (SD)	-0.1340 (0.1019)	-0.3392 (0.3446)	-0.1008 (0.1808)	-0.1043 (0.2954)	0.2510 (0.3419)	-0.6317 (0.1999)	0.7709 (0.2118)
$\beta_1 = 3$ Bias (SD)	-0.1210 (0.2087)	-0.0508 (0.6463)	-0.0631 (0.2187)	-0.2116 (0.4516)	0.2857 (0.4334)	0.0015 (0.1012)	-0.2658 (0.1193)

Table 5.6. A: Performances of Different Estimators

censoring rate $= 29.7\%$, $\varepsilon \sim t(2)$, ***n = 100***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	$\tau = 0.5$		
$\beta_0 = -2$							
Bias	-0.5400	-0.1159	-0.6415	-0.0385	0.2041	0.1856	0.4297
(SD)	(0.3301)	(0.6246)	(0.4430)	(0.4368)	(0.4905)	(0.5115)	(0.4994)
$\beta_1 = 3$							
Bias	-0.4040	-1.1075	-0.4281	-0.1335	0.2147	0.0114	-0.5068
(SD)	(0.6289)	(1.1656)	(0.5313)	(0.6987)	(0.5238)	(0.2035)	(0.2582)

Table 5.6. B: Performances of Different Estimators

censoring rate $= 28.1\%$, $\varepsilon \sim t(2)$, ***n = 200***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	$\tau = 0.5$		
$\beta_0 = -2$							
Bias	-0.5073	-0.5717	-0.6300	-0.1108	0.0810	-0.1675	0.4141
(SD)	(0.2982)	(0.8982)	(0.3844)	(0.3013)	(0.2880)	(0.4631)	(0.4594)
$\beta_1 = 3$							
Bias	-0.4413	-1.6455	-0.4163	-0.2581	0.0744	0.0020	0.5124
(SD)	(0.6174)	(1.3123)	(0.5001)	(0.5995)	(0.4775)	(0.1437)	(0.1848)

Table 5.7. A: Performances of Different Estimators

censoring rate = 26.2%, $\varepsilon \sim Gumbel^{+}(0,5)$, ***n = 100***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	τ = 0.37		
β_0 = -2 Bias (SD)	1.1859 (0.5508)	4.6696 (0.5944)	1.4614 (0.9480)	1.6400 (0.7482)	0.1379 (0.6539)	0.1874 (0.4420)	0.8057 (0.3802)
β_1 = 3 Bias (SD)	-0.7790 (0.5270)	-2.4757 (0.5442)	-0.1113 (1.1291)	-0.4304 (0.7822)	0.0519 (0.6867)	-0.0138 (0.5920)	-1.1260 (0.5788)

Table 5.7. B: Performances of Different Estimators

censoring rate = 26.9%, $\varepsilon \sim Gumbel^{+}(0,5)$, ***n = 200***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	τ = 0.37		
β_0 = -2 Bias (SD)	1.2861 (0.3702)	5.1294 (0.3758)	1.5220 (0.7198)	1.7736 (0.5550)	0.0815 (0.4789)	0.1817 (0.3304)	0.8041 (0.2779)
β_1 = 3 Bias (SD)	-0.8018 (0.4247)	-2.9283 (0.3128)	-0.0633 (0.8660)	-0.2357 (0.6233)	0.0424 (0.4947)	0.0007 (0.4154)	-1.1398 (0.4119)

Table 5.8.A: Performances of Different Estimators

censoring rate = 26.2%, $\varepsilon \sim Gumbel^{-}(0,5)$, ***n = 100***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	$\tau = 0.63$		
$\beta_0 = -2$							
Bias	-4.2216	-0.4363	-5.1710	-2.2504	-4.5248	-3.9803	-2.9932
(SD)	(0.8235)	(1.2551)	(0.6908)	(0.7482)	(0.8694)	(0.7141)	(0.7200)
$\beta_1 = 3$							
Bias	-1.0619	-2.3659	-1.4866	-0.6683	0.1208	0.0448	1.8801
(SD)	(0.7909)	(0.7207)	(0.7472)	(0.7100)	(0.8777)	(0.7182)	(0.5088)

Table 5.8.B: Performances of Different Estimators

censoring rate = 25.8%, $\varepsilon \sim Gumbel^{-}(0,5)$, ***n = 200***

	LS			LAD	Quantile	Linear rank	Log -rank
	Objective function	imputation	weighting	(Median)	$\tau = 0.63$		
$\beta_0 = -2$							
Bias	-4.0744	-0.9522	-5.1669	-2.1398	-4.5600	-4.0065	-3.0066
(SD)	(0.6259)	(0.3517)	(0.4993)	(0.5370)	(0.6116)	(0.5334)	(0.5083)
$\beta_1 = 3$							
Bias	-1.1297	-2.4377	-1.4950	-0.5732	0.0320	0.0200	1.9184
(SD)	(0.7273)	(0.2811)	(0.5319)	(0.5040)	(0.6220)	(0.4965)	(0.3483)