2012-07-24

Jackknife empirical likelihood for smoothed weighted rank regression with censored data

Huang, Longlong

UNIVERSITY OF CALGARY

Jackknife empirical likelihood for smoothed weighted rank regression with censored data

by

Longlong Huang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

July, 2012

Abstract

Rank regression is a highly-efficient and robust approach to estimate regression coefficients and to make inference in the presence of outlying survival times. Heller (2007) developed a smoothed weighted rank regression function, which is used to estimate the regression parameter vector in an accelerated failure time model with right censored data. This function can be expressed as a U-statistic. However, since inference is based on a normal approximation approach, it could perform poorly when sample sizes are small and censoring rates are high. To increase inference accuracy and robustness, we propose a jackknife empirical likelihood method for the U-statistic obtained from the estimating function of Heller. The jackknife empirical likelihood ratio is shown to be a standard Chi-squared statistic. Simulations were conducted to compare the proposed method with the normal approximation method. As expected, the new method gives better coverage probability for small samples with high censoring rates. The Stanford Heart Transplant Data, Veterans Administration Lung Cancer Data and Multiple Myeloma Data sets are used to illustrate the proposed method.

Acknowledgements

I would like to express my deepest gratitude and sincere appreciation to my advisors, Dr. Karen Kopciuk and Dr. Xuewen Lu, for their insightful guidance, generous support and kind encouragement, which are highly appreciated and always remembered. They are talented researchers and patient teachers in statistics. Thanks for sharing their great wealth knowledge with me. Working with them is a real pleasure.

I would like to thank my committee members Dr. Murray Burke and Dr. Mingfu Liu for taking some valuable time out of their schedules to review my thesis and for giving me some critical suggestions in my thesis research.

A lot of thanks go to all faculty members in the Department of Mathematics and Statistics in the University of Calgary who have taught and enlightened me along the way. I'm grateful to the financial support of the department.

My special thanks also go to my families, without their selfless love and helpful support I wouldn't make any achievement.

Finally, I will thank all the people who have supported me in my study.

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter 1

Introduction

1.1 An Overview of Regression Methods in Survival Analysis

Survival analysis is widely used in the areas of health, ecology, sociology, economics, insurance, etc. The primary interest of survival analysis is often to understand the relationship between survival times and covariates measured on study participants, such as physical and biological measurements and medical conditions. Examples of survival times include time to death, time until tumor recurrence, time at which a five-year term insurance policy terminates, time until stockmarket crash, time until a machine part fails, and so on. Typically, survival data are not fully observed on all subjects, but rather some values are censored. For example, there may be subjects who choose to quit participating, who move too far away to be followed, or who die from some unrelated event.

For $i = 1, \ldots, n$, let T_i represent the survival time for the i^{th} subject, $\mathbf{X}_i$ be the associated p-dimensional vector of covariates, C_i denote the censoring time for the i^{th} subject and δ_i denote the event indicator, i.e., $\delta_i = I(T_i \leq C_i)$, which takes value 1 if the event time is observed, or 0 if the event time is censored. We define Y_i as the minimum of the survival time and the censoring time, i.e., $Y_i = \min(T_i, C_i)$. Then, the observed data are in the form $(Y_i, \delta_i, \mathbf{X}_i)$, $i = 1, 2, \ldots, n$, which are assumed to be an independent and identically distributed (i.i.d.) sample from $(Y, \delta, \mathbf{X})$. Survival analysis focuses on the distribution of survival times and the association between survival time and risk factors or covariates. The survival function at time t conditional on $\mathbf{X}$ is defined as

$$S(t|\mathbf{X}) = P(T \geq t|\mathbf{X}).$$

The Cox proportional hazards (PH) model is the most prominent regression model used in survival analysis, especially when possible censoring exists. The conditional hazard function for a subject with a p-dimensional covariate vector $\mathbf{X}$ is

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{X}),$$

where $\boldsymbol{\beta}$ is a p-dimensional regression parameter vector and $\lambda_0(t)$ is the baseline hazard, which is usually left unspecified. The semiparametric approach taken in the Cox PH model allows for no assumptions to be made about the functional form of the distribution of survival times. But it does have the proportional hazards assumption, which is the hazards are proportional, or the hazards ratio is assumed constant over the observed survival times.

Estimation and inference of the regression parameters from the Cox PH model are based on the score function

$$\widetilde{Q}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n}\delta_i\left\{\boldsymbol{X}_i - \frac{\sum_{j=1}^{n}\boldsymbol{X}_j I(Y_j \geq Y_i)\exp\left[\boldsymbol{\beta}^T\boldsymbol{X}_j\right]}{\sum_{j=1}^{n} I(Y_j \geq Y_i)\exp\left[\boldsymbol{\beta}^T\boldsymbol{X}_j\right]}\right\},$$

which is derived from the partial likelihood under the proportional hazards assumption. The parameter estimates, $\widehat{\boldsymbol{\beta}}$, are computed as the zero solution to the score equation, $\widetilde{Q}_n(\boldsymbol{\beta}) = 0$.

When the proportional hazards assumption is not satisfied, however, the Cox PH model can produce incorrect regression parameter estimates. The most common alternative approach to the Cox PH model for survival times is the accelerated failure time (AFT) model defined as

$$\log(T_i) = \boldsymbol{\beta}^T\boldsymbol{X}_i + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where the ε_i's are independent identically distributed random errors with an unknown distribution function, and $\boldsymbol{\beta}$ is the regression parameter vector to be estimated. The log survival from the regression residual $\varepsilon^{\boldsymbol{\beta}} = \log(T) - \boldsymbol{\beta}^T\boldsymbol{X}$ can be very large for small failure times, which is an indication that estimation and inference are sensitive to small failure times. Rank regression is one approach to regain robustness with respect to the outlying log survival times.

The following sections of the introduction will review some existing estimation and inference approaches for the AFT model based on rank regression methods.

1.2 Rank Estimation and Inference Methods

Prentice (1978) proposed a linear log-rank test statistic to test the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ is the true value of the parameter of $\boldsymbol{\beta}$. Tsiatis (1990) provided an estimating equation based on linear rank tests, where the observed survival times in the log-rank statistic are replaced by the observed residuals $r_i^{\boldsymbol{\beta}} = \log(Y_i) - \boldsymbol{\beta}^T \boldsymbol{X}_i$. The regression estimate of $\boldsymbol{\beta}$ is determined from the zero crossing of the estimating equation

$$\widetilde{\widetilde{Q}}_n(\boldsymbol{\beta}) = n^{-1/2} \sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{X}_i - \frac{\sum_{j=1}^{n} \boldsymbol{X}_j I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})}{\sum_{j=1}^{n} I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})} \right\}.$$

When $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, this rank estimating function, $\widetilde{\widetilde{Q}}_n(\boldsymbol{\beta})$, is asymptotically normally distributed with mean zero.

Later, Tsiatis (1990) and Ying (1993) extended this function to a weighted rank estimating function

$$\widetilde{\widetilde{Q}}_n(\boldsymbol{\beta}; w) = n^{-1/2} \sum_{i=1}^{n} \delta_i w(r_i^{\boldsymbol{\beta}}) \left\{ \boldsymbol{X}_i - \frac{\sum_{j=1}^{n} \boldsymbol{X}_j I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})}{\sum_{j=1}^{n} I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})} \right\},$$

where weights $w(r_i^{\boldsymbol{\beta}})$ can be chosen to increase the efficiency of the estimator $\widehat{\boldsymbol{\beta}}$. The choice of $w(r_i^{\boldsymbol{\beta}}) = 1$ corresponds to the log-rank type of weights, which is asymptotically efficient when the error distribution is an extreme value distribution.

Fygenson and Ritov (1994) selected the weight function to be $w(r_i^{\boldsymbol{\beta}}) = n^{-1} \sum_{j=1}^{n} I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})$ to produce a monotone rank estimating function

$$\widetilde{S}_n(\boldsymbol{\beta}) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i (\boldsymbol{X}_i - \boldsymbol{X}_j) \left[1 - I(r_j^{\boldsymbol{\beta}} > r_i^{\boldsymbol{\beta}}) \right].$$

Because of the indicator function, $I(r_j^{\boldsymbol{\beta}} > r_i^{\boldsymbol{\beta}})$, this estimating function is not continuous in $\boldsymbol{\beta}$. This discontinuity creates difficulties in the derivation of the asymptotic distribution and computation of the estimator $\widehat{\boldsymbol{\beta}}$.

To overcome these difficulties, recently, Heller (2007) developed a smoothed rank estimating function, which is monotone and continuous with respect to the parameter vector.

He established an inference procedure based on a normal approximation (NA) method. To reduce the influence of outlying covariate values, he introduced a weight function in the smoothed rank estimating function. His weighted estimating function is what we will use to develop a new inference method in this thesis, which will be introduced in next subsection.

1.2.1 Smoothed Weighted Rank Regression with Censored Data

Heller's smoothed weighted rank estimating function for estimating $\boldsymbol{\beta}$ is given by

$$S_n(\boldsymbol{\beta}; w) = (S_{n1}(\boldsymbol{\beta}; w), \ldots, S_{np}(\boldsymbol{\beta}; w))^T,$$

where the k^{th} component is satisfied as

$$S_{nk}(\boldsymbol{\beta}; w) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i (X_{ik} - X_{jk}) w_{ij} \times \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right], \quad k = 1, \ldots, p. \tag{1.1}$$

Here w_{ij} is a weight function defined by

$$w_{ij} = \min\left\{ 1, \frac{1}{\max_k (X_{ik} - X_{jk})^2} \right\},$$

which is symmetric and chosen to reduce the influence of outlying covariate values on the estimator of $\boldsymbol{\beta}$ and its asymptotic variance. The local cumulative distribution function, $\Phi(\cdot)$, is a smooth approximation to the indicator function obtained by choosing a bandwidth, h, to converge to zero; it is usually taken to be the standard normal distribution, which ensures that $S_n(\boldsymbol{\beta}; w)$ is differentiable in $\boldsymbol{\beta}$ and has bounded influence. The bandwidth, h, is used for smoothing purposes and is chosen such that as $n \to \infty$, $nh \to \infty$, and $nh^4 \to 0$. Suppose the regression estimator, $\widehat{\boldsymbol{\beta}}$, is the zero solution of this estimating equation. In practice, as suggested by Heller (2007), h can be set equal to $\hat{\sigma} n^{-.26}$, where $\hat{\sigma}$ is the sample standard deviation of the residuals, $r_i^{\hat{\boldsymbol{\beta}}}$, from uncensored observations. The exponent, -0.26, of n provides the quickest rate of convergence while satisfying the bandwidth constraint of $nh^4 \to 0$.

The estimating function given in (1.1) is monotone and continuous with respect to $\boldsymbol{\beta}$. Heller (2007) obtained the asymptotic normal distribution of $\widehat{\boldsymbol{\beta}}$. Computation of $\widehat{\boldsymbol{\beta}}$ becomes much easier than that of an estimator derived from a non-smoothed estimating equation, and it may be implemented through the standard Newton-Raphson algorithm.

Heller (2007) assumed four regularity conditions (C1-C4) for the proof of the main theorems in his paper. We modify these conditions and use them in our new theorem.

C1. The parameter vector $\boldsymbol{\beta}$ lies in a p-dimensional bounded rectangle $\mathcal{B}$ and the covariate vector, $\mathbf{X}$, $E(\boldsymbol{X}\boldsymbol{X}^T) < M < \infty$.

C2. The term, $n^{-1/2}S_n(\boldsymbol{\beta}; w)$, has a bounded first derivative, $n^{-1/2}A_n(\boldsymbol{\beta}; w)$, in a compact neighborhood of $\boldsymbol{\beta}_0$, with $n^{-1/2}A_n(\boldsymbol{\beta}; w)$ nonzero in that neighborhood.

C3. The local distribution function $\Phi(z)$ is continuous and its derivative $\phi(z) = \partial\Phi(z)/\partial z$ is symmetric about zero with $\int z^2\phi(z) < \infty$.

C4. The bandwidth, h, is chosen such that as $n \to \infty$, both $nh \to \infty$, and $nh^4 \to 0$.

Theorem 1. *[See Heller (2007), Theorem 2.] For the AFT model, under conditions C1-C4, the weighted rank estimating function vector $S_n(\boldsymbol{\beta}; w)$ is a monotone field, is differentiable in $\boldsymbol{\beta}$, and has bounded influence. Then $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}(w)V(w)A^{-T}(w))$, where*

$$A(w) = \lim_{n\to\infty} E\left\{n^{-1/2}\partial S_n(\boldsymbol{\beta}; w)/\partial\boldsymbol{\beta}\right\}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

and

$$V(w) = \lim_{n\to\infty} n^{-1}\,Var\left\{S_n(\boldsymbol{\beta}_0; w)\right\}.$$

The estimated variance-covariance matrix is given by $\Sigma_n = A_n^{-1}(\widehat{\boldsymbol{\beta}}; w)V_n(\widehat{\boldsymbol{\beta}}; w)A_n^{-T}(\widehat{\boldsymbol{\beta}}; w)$, where the (l,m) element of the second derivative matrix $A_n(\boldsymbol{\beta}; w)$ is

$$A_{n(l,m)}(\boldsymbol{\beta}; w) = n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i w_{ij} h^{-1}(X_{il} - X_{jl})(X_{im} - X_{jm})\phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h}\right),$$

and the (l,m) element of $V_n(\boldsymbol{\beta}; w)$ is

$$V_{n(l,m)}(\boldsymbol{\beta}; w) = n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n}(X_{il} - X_{jl})(X_{im} - X_{jm})w_{ij}(e_{ij}^{\boldsymbol{\beta}} - e_{ji}^{\boldsymbol{\beta}})w_{ik}(e_{ik}^{\boldsymbol{\beta}} - e_{ki}^{\boldsymbol{\beta}}),$$

where

$$e_{ij}^{\boldsymbol{\beta}} = \delta_i \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right].$$

So the $(1-\alpha)$-level confidence region for $\boldsymbol{\beta}$ is

$$R_{NA} = \left\{ \boldsymbol{\beta} : (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\Sigma_n/n)^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le \chi^2_{\alpha,p} \right\},$$

where $\chi^2_{\alpha,p}$ is the upper quantile of the Chi-squared distribution with p degrees of freedom. When $p = 1$, the confidence region, R_{NA}, becomes the confidence interval given by

$$CI_{NA} = \left\{ \boldsymbol{\beta} : \widehat{\boldsymbol{\beta}} - Z_{\alpha/2}\sqrt{\Sigma_n/n} \le \boldsymbol{\beta} \le \widehat{\boldsymbol{\beta}} + Z_{\alpha/2}\sqrt{\Sigma_n/n} \right\},$$

where $Z_{\alpha/2}$ is the $(\alpha/2)^{\text{th}}$ upper quantile of the standard normal distribution.

Taking $w_{ij} \equiv 1$, we obtain the smoothed unweighted estimating function $S_n(\widehat{\boldsymbol{\beta}})$ given by Heller (2007) in the following

$$S_n(\boldsymbol{\beta}) = n^{-2/3} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i (\boldsymbol{X}_i - \boldsymbol{X}_j) \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right].$$

The estimator of $\boldsymbol{\beta}$ is the zero solution to this estimating equation and Theorem 1 in Heller (2007) provides the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$.

By imposing weights in the smoothed rank estimating function, Heller (2007) showed that it produced bounded influence in estimation, and the bounded influence provided stability to the regression estimate $\widehat{\boldsymbol{\beta}}$ in the presence of outlying survival times and covariate values.

1.3 Inference with U-statistics and Jackknife Empirical Likelihood Method

In contrast to the NA method, the empirical likelihood (EL) method is an attractive alternative approach to obtain confidence regions without requiring a variance calculation. The EL method has many other nice features: it combines the reliability of nonparametric methods with the effectiveness of the likelihood approach, it employs a simple and efficient algorithm for the constrained maximization problem and confidence regions are invariant under transformations. Its application can be found in many publications. Owen (1988, 1990) first introduced the EL method; Qin and Lawless (1994) established Wilks' theorem for EL in an estimating equation approach; Owen (1991) considered the empirical likelihood method for linear regression; Zhou (2005) considered the EL in an AFT model and Zhao (2011) studied the EL based on Fygenson and Ritov's (1994) estimating equation for the AFT model. Recently, Jing *et al.* (2009) proposed a jackknife empirical likelihood (JEL) method, which combines the jackknife and the empirical likelihood. The most important property of the JEL method is its simplicity, which overcomes computational difficulty in an optimization problem with many nonlinear equations when the sample size gets large. In this thesis, I will develop the JEL method for the AFT model and use this method to analyze the smoothed weighted or unweighted rank regression with censored data.

In the following subsections, we will give an overview about U-statistics, the empirical likelihood and the jackknife method.

1.3.1 U-statistics

In nonparametric problems, U-statistics are often uniformly minimum-variance unbiased estimators, so the use of U-statistics is an effective way of obtaining unbiased estimators.

The basic theory of U-statistics was developed by Hoeffding (1948). Let $X_1, \ldots, X_n$ be a random sample from a distribution function F. Let U_n be a U-statistic with degree m defined by

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} k(X_{i_1}, \ldots, X_{i_m}), \tag{1.2}$$

where the kernel k is symmetric in $X_{i_1}, \ldots, X_{i_m}$. Suppose $\theta = E[k(X_{i_1}, \ldots, X_{i_m})]$ is the parameter of interest. An obvious property of U-statistics is that U_n is an unbiased estimate of θ. For $l = 1, \ldots, m$, let

$$\begin{aligned} k_l(x_1, \ldots, x_l) &= E\left[k(X_1, \ldots, X_m) | X_1 = x_1, \ldots, X_l = x_l\right] \\ &= E\left[k(x_1, \ldots, x_l, X_{l+1}, \ldots, X_m)\right]. \end{aligned}$$

Note that $k_m = k$. Define

$$\tilde{k}_l = k_l - E\left[k(X_1, \ldots, X_m)\right] = k_l - \theta,$$

and $\tilde{k} = \tilde{k}_m$. Then, for any U_n defined by (1.2),

$$U_n - E(U_n) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} \tilde{k}(X_{i_1}, \ldots, X_{i_m}).$$

By Hoeffding's theorem, the variance of the U-statistic given by (1.2) is

$$\mathrm{Var}(U_n) = \binom{n}{m}^{-1} \sum_{l=1}^{m} \binom{m}{l} \binom{n-m}{m-l} \sigma_l^2,$$

where

$$\begin{aligned} \sigma_l^2 &= \mathrm{Var}\left[k_l(X_1, \ldots, X_l)\right] \\ &= E\left[\tilde{k}_l(X_1, \ldots, X_l)\right]^2. \end{aligned}$$

Several properties have been proven and can be found in the literature, such as when $n \to \infty$,

$$\mathrm{Var}(U_n) \sim \frac{m^2 \sigma_1^2}{n},$$

and

$$\sqrt{n}(U_n - \theta) \to N(0, m^2 \sigma_1^2),$$

where $\sigma_1^2 = \mathrm{Var}\left[k_1(X_1)\right] = E\left[\tilde{k}_1(X_1)\right]^2$.

1.3.2 Empirical Likelihood

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from a distribution function F. Suppose the $\mathbf{X}_i$'s are p-dimensional vectors. Empirical likelihood is a non-parametric statistical analysis approach which does not impose any parametric assumptions on the common distribution F. Let

$$F(\{\mathbf{x}_i\}) = P(\mathbf{X}_i = \mathbf{x}_i) = F(\mathbf{x}_i) - F(\mathbf{x}_i -), \qquad i = 1, \ldots, n,$$

where $F(\cdot)$ is right continuous and $\mathbf{x}_i$ is the observed value of $\mathbf{X}_i$. Denote $p_i = F(\{\mathbf{x}_i\})$. Since the $\mathbf{X}_i$'s are assumed to be independent, the empirical likelihood function becomes

$$\begin{aligned} L_n(F) &= \prod_{i=1}^{n} F(\{\mathbf{x}_i\}) \\ &= \prod_{i=1}^{n} p_i, \end{aligned} \tag{1.3}$$

with constraints $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$. Notice that the empirical likelihood function attains its maximum, $L_n(F_n) = n^{-n}$, at $p_i = n^{-1}$. F_n is the empirical cumulative distribution function of the sample data. Then the empirical likelihood ratio is given by

$$\begin{aligned} R_n(F) &= \frac{L_n(F)}{L_n(F_n)} \\ &= \prod_{i=1}^{n}(np_i). \end{aligned}$$

Usually, the targeted applications of the empirical likelihood are inferences on parameters in the form of some functionals of the population distribution F, say $\theta = Q(F)$. Here F is known to be a member of nonparametric distribution family $\mathcal{F}$, where, for example, F could be a class of any distributions and θ could be the population mean. To make inference about θ using a likelihood approach, a likelihood value at θ is needed. The idea behind profile likelihood is to find the value of F at which the empirical likelihood attains the maximum among the set of $Q(F) = \theta$. The profile likelihood function for θ is defined as

$$\begin{aligned} L_n(\theta) &= \sup\left\{ L_n(F) | F(\mathbf{x}) = \sum_{i=1}^{n} p_i I(\mathbf{x}_i \leq \mathbf{x}), Q(F) = \theta, F \in \mathcal{F} \right\} \\ &= \max\left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \mathbf{x}_i = \theta, p_i \geq 0 \right\}. \end{aligned} \tag{1.4}$$

Then the likelihood ratio function evaluated at θ is

$$\begin{aligned} R_n(\theta) &= \frac{L_n(\theta)}{L_n(F_n)} \\ &= \frac{L_n(\theta)}{n^{-n}} \\ &= \max\left\{ \prod_{i=1}^{n}(np_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \mathbf{x}_i = \theta, p_i \geq 0 \right\}. \end{aligned} \tag{1.5}$$

For computation of the profile likelihood, since the log is a monotone transformation, take the logarithms of (1.3) and (1.4), and we obtain

$$l_n(F) = \sum_{i=1}^{n} \log p_i,$$

$$l_n(\theta) = \max\left\{\sum_{i=1}^{n} \log p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \mathbf{x}_i = \theta, p_i \geq 0\right\}.$$

It is much more convenient to work with the log-empirical likelihood function. In fact, for each given value of θ, $l_n(\theta)$ is the maximum of $l_n(F)$ for all F such that $Q(F) = \theta$. Then to compute $l_n(\theta)$, the numerical problem becomes:

$$\begin{aligned} \text{maximize :} \quad & \sum_{i=1}^{n} \log p_i \\ \text{subject to :} \quad & 0 < p_i < 1, \\ & \sum_{i=1}^{n} p_i = 1, \\ & \sum_{i=1}^{n} p_i \mathbf{x}_i = \theta, \quad i = 1, \ldots, n. \end{aligned}$$

In order to solve this maximization problem, we use the Lagrange multiplier method. Let

$$f(\eta, \lambda) = \sum_{i=1}^{n} \log p_i + \eta(\sum_{i=1}^{n} p_i - 1) - n\lambda^T(\sum_{i=1}^{n} p_i \mathbf{x}_i - \theta),$$

where η and $\lambda = (\lambda_1, \ldots, \lambda_p)^T$ are Lagrange multipliers. After taking derivatives with respect to p_i, η and λ respectively and setting them equal to zero, we can get three equations to determine the values of p_i, η and λ. Then we have

$$\eta = n\lambda^T\theta - n,$$

$$p_i = \frac{1}{n\left\{1 + \lambda^T(\mathbf{x}_i - \theta)\right\}},$$

with λ satisfying

$$f(\lambda) = \sum_{i=1}^{n} \frac{\mathbf{x}_i - \theta}{1 + \lambda^T(\mathbf{x}_i - \theta)} = 0.$$

If we plug p_i back into (1.5), $R_n(\theta)$ can be written as

$$R_n(\theta) = \prod_{i=1}^{n} \frac{1}{1 + \lambda^T(\mathbf{x}_i - \theta)}.$$

If we then take the logarithm of $R_n(\theta)$ and multiply by -2, we obtain the empirical likelihood statistics

$$-2 \log R_n(\theta) = 2 \sum_{i=1}^{n} \log \left\{1 + \lambda^T(\mathbf{x}_i - \theta)\right\}.$$

1.3.3 Jackknife

In statistical inference, we usually need to obtain variances of estimators to construct confidence regions and to test hypotheses. But it is often difficult or impossible to determine the distribution of these statistics, and to obtain estimates of the variances. Resampling methods turn out to be useful in these settings, which enable inference across a wide range of statistics under very general conditions. The jackknife method is one resampling approach to assess the variability of statistics or estimators. It was introduced by Quenouille (1956) to construct a bias estimator that could be used in very general situations. Tukey (1958) suggested that the jackknife estimates be obtained by removing data and then recalculating the estimator. The jackknife method provides a general purpose statistical tool that is easy to implement.

Let D_n be an estimator of θ based on n independent random variables $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ for some function f, denoted as

$$D_n = f(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{X}_i, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n).$$

Generate a jackknife sample $\mathbf{X}^{-i} = \{\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n\}$ by leaving out the i^{th} observation. Then calculate the statistic D_{n-1}^{-i} by applying the estimation process to the sample of $(n-1)$ variables formed from the original data set,

$$D_{n-1}^{-i} = f(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n).$$

Define the jackknife pseudo-values by

$$\widehat{V}_i = nD_n - (n-1)D_{n-1}^{-i}.$$

These pseudo-values assume the same role as the $\mathbf{X}_i$'s in estimating θ, hence the jackknife estimate of θ is given by the average of the pseudo-values

$$D_{n,jack} = \frac{1}{n}\sum_{i=1}^{n} \widehat{V}_i.$$

Shi (1984) showed that the pseudo-values $\widehat{V}_i$ are asymptotically independent under mild conditions. Then the jackknife estimator $D_{n,jack}$ can be viewed as a sample average of approximately independent random variables $\widehat{V}_i$'s. We can then use empirical likelihood to make inference about the population parameter, θ.

1.3.4 JEL Method

Since the jackknife estimator $D_{n,jack}$ of θ is a sample average of approximately independent random variables, $\widehat{V}_i$'s, then the empirical likelihood evaluated at θ is

$$L_n(\theta) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\widehat{V}_i = \theta, p_i \geq 0\right\}.$$

Note that $\prod_{i=1}^{n} p_i$ still attains its maximum at $p_i = n^{-1}$. So we define the jackknife empirical likelihood ratio at θ by

$$R_n(\theta) = \max\left\{\prod_{i=1}^{n} (np_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\widehat{V}_i = \theta, p_i \geq 0\right\}.$$

Using Lagrange multipliers, when

$$\min_{1\leq i\leq n} \widehat{V}_i < \theta < \max_{1\leq i\leq n} \widehat{V}_i, \tag{1.6}$$

we have

$$p_i = \frac{1}{n\left\{1 + \lambda^T(\widehat{V}_i - \theta)\right\}},$$

where λ satisfies

$$f(\lambda) = \sum_{i=1}^{n} \frac{\widehat{V}_i - \theta}{1 + \lambda^T(\widehat{V}_i - \theta)} = 0.$$

The jackknife empirical log-likelihood ratio at θ becomes

$$\log R(\theta) = -\sum_{i=1}^{n} \log\left\{1 + \lambda^T(\widehat{V}_i - \theta)\right\}.$$

When the dimension of θ is 1, i.e., θ is a scalar, under some regularity conditions, Jing *et al.* (2009) showed that $-2\log R(\theta)$ converged to the standard Chi-squared distribution with one degree of freedom. We postulate that their results also hold for a vector θ of dimension p ($p > 1$).

Chapter 2

Methodology

In this chapter, we will express Heller's (2007) smoothed weighted rank estimation function as a U-statistic, then apply the JEL to the U-statistic and then derive and prove a new theorem to show that the JEL ratio is a standard Chi-squared statistic, from which we are able to calculate the confidence region of the unknown parameter vector $\boldsymbol{\beta}$.

2.1 Smoothed Weighted Rank Estimating Function in the Form of U-statistics

Let $\mathbf{Z}_i = (Y_i, \delta_i, \mathbf{X}_i)$, $i = 1, \ldots, n$. We re-express Heller's (2007) smoothed weighted rank estimating function $S_n(\boldsymbol{\beta}; w)$ as a U-statistic with a symmetric kernel function,

$$
\begin{aligned}
& S_n(\boldsymbol{\beta}; w) \\
= \ & n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i (\boldsymbol{X}_i - \boldsymbol{X}_j) w_{ij} \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right] \\
= \ & n^{-3/2} \left\{ \sum_{i<j} \delta_i (\boldsymbol{X}_i - \boldsymbol{X}_j) w_{ij} \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right] + \sum_{j<i} \delta_i (\boldsymbol{X}_i - \boldsymbol{X}_j) w_{ij} \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right] \right\} \\
= \ & n^{-3/2} \left\{ \sum_{i<j} \delta_i (\boldsymbol{X}_i - \boldsymbol{X}_j) w_{ij} \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right] + \sum_{i<j} \delta_j (\boldsymbol{X}_j - \boldsymbol{X}_i) w_{ji} \left[1 - \Phi\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{h} \right) \right] \right\} \\
= \ & n^{-3/2} \sum_{i<j} \left\{ (\boldsymbol{X}_i - \boldsymbol{X}_j) w_{ij} \left\{ \delta_i \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right] + \delta_j \left[1 - \Phi\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{h} \right) \right] \right\} \right\} \\
= \ & \left[n^{-3/2} \binom{n}{2} \right] \left[\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta}) \right] \\
\equiv \ & \frac{n-1}{2n^{1/2}} S_n^*(\boldsymbol{\beta}; w), \qquad (2.1)
\end{aligned}
$$

where $S_n^*(\boldsymbol{\beta}; w)$ is a U-statistic of degree 2

$$
S_n^*(\boldsymbol{\beta}; w) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta}) \equiv U_n(\boldsymbol{\beta}) \qquad (2.2)
$$

with the kernel function

$$k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta}) = (\boldsymbol{X}_i - \boldsymbol{X}_j)w_{ij}\left\{\delta_i\left[1-\Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h}\right)\right] + \delta_j\left[1-\Phi\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{h}\right)\right]\right\}. \quad (2.3)$$

The kernel has zero expectation, i.e., $\theta = E[k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta})] = 0$.

2.2 JEL Method for the Smoothed Weighted Estimating Function

To apply JEL to this U-statistic, the jackknife pseudo-values become

$$\widehat{V}_i(\boldsymbol{\beta}) = nU_n(\boldsymbol{\beta}) - (n-1)U_{n-1}^{(-i)}(\boldsymbol{\beta}), \quad (2.4)$$

where $U_{n-1}^{(-i)}(\boldsymbol{\beta}) = U(\mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1}, \mathbf{Z}_{i+1}, \ldots, \mathbf{Z}_n)$ is the statistic $U_{n-1}(\boldsymbol{\beta})$ computed on the sample of $n-1$ variables formed from the original data set by deleting the i^{th} data value. The population parameter of interest is $\boldsymbol{\beta}$, satisfying $E[k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta})] = 0$.

Taking the expectation of (2.4), we get

$$\begin{aligned}
E\left[\widehat{V}_i(\boldsymbol{\beta})\right] &= E\left[nU_n(\boldsymbol{\beta}) - (n-1)U_{n-1}^{(-i)}(\boldsymbol{\beta})\right] \\
&= E\left[nU_n(\boldsymbol{\beta})\right] - E\left[(n-1)U_{n-1}^{(-i)}(\boldsymbol{\beta})\right] \\
&= nE\left[U_n(\boldsymbol{\beta})\right] - (n-1)E\left[U_{n-1}^{(-i)}(\boldsymbol{\beta})\right] \\
&= n\binom{n}{2}^{-1}\sum_{1\le i<j\le n} E\left[k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta})\right] - \\
&\quad (n-1)\binom{n-1}{2}^{-1}\sum_{1\le i<j\le n-1} E\left[k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta})\right] \\
&= n\times 0 - (n-1)\times 0 \\
&= 0.
\end{aligned}$$

Then the jackknife estimator of $\theta \equiv E\left[k(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta})\right] = 0$ is defined as

$$U_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i(\boldsymbol{\beta}).$$

As previously mentioned in Chapter 1, the log-rank test based on the residuals, $r_i^{\boldsymbol{\beta}} = \log(Y_i) - \boldsymbol{\beta}^T\boldsymbol{X}_i$, is obtained from these hypotheses:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{vs.} \quad H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

In our smoothed weighted rank estimation problem, this test is equivalent to the test

$$H_0 : E\left[\widehat{V}_i(\boldsymbol{\beta})\right]\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = 0 \quad \text{vs.} \quad H_a : E\left[\widehat{V}_i(\boldsymbol{\beta})\right]\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \neq 0.$$

Now apply the empirical likelihood method to $\widehat{V}_i(\boldsymbol{\beta})$'s to obtain the empirical likelihood statistics at the value of $\boldsymbol{\beta}$. The empirical likelihood function at the value of $\boldsymbol{\beta}$ is given by,

$$L(\boldsymbol{\beta}) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \widehat{V}_i(\boldsymbol{\beta}) = 0, p_i \geq 0\right\}.$$

Then define the JEL ratio at $\boldsymbol{\beta}$ by

$$\begin{aligned} R(\boldsymbol{\beta}) &= \frac{L(\boldsymbol{\beta})}{n^{-n}} \\ &= \max\left\{\prod_{i=1}^{n}(np_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \widehat{V}_i(\boldsymbol{\beta}) = 0, p_i \geq 0\right\}. \end{aligned}$$

Using Lagrange multipliers, when 0 is contained in the convex hull of $\widehat{V}_i(\boldsymbol{\beta})$'s, we have

$$p_i = \frac{1}{n\left\{1 + \lambda^T \widehat{V}_i(\boldsymbol{\beta})\right\}} \quad , \tag{2.5}$$

where λ satisfies

$$f(\lambda) = \sum_{i=1}^{n} \frac{\widehat{V}_i(\boldsymbol{\beta})}{1 + \lambda^T \widehat{V}_i(\boldsymbol{\beta})} = 0. \tag{2.6}$$

The jackknife empirical log-likelihood ratio at $\boldsymbol{\beta}$ becomes

$$\log R(\boldsymbol{\beta}) = -\sum_{i=1}^{n} \log\left\{1 + \lambda^T \widehat{V}_i(\boldsymbol{\beta})\right\},$$

then we get

$$-2\log R(\boldsymbol{\beta}) = 2\sum_{i=1}^{n} \log\left\{1 + \lambda^T \widehat{V}_i(\boldsymbol{\beta})\right\}. \tag{2.7}$$

Jing *et al.* (2009) proposed a JEL for one-dimensional U-statistics and obtained asymptotic results. We first extend their method to p-dimensional U-statistics and then apply JEL for the smoothed weighted rank regression estimating function.

Define

$$g(\mathbf{z}; \boldsymbol{\beta}) = E\left[k(\mathbf{z}, \mathbf{Z}_2; \boldsymbol{\beta})\right] - \theta = E\left[k(\mathbf{z}, \mathbf{Z}_2; \boldsymbol{\beta})\right] - 0 = E\left[k(\mathbf{z}, \mathbf{Z}_2; \boldsymbol{\beta})\right]$$

and

$$\sigma_g^2 = \mathrm{Var}\left[g(\mathbf{Z}_1; \boldsymbol{\beta})\right].$$

Theorem 2. *Assume that* $-\infty < E\left[k^2(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \boldsymbol{\beta})\right] < \infty$ *and* $\sigma_g^2 > 0$*, under the conditions of Heller (2007) for the normal approximation of the distribution of* $\widehat{\boldsymbol{\beta}}$*, when* $\boldsymbol{\beta} = \boldsymbol{\beta}_0$*, we have*

$$-2\log R(\boldsymbol{\beta}) \xrightarrow{d} \chi_p^2,$$

as $n \to \infty$*,* $\xrightarrow{d}$ *denotes convergence in distribution.*

Based on this result, a $(1-\alpha)$-level confidence region for $\boldsymbol{\beta}$ is

$$R_{JEL} = \left\{\boldsymbol{\beta} : -2\log R(\boldsymbol{\beta}) \leq \chi^2_{\alpha,p}\right\},$$

where $\chi^2_{\alpha,p}$ is the upper quantile of Chi-squared distribution with p degrees of freedom. Notice that when constructing this confidence region, there is no need to solve any estimating equation nor to estimate the variance matrix.

Proof. We will give a detailed proof under the case of one-dimensional β first, and later, we will provide an outline for the proof of multi-dimensional case. We need several technical lemmas to show the results, following Jing *et al.*'s (2009) arguments, we show them first.

Lemma 1. *Suppose that* $E\left[k^2(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \beta)\right] < \infty$ *and* $\sigma_g^2 > 0$*. Then as* $n \to \infty$*, we have* $P(\min_{1\leq i\leq n} \widehat{V}_i(\beta) < 0 < \max_{1\leq i\leq n} \widehat{V}_i(\beta)) \longrightarrow 1.$

Proof. First check $E\left[k^2(\mathbf{Z}_1, \mathbf{Z}_2; \beta)\right] < \infty$. Since $0 < w_{ij} \leq 1$, $\delta_i = 0$ or 1 and $0 \leq \Phi(\cdot) \leq 1$, from (2.3), we have

$$\begin{aligned}
k^2(\mathbf{Z}_i, \mathbf{Z}_j; \beta) &= (X_i - X_j)^2 w_{ij}^2 \left\{ \delta_i \left[1 - \Phi\left(\frac{r_i^\beta - r_j^\beta}{h} \right) \right] + \delta_j \left[1 - \Phi\left(\frac{r_j^\beta - r_i^\beta}{h} \right) \right] \right\}^2 \\
&\le (X_i - X_j)^2 \times 1 \times \left\{ \delta_i \left[1 - \Phi\left(\frac{r_i^\beta - r_j^\beta}{h} \right) \right] + \delta_j \left[1 - \Phi\left(\frac{r_j^\beta - r_i^\beta}{h} \right) \right] \right\}^2 \\
&\le (X_i - X_j)^2 \{1 + 1\}^2 \\
&= 4(X_i - X_j)^2 \\
&\le 8(X_i^2 + X_j^2).
\end{aligned}$$

Taking the expectation of $k^2(\mathbf{Z}_i, \mathbf{Z}_j; \beta)$,

$$\begin{aligned}
E\left[k^2(\mathbf{Z}_i, \mathbf{Z}_j; \beta)\right] &\le 8E(X_i^2 + X_j^2) \\
&= 8 \times 2E(X_1^2) \\
&= 16E(X_1^2) < \infty.
\end{aligned}$$

By some similar arguments and tedious calculation, we can show that $\sigma_g^2 > 0$.

When $p = 1$, X_i is scalar. It suffices to show that $P(\min_{1\le i\le n} \widehat{V}_i(\beta) \ge 0) \to 0$ and $P(\max_{1\le i\le n} \widehat{V}_i(\beta) \le 0) \to 0$. We will only prove $P(\max_{1\le i\le n} \widehat{V}_i(\beta) \le 0) \to 0$ since the proof of $P(\min_{1\le i\le n} \widehat{V}_i(\beta) \ge 0) \to 0$ can be done similarly.

Let $\xi_{ni} = \psi(\widehat{V}_i(\beta))$, where $\psi(s)$ is a nondecreasing, twice differentiable function with bounded first and second derivatives such that

$$\psi(s) = \begin{cases} 0, & \text{if } s \le 0 \\ a(s), & \text{if } 0 < s < \epsilon \\ 1, & \text{if } s \ge \epsilon, \end{cases} \tag{2.8}$$

with $0 < a(s) < 1$ for $0 < s < \epsilon$. Then,

$$
\begin{aligned}
P(\max_{1\le i\le n} \widehat{V}_i(\beta) \le 0) &= P(\widehat{V}_1(\beta) \le 0, \ldots, \widehat{V}_n(\beta) \le 0) \\
&= P(\xi_{n1} = 0, \ldots, \xi_{nn} = 0) \\
&= P\left(\sum_{i=1}^{n} \xi_{ni} = 0\right) \\
&= P\left(\sum_{i=1}^{n} (\xi_{ni} - E\xi_{n1}) = -nE\xi_{n1}\right) \\
&\le P\left(\left|\sum_{i=1}^{n} (\xi_{ni} - E\xi_{n1})\right| \ge nE\xi_{n1}\right) \\
&\le \frac{E\left[\sum_{i=1}^{n} (\xi_{ni} - E\xi_{n1})\right]^2}{n^2 (E\xi_{n1})^2} \\
&\le \frac{n\mathrm{Var}(\xi_{n1}) + n(n-1)\mathrm{Cov}(\xi_{n1}, \xi_{n2})}{n^2 (E\xi_{n1})^2}.
\end{aligned}
$$

It suffices to show that

(a) $\mathrm{Var}(\xi_{n1}) \le 1$,

(b) $lim_{n\to\infty} E\xi_{n1} \ge c > 0$ for some constant c,

(c) $\mathrm{Cov}(\xi_{n1}, \xi_{n2}) \to 0$.

Proof of (a). $\mathrm{Var}(\xi_{n1}) \le E\xi_{n1}^2 \le 1$.

Proof of (b). By the Hoeffding decomposition, $U_n(\beta) = 2n^{-1} \sum_{i=1}^{n} g(\mathbf{Z}_i; \beta) + \binom{n}{2}^{-1} \sum_{i<j}^{n} \varphi(\mathbf{Z}_i, \mathbf{Z}_j; \beta)$, where $g(\mathbf{z}; \beta) = E\left[k(\mathbf{z}, \mathbf{Z}_1; \beta)\right]$ and $\varphi(\mathbf{z}, \mathbf{y}; \beta) = k(\mathbf{z}, \mathbf{y}; \beta) - g(\mathbf{z}; \beta) - g(\mathbf{y}; \beta)$. Then after some algebraic calculations, we have

$$
\begin{aligned}
\widehat{V}_i(\beta) &= 2g(\mathbf{Z}_i; \beta) + \frac{2}{n-2} \sum_{l=1, l\ne i}^{n} \varphi(\mathbf{Z}_i, \mathbf{Z}_l; \beta) - \binom{n-1}{2}^{-1} \sum_{i<j}^{n} \varphi(\mathbf{Z}_i, \mathbf{Z}_j; \beta) \\
&= 2g(\mathbf{Z}_i; \beta) + R_{ni},
\end{aligned}
$$

where R_{ni} is the remainder term. Using a Taylor series expansion

$$
\begin{aligned}
\xi_{ni} &= \psi(\widehat{V}_i(\beta)) \\
&= \psi\left[2g(\mathbf{Z}_i; \beta) + R_{ni}\right] \\
&= \psi\left[2g(\mathbf{Z}_i; \beta)\right] + \psi'\left[2g(\mathbf{Z}_i; \beta)\right] R_{ni} + \eta_i R_{ni}^2,
\end{aligned}
$$

where $|\eta_i| < M$ for some constant $M > 0$. Here and after, M stands for some generic constant, which could be different for each occasion. Since

$$E\left[R_{ni}^2\right] \leq Mn^{-1}E\varphi^2(\mathbf{Z}_1, \mathbf{Z}_2; \beta) + Mn^{-2}E\varphi^2(\mathbf{Z}_1, \mathbf{Z}_2; \beta) \longrightarrow 0, \tag{2.9}$$

this implies that $R_{ni} \to 0$ in probability. This implies that

$$\begin{aligned} E\xi_{ni} &= E\left\{\psi\left[2g(\mathbf{Z}_i; \beta)\right]\right\} + E\left\{\psi'\left[2g(\mathbf{Z}_i; \beta)\right] R_{ni}\right\} + E\left[\eta_i R_{ni}^2\right] \\ &\to E\left\{\psi\left[2g(\mathbf{Z}_i; \beta)\right]\right\}. \end{aligned} \tag{2.10}$$

Notice that $E\left[g(\mathbf{Z}_i; \beta)\right] = 0$ and $\sigma_g^2 > 0$, we get $P(g(\mathbf{Z}_i; \beta) > 0) > 0$, which implies that $E\left\{\psi\left[2g(\mathbf{Z}_i; \beta)\right]\right\} > 0$. This proves (b).

Proof of (c). Note that $\mathrm{Cov}(\xi_{n1}, \xi_{n2}) = E(\xi_{n1}\xi_{n2}) - E(\xi_{n1})E(\xi_{n2})$. By a Taylor series expansion, we have

$$\xi_{ni} = \psi(\widehat{V}_i(\beta)) = \psi\left[2g(\mathbf{Z}_i; \beta) + R_{ni}\right] = \psi\left[2g(\mathbf{Z}_i; \beta)\right] + \lambda_i R_{ni},$$

where $|\lambda_i| < M$ for some constant $M > 0$. Therefore,

$$\begin{aligned} E(\xi_{n1}\xi_{n2}) &= E\left\{(\psi\left[2g(\mathbf{Z}_1; \beta)\right] + \lambda_1 R_{n1})(\psi\left[2g(\mathbf{Z}_2; \beta)\right] + \lambda_2 R_{n2})\right\} \\ &\to \left(E\left\{(\psi\left[2g(\mathbf{Z}_1; \beta)\right]\right\}\right)^2, \end{aligned}$$

since

$$|E\left\{R_{n1}\psi\left[2g(\mathbf{Z}_2; \beta)\right]\right\}| = |E\left\{R_{n2}\psi\left[2g(\mathbf{Z}_1; \beta)\right]\right\}| \leq ME\left|R_{ni}\right| \leq M(ER_{ni}^2)^{1/2} \to 0$$

and $|ER_{n1}R_{n2}| \leq ER_{n1}^2 \to 0$ from (2.9). Thus (c) follows from this and (2.10). □

Lemma 2. *[See Hoeffding (1948).] If $E\left[k^2(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \beta)\right] < \infty$, we have $\sqrt{n}U_n(\beta)/(2\sigma_g) \xrightarrow{d} N(0, 1)$.*

Lemma 3. *Let $G = n^{-1}\sum_{i=1}^n \widehat{V}_i^2(\beta)$, if $E\left[k^2(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \beta)\right] < \infty$, then we have $G = 4\sigma_g^2 + o(1)$ with probability one.*

Proof. Since $E\left[k^2(\mathbf{Z}_1, \mathbf{Z}_2; \beta)\right] < \infty$, notice that

$$
\begin{aligned}
G &= \frac{1}{n}\sum_{i=1}^{n} \widehat{V}_i^2(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{V}_i(\beta) - U_n(\beta) + U_n(\beta)\right]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{V}_i(\beta) - U_n(\beta)\right]^2 + U_n^2(\beta).
\end{aligned}
$$

From Theorem 1.3 of Lee (1990), we have $\mathrm{Var}\left[U_n(\beta)\right] = n^{-1}\left[4\sigma_g^2 + O(n^{-1})\right]$. Denote the jackknife estimate of $\mathrm{Var}(U_n(\beta))$ by $\widehat{\mathrm{Var}}(JACK)$. Then we have

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{V}_i(\beta) - U_n(\beta)\right]^2 &= (n-1)\widehat{\mathrm{Var}}(JACK) = (n-1)\left[\mathrm{Var}(U_n(\beta)) + o(n^{-1})\right] \\
&= 4\sigma_g^2 + o(1)
\end{aligned}
$$

with probability one. In addition, the strong law of large number for U-statistics results in $U_n(\beta) = o(1)$ a.s. Therefore, $G = 4\sigma_g^2 + o(1)$ a.s, which completes the proof. □

Lemma 4. *Let $H_n = \max_{1\le i\ne j\le n}|k(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \beta)|$, if $E\left[k^2(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \beta)\right] < \infty$, then $H_n = o(n^{1/2})$ with probability one.*

Proof. By a chaining argument, it suffices to prove that $2^{-n/2}\max_{1\le j<2^n}|k(\mathbf{Z}_j, \mathbf{Z}_{2^n}; \beta)| \to 0$ a.s. For each $\epsilon > 0$, we have

$$
\begin{aligned}
&\sum_{n=1}^{\infty} P \left\{\max_{1\le j<2^n}|k(\mathbf{Z}_j, \mathbf{Z}_{2^n}; \beta)| \ge \epsilon 2^{n/2}\right\} \\
&\le \sum_{n=1}^{\infty} 2^n P\left\{|k(\mathbf{Z}_j, \mathbf{Z}_{2^n}; \beta)| \ge \epsilon 2^{n/2}\right\} \\
&= \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} 2^n P\left\{2^{(m+1)/2} > \epsilon^{-1}|k(\mathbf{Z}_1, \mathbf{Z}_2; \beta)| \ge 2^{m/2}\right\} \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{m} 2^n P\left\{2^{(m+1)/2} > \epsilon^{-1}|k(\mathbf{Z}_1, \mathbf{Z}_2; \beta)| \ge 2^{m/2}\right\} \\
&\le \sum_{m=1}^{\infty} 2^{m+1} P\left\{2^{(m+1)/2} > \epsilon^{-1}|k(\mathbf{Z}_1, \mathbf{Z}_2; \beta)| \ge 2^{m/2}\right\} \\
&\le 2\epsilon^{-2} E\left|k(\mathbf{Z}_1, \mathbf{Z}_2; \beta)\right|^2 \\
&< \infty.
\end{aligned}
$$

By the Borel-Cantelli lemma, we obtain the result. □

Corollary 1. *Let* $W_n = \max_{1\le i\le n} \left|\widehat{V}_i(\beta)\right|$*, if* $E\left[k^2(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \beta)\right] < \infty$*, then* $W_n = o(n^{1/2})$ *and* $n^{-1}\sum_{i=1}^n \left|\widehat{V}_i(\beta)\right|^3 = o(n^{1/2})$.

Proof. It is easy to check that

$$
\begin{aligned}
U_n(\beta) &= \frac{1}{n(n-1)} \sum_{l=1}^n \sum_{j=1, j\neq l}^n k(\mathbf{Z}_l, \mathbf{Z}_j; \beta) \\
&= \frac{2}{n(n-1)} \sum_{j=1, j\neq i}^n k(\mathbf{Z}_i, \mathbf{Z}_j; \beta) + \frac{n-2}{n} U_{n-1}^{(-i)}(\beta).
\end{aligned}
$$

Then for any $1 \le i \le n$,

$$
\begin{aligned}
\left|\widehat{V}_i(\beta)\right| &= \left|\frac{2}{n-1} \sum_{j=1, j\neq i}^n k(\mathbf{Z}_i, \mathbf{Z}_j; \beta) - U_{n-1}^{(-i)}(\beta)\right| \\
&\le 3 \max_{1\le i\neq j\le n} \left|k(\mathbf{Z}_i, \mathbf{Z}_j; \beta)\right| = 3H_n.
\end{aligned}
$$

By Lemma 4, $W_n = o(n^{1/2})$ a.s. From this and Lemma 3, we have

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \left|\widehat{V}_i(\beta)\right|^3 &\le W_n \times \frac{1}{n}\sum_{i=1}^n \widehat{V}_i^2(\beta) \\
&= o(n^{1/2}) \times (4\sigma_g^2 + o(1)) = o(n^{1/2}).
\end{aligned}
$$

□

Proof of Theorem 2. Lemma 1 has already showed that when $\min_{1\le i\le n} \widehat{V}_i(\beta) < 0 < \max_{1\le i\le n} \widehat{V}_i(\beta)$ with probability one, the solution of (2.5) and (2.6) exists and is unique. We now show that the root of (2.6) satisfies $|\lambda| = O_p(n^{-1/2})$.

$$
\begin{aligned}
0 = |f(\lambda)| &= \frac{1}{n}\left|\sum_{i=1}^n \widehat{V}_i(\beta) - \lambda \sum_{i=1}^n \frac{\widehat{V}_i^2(\beta)}{1+\lambda \widehat{V}_i(\beta)}\right| \\
&\ge \frac{|\lambda|}{n}\sum_{i=1}^n \frac{\widehat{V}_i^2(\beta)}{1+\lambda \widehat{V}_i(\beta)} - \frac{1}{n}\left|\sum_{i=1}^n \widehat{V}_i(\beta)\right| \\
&\ge \frac{|\lambda|\, G}{1+|\lambda|\, W_n} - \left|\frac{1}{n}\sum_{i=1}^n \widehat{V}_i(\beta)\right|.
\end{aligned}
$$

By Lemma 2, the second term is $O_p(n^{-1/2})$. From Lemma 3, $G = 4\sigma_g^2 + o(1)$ a.s., it follows that $|\lambda|/(1+|\lambda|\, W_n) = O_p(n^{-1/2})$, and hence from Corollary 1,

$$
|\lambda| = O_p(n^{-1/2}). \tag{2.11}
$$

Write $\gamma_i = \lambda \widehat{V}_i(\beta)$. Then from Corollary 1 and (2.11),

$$\max_{1 \le i \le n} |\gamma_i| = O_p(n^{-1/2}) o(n^{1/2}) = o_p(1). \tag{2.12}$$

Expanding (2.6) by a Taylor series expansion,

$$\begin{aligned} 0 &= f(\lambda) = \frac{1}{n} \sum_{i=1}^n \widehat{V}_i(\beta)(1 - \gamma_i + \gamma_i^2/(1+\gamma_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \widehat{V}_i(\beta) - G\lambda + \frac{1}{n} \sum_{i=1}^n \widehat{V}_i(\beta) \gamma_i^2/(1+\gamma_i), \end{aligned}$$

where the last term is bounded by

$$\frac{1}{n} \sum_{i=1}^n \left| \widehat{V}_i(\beta) \right|^3 \lambda^2 |1+\gamma_i|^{-1} = o(n^{1/2}) O_p(n^{-1}) O_p(1) = o_p(n^{-1/2}).$$

Therefore we can write

$$\begin{aligned} \lambda &= G^{-1} \left(\frac{1}{n} \sum_{i=1}^n \widehat{V}_i(\beta) \right) + \rho \\ &= G^{-1} U_n(\beta) + \rho, \end{aligned} \tag{2.13}$$

where $|\rho| = o_p(n^{-1/2})$. Using a Taylor series expansion,

$$\log(1+\gamma_i) = \gamma_i - \gamma_i^2/2 + \eta_i,$$

here for some finite $B > 0$, we have $P(|\eta_i| \le B |\gamma_i|^3, 1 \le i \le n) \to 1$ as $n \to \infty$. Then

$$\begin{aligned} -2 \log R(\beta) &= -2 \sum_{i=1}^n \log(np_i) = 2 \sum_{i=1}^n \log(1+\gamma_i) \\ &= 2 \sum_{i=1}^n \gamma_i - \sum_{i=1}^n \gamma_i^2 + 2 \sum_{i=1}^n \eta_i \\ &= 2n\lambda U_n(\beta) - nG\lambda^2 + 2 \sum_{i=1}^n \eta_i \\ &= \frac{nU_n^2(\beta)}{G} - nG\rho^2 + 2 \sum_{i=1}^n \eta_i. \end{aligned} \tag{2.14}$$

For the first term, we have

$$\frac{nU_n^2(\beta)}{G} \xrightarrow{d} \chi_1^2$$

by Lemmas 2 and 3. For the second term, it follows from Lemma 3 and (2.13), that

$$\left|nG\rho^2\right| = n(4\sigma_g^2 + o(1))o_p(n^{-1}) = o_p(1).$$

For the last term, from Corollary 1 and (2.11), we have

$$\begin{aligned}\left|\sum_{i=1}^{n}\eta_i\right| &\le \sum_{i=1}^{n}|\eta_i| \le B\,|\lambda|^3\sum_{i=1}^{n}\left|\widehat{V}_i(\beta)\right|^3 \\ &= O_p(n^{-3/2})o(n^{3/2}) = o_p(1).\end{aligned}$$

Then it follows that $-2\log R(\beta) \xrightarrow{d} \chi_1^2$.

For the p-dimensional $\boldsymbol{\beta}$ case, using arguments similar to those in Owen (2001), we have

$$\|\lambda\| = O_p(n^{-1/2})$$

and

$$\max_{1\le i\le n}\left\|\widehat{V}_i(\boldsymbol{\beta})\right\| = o(n^{1/2}).$$

Equation (2.14) can be expanded as

$$\begin{aligned}-2\log R(\boldsymbol{\beta}) &= -2\sum_{i=1}^{n}\log(np_i) = 2\sum_{i=1}^{n}\log(1+\gamma_i) \\ &= 2\sum_{i=1}^{n}\gamma_i - \sum_{i=1}^{n}\gamma_i^2 + 2\sum_{i=1}^{n}\eta_i \\ &= 2n\lambda^T U_n(\boldsymbol{\beta}) - n\lambda^T G\lambda + 2\sum_{i=1}^{n}\eta_i \\ &= nU_n^T(\boldsymbol{\beta})G^{-1}U_n(\boldsymbol{\beta}) - n\rho^T G^{-1}\rho + 2\sum_{i=1}^{n}\eta_i.\end{aligned}$$

For the first term, in the limit as $n\to\infty$,

$$nU_n^T(\boldsymbol{\beta})G^{-1}U_n(\boldsymbol{\beta}) \xrightarrow{d} \chi_p^2.$$

For the second term,

$$n\rho^T G^{-1}\rho = no_p(n^{-1/2})O_p(1)o_p(n^{-1/2}) = o_p(1).$$

For the last term,

$$\begin{aligned}\left|\sum_{i=1}^{n}\eta_i\right| &\leq \sum_{i=1}^{n}|\eta_i| \leq B\left\|\lambda\right\|^3\sum_{i=1}^{n}\left\|\widehat{V}_i(\boldsymbol{\beta})\right\|^3 \\ &= O_p(n^{-3/2})o(n^{3/2}) = o_p(1).\end{aligned}$$

Therefore, $-2\log R(\boldsymbol{\beta}) \xrightarrow{d} \chi_p^2$.

$\square$

Chapter 3

Simulation Studies

In this chapter, simulations are conducted to compare the performance of the JEL method and the NA method for conducting inference for $\boldsymbol{\beta}$. In the first simulation study, we will fit a model with a one-dimensional continuous covariate, and then compare the coverage probability and average length of the confidence intervals of these two different methods. In the second simulation study, a model with two-dimensional covariate is fit, and the performance of the proposed JEL method is compared with the NA method in terms of joint coverage probabilities.

In the case of one-dimensional parameter β, each sample $\mathbf{Z}_1^{(j)}, \ldots, \mathbf{Z}_n^{(j)}$, $j = 1, \ldots, B$, is drawn from some underlying population $\mathbf{Z} = (Y, \delta, X)$. We calculate the $(1-\alpha)$-level confidence intervals $\widehat{CI}_j = \left\{\beta : -2\log R(\beta) \leq \chi^2_{\alpha,1}\right\}$, $j = 1, \ldots, B$, and denote the length of $\widehat{CI}_j$ by $\left|\widehat{CI}_j\right|$. Define the indicator function

$$I\left\{\beta \in \widehat{CI}_j\right\} = \begin{cases} 1 & \text{if the parameter value } \beta \text{ falls in the } (1-\alpha)\text{-level confidence interval,} \\ 0 & \text{otherwise.} \end{cases}$$

The coverage probability and average length are calculated by $B^{-1}\sum_{j=1}^{B} I(\beta \in \widehat{CI}_j)$ and $B^{-1}\sum_{j=1}^{B}\left|\widehat{CI}_j\right|$, respectively. In the case of the p-dimensional ($p > 1$) parameter vector $\boldsymbol{\beta}$, the $(1-\alpha)$-level confidence regions are calculated by $\widehat{CR}_j = \left\{\boldsymbol{\beta} : -2\log R(\boldsymbol{\beta}) \leq \chi^2_{\alpha,p}\right\}$. Then the joint coverage probability is given by $B^{-1}\sum_{j=1}^{B} I(\boldsymbol{\beta} \in \widehat{CR}_j)$.

Using the NA approach, the $(1-\alpha)$-level confidence interval for β is calculated as $\left(\hat{\beta} - Z_{\alpha/2}se(\hat{\beta}), \hat{\beta} + Z_{\alpha/2}se(\hat{\beta})\right)$, where $se(\hat{\beta})$ is the standard error of $\hat{\beta}$. Denote the indicator function

$$
\begin{aligned}
& I \quad \left\{\beta \in \left(\hat{\beta}_i - Z_{\alpha/2} se(\hat{\beta}_i), \hat{\beta}_i + Z_{\alpha/2} se(\hat{\beta}_i)\right)\right\} \\
= & \begin{cases} 1 & \text{if the parameter value } \beta \text{ falls in the } (1-\alpha)\text{-level confidence interval,} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

So the average length of the $(1-\alpha)$-level confidence interval is given by $B^{-1} \sum_{i=1}^{B} 2Z_{\alpha/2} se(\hat{\beta}_i)$, and the coverage probability is $B^{-1} \sum_{i=1}^{B} I \left\{\beta \in \left(\hat{\beta}_i - Z_{\alpha/2} se(\hat{\beta}_i), \hat{\beta}_i + Z_{\alpha/2} se(\hat{\beta}_i)\right)\right\}$.

3.1 JEL and NA based on One-dimensional Covariate in Smoothed Weighted Rank Regression with Censored Data

In the following AFT model with only a one-dimensional parameter and covariate

$$\log(T_i) = \beta X_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

we set the true regression coefficient, β, to two. The censoring time C_i was generated from a uniform distribution $U(0, \tau)$, where τ determines the censoring rate (cr). In the simulations we chose τ to produce cr $= \{0, 25\%, 50\%, 75\%\}$.

In this simulation set, four different simulation scenarios are introduced. In the first scenario, the univariate covariate X_i was generated from the standard normal distribution $N(0, 1)$. The error term ε_i was generated from the normal distribution with mean 1 and standard deviation σ. Here, σ represents the strength of the relationship between the covariate X_i and the survival time T_i, which ranged from 1 to 4. These simulations were implemented to examine the performance of the regression estimator when the error distribution is symmetric. The second scenario was similarly structured except the error term ε_i was generated from a Chi-squared distribution with 6 degrees of freedom. These simulations were run to explore the properties of the regression estimator when the error distribution is asymmetric. In the third scenario, the covariate X_i was also generated from the standard normal distribution $N(0,1)$, but the error term ε_i was generated from a contaminated normal distribution. The contaminated normal distribution was obtained by generating 95% of the errors, ε_i, from the normal distribution with mean 1 and standard deviation σ and 5% of the errors, ε_i, was generated from the normal distribution with mean 1 and standard deviation 2σ. In the fourth scenario, a data set with a contaminated covariate and a contaminated normal error distribution were generated to test the robustness of the rank based regression estimates. Ninety-five percent of the values of the covariate X_i were generated from the standard normal distribution $N(0,1)$ and 95% of error term ε_i values were generated from the normal distribution with mean 1 and standard deviation σ. The other 5% of the covariate X_i values were generated from the normal distribution $N(-5,1)$ and 5% of the error term ε_i values were generated from a normal distribution with mean 1 and standard deviation 2σ.

Two different sample sizes were considered: $n = 50$ and 100, and there were $B = 5000$ replications for each simulation setting. Weighted and unweighted rank regression estimates were investigated for both uncontaminated and contaminated data. The results are reported in Tables 3.1 - 3.4.

Heller (2007) has shown that the NA based on the smoothed weighted rank estimating equation can reduce the influence of outlying covariate values on $\hat{\beta}$ and its asymptotic variance. Without losing the robustness of the NA approach, in Tables 3.1 - 3.4, we demonstrate that the coverage probabilities of JEL are greater and closer to the nominal level 95% than those of NA in most cases, including contaminated data. This is particularly true for smaller sample sizes (e.g. $n = 50$) and higher censoring rates (e.g. cr = 75%). This implies that JEL generally has better coverage than NA. On the other hand, the average length of JEL is slightly longer than that of NA. Also, when the sample size increases or the censoring rate decreases, the average length shortens. This may be a result of a larger sample size and lower censoring rate, where less information is susceptible to being lost. The confidence interval based on the NA is symmetric whereas the confidence interval based on the JEL is

Table 3.1: Coverage probabilities and average lengths (in parentheses) of the 95% confidence intervals for the regression parameter, β. Here, $\beta = 2$ when $\varepsilon \sim N(1, \sigma^2)$, τ is the censoring parameter, cr is the censoring rate, the sample size, n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.943(0.620)	0.953(0.646)	0.942(0.624)	0.956(0.639)
	36.9	25	0.936(0.771)	0.950(0.813)	0.930(0.721)	0.944(0.749)
	7.05	50	0.936(1.009)	0.950(1.084)	0.908(0.915)	0.931(0.975)
	1.4	75	0.898(1.510)	0.924(1.829)	0.867(1.327)	0.912(1.783)
2	∞	0	0.942(1.239)	0.952(1.292)	0.933(1.139)	0.946(1.171)
	54.5	25	0.940(1.424)	0.953(1.504)	0.922(1.313)	0.937(1.364)
	7.12	50	0.929(1.724)	0.941(1.846)	0.911(1.563)	0.926(1.646)
	0.965	75	0.901(2.368)	0.915(2.626)	0.865(2.048)	0.894(2.365)
4	∞	0	0.942(2.477)	0.952(2.584)	0.929(2.275)	0.942(2.339)
	163	25	0.942(2.696)	0.954(2.857)	0.927(2.485)	0.945(2.591)
	7.25	50	0.937(3.075)	0.948(3.306)	0.918(2.800)	0.935(2.951)
	0.333	75	0.916(3.958)	0.925(4.336)	0.878(3.442)	0.895(3.764)
			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.941(0.429)	0.946(0.437)	0.945(0.429)	0.956(0.435)
	36.9	25	0.942(0.534)	0.949(0.547)	0.935(0.502)	0.943(0.512)
	7.05	50	0.940(0.698)	0.947(0.724)	0.928(0.648)	0.939(0.669)
	1.4	75	0.930(1.052)	0.941(1.119)	0.909(0.957)	0.925(1.018)
2	∞	0	0.941(0.857)	0.946(0.874)	0.938(0.800)	0.947(0.812)
	54.5	25	0.941(0.980)	0.948(1.005)	0.935(0.918)	0.941(0.937)
	7.12	50	0.944(1.192)	0.951(1.233)	0.933(1.107)	0.940(1.140)
	0.965	75	0.928(1.645)	0.938(1.736)	0.910(1.493)	0.918(1.554)
4	∞	0	0.941(1.715)	0.946(1.749)	0.938(1.600)	0.947(1.624)
	163	25	0.942(1.848)	0.947(1.890)	0.933(1.729)	0.941(1.769)
	7.25	50	0.941(2.113)	0.946(2.186)	0.930(1.967)	0.936(2.024)
	0.333	75	0.930(2.697)	0.942(2.852)	0.913(2.469)	0.927(2.562)

Table 3.2: Coverage probabilities and average lengths (in parentheses) of the 95% confidence intervals for the regression parameter, β. Here, $\beta = 2$ with $\varepsilon \sim \chi^2_{df=6}$, τ is the censoring parameter, cr is the censoring rate, the sample size, n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.949(0.550)	0.955(0.576)	0.934(0.585)	0.940(0.611)
	36.2	25	0.942(0.638)	0.953(0.672)	0.938(0.604)	0.949(0.631)
	6.96	50	0.935(0.796)	0.947(0.854)	0.917(0.716)	0.936(0.770)
	1.407	75	0.905(1.156)	0.939(1.514)	0.879(1.027)	0.930(1.472)
2	∞	0	0.944(1.097)	0.950(1.149)	0.939(1.016)	0.947(1.061)
	47.7	25	0.944(1.153)	0.954(1.209)	0.937(1.041)	0.948(1.080)
	6.25	50	0.934(1.297)	0.947(1.367)	0.915(1.143)	0.930(1.205)
	0.94	75	0.906(1.666)	0.931(1.941)	0.879(1.450)	0.914(1.811)
4	∞	0	0.944(2.194)	0.950(2.299)	0.933(2.025)	0.941(2.116)
	103	25	0.944(2.198)	0.954(2.308)	0.928(1.990)	0.943(2.066)
	7.46	50	0.943(2.231)	0.956(2.336)	0.931(1.975)	0.942(2.050)
	0.293	75	0.912(2.482)	0.928(2.688)	0.882(2.119)	0.904(2.389)
			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.950(0.379)	0.954(0.385)	0.934(0.397)	0.938(0.404)
	36.2	25	0.951(0.440)	0.956(0.448)	0.948(0.413)	0.953(0.419)
	6.96	50	0.946(0.547)	0.956(0.561)	0.934(0.497)	0.945(0.509)
	1.407	75	0.933(0.788)	0.949(0.837)	0.916(0.712)	0.935(0.766)
2	∞	0	0.949(0.757)	0.953(0.769)	0.942(0.708)	0.948(0.720)
	47.7	25	0.950(0.795)	0.956(0.810)	0.941(0.724)	0.946(0.733)
	6.25	50	0.943(0.892)	0.951(0.910)	0.934(0.798)	0.941(0.811)
	0.94	75	0.931(1.154)	0.942(1.196)	0.913(1.024)	0.929(1.068)
4	∞	0	0.949(1.514)	0.953(1.538)	0.942(1.416)	0.947(1.440)
	103	25	0.951(1.515)	0.956(1.548)	0.940(1.391)	0.950(1.411)
	7.46	50	0.952(1.533)	0.956(1.564)	0.936(1.375)	0.947(1.394)
	0.293	75	0.929(1.724)	0.940(1.770)	0.916(1.503)	0.925(1.544)

Table 3.3: Coverage probabilities and average lengths (in parentheses) of the 95% confidence intervals for the regression parameter, β. Here, $\beta = 2$ based on the model with a contaminated normal error distribution, τ is the censoring parameter, cr is the censoring rate, the sample size, n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.945(0.640)	0.955(0.670)	0.944(0.638)	0.955(0.658)
	43	25	0.937(0.781)	0.949(0.826)	0.928(0.753)	0.943(0.790)
	7.1	50	0.935(1.042)	0.946(1.130)	0.909(0.953)	0.928(1.021)
	1.22	75	0.895(1.661)	0.914(2.259)	0.857(1.457)	0.903(1.999)
2	∞	0	0.944(1.279)	0.954(1.339)	0.933(1.177)	0.944(1.218)
	63	25	0.941(1.455)	0.951(1.543)	0.928(1.345)	0.939(1.409)
	7.18	50	0.930(1.780)	0.942(1.926)	0.913(1.629)	0.928(1.734)
	0.84	75	0.898(2.583)	0.913(2.893)	0.863(2.244)	0.889(2.623)
4	∞	0	0.944(2.557)	0.954(2.678)	0.930(2.350)	0.940(2.433)
	187.5	25	0.943(2.766)	0.955(2.943)	0.929(2.554)	0.945(2.682)
	7.28	50	0.941(3.174)	0.951(3.444)	0.921(2.906)	0.938(3.103)
	0.287	75	0.914(4.224)	0.924(4.730)	0.881(3.729)	0.895(4.128)

			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.943(0.445)	0.948(0.455)	0.953(0.440)	0.960(0.448)
	43	25	0.942(0.543)	0.948(0.557)	0.938(0.512)	0.945(0.525)
	7.1	50	0.940(0.724)	0.947(0.755)	0.929(0.678)	0.939(0.706)
	1.22	75	0.929(1.159)	0.931(1.254)	0.906(1.066)	0.916(1.143)
2	∞	0	0.943(0.891)	0.948(0.910)	0.939(0.832)	0.946(0.847)
	63	25	0.942(1.005)	0.948(1.032)	0.936(0.944)	0.942(0.967)
	7.18	50	0.944(1.234)	0.951(1.282)	0.930(1.155)	0.940(1.202)
	0.84	75	0.933(1.786)	0.936(1.929)	0.914(1.654)	0.919(1.748)
4	∞	0	0.943(1.781)	0.948(1.819)	0.939(1.664)	0.946(1.695)
	187.5	25	0.943(1.906)	0.947(1.962)	0.937(1.786)	0.943(1.823)
	7.28	50	0.939(2.187)	0.944(2.270)	0.930(2.043)	0.936(2.119)
	0.287	75	0.933(2.868)	0.944(3.095)	0.914(2.661)	0.928(2.830)

Table 3.4: Coverage probabilities and average lengths (in parentheses) of the 95% confidence intervals for the regression parameter, β. Here, $\beta = 2$ based on the model with a contaminated covariate and a contaminated normal error distribution, τ is the censoring parameter, cr is the censoring rate, the sample size n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.947(0.593)	0.954(0.617)	0.886(0.779)	0.890(0.754)
	33.06	25	0.944(0.736)	0.956(0.778)	0.873(0.836)	0.921(0.770)
	5.98	50	0.933(0.951)	0.942(1.051)	0.809(0.844)	0.895(0.824)
	1.03	75	0.892(1.328)	0.919(1.637)	0.702(0.832)	0.889(1.419)
2	∞	0	0.946(1.185)	0.953(1.233)	0.896(1.174)	0.928(1.112)
	47.7	25	0.941(1.368)	0.954(1.445)	0.871(1.325)	0.926(1.245)
	5.85	50	0.938(1.661)	0.949(1.838)	0.824(1.442)	0.908(1.409)
	0.66	75	0.886(2.283)	0.906(2.913)	0.717(1.411)	0.888(2.067)
4	∞	0	0.946(2.370)	0.953(2.466)	0.868(2.228)	0.905(2.128)
	133.5	25	0.948(2.598)	0.955(2.750)	0.868(2.434)	0.923(2.329)
	5.33	50	0.934(2.986)	0.950(3.267)	0.833(2.694)	0.913(2.565)
	0.185	75	0.902(3.942)	0.915(4.810)	0.741(2.708)	0.894(3.311)
			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.945(0.406)	0.950(0.414)	0.913(0.515)	0.937(0.557)
	33.06	25	0.945(0.500)	0.952(0.514)	0.932(0.562)	0.951(0.586)
	5.98	50	0.942(0.638)	0.947(0.668)	0.895(0.632)	0.919(0.625)
	1.03	75	0.908(0.890)	0.916(1.002)	0.825(0.744)	0.871(0.712)
2	∞	0	0.949(0.812)	0.950(0.830)	0.930(0.812)	0.944(0.837)
	47.7	25	0.947(0.931)	0.950(0.954)	0.911(0.914)	0.927(0.941)
	5.85	50	0.940(1.124)	0.947(1.171)	0.876(1.073)	0.913(1.058)
	0.66	75	0.909(1.541)	0.920(1.730)	0.780(1.304)	0.841(1.217)
4	∞	0	0.949(1.625)	0.950(1.656)	0.906(1.586)	0.920(1.632)
	133.5	25	0.947(1.766)	0.952(1.812)	0.901(1.707)	0.923(1.768)
	5.33	50	0.941(2.025)	0.949(2.097)	0.881(1.939)	0.914(1.951)
	0.185	75	0.921(2.661)	0.932(2.904)	0.783(2.437)	0.863(2.216)

asymmetric, because the latter method constructed from the data set instead of a specified distribution. These results indicate that JEL does not require estimation of the variance and can automatically adjust the orientation of the confidence region by contouring a log likelihood ratio. At the same time, JEL achieves better coverage probability than NA at a cost of having longer confidence interval. In the cases of uncontaminated covariates, using weights or not using does not affect results too much when the censoring rate is not too high. However, in the cases of contaminated covariates, the weighted methods perform much better than the unweighted methods.

3.2 JEL and NA based on Two-dimensional Covariate in Smoothed Weighted Rank Regression with Censored Data

In this set of simulation studies, we simulated the model with a two-dimensional covariate as

$$\begin{aligned}\log(T_i) &= \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i \\ &= \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \qquad i = 1, \ldots, n.\end{aligned}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = (2, -1)^T$. Similar to the one-dimensional simulations, this simulation set also has four scenarios. In the first scenario, X_{1i} and X_{2i} were generated from the standard normal distribution $N(0, 1)$, X_{1i} and X_{2i} are independent; the error term ε_i was generated from the normal distribution $N(1, \sigma^2)$. In the second scenario, the independent X_{1i} and X_{2i} were also generated from the standard normal distribution $N(0, 1)$; the error term ε_i was generated from a Chi-squared distribution with degree freedom of 6. In the third scenario, covariates X_{1i} and X_{2i} were generated the same as previous two scenarios, however now the error term ε_i was generated from a contaminated normal distribution. Ninety-five percent of ε_i values were generated from the normal distribution with mean 1 and standard deviation σ, 5% of ε_i was generated from the normal distribution with mean 1 and standard deviation 2σ.

In the fourth scenario, 95% of the covariates X_{1i} and X_{2i} were generated from the standard normal distribution $N(0, 1)$, and 95% of ε_i was generated from the normal distribution with mean 1 and standard deviation σ; the other 5% of the X_{1i} and X_{2i} were generated from the normal distribution $N(-5, 1)$, and 5% of ε_i was generated from the normal distribution with mean 1 and standard deviation 2σ. All other information was the same as that in the one-dimensional case.

We report the joint coverage probability from NA and JEL respectively in Tables 3.5 - 3.8. The results show that in most cases, JEL has better performance than NA. The joint coverage probabilities under NA are smaller than those under JEL. When the sample size increases, from 50 to 100, the joint coverage probability increases as well. Especially when the censoring rate gets large, the coverage probability based on JEL is much closer to the nominal level than the NA. This again indicates that JEL has greater inference precision than the NA approach. When the covariates distributions are contaminated, the weighted NA and JEL methods work better than the corresponding unweighted methods. The weighted methods perform better even when there is no contamination in the covariates and when the censoring rate is high.

3.3 Summary of Simulation Studies

Based on the simulation studies, we summarize the properties of the proposed JEL method in comparison to NA method as follows. Firstly, when the sample size n increases, the coverage probability becomes closer to the nominal level and the average length becomes shorter under both the JEL and NA approaches. Secondly, as the censoring rate increases, the performance of both the JEL and NA methods deteriorates. However, the JEL method apparently has better coverage probability than NA. Thirdly, like the NA method, the JEL method based on a smoothed weighted rank estimating equation, in contrast to a smoothed unweighted ($w \equiv 1$) rank estimating equation, can reduce the influence of outlying covariate values on

Table 3.5: Joint coverage probabilities of the 95% confidence region for the regression parameter, $\boldsymbol{\beta}$. Here, $\boldsymbol{\beta} = (2, -1)^T$ when $\varepsilon \sim N(1, \sigma)$, τ is the censoring parameter, cr is the censoring rate, the sample size, n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.942	0.958	0.924	0.943
	42.04	25	0.930	0.950	0.912	0.940
	7.06	50	0.918	0.939	0.874	0.912
	1.225	75	0.856	0.931	0.782	0.921
2	∞	0	0.926	0.946	0.917	0.939
	60.5	25	0.931	0.951	0.911	0.932
	7.13	50	0.922	0.938	0.885	0.913
	0.865	75	0.873	0.909	0.812	0.879
4	∞	0	0.926	0.946	0.907	0.929
	171	25	0.932	0.949	0.906	0.928
	7.23	50	0.934	0.946	0.892	0.922
	0.309	75	0.886	0.900	0.838	0.867

			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.943	0.954	0.920	0.935
	42.04	25	0.933	0.948	0.926	0.942
	7.06	50	0.927	0.945	0.908	0.929
	1.225	75	0.905	0.930	0.869	0.911
2	∞	0	0.936	0.947	0.924	0.941
	60.5	25	0.936	0.947	0.923	0.934
	7.13	50	0.929	0.943	0.916	0.928
	0.865	75	0.913	0.923	0.883	0.898
4	∞	0	0.936	0.947	0.923	0.940
	171	25	0.941	0.950	0.923	0.939
	7.23	50	0.932	0.943	0.915	0.930
	0.309	75	0.909	0.920	0.882	0.897

Table 3.6: Joint coverage probabilities of 95% confidence region for the regression parameter $\boldsymbol{\beta}$. Here, $\boldsymbol{\beta} = (2, -1)^T$ with $\varepsilon \sim \chi^2_{df=6}$, τ is the censoring parameter, cr is the censoring rate, the sample size, n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.954	0.965	0.919	0.947
	41.7	25	0.943	0.957	0.927	0.947
	6.93	50	0.919	0.948	0.881	0.928
	1.225	75	0.860	0.954	0.794	0.946
2	∞	0	0.933	0.951	0.922	0.943
	54.3	25	0.932	0.949	0.915	0.944
	6.26	50	0.925	0.950	0.892	0.926
	0.832	75	0.869	0.931	0.808	0.912
4	∞	0	0.933	0.950	0.907	0.931
	111.3	25	0.931	0.953	0.901	0.933
	4.36	50	0.924	0.946	0.896	0.925
	0.271	75	0.878	0.908	0.830	0.886
			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.956	0.961	0.933	0.942
	41.7	25	0.946	0.956	0.933	0.947
	6.93	50	0.937	0.951	0.920	0.943
	1.225	75	0.916	0.948	0.881	0.931
2	∞	0	0.946	0.953	0.931	0.945
	54.3	25	0.941	0.952	0.926	0.939
	6.26	50	0.938	0.952	0.916	0.939
	0.832	75	0.913	0.935	0.882	0.917
4	∞	0	0.946	0.953	0.930	0.943
	111.3	25	0.939	0.949	0.922	0.941
	4.36	50	0.939	0.953	0.922	0.943
	0.271	75	0.916	0.935	0.883	0.911

Table 3.7: Joint coverage probabilities of the 95% confidence region for the regression parameter $\boldsymbol{\beta}$. Here, $\boldsymbol{\beta} = (2, -1)^T$ based on the model with a contaminated normal error distribution, τ is the censoring parameter, cr is the censoring rate, the sample size, n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.941	0.956	0.929	0.945
	43	25	0.929	0.950	0.913	0.940
	7.08	50	0.919	0.937	0.878	0.912
	1.22	75	0.863	0.924	0.790	0.913
2	∞	0	0.927	0.945	0.916	0.934
	63	25	0.930	0.950	0.912	0.930
	7.15	50	0.926	0.937	0.892	0.912
	0.842	75	0.879	0.907	0.818	0.878
4	∞	0	0.927	0.945	0.907	0.926
	187.5	25	0.932	0.950	0.907	0.928
	7.3	50	0.936	0.946	0.898	0.921
	0.287	75	0.897	0.902	0.840	0.870
			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.943	0.954	0.931	0.940
	43	25	0.936	0.947	0.926	0.941
	7.08	50	0.930	0.944	0.910	0.927
	1.22	75	0.908	0.925	0.874	0.903
2	∞	0	0.937	0.949	0.925	0.939
	63	25	0.940	0.947	0.922	0.935
	7.15	50	0.934	0.942	0.921	0.930
	0.842	75	0.918	0.921	0.890	0.897
4	∞	0	0.937	0.949	0.924	0.938
	187.5	25	0.939	0.950	0.923	0.939
	7.3	50	0.935	0.945	0.920	0.930
	0.287	75	0.921	0.921	0.892	0.896

Table 3.8: Joint coverage probabilities of the 95% confidence region for the regression parameter $\boldsymbol{\beta}$. Here, $\boldsymbol{\beta} = (2, -1)^T$ based on the model with contaminated covariates and a contaminated normal error distribution, τ is the censoring parameter, cr is the censoring rate, the sample size, n, is 50 or 100 and the number of replications, B, is 5000.

			$n = 50$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.947	0.961	0.858	0.899
	38.3	25	0.928	0.947	0.833	0.901
	6.09	50	0.915	0.933	0.780	0.889
	0.955	75	0.857	0.927	0.659	0.905
2	∞	0	0.933	0.951	0.859	0.909
	55.2	25	0.935	0.950	0.840	0.908
	6.23	50	0.924	0.936	0.791	0.896
	0.685	75	0.868	0.896	0.682	0.876
4	∞	0	0.933	0.951	0.828	0.890
	162	25	0.936	0.953	0.832	0.906
	6.37	50	0.937	0.945	0.802	0.904
	0.24	75	0.887	0.896	0.716	0.864

			$n = 100$			
			weighted		unweighted	
σ	τ	cr(%)	NA	JEL	NA	JEL
1	∞	0	0.944	0.954	0.882	0.913
	38.3	25	0.935	0.947	0.905	0.929
	6.09	50	0.936	0.943	0.869	0.906
	0.955	75	0.911	0.924	0.789	0.862
2	∞	0	0.938	0.949	0.892	0.920
	55.2	25	0.938	0.952	0.888	0.916
	6.23	50	0.933	0.940	0.867	0.901
	0.685	75	0.914	0.916	0.795	0.862
4	∞	0	0.938	0.949	0.877	0.905
	162	25	0.942	0.952	0.876	0.915
	6.37	50	0.937	0.944	0.862	0.904
	0.24	75	0.920	0.920	0.799	0.869

the estimator of unknown parameter when contaminated covariate values are present; see Table 3.4 and Table 3.8. Fourthly, when covariates are not contaminated, but the error values are contaminated, the NA and JEL methods based on smoothed weighted rank estimation equations also perform better than those based on smoothed unweighted estimation equations in most cases; see Table 3.3 and Table 3.7.

Chapter 4

Application to Case Studies

In this chapter, three real data sets were analyzed to illustrate our proposed method and to make comparisons with other methods. The data sets include the Stanford Heart Transplant Data, the Veterans Administration Lung Cancer Data and the Multiple Myeloma Data. We considered a single continuous covariate in the first two data sets and two continuous covariates in the third data set. The weighted and unweighted methods were used in all the cases.

4.1 Stanford Heart Transplant Data

The Stanford Heart Transplant data can be found in Miller and Halpern (1982), and is obtained by using `attach(stanford2)` inside the R `survival` package. In short, the Stanford heart transplant program began in October 1967 and a total of 184 patients received heart transplants. The information in the data set includes: survival time in days; an indicator of whether the patient was dead or alive by February 1980; the age of the patient in years at the time of transplant; and the T5 mismatch score, which made a distinction between deaths primarily due to rejection of the donor heart by the recipient's immune system and non-rejection related deaths. For 27 of the 184 transplant patients, the T5 mismatch scores were missing because the tissue typing was never completed. The 5 patients with survival times less than 10 days were deleted. In the end, there were 152 cases with a complete data record, which we used to fit this model

$$\log_{10}(T_i) = \beta X_i + \varepsilon_i,$$

where T_i is the survival time, and X_i is the age of the i^{th} patient at heart transplant. In this data set, the censoring rate was cr $= 36\%$ with 55 people still alive at the end of the observation period and 97 deceased individuals.

We used box plots to check the outliers of the observed response Y_i and covariate X_i. An outlier is an element of a data set that distinctly stands out from the rest of the data. Figure 4.1 clearly shows that there were some outliers of the observed survival time and the age of the patients. In the box plot of the observed survival time, the outliers are identified as the largest couple of values in the data set, and appear as the circles to the top of the box plot. And in the box plot of the age of the patients at transplant, the outliers are identified as some small values in the data set, and appear as the circles to the bottom of the box plot. As we discussed in the previous chapters, rank regression is robust to outlying survival times in estimating regression coefficients, and the weight function can reduce the influence of outlying covariate values on the estimator. We used the NA and JEL based methods and both the weighted and unweighted rank estimation equations to obtain results.

Using weighted rank estimation equation, the regression parameter estimate was $\hat{\beta} = -0.0537$ with $se(\hat{\beta}) = 0.0171$. The 95% confidence interval for β is $(-0.0890, -0.0210)$ by JEL and $(-0.0872, -0.0202)$ by NA, respectively, which indicates, by both the NA and JEL methods, a significant negative association exists between age and survival time in this patient population.

Using the unweighted method, we obtained $\hat{\beta} = -0.0255$ with $se(\hat{\beta}) = 0.0107$. The 95% JEL confidence interval for β was $(-0.0436, -0.0035)$ and 95% NA confidence interval for β was $(-0.0465, -0.0044)$. These results were quite different from those based on weighted method. However, the result obtained from the unweighted method is very similar to that of Zhou (2005): $\hat{\beta} = -0.0253$ with a 95% confidence interval based on the NA method of $(-0.0446, -0.0030)$. This is not surprising, since Zhou's method does not take outliers into consideration and it is an unweighted method as well. In this data set, the box plots show that the covariate Age at transplant has outliers, so the result from the weighted method is preferred.

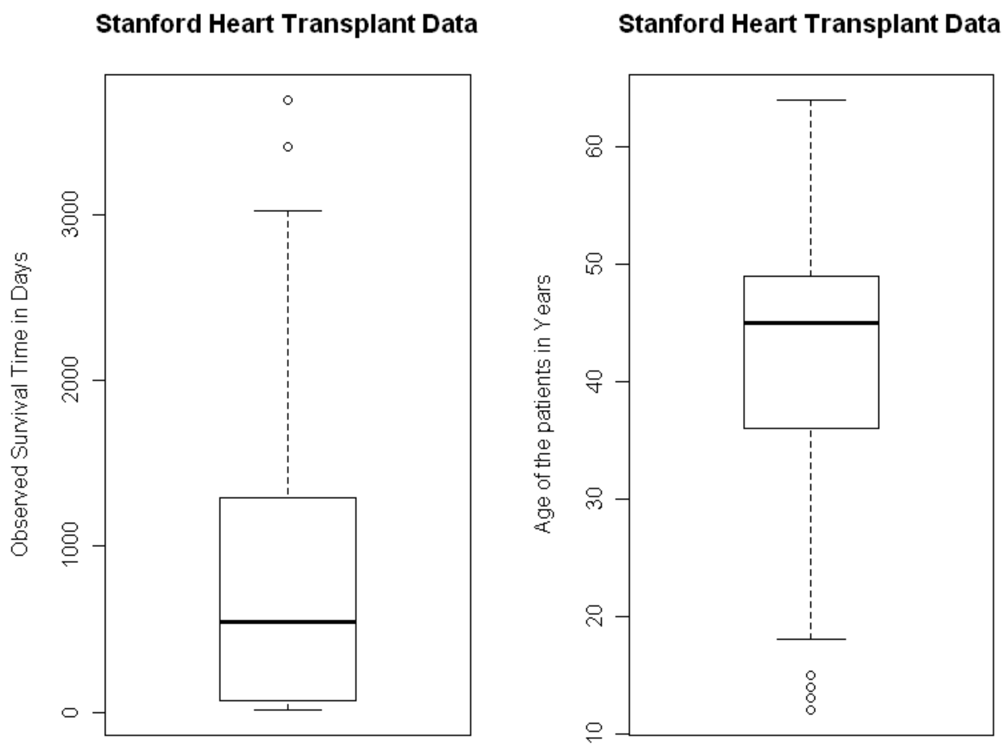

Figure 4.1: Box Plots of the Observed Survival Time and the Age of the Patients in the Stanford Heart Transplant Data set.

4.2 Veterans Administration Lung Cancer Data

The data were derived from a clinical trial of 137 male patients with advanced inoperable lung cancer. The end point was time to death and there were eight variables measured at randomization: treatment, cell type, survival time, censoring status, Karnofsky performance status, age in years, time in months from diagnosis to the start of therapy, and prior therapy. The data set can be found in Kalbfleisch and Prentice (1980), and also can be retrieved by `attach(veteran)` inside the R `survival` package. Nine of the 137 observations were censored, yielding a censoring rate of 6.57%.

We also drew box plots of the observed survival time and the covariate, Karnofsky performance status. Figure 4.2 displays that there were outliers among the observed survival time, while Karnofsky performance status didn't appear to have any outliers. We used both weighted and unweighted methods to estimate the regression coefficient of the Karnofsky performance status.

Based on the weighted method, Heller (2007) examined the relationship between Karnofsky performance status and time to death using the AFT model with one covariate. He obtained $\hat{\beta} = 0.0383$ with $se(\hat{\beta}) = 0.0046$. The 95% JEL confidence interval we calculated was (0.0292,0.0478), which is quite close to the 95% NA confidence interval (0.0293,0.0473). These results imply that higher Karnofsky performance status could prolong the life time of the patients using either method.

Based on the unweighted method, we obtained $\hat{\beta} = 0.0397$ with $se(\hat{\beta}) = 0.0041$, the Karnofsy performance status appears to be associated with survival time. The ninety-five percent confidence interval for β was (0.0316,0.0483) by JEL and (0.0318,0.0477) by NA. The results of weighted and unweighted estimation equations did not differ by much. This makes sense, because the weight function is used to reduce the influence of outlying covariate values on the estimator. If the covariate did not include outliers, we wouldn't expect a big difference between these two methods.

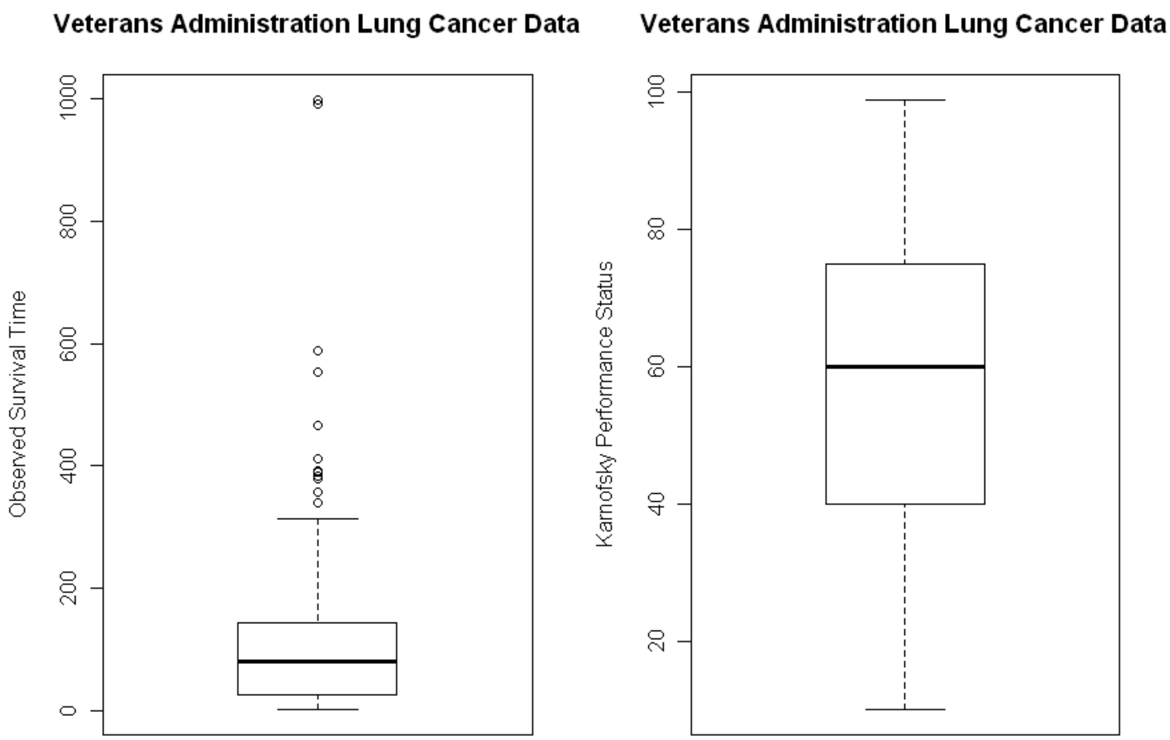

Figure 4.2: Box Plots of the Observed Survival Time and Karnofsky Performance Status in the Veterans Administration Lung Cancer Data set.

4.3 Multiple Myeloma Data

The Multiple Myeloma data were reported by Krall *et al.* (1975), and also can be obtained from *SAS/STAT User's Guide (1999): HW 14.25 data.* The data contained information for two response variables and nine covariates: survival time, censoring status, logarithm of blood urea nitrogen (LogBUN), haemoglobin (HGB), Platelet, Age, LogWBC, Frac, LogPBM, Protein, SCalc. Out of total of 65 observations, 17 were censored.

Following Jin *et al.* (2003), we fit the model with these two covariates: the logarithm of blood urea nitrogen and haemoglobin. As Jin *et al.* (2003) suggested, to improve numerical efficiencies, we standardized these two covariates Log(BUN) and HGB in our analysis. This standardization will transform the original covariates to have zero mean and unit variance.

The fitted model is

$$\log(T_i) = \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i,$$

where X_{1i} is the standardized score of Log(BUN), and X_{2i} is the standardized score of HGB.

From box plots 4.3 we noticed that the standardized score of HGB didn't show any outliers, while the observed survival time and the standardized score of Log(BUN) both had outliers. Again, we used the weighted and unweighted methods in our model estimation.

The smoothed weighted rank estimates of regression coefficients were $(\hat{\beta}_1, \hat{\beta}_2) = (-0.4622, 0.2714)$ with estimated standard errors $(0.1626, 0.1753)$. The 95% Chi-square test statistic with 2 degrees of freedom for (β_1, β_2) based on JEL method and NA method were 11.34 and 11.21, and the corresponding p-values are 0.003 and 0.004, respectively, which are very close to each other, indicating a jointly significant effect of Log(BUN) and HGB.

The unweighted rank estimates of regression coefficients were $(\hat{\beta}_1, \hat{\beta}_2) = (-0.5142, 0.2839)$ with estimated standard errors $(0.1399, 0.1732)$. The 95% Chi-square test statistic with 2 degrees of freedom for (β_1, β_2) based on JEL method and NA method were 16.909 and 16.666, and the corresponding p-values are 0.00021 and 0.00024, respectively. The estimated regression coefficients are similar to Jin *et al.*'s (2003) results which did not consider

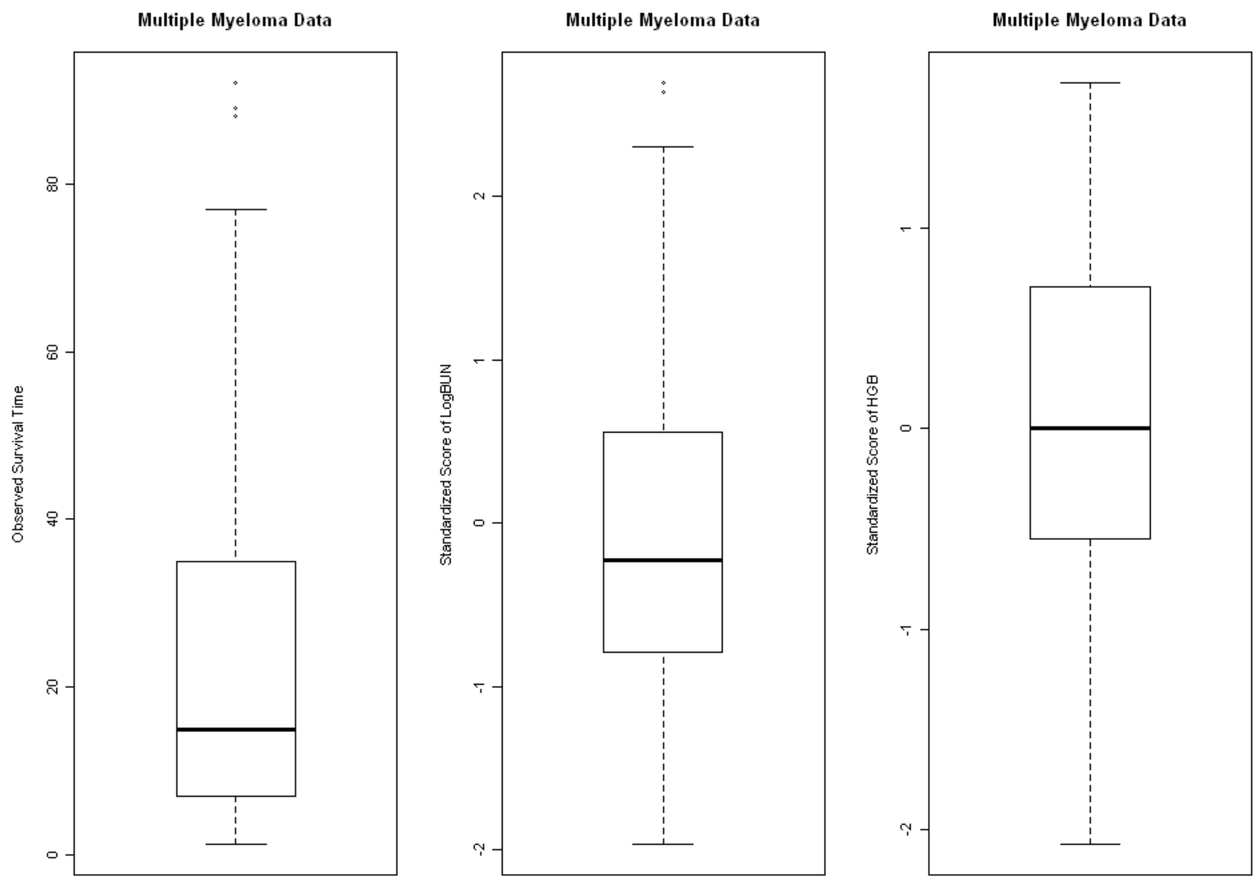

Figure 4.3: Box Plots of the Observed Survival Time, the Standardized Score of LogBUN and the Standardized Score of HGB in the Multiple Myeloma Data set.

outliers: their parameters estimates were $(-0.532, 0.292)$ with estimated standard errors $(0.146, 0.169)$. These results once again substantiate our claim that when covariates have outliers (e.g. standardized score of LogBUN), the rank estimates of regression coefficients under the weighted and unweighted methods may be quite different, and the result under the weighted method is recommended, since it takes the outliers in the covariates into account.

Chapter 5

Conclusion and Discussion

In this thesis, we applied the JEL to Heller (2007)'s smoothed weighted rank regression with censored data. Empirical likelihood and jackknife have been integrated to yield the new method. Our purpose is to conduct inference for the regression parameter $\boldsymbol{\beta}$ in the AFT model while maintaining the robustness of smoothed weighted rank regression and the simplicity of JEL. The key idea of this thesis is to turn the U-statistic of interest into a sample mean based on jackknife pseudo-values. We proved a new theorem, Theorem 2, which can be widely and effectively used in a general class of AFT models in survival analysis. The proposed JEL method preserves some important features of empirical likelihood including the Wilks property of Chi-squared limiting distribution. When using the JEL to carry out inference, there is no need to solve estimating equations or to estimate covariance matrices. Implementation becomes easy in a standard software environment, for example, R package such as emplik already exists to maximize the empirical likelihood functions and obtain the values of the test statistics.

Simulation studies indicate that the proposed JEL method not only carries over superior properties of the NA method such as easy computation and robustness to covariate outliers, but it also improves the accuracy of inferences for regression parameters. This is especially evident when sample size gets smaller or censoring rate gets higher, the JEL method provides more accurate coverage probability than the NA method. Moreover, the JEL method possesses the desirable feature that the shape and orientation of the resulting confidence regions are determined by the data.

There is still much to explore for extending the empirical likelihood method to other types of data such as current status data and interval censored data. It is also likely worthwhile to apply the JEL method to other types of models such as transformation models. It is also an interesting possibility to apply the adjusted empirical likelihood proposed by Chen *et al.* (2008) to further improve the accuracy and reduce the computation burden of the JEL method. All of these topics will be of interest in our future research.

Bibliography

[1] Chen, J., Variyath, A.M. and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist*, 17, 426-443.

[2] Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *The Annals of Statistics*, 22, 732-746.

[3] Heller, G. (2007). Smoothed rank regression with censored data. *Journal of the American Statistical Association*, 102, 552-559.

[4] Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *The Annals of Statistics*, 19, 293-325.

[5] Jin, Z., Lin, D.Y., Wei, L.J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, 90, 341-353.

[6] Jing, B.Y., Yuan, J.Y. and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*, 104, 1224-1232.

[7] Kalbfleisch, J.D. and Prentice, R.L. (1980). *The statistical analysis of failure time data.* New York: Wiley.

[8] Krall, J.M., Uthof, V.A. and Harley, J.B. (1975). A set-up procedure for selecting variables associated with survival. *Biometrics*, 31, 49-57.

[9] Lee, A.J. (1990). *U-Statistics, Theory and Practice*. New York: Marcel Dekker.

[10] Miller, R.G., and Halpern, J. (1982). Regression with censored data. *Biometrika*, 69, 521-531.

[11] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75, 237-249.

[12] Owen, A.B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18, 90-120.

[13] Owen, A.B. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 19, 1725-1747.

[14] Owen, A.B. (2001). *Empirical likelihood*. London: Chapman and Hall.

[15] Prentice, R. (1978). Linear rank tests with right censored data. *Biometrika*, 65, 167.

[16] Qin, J. and Lawless, J.F. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22, 300-325.

[17] Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.

[18] Ritov, Y. (1988). Tightness of monotone random fields. *Journal of the Royal Statistical Society*, Ser.B, 49, 331-333.

[19] SAS Institute Inc. (1999). SAS/STAT *User's Guide*, Version 8. Cary, NC: SAS Institute Inc.

[20] Shi, X. (1984). The approximate independence of jackknife pseudo-values and bootstrap methods. *Journal of Wuhan Institute Hydra-Electric Engineering*, 2, 83-90.

[21] Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, 18, 303-328.

[22] Tukey, J.W. (1958). Bias and confidence in not-quite large samples. *The Annals of Mathematical Statistics*, 29, 614.

[23] Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, 21, 76-99.

[24] Zhao, Y. (2011). Empirical likelihood inference for the accelerated failure time model. *Statistics and Probability Letters*, 81, 603-610.

[25] Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika*, 92, 492-498.