ScholarWorks@GSU

Empirical Likelihood Inferences in Survival Analysis

Authors	Yu, Xue
Citation	Yu, Xue. "Empirical Likelihood Inferences in Survival Analysis". Dissertation. Georgia State University, 2018. https://doi.org/10.57709/12486489
DOI	https://doi.org/10.57709/12486489
Download date	2026-05-12 00:24:55
Link to Item	https://hdl.handle.net/20.500.14694/10333

EMPIRICAL LIKELIHOOD INFERENCES IN SURVIVAL ANALYSIS

by

XUE YU

Under the Direction of Yichuan Zhao, Phd

ABSTRACT

In survival analysis, different regression models are used to estimate the effects of covariates on the survival time. The proportional hazards model is commonly applied. However, the proportional hazards model does not always give good fit in the real life. Other models, such as proportional odds models, additive hazards models are useful alternative. Motivated by this limitation, we investigate empirical likelihood method and make inference for semi-parametric transformation models and accelerated failure time models in this dissertation. The proposed empirical likelihood methods can solve several challenging and open problems.

These interesting problems include semiparametirc transformation model with length-biased sampling, semiparametric analysis based on weighted estimating equations with missing covariates. In addition, a more computationally efficient method called jackknife empirical likelihood (JEL) is proposed, which can be applied to make statistical inference for the accelerated failure time model without computing the limiting variance. We show that under certain regularity conditions, the empirical log-likelihood ratio test statistic converges to a standard chi-squared distribution.

Finally, computational algorithms are developed for utilizing the proposed empirical likelihood and jackknife empirical likelihood methods. Extensive simulation studies on coverage probabilities and average lengths of confidence intervals for the regression parameters for those topics indicate good finite samples performance under various settings. Furthermore, for each model, real data sets are analyzed for illustration of the proposed methods.

INDEX WORDS: Empirical likelihood, Semiparametirc transformation model, Jackknife, Length-biased sampling, Missing covariates, Accelerated failure time model.

EMPIRICAL LIKELIHOOD INFERENCES IN SURVIVAL ANALYSIS

by

XUE YU

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

Copyright by
Xue Yu
2018

EMPIRICAL LIKELIHOOD INFERENCES IN SURVIVAL ANALYSIS

by

XUE YU

Committee Chair: Yichuan Zhao

Committee: Xin Qi

Jing Zhang

Yichen Cheng

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2018

DEDICATION

This dissertation is dedicated to my family, my advisor, and all my dear friends.

ACKNOWLEDGEMENTS

I would like to give my profound thanks and sincere gratitude to everyone for helping me to complete this dissertation.

Foremost, I would like to express my sincere gratitude to my advisor, Dr. Yichuan Zhao, for the tremendous and continuous support of my Ph.D. study and research, for his motivation, immense knowledge and rigorous teaching and research spirit. His guidance helped me in all the time of study and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D. study. It is an excellent experience in my life, which would be a fortune to my work in the future.

Besides my advisor, I would like to thank the rest of my dissertation committee: Dr. Xin Qi, Dr. Jing Zhang, and Dr. Yichen Cheng, who kindly agreed to serve on my dissertation committee and have given me continuous help, encouragement and insightful comments. Without their precious support, it would not be possible to conduct this research.

Also, I would like to thank other professors from Mathematics & Statistics Department for educating me in various courses. Dr. Guantao Chen and Dr. Jeff Qin, who gave me support and help during my graduate study. I want to say thank you to Dr. Leslie Meadows, who agreed to be my reference, and helped me when I was a lab assistant in Commons MILE, guided me to be a better instructor when I was teaching Math 1070.

Furthermore, I could not forget to show the broad appreciation to my friends, Adil Virani, Shuenn Siang Ng and Tingting Yuan, who are also my lovely family. Adil is at the University of Maryland, who is the smartest and most caring person I have ever met. I will miss the laughing in the department from Siang, who supports me unconditionally.

I want to thank my schoolmates, Bing Liu, Faysal Satter, Kangni Alemdjrodo, Husneara Rahman, Yueheng An, Ali Jinnah, Maxime Bouadoumou and Che Hao Yeh. We had such a memorable time during the last five years. It is my honor to study with these hard-working people.

Last but not the least, I want to thank everyone who has helped me throughout my graduate life, especially my immediate family to whom this dissertation is dedicated to. I would like to thank my dearest parents, who provide me with the solid support. My husband Cai Huang, who will also get his Ph.D. in Bioinformatics this summer. We take care of each other, and we grow up together. It is a wonderful achievement for our family. I would also express many thanks to all my family members in China, for supporting me spiritually throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

- EL - Empirical likelihood
- JEL - Jackknife empirical likelihood
- AEL - Adjusted empirical likelihood
- NA - Normal approximation
- AFT - Accelerated failure time

CHAPTER 1

INTRODUCTION

1.1 Survival analysis

Survival analysis is a collection of statistical approaches for data analysis, for which the outcome variable of interest is the time until an event occurs (Kleinbaum (1998)). In survival analysis, we refer the time variable as a survival time. The time to event can be measured in hours, days, weeks, years from the beginning of follow-up of an individual until an event occurs. Furthermore, the event of interest can be various things, such as death, an occurrence of a disease, relapse from remission, a machine part breaks down and so on. Although the terminology survival analysis was initially developed by biostatisticians to analyze the occurrence of deaths in medical science, the methods are then applied in different fields including engineering, economics, actuarial science, etc.

In survival analysis, the observations are censored, which is a unique feature. In essence, censoring occurs when we have some information about individual survival time, but not all subjects' survival time are fully observed. Thus, the data about their survival time is incomplete. For instance, in a medical study, censoring can occur if a subject chooses to quit participating in the study, or dies from some unrelated events, or when there is a loss of follow-up. Most survival data are right-censored, meaning that the data is censored at the right side of the observed survival time interval.

1.2 Empirical likelihood

In statistics, the empirical function is the cumulative distribution function (CDF) associated with the empirical measure of the sample. Let $X_1, X_2, ..., X_n$ be independent random vectors in $\mathbb{R}^p$ and for $p \geq 1$ with common distribution function F_0, δ_x be a point mass at x.

The empirical distribution is defined as

$$F_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}, \tag{1.1}$$

where F_n is the nonparametric maximum likelihood estimate of F_0 based on $X_1, X_2, ..., X_n$.

Empirical likelihood (EL) is a statistical approach for nonparametric inference. To make inference for parameters, it adds weights in an estimating equation, which results in a new objective function containing weights and depending on the parameters. The empirical likelihood is a different approach to other non-parametric methods, which has sampling properties that are similar to the bootstrap. The basic idea of bootstrap is that inference about a population can be modeled by resampling the sample data and performing deduction on them.

The classical empirical likelihood (EL) was proposed by Thomas and Grunkemeier (1975), who inverted a nonparametric likelihood ratio test to obtain confidence intervals for the survival probability for right censored data. In the empirical likelihood (EL) theory, we can estimate an unknown parameter vector by maximizing the empirical likelihood under constraints.

Based on this idea, Owen (1988) developed the empirical likelihood method. EL is very appealing because by using EL method, researchers can respect the shape of confidence regions without having to specify a parametric family for the data. EL methods are more general than usual parametric likelihood method, and EL can be applied to many estimating equations as well. Many researchers have implemented this method in many interesting research fields. For example, Qin and Lawless (1994) linked estimating equations and empirical likelihood, they also developed ways of combining information about parameters. Zhao and Jinnah (2011) applied a variant of plug-in empirical likelihood by calculating the cumulative baseline hazard function and made inference for Cox regression models. Zhao and Chen (2008) made empirical likelihood inference for censored median regression models via nonparametric kernel estimation, and the linear transformation model with interval-censored failure time data was also studied in Zhang and Zhao (2013), among others.

More recently, Chen et al. (2015) proved the asymptotic normality of the log empirical likelihood-ratio statistic when the sample size and the data dimension are large. Huang and Zhao (2018) developed empirical likelihood confidence intervals for the bivariate survival function in the presence of univariate censoring. Other research works include Yang and Zhao (2012b), Yang and He (2012), Wang et al. (2013), Wang et al. (2017). It is noticeable that EL has been recognized as a useful tool in statistical sciences. Moreover, we can improve EL by reducing the numerical difficulties coming from the constrained optimization, which leads to a closed-form expression as a function of the parameters from the constraints and the Lagrangian multipliers.

1.3 Jackknife empirical likelihood

Although empirical likelihood approach shows attractive properties, the computational cost might be expensive when dealing with more complicated problems. A simpler and more computationally reliable method, jackknife empirical likelihood (JEL), has been widely used. Quenouille (1956) invented the jackknife as a resampling method, and it has been shown that jackknife is still useful when the sample size n is small. The JEL method is proposed by Jing et al. (2009), which combines EL and jackknife resampling method. “The jackknife empirical likelihood (JEL) is the combined version of jackknife and empirical likelihood method. The fundamental idea of the JEL method is to turn the statistic of interest into a sample mean based on the jackknife pseudo-values“ (see Jing et al. (2009)).

JEL enables us to construct confidence regions by introducing jackknife pseudo-values into the EL method. A significant advantage of the JEL method is its simplicity, and it is a natural application of empirical likelihood to the sample mean of jackknife pseudo-values. Let $Z_1, ..., Z_n$ be independent random variables, and let

$$T_n = T\left(Z_1, ..., Z_n\right)$$

be the estimator of the parameter θ. The jackknife pseudo-values is defined as

$$\widehat{V}_i = nT_n - (n-1)T_{n-1}^{(-i)},\ i = 1, ..., n$$

where $T_{n-1}^{(-i)} := T(Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_n)$. Actually, $T_{n-1}^{(-i)}$ is computed on the sample $n-1$ variables formed from the original data set by deleting the ith observation. Thus, the jackknife estimator of θ is the average of all the pseudo-values

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

Since the empirical likelihood is an easy tool, while calculating the sample mean, empirical likelihood is applied to the jackknife pseudo-values. Let $G_p(x) = \sum_{i=1}^{n} p_i I\left\{\hat{V}_i \leq x\right\}$ and $\theta_p = \sum_{i=1}^{n} p_i E\hat{V}_i$. Then the empirical likelihood $L(\theta)$ is evaluated at θ,

$$L(\theta) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i\hat{V}_i = \theta_p, \sum_{i=1}^{n} p_i = 1\right\}.$$

The jackknife empirical likelihood ratio at θ is

$$R(\theta) = \frac{L(\theta)}{n^{-n}} = \max\left\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i\hat{V}_i = \theta_p, \sum_{i=1}^{n} p_i = 1, p_i \geq 0\right\}.$$

Using Lagrange multipliers, when

$$\min_{1\leq i\leq n}\widehat{V}_i < \theta_p < \max_{1\leq i\leq n}\widehat{V}_i,$$

we can obtain that

$$p_i = \frac{1}{n}\frac{1}{1+\lambda(\widehat{V}_i - \theta_p)},$$

where λ satisfies

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

Then we get the jackknife empirical log-likelihood ratio

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

According to Jing et al. (2009), the Wilks' theorem holds

$$-2logR(\theta) \xrightarrow{d} \chi_1^2.$$

Moreover, JEL is very appealing because it is more general than usual parametric likelihood as it can be applied to test complicated hypotheses. For instance, Zhao et al. (2015b) proposed using JEL to study the mean absolute deviation, and Lin et al. (2017) developed JEL for the error variance in a linear regression model. Furthermore, JEL is commonly used in survival analysis. It not only provides efficient evaluation over survival functions regardless of complete or censored data but also can be applied in different models, such as accelerated failure time models (Bouadoumou et al. (2015)) and linear transformation models (Yang et al. (2016)). JEL has other applications in clinical experiments, for example, receiver operating characteristic (ROC) curve, a widely used graphical plot evaluating the discriminating power of a diagnostic test. Liu and Zhao (2012) proposed semi-empirical likelihood-based confidence intervals for ROC curves of two populations with missing data. Then An and Zhao (2017) extended to the difference of two volumes under ROC surfaces. Yang and Zhao (2013) constructed smoothed jackknife EL confidence intervals for the difference of ROC curves. Yang and Zhao (2015) made smoothed JEL inference for ROC curves with missing data. Furthermore, Yang et al. (2017) made JEL inference for the partial area under ROC curves. In addition, there are more research works about the difference of quantiles and differences of two Gini indices, like Wang and Zhao (2016), Yang and Zhao (2017), Yang and Zhao (2018), etc. These research works indicate that the JEL based methods are employed in many biostatistical fields, and they can be handy in dealing with more general statistics beyond classical U-statistics. Notably, JEL methods can be easily implemented in a standard software environment, and it will ease the computational burden for practical use.

1.4 Linear transformation models

Counting processes have been used to describe event history data. The traditional survival data can be characterized as a counting process with a single jump at the survival time. Some statistical models were proposed to formulate the effects of covariates on counting processes (see Andersen et al. (2012)).

In survival analysis, the data can be represented as counting process notation, $N(t)$, which is the count of the number of events observed in a time interval $[0, t]$. Let $Z(t)$ be a vector of time-varying covariates. Define T as the survival time. The proportional intensity model has the form that

$$\Lambda_Z(t) = \int_0^t Y(u) e^{\beta^T Z(u)} d\Lambda(u), \tag{1.2}$$

where $Y(t) = I(T \geq t)$, $\Lambda(\cdot)$ is an unspecified increasing function, and β is a vector of unknown regression parameters (Zeng and Lin (2006)).

For survival data, model (1.2) becomes the classical proportional hazards model (Cox (1972)). The proportional hazards model estimates the relative risk of experiencing an event of interest between two groups of subjects and assumes the hazard ratio to be constant. One of the advantages of the proportional hazards model is the estimation of the regression parameters does not depend on the unspecified baseline hazard function. Cox (1975) proposed a partial likelihood estimation technique to make inference about the regression parameters. Other research work includes Tsiatis (1981), Andersen and Gill (1982), and Lin and Ying (1993), etc.

One of the important assumptions of the proportional hazards model is that the hazards ratio is constant over the observed survival times. However, when this assumption is not met, a useful alternative model, is proposed, the proportional odds model (Bennett (1983); Pettitt (1984)). The proportional odds model assumes the ratio of the odds of survival associated with two sets of covariate values to be constant over time. This model was studied by many researchers, for example, Aalen (1980), Buckley (1984), Pierce and Preston (1984), Huffer and McKeague (1991), among others.

The proportional hazards and proportional odds models belong to the class of semiparametric linear transformation models (Dabrowska and Doksum (1988)). The transformation models provide many other potential choices in survival analysis. This class of models were studied in Dabrowska and Doksum (1988), Cheng et al. (1995), Chen et al. (2002), Huang and Wang (2010), Wang and Wang (2015), etc.

1.5 Summary

For the complete and censored data, the traditional EL procedures have been developed to make inference for linear transformation models. However, not many research works studied length-biased data. Therefore, we consider EL inference for semiparametric transformation models with length-biased sampling. The biased sampling occurs when the sampling distribution is different from the population distribution, commonly seen in survey studies. Qin (2017) pointed out that when the sampling plan is adopted, this bias happens because not every unit in the population has an equal chance to be sampled. Moreover, length-biased sampling appears when the probability of selecting an interval is proportional to the length of the interval. We will explore this problem in Chapter 2.

Missing data are a frequently encountered problem in epidemiology and biostatistics. The complete-case analysis is one of the most commonly used methods, which is only including those participants without missing observations. However, when dealing with missing covariates, this method will naturally reduce the statistical power and produce biased estimates. An idea based on weighted estimating equations has been proposed and it shows an advantage of dealing with missing covariates. We also introduced EL method to make a reliable inference for semiparametric transformation models with missing covariates. We will discuss this topic in Chapter 3.

An alternative to the commonly used proportional hazards model is the accelerated failure time (AFT) model. An AFT model assumes that the effect of a covariate is to accelerate the life course of a disease by some constant, while the proportional hazards model assumes that the effect of a covariate is to multiply the hazard rate by some constant. A critical estimation procedure for the accelerated failure time model is the rank-based estimating equations with Gehan-type weight (see Fygenson and Ritov (1994)). However, when the rank-based estimating equation is not smoothed, it is difficult to compute the estimator of regression parameters. To overcome that difficulty, Brown and Wang (2007) and Heller (2007) used an induced smoothing approach that smoothed the estimating functions to obtain point and variance estimators. We proposed a more computationally efficient method (jackknife empirical likelihood) to make statistical inference for the AFT model without computing the limiting variance, which will be discussed in Chapter 4.

The rest of the dissertation is organized as follows. We describe empirical likelihood inference for semiparametric transformation models with length-biased sampling in Chapter 2. In Chapter 3, we discuss empirical likelihood inference for semiparametric transformation models with missing covariates. In Chapter 4, we develop the jackknife empirical likelihood inference for the accelerated failure time model. Summary and future research directions are discussed in Chapter 5. Chapters 2, 3 and 4 are written in manuscript style and have been submitted to statistical journals. Moreover, Chapter 2 is under minor revision, and Chapter 4 has been accepted for publication.

CHAPTER 2

EMPIRICAL LIKELIHOOD INFERENCE FOR SEMIPARAMETRIC TRANSFORMATION MODELS WITH LENGTH-BIASED SAMPLING

2.1 Background

A general class of semiparametric regression models, the so-called linear transformation models, have been explored by many authors, who proposed various estimation approaches and made statistical inferences. The transformation models are defined in Chen et al. (2002) as

$$H(T) = -\beta' Z + \varepsilon,$$

where H is an unknown monotone function, Z is a vector of covariates, β is an $p \times 1$ unknown vector of regression parameters of interest, ε's are the random variables with an unspecified distribution, and we assume that ε's are independent of Z. The proportional hazards model is a special case with ε following the extreme value distribution, and if ε follows the standard logistic distribution, it becomes a proportional odds model.

The empirical likelihood approach was introduced by Owen (1988, 1990, 2001) based on the original idea proposed by Thomas and Grunkemeier (1975). The empirical likelihood method provides a way to construct confidence regions of regression estimators, an empirical log-likelihood ratio test statistic is developed. In addition, Owen (1990) showed that under certain regularity conditions, the Wilk's theorem (Wilks (1938)) of chi-squired limiting distribution of log-likelihood ratio still holds. This approach offers the advantages of eliminating the need to specify a distribution of the data, and often yields more efficient estimates of the parameters than many common estimators.

As censored data are very common, empirical likelihood approach has been extended to the area of survival analysis as well. Wang and Jing (2001) applied the empirical likelihood method to a class of functionals of the survival function, and showed that it follows a chi-squared distribution. Zhou (2005) used the empirical likelihood method to make a confidence interval based on the rank estimators of the regression coefficient in the accelerated failure time model. Zhao (2010) applied an empirical likelihood ratio method and derived its limiting distribution via U-statistics. Lu and Liang (2006) showed that the limiting distribution of the empirical likelihood ratio is a weighted sum of standard chi-squared distribution. Subsequently, Yu et al. (2011) appropriately modified some constructions based on Lu and Liang (2006), thereby deducing that the limiting distribution follows a standard chi-squared distribution. More recently, in the light of Owen's work, Chen et al. (2008) developed the adjusted empirical likelihood for general estimating equations, which has been interpreted in various research paper, such as Wang et al. (2016), Zhao et al. (2015b) and Lin et al. (2017).

In recent decades, not only have censored data appeared in the survival analysis, but also an abundance of length-biased data have been identified. Length time bias is a form of selection bias, and length-biased data are left-truncated and right censored data under the stationary assumption, that states, the initial times follow a stationary Poisson process. As a matter of fact, length time bias is often discussed in the context of observational studies. More accurate cancer screenings is just one of the most common benefits that this methodology has provided. Numerous more examples can be found in Shen et al. (2009). The observed samples are not randomly selected from the population of interest, but with some probability proportional to their lengths. Under the length-biased sampling, the subjects who have been at risk before entering the study might have longer observed time intervals from initiation to failure than those from the underlying distribution of the general population.

Length-biased data have been studied extensively. Notably, Shen et al. (2009) made inferences for semiparametric transformation models with length-biased sampling based on the ranks of observed failure times, while Wang and Wang (2015) obtained the estimators from counting process-based unbiased estimating equations. The crucial step of the latter method was to construct martingale estimating equations. However, having discussed the advantages of the Normal Approximation (NA) method (Wang and Wang (2015)), it is important to note that the computational cost of NA method is not only costly due to complicated variance estimation, but also may introduce a bias. In this Chapter, we endeavor to overcome these cost issues by proposing empirical likelihood (EL) inferences for semiparametric transformation models. Nevertheless, we will still consider the martingale type estimation equations proposed by Wang and Wang (2015), and derive a empirical log-likelihood ratio test statistic that has a standard chi-squared limiting distribution.

The remainder of this chapter is organized as follows. In Section 2.2, empirical likelihood (EL) method and adjusted empirical likelihood (AEL) method inference procedure will be introduced. In Section 2.3, simulation studies are carried out to demonstrate the performance of the proposed EL and AEL methods. Furthermore, a real data analysis is shown in Section 2.4. We will be discussing about our findings, along with our further work will be covered in Section 2.5. Proofs are provided in the Appendix.

2.2 Main results

2.2.1 Notation

We adopted the notations from Wang and Wang (2015). Assume that T_0 is the time measured from the initiating event to failure, A is the truncation variable, which measures the time from the initiating event to exam. The residual censoring time is denoted by C, T is the observed failure time satisfying $T_0 > A$, and $V = T - A$. We also define $\tilde{T} = \min(T, A + C)$ and $\delta = I(V \leq C)$. We assume that C and (A, V) are independent given Z.

2.2.2 Normal approximation method

Let (β_0, H_0) be the true values of (β, H), $\lambda_\varepsilon(\cdot)$ and $\Lambda_\varepsilon(\cdot)$ be the hazard and cumulative hazard functions of ε. For $i = 1, 2, ..., n$, we have the following counting process notations:

$$Y_i(t) = I(\tilde{T}_i \geq t), N_i(t) = \delta_i I(\tilde{T}_i \leq t),$$

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) r(t, \tilde{T}_i, \delta_i) d\Lambda_\varepsilon \left\{ {\beta_0}' Z_i + H_0(u) \right\}.$$

Here $r(t, \tilde{T}, \delta) = \delta\omega(t)/\omega(\tilde{T})$, where $\omega(t) = \int_0^t S_C(u)du$ and $S_C(\cdot)$ is the survival function for the residual censoring time C. Under certain filtration (Fleming and Harrington (2011)), $M(t)$ is a martingale process.

Wang and Wang (2015) proposed the following estimation equations:

$$\sum_{i=1}^{n} \left[dN_i(t) - Y_i(t)\hat{r}(t, \tilde{T}_i, \delta_i) d\Lambda_\varepsilon\{\beta' Z_i + H(t)\} \right] = 0, (0 \le t \le \tau) \tag{2.1}$$

$$\sum_{i=1}^{n} \int_0^\infty Z_i \left[dN_i(t) - Y_i(t)\hat{r}(t, \tilde{T}_i, \delta_i) d\Lambda_\varepsilon\{\beta' Z_i + H(t)\} \right] = 0. \tag{2.2}$$

where $\tau = \inf\{t : P(\tilde{T} > t) = 0\}$, $\hat{r}(t, \tilde{T}, \delta) = \delta\hat{\omega}(t)/\hat{\omega}(\tilde{T})$, $\hat{\omega}(t) = \int_0^t \hat{S}_C(u)du$, $\hat{S}_C$ is the Kaplan-Meier estimator of the censoring time C.

They used the solution of equations (2.1) and (2.2), which is denoted by $(\hat{\beta}, \hat{H})$ to be the estimator of (β_0, H_0). For any u, $t \in [0, \tau]$, define

$$B_1(t) = E[Y(t) r(t, \tilde{T}, \delta) \dot{\lambda}_\varepsilon\{\beta_0' Z + H_0(t)\}],$$

$$B_2(t) = E[Y(t) r(t, \tilde{T}, \delta) \lambda_\varepsilon\{\beta_0' Z + H_0(t)\}],$$

$$B_1^Z(t) = E[Z Y(t) r(t, \tilde{T}, \delta) \dot{\lambda}_\varepsilon\{\beta_0' Z + H_0(t)\}],$$

$$B_2^Z(t) = E[Z Y(t) r(t, \tilde{T}, \delta) \lambda_\varepsilon\{\beta_0' Z + H_0(t)\}],$$

$$B(t, u) = \exp\left\{ \int_u^t \left\{ B_2^{-1}(u) B_1(u) \right\} dH_0(u) \right\},$$

$$z(t) = B_2^{-1}(t) \left[B_2^Z(t) + \int_t^\tau \left\{ B_1^Z(u) - B_2^{-1}(u) B_2^Z(u) B_1(u) \right\} B(u, t) dH_0(u) \right], \tag{2.3}$$

and

$$a(t) = \frac{\int_0^\tau E\left[\frac{Z\delta Y(u)}{\omega(\tilde{T})}\left(I(t \le u)\int_t^u S_C(u)du - \frac{\omega(u)}{\omega(\tilde{T})}I(t \le \tilde{T})\int_t^{\tilde{T}} S_C(u)du\right) \times d\Lambda_\varepsilon\left\{\beta_0' Z_i + H_0(u)\right\}\right]}{\pi(t)}. \tag{2.4}$$

Denote

$$\Sigma_*(\beta_0) = E\left[\int_0^\tau \{Z - z(t)\}Z'Y(t)r(t,\tilde{T},\delta)\dot{\lambda}_\varepsilon\{\beta_0' Z + H_0(t)\}dH_0(t)\right],$$

$$\Sigma^*(\beta_0) = E\bigg[\int_0^\tau (\{Z - z(t)\}dM(t) + a(t)dM_C(t))\bigg]^{\otimes 2},$$

where $\dot{\lambda}_\varepsilon = d\lambda_\varepsilon(t)/dt$, $M_C(t) = I(V \wedge C \le t, \delta = 0) - \int_0^t I(V \wedge C \ge u)d\Lambda_C(u)$, $\Lambda_C(t)$ is the cumulative hazard function of C, and $\pi(t) = P(\tilde{T} - A \ge t)$.

Under the conditions (D.1)-(D.5) given in the Appendix, Wang and Wang (2015) proved that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, {\Sigma_*}^{-1}\Sigma^*({\Sigma_*}^{-1})').$$

To establish the asymptotic properties of $\hat{\beta}$, some definitions are given first. Define

$$\hat{B}_1(\beta,t) = n^{-1}\sum_{i=1}^n Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)\dot{\lambda}_\varepsilon\{\beta' Z_i + \hat{H}(\beta,t)\},$$

$$\hat{B}_2(\beta,t) = n^{-1}\sum_{i=1}^n Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)\lambda_\varepsilon\{\beta' Z_i + \hat{H}(\beta,t)\},$$

$$\hat{B}_1^Z(\beta,t) = n^{-1}\sum_{i=1}^n Z_i Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)\dot{\lambda}_\varepsilon\{\beta' Z_i + \hat{H}(\beta,t)\},$$

$$\hat{B}_2^Z(\beta,t) = n^{-1}\sum_{i=1}^n Z_i Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)\lambda_\varepsilon\{\beta' Z_i + \hat{H}(\beta,t)\},$$

$$\hat{B}(\beta,t,u) = \exp\left\{\int_u^t \left\{\hat{B}_2^{-1}(\beta,u)\hat{B}_1(\beta,u)\right\} d\hat{H}(\beta,u)\right\},$$

$$\hat{z}(\beta,t)=\hat{B}_2^{-1}(\beta,t)\left[\hat{B}_2^Z(\beta,t)+\int_t^\tau\left\{\hat{B}_1^Z(\beta,u)-\hat{B}_2^{-1}(\beta,u)\hat{B}_2^Z(\beta,u)\hat{B}_1(\beta,u)\right\}\hat{B}(\beta,u,t)d\hat{H}(\beta,u)\right], \tag{2.5}$$

and

$$\hat{a}(\beta,t)=\int_0^\tau n^{-1}\sum_{i=1}^n\left[\frac{Z_i\delta_iY_i(u)}{\hat{\omega}(\tilde{T}_i)}\left(I(t\le u)\int_t^u\hat{S}_C(u)du-\frac{\hat{\omega}(u)}{\hat{\omega}(\tilde{T}_i)}I(t\le\tilde{T}_i)\int_t^{\tilde{T}_i}\hat{S}_C(u)du\right)\right.$$
$$\left.\times d\Lambda_\varepsilon\left\{\beta'Z_i+\hat{H}(u)\right\}\right]\Big/\left[n^{-1}\sum_{i=1}^n I(\tilde{T}_i-A_i\ge t)\right]. \tag{2.6}$$

Furthermore, one can also define the $\hat{\Sigma}_*(\beta)$ and $\hat{\Sigma}^*(\beta)$, the consistent estimators of Σ_* and Σ^* as follows, respectively.

$$\hat{\Sigma}_*(\beta)=\sum_{i=1}^n\left[\int_0^\tau\{Z_i-\hat{z}(\beta,t)\}Z_i'Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)\dot{\lambda}_\varepsilon\{\beta'Z_i+\hat{H}(\beta,t)\}d\hat{H}(\beta,t)\right]\Big/n,$$

$$\hat{\Sigma}^*(\beta)=\sum_{i=1}^n\left[\int_0^\tau\left(\{Z_i-\hat{z}(\beta,t)\}d\hat{M}_i(\beta,t)+\hat{a}(\beta,t)d\hat{M}_{Ci}(t)\right)\right]^{\otimes 2}\Big/n,$$

where

$$\hat{M}_i(\beta,t)=\hat{N}_i(t)-\int_0^t Y_i(u)\hat{r}(u,\tilde{T}_i,\delta_i)d\Lambda_\varepsilon\{\beta'Z_i+\hat{H}(\beta,u)\}, \tag{2.7}$$

$$\hat{M}_{C_i}(t)=I(V_i\wedge C_i\le t,\delta_i=0)-\int_0^t I(V_i\wedge C_i\ge u)d\hat{\Lambda}_C(u), \tag{2.8}$$

and $\hat{\Lambda}_C(t)$ is the Nelson-Aalen estimator of the residual censoring time C.

Thus, the $100(1-\alpha)\%$ NA-based confidence region for β can be established as

$$R_\alpha^{NA}=\left\{\beta:n(\hat{\beta}-\beta)'\hat{\Sigma}_*(\hat{\beta})(\hat{\Sigma}^*(\hat{\beta}))^{-1}(\hat{\Sigma}_*(\hat{\beta}))'(\hat{\beta}-\beta)\le\chi_p^2(\alpha)\right\},$$

where $\chi_p^2(\alpha)$ is the upper α-quantile of distribution of χ_p^2.

2.2.3 Empirical likelihood method

As can be seen, Wang and Wang (2015) put forward a good concept to make inferences of β. However, it does not bypass perplexed variance computations. In this subsection, an empirical likelihood method for the length-biased data are proposed, and the asymptotic chi-squared distribution of the empirical log-likelihood ratio and the confidence intervals of regression parameters are developed.

Motivated by the estimating equation (2.2), for $i = 1, ..., n$, we propose

$$W_{ni}(\beta) = \int_0^\infty \{Z_i - \hat{z}(\beta, t)\} \, d\hat{M}_i(\beta, t) + \hat{a}(\beta, t) d\hat{M}_{C_i}(t),$$

where $\hat{z}(\beta, t)$, $\hat{a}(\beta, t)$, $\hat{M}_i(\beta, t)$ and $\hat{M}_{C_i}(t)$ are defined in equations (2.5) to (2.8).

In addition to the standard unit total probability constraints, another constraint $\sum_{i=1}^{n} p_i W_{ni}(\beta) = 0$ is added. Subsequently, the following empirical likelihood ratio is developed as

$$R_n(\beta) = \sup \left\{ \prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i W_{ni}(\beta) = 0, \sum_{i=1}^{n} p_i = 1, p_i \geq 0, i = 1, ..., n \right\}.$$

Also, we can define the empirical log-likelihood ratio $l_n(\beta) = -2 \log R_n(\beta)$, therefore, the empirical log-likelihood ratio can be expressed as

$$l_n(\beta) = -2\text{sup} \left\{ \sum_{i=1}^{n} \log(np_i) : \sum_{i=1}^{n} p_i W_{ni}(\beta) = 0, \sum_{i=1}^{n} p_i = 1, p_i \geq 0, i = 1, ..., n \right\}.$$

Using Lagrange multipliers, it can be shown that

$$l_n(\beta) = 2 \sum_{i=1}^{n} \log \left(1 + (\theta(\beta))' W_{ni}(\beta)\right),$$

where $\theta(\beta)$ is the solution to the below equation

$$\frac{1}{n} \sum_{i=1}^{n} \frac{W_{ni}(\beta)}{1 + (\theta(\beta))' W_{ni}(\beta)} = 0. \tag{2.9}$$

After all, at the true value β_0 of β, the proposed empirical log-likelihood ratio test statistic can be shown following a standard chi-squared limiting distribution. In order to derive the Wilks theorem, some results are needed first.

Theorem 2.1. *Under the conditions (D.1)-(D.5) given in the Appendix, as $n \longrightarrow \infty$, one has that*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta_0) \xrightarrow{d} N(0, \Sigma^*(\beta_0)).$$

Moreover, as $n \longrightarrow \infty$, one has that

$$\frac{1}{n} \sum_{i=1}^{n} W_{ni}(\beta_0) W_{ni}(\beta_0)' \xrightarrow{p} \Sigma^*(\beta_0).$$

By using the results of Theorem 2.1, the limiting distribution of the estimated empirical log-likelihood ratio can be shown in the following theorem.

Theorem 2.2. *Assume that the same regularity conditions given in Theorem 2.1 hold. As $n \longrightarrow \infty$, one has that*

$$l_n(\beta_0) \xrightarrow{d} \chi_p^2,$$

where χ_p^2 is a standard chi-squared random variable with p degrees of freedom.

Hence, the $100(1-\alpha)\%$ empirical likelihood confidence region for β can be established as

$$R_\alpha^{EL} = \left\{ \beta : l_n(\beta) \le \chi_p^2(\alpha) \right\},$$

where $\chi_p^2(\alpha)$ is the upper α-quantile of distribution of χ_p^2.

In practice, if we are only interested in a part of parameters. Define $\beta = (\beta_1', \beta_2')'$, where $\beta_1 \in R^q$ and $\beta_2 \in R^{p-q}$. We construct the empirical likelihood confidence region for β_1. The above proposed procedure can be used, and the profile empirical likelihood ratio is defined as

$$l_n^*(\beta_1) = \inf_{\beta_2} l_n(\beta_1', \beta_2')'.$$

We obtain Theorem 2.3 for the proposed profile log-empirical likelihood ratio $l_n^*(\beta_1)$.

Theorem 2.3. *Assume that the same regularity conditions given in Theorem 2.1 hold. As*

$n \longrightarrow \infty$, one has that

$$l_n^*(\beta_{10}) \xrightarrow{d} \chi_q^2,$$

where β_{10} is the true value of the parameter of interest β_1, and χ_q^2 is a standard chi-squared random variable with q degrees of freedom.

Thus, the $100(1-\alpha)\%$ profile EL confidence region for β_1 can be established as

$$R_\alpha^{EL*} = \left\{\beta_1 : l_n^*(\beta_1) \leq \chi_q^2(\alpha)\right\},$$

where $\chi_q^2(\alpha)$ is the upper α-quantile of distribution of χ_q^2.

2.2.4 Adjusted empirical likelihood method

Afterwards, Chen et al. (2008) developed the adjusted empirical likelihood (AEL) method to improve the performances of the empirical likelihood methods in terms of coverage probability. The key idea of the AEL method is to add one more value to $W_{ni}(\beta)$, which is $W_{n,n+1}(\beta) = -(a_n/n)\sum_{i=1}^{n} W_{ni}(\beta)$, where $a_n = \max(1, \log(n)/2)$. Then $W_{ni}^{ad}(\beta) = W_{ni}(\beta)$, $i = 1, ..., n+1$.

Motivated by their idea, we introduce the adjusted empirical likelihood ratio function as follows

$$R_n^{ad}(\beta) = \sup\left\{\prod_{i=1}^{n+1}(n+1)p_i : \sum_{i=1}^{n+1} p_i W_{ni}^{ad}(\beta) = 0, \sum_{i=1}^{n+1} p_i = 1, p_i \geq 0, i = 1, ..., n+1\right\}.$$

Therefore, define the adjusted empirical log-likelihood ratio to be

$$\log R^{ad}(\beta) = -\sum_{i=1}^{n+1} \log\left(1 + (\theta^{ad}(\beta))' W_{ni}^{ad}(\beta)\right),$$

by the same arguments that in subsection 2.2.3, $\theta^{ad}(\beta)$ satisfies

$$\frac{1}{n+1}\sum_{i=1}^{n+1} \frac{W_{ni}^{ad}(\beta)}{1 + (\theta^{ad}(\beta))' W_{ni}^{ad}(\beta)} = 0. \tag{2.10}$$

As in Chen et al. (2008), the following result holds for the AEL procedure.

Theorem 2.4. *Assume that the same regularity conditions given in Theorem 2.1 hold. Define* $l_n^{ad}(\beta_0) = -2\log R^{ad}(\beta_0)$*. As* $n \longrightarrow \infty$*, one has that*

$$l_n^{ad}(\beta_0) \xrightarrow{d} \chi_p^2.$$

Thus, the $100(1-\alpha)\%$ adjusted empirical likelihood confidence region for β can be constructed as

$$R_\alpha^{ad} = \left\{\beta : l_n^{ad}(\beta) \leq \chi_p^2(\alpha)\right\}.$$

Similarly, we can construct the profile adjusted empirical likelihood confidence region for β_1. The profile adjusted empirical likelihood ratio is defined as

$$l_n^{*,ad}(\beta_1) = \inf_{\beta_2} l_n^{ad}(\beta_1', \beta_2')'.$$

One can get Theorem 2.5 for the proposed profile adjusted log-empirical likelihood ratio $l_n^{*,ad}(\beta_1)$. Combined Theorems 2.3 and 2.4, the proof of Theorem 2.5 can be obtained.

Theorem 2.5. *Assume that the same regularity conditions given in Theorem 2.1 hold. As* $n \longrightarrow \infty$*, one has that*

$$l_n^{*,ad}(\beta_{10}) \xrightarrow{d} \chi_q^2,$$

where β_{10} *is the true value of the parameter of interest* β_1*, and* χ_q^2 *is a standard chi-squared random variable with* q *degrees of freedom.*

Thus, the $100(1-\alpha)\%$ profile adjusted empirical likelihood confidence region for β_1 can be established as

$$R_\alpha^{*,ad} = \left\{\beta_1 : l_n^{*,ad}(\beta_1) \leq \chi_q^2(\alpha)\right\},$$

where $\chi_q^2(\alpha)$ is the upper α-quantile of distribution of χ_q^2.

2.3 Simulation studies

In this section, a series of simulation studies are conducted to assess our proposed EL and AEL methods, and to compare the relative performances with NA method proposed by Wang and Wang (2015).

In order to compare EL, AEL and NA methods, we adopted as the same simulation settings as that in Wang and Wang (2015). Like Wang and Wang (2015), failure times T_0 were simulated from the transformation model

$$H(T_0) = -\beta' Z + \varepsilon,$$

where $H(t) = 2\log(t)$, $\beta = (\beta_1,\ \beta_2)'$, $Z = (Z_1,\ Z_2)'$, and Z_1, Z_2 are drawn from a jointly normal distribution with each mean 0, standard deviation 1, and correlation 0.5. The truncation variable A was generated from a uniform distribution $(0, \tau)$, where τ was chosen such that it is larger than the upper bound of T_0. We only kept paired data (A_i, T_{0i}) satisfying $A_i \leq T_{0i}$ $(i = 1, ..., n)$.

Moreover, we set $\beta_0 = (0,\ 0)'$ or $(1,\ 1)'$. The residual censoring variable C was generated independently from the uniform distribution on $(0,\ c)$, where the constant c was chosen to yield three different censoring rates (CR): 0%, 10%, and 30%. In addition, the sample sizes were 100 and 200, and all the simulations were based on 1000 replications.

Proportional hazards model and proportional odds model are considered in our simulations. Under proportional hazards model , we generated ε from extreme value distribution. Then the 95% coverage probabilities (CP) and average lengths (AL) of confidence intervals for the estimators of $\beta = (\beta_1,\ \beta_2)'$ were calculated .

From Table 2.1, we can see that the coverage probabilities of estimators for all methods improve as the sample sizes increase from 100 to 200. For most scenarios, we lose more information as the censoring rates increase, which causes the coverage probability to decrease and average lengths of the confidence intervals to get wider. For all the three methods, when the sample sizes become larger, the lengths of confidence intervals become shorter. Moreover, AEL preforms slightly better than EL in terms of coverage probability. For instance, when the sample size $n = 100$, $(\beta_1,\ \beta_2) = (1,\ 1)$, and CR = 30%, the coverage probability of EL confidence intervals is (0.924, 0.928), while the coverage probability of AEL confidence intervals is (0.931, 0.930). This result is expected because AEL procedure can reduce the amount of deviation. Under proportional odds model, ε was simulated from standard logistic distribution. Table 2.2 shows as the same patterns as the results in Table 2.1. Generally speaking, our results are consistent with those in Wang and Wang (2015) for all cases. Notably , all the coverage probabilities are reasonably close to the nominal level 95% and AEL method outperforms in most cases especially when the censoring rate is 30%.

2.4 Application to real data

In this section, dementia data were analyzed to illustrate the proposed EL and AEL methods. The data chosen are acquired from a multi-center epidemiological study, that is, the Canadian Study of Health and Aging (CSHA). Dementia is a chronic or persistent disorder of the mental processes caused by brain disease or injury, which is a progressive degenerative medical condition. In the United States and Canada, dementia is one of the leading causes of all deaths. (See Shen et al. (2009))

In the study, more than 14,000 subjects were 65 years or older. They were randomly chosen to be invited for a health survey in Canada. A total of 10,263 subjects agreed to participate. The participants were screened for dementia in 1991, and 1132 of 10263 people were identified as having dementia. All those patients had been followed until 1996, and their dates of death or last follow-up were recorded from the time of screening. Besides medical records, the data included three dementia categories, which were probable Alzheimer's disease, possible Alzheimer's disease and vascular dementia. Other than that, date of screening for dementia, date of death or censoring and death indicators were recorded. Patients with worse prognosis of dementia had higher chances to die before the study recruitment. Thus, Shen et al. (2009) pointed out that by excluding missing data of disease onset or dementia type, the rest of 818 participants are length-biased with stationary assumption. It was also validated in Addona and Wolfson (2006).

Among all patients, 393 had probable Alzheimer's disease, 252 had possible Alzheimer's disease, and 173 had vascular dementia. Since 638 out of 818 patients died at the end of study, others were determined as right censored data. The goal of the study is to evaluate if three dementia types had different impacts on patients' survival time. We chose one of the three types named probable Alzheimer's disease as the baseline variable and other two (Alzheimer's disease and vascular dementia) as indicators. Under proportional hazards and proportional odds models, lengths of confidence intervals of covariates for all three methods (NA, EL and AEL) are compared.

In Table 2.3, we reported the lower bound (LB) and upper bound (UB) of the parameters as well as the 95% confidence interval lengths by using NA, EL and AEL methods. Similar results are shown by three methods indicate our proposed EL and AEL methods do provide valid inferences. In conclusion, there are little associations between the type of dementia and patients' survival time.

2.5 Discussion

In this chapter, we propose empirical likelihood and adjusted empirical likelihood methods to make statistical inferences for the semiparametric transformation models under length-biased sampling. Motivated by Wang and Wang (2015), who constructed unbiased estimating equations based on counting processes. An empirical log-likelihood ratio is developed. Moreover, the test statistic put forward is proven to follow the standard chi-squared distribution. Therefore, when we make statistical inferences about the regression parameters of interest, complicated variance covariance matrix estimations can be avoided. Our approaches offer the advantages of easy implementation, and can be applied in other regression models with length-biased data. This will be our future research topics.

Table 2.1 Simulation results for the proportional hazards model

	NA		EL		AEL	
CR	CP	AL	CP	AL	CP	AL
$n = 100$ (β_1, β_2)=(0, 0)						
0%	(0.956, 0.958)	(1.117, 0.919)	(0.947, 0.946)	(1.284, 1.139)	(0.957, 0.954)	(1.471, 1.404)
10%	(0.938, 0.917)	(1.126, 1.297)	(0.933, 0.930)	(1.352, 1.298)	(0.940, 0.939)	(1.467, 1.621)
30%	(0.920, 0.896)	(2.952, 3.518)	(0.924, 0.901)	(3.377, 3.492)	(0.931, 0.928)	(3.567, 3.661)
$n = 100$ (β_1, β_2)=(1, 1)						
0%	(0.939, 0.940)	(0.820, 0.868)	(0.941, 0.940)	(1.336, 1.284)	(0.942, 0.942)	(1.564, 1.603)
10%	(0.937, 0.935)	(0.994, 0.839)	(0.939, 0.935)	(1.400, 1.369)	(0.940, 0.942)	(1.783, 1.821)
30%	(0.918, 0.924)	(1.359, 1.385)	(0.924, 0.928)	(2.081, 2.304)	(0.931, 0.930)	(2.640, 2.357)
$n = 200$ (β_1, β_2)=(0, 0)						
0%	(0.955, 0.949)	(0.832, 0.715)	(0.949, 0.948)	(1.128, 1.302)	(0.949, 0.949)	(1.288, 1.312)
10%	(0.939, 0.924)	(1.101, 1.413)	(0.943, 0.932)	(1.227, 1.204)	(0.945, 0.943)	(1.358, 1.421)
30%	(0.935, 0.937)	(2.075, 2.014)	(0.934, 0.935)	(2.493, 2.361)	(0.941, 0.939)	(2.699, 2.853)
$n = 200$ (β_1, β_2)=(1, 1)						
0%	(0.954, 0.948)	(0.726, 0.823)	(0.942, 0.942)	(1.179, 1.201)	(0.944, 0.943)	(1.307, 1.249)
10%	(0.941, 0.936)	(0.814, 1.248)	(0.943, 0.936)	(1.289, 1.290)	(0.944, 0.942)	(1.583, 1.697)
30%	(0.921, 0.930)	(1.893, 1.582)	(0.925, 0.939)	(2.364, 2.461)	(0.938, 0.938)	(2.011, 2.176)

Table 2.2 Simulation results for the proportional odds model

	NA		EL		AEL	
CR	CP	AL	CP	AL	CP	AL
$n = 100$ (β_1, β_2)=(0, 0)						
0%	(0.944, 0.948)	(1.106, 1.109)	(0.940, 0.947)	(1.224, 1.351)	(0.948, 0.947)	(1.508, 1.473)
10%	(0.931, 0.926)	(1.130, 1.116)	(0.935, 0.933)	(1.405, 1.398)	(0.938, 0.940)	(1.664, 1.790)
30%	(0.922, 0.901)	(2.014, 2.015)	(0.934, 0.928)	(2.904, 2.887)	(0.932, 0.928)	(3.081, 3.114)
$n = 100$ (β_1, β_2)=(1, 1)						
0%	(0.941, 0.942)	(1.096, 1.066)	(0.936, 0.939)	(1.335, 1.374)	(0.941, 0.941)	(1.614, 1.409)
10%	(0.934, 0.914)	(1.227, 1.368)	(0.931, 0.937)	(1.621, 1.503)	(0.940, 0.940)	(1.881, 1.892)
30%	(0.932, 0.936)	(2.121, 2.242)	(0.934, 0.935)	(2.228, 2.404)	(0.935, 0.936)	(2.399, 2.514)
$n = 200$ (β_1, β_2)=(0, 0)						
0%	(0.951, 0.948)	(0.784, 0.800)	(0.948, 0.947)	(1.077, 0.987)	(0.951, 0.951)	(1.410, 1.427)
10%	(0.941, 0.952)	(0.921, 0.976)	(0.943, 0.950)	(1.114, 1.281)	(0.946, 0.944)	(1.523, 1.408)
30%	(0.944, 0.951)	(1.639, 1.564)	(0.938, 0.947)	(2.003, 2.186)	(0.942, 0.942)	(2.245, 2.379)
$n = 200$ (β_1, β_2)=(1, 1)						
0%	(0.945, 0.945)	(0.729, 0.702)	(0.943, 0.932)	(1.229, 1.208)	(0.949, 0.947)	(1.482, 1.411)
10%	(0.942, 0.943)	(1.317, 1.333)	(0.949, 0.950)	(1.504, 1.397)	(0.947, 0.947)	(1.704, 1.693)
30%	(0.937, 0.940)	(1.635, 2.164)	(0.932, 0.944)	(1.881, 1.924)	(0.932, 0.934)	(1.995, 1.908)

Table 2.3 Interval lengths (LB, UB) of regression parameters for dementia data

Method	Model	Vascular dementia		Alzheimer's disease	
EL	Proportional Hazards	0.312	(-0.108, 0.204)	0.183	(-0.156, 0.027)
AEL		0.391	(-0.213, 0.178)	0.248	(-0.245, 0.003)
NA		0.249	(-0.092, 0.157)	0.189	(-0.193, -0.004)
EL	Proportional odds	0.729	(-0.235, 0.494)	0.550	(-0.582, -0.032)
AEL		0.751	(-0.178, 0.573)	0.607	(-0.603, 0.004)
NA		0.564	(0.084, 0.648)	0.397	(-0.576, -0.179)

CHAPTER 3

EMPIRICAL LIKELIHOOD INFERENCE FOR SEMIPARAMETRIC TRANSFORMATION MODELS WITH MISSING COVARIATES

3.1 Background

The proportional hazards model was proposed by Cox (1972), which is the most popular approach used in survival analysis. The proportional odds model is an natural alternative method. As a generalization of these two well-known models, the semiparametric transformation model provides many other choices. Let T be the failure time; Z be a vector of covariates; β, a $p \times 1$ unknown vector of regression parameters of interest. Then, the transformation model is defined as follows (see Chen et al. (2002))

$$H(T) = -\beta' Z + \varepsilon,$$

where H is an unknown monotone increasing transformation function, and ε's are the random error components with an unspecified distribution. Moreover, ε's are assumed independent of Z.

Empirical likelihood (EL) is one of the most notable methodologies for nonparametric inference, which is based on a data-driven likelihood ratio function. Thomas and Grunkemeier (1975) proposed the original idea, and then Owen (1988, 1990, 2001) completely summarized it. Moreover, Owen (1988) showed that EL approach has the advantage of eliminating the requirement to specify a distribution of the data, and the log-likelihood ratio follows a chi-squared limiting distribution. The approach has rapidly drawn many attentions and since then been extended to various cases. For instance, it has been applied to generalized linear models (Kolaczyk (1994)), linear regression with censored data (Zhou and Li (2008)), and general estimating equations introduced by Qin and Lawless (1994). More recently, Zhao (2010) applied an empirical likelihood ratio method and derived its limiting distribution via U-statistics. Yang and Zhao (2012b) developed EL confidence regions for regression parameters of the survival rate. More research works are developed, including Zhao et al. (2015a), Yang et al. (2016), Wang et al. (2016), Wang and Zhao (2016), Wang et al. (2017), Yang et al. (2017), Lin et al. (2017), An and Zhao (2017), Huang and Zhao (2018), Yang and Zhao (2018), among others.

In recent decades, as missing data are frequently seen in biostatistics and epidemiological studies, it is essential to select efficient approaches to deal with the missing covariates. Rubin (1976) defined a missing data process called missing at random (MAR). MAR implies that the missing data process depends only on the observed covariates and the outcome. In this chapter, we assume MAR assumption holds.

As a matter of fact, missing covariates have been studied extensively. For instance, one of the heavily used methods is complete case analysis, that is, one would rule out the subjects with missing covariates from the dataset, then perform the analysis. However, complete case analysis is biased and inefficient, especially when the data are not missing completely at random. Other research work include Lin and Ying (1993), Chen and Little (1999), and Qi et al. (2005), etc. Notably, Qi et al. (2005) proposed the weighted estimating equations for proportional hazards models with missing covariates by using the inverse probability weighted idea. However, having discussed the advantages of the Cox regression model, it is important to note that the Cox model has some limitations in practice. Considering other models, such as additive hazards model (Qi et al. (2018)), or more general semiparametric transformation model is necessary. Huang and Wang (2010) proposed inverse probability weighted estimators for linear transformation models with missing covariates, and furthermore proved that the estimators are consistent and asymptotically normal. Notwithstanding, the methodology of Huang and Wang (2010) involves complicated covariance matrix estimations. Motivated by Yang and Zhao (2012a) and Zhao and Yang (2012), we proposed the empirical likelihood method for the transformation model to avoid the variance matrix calculation.

The remainder of this chapter is organized as follows. In Section 3.2, an empirical likelihood (EL) based inference procedure are introduced. Extensive simulation studies are carried out to demonstrate the performance of proposed EL method in Section 3.3. In Section 3.4, we conduct a real data analysis. Furthermore, Section 3.5 presents a discussion of our findings and some recommendations. Proofs of theorems are provided in the Appendix.

3.2 Main results

Throughout the paper, we adopt the same notations as Huang and Wang (2010). Suppose some components of Z are missing. Define Z^c to be the covariates that are available and Z^m to be the covariates that are sometimes missing. Therefore, one can express the covariates as $Z = (Z^c, Z^m)$. Assume the censoring time is denoted by C, X is the observed failure time satisfying $X = \min(T, C)$, and $\delta = I(T \leq C)$. We assume that T and C are independent given Z.

3.2.1 Normal approximation method

For $i = 1, 2, ...n$, we introduce the following counting process notations:

$$Y_i(t) = I(X_i \geq t), N_i(t) = \delta_i I(X_i \leq t),$$

$$M_i(\beta_0, t) = N_i(t) - \int_0^t Y_i(u) d\Lambda \left\{ \beta_0' Z_i + H_0(u) \right\},$$

where (β_0, H_0) is the true value of (β, H), $\Lambda(\cdot)$ is the cumulative hazard functions of ε. Moreover, consider a non-missingness indicator V, if Z^m is available, $V = 1$; if Z^m is missing, $V = 0$. Under the assumption MAR, define the non-missingness probability

$$\pi_i = P(V_i = 1 \,|X_i, \delta_i, Z_i^c, Z_i^m) = P(V_i = 1 \,|X_i, \delta_i, Z_i^c).$$

Define $\tau = \inf\{t : P(X > t) = 0\}$. If π is known, Chen et al. (2002) proposed a unified estimation process to analyze the linear transformation model. Combining with the inverse probability weighted idea, Huang and Wang (2010) developed the following estimation equations:

$$\sum_{i=1}^{n} \frac{V_i}{\pi_i} [dN_i(t) - Y_i(t) d\Lambda\{\beta' Z_i + H(t)\}] = 0 \ (0 \leq t \leq \tau), \tag{3.1}$$

$$\sum_{i=1}^{n} \int_0^{\infty} \frac{V_i}{\pi_i} Z_i [dN_i(t) - Y_i(t) d\Lambda\{\beta' Z_i + H(t)\}] = 0. \tag{3.2}$$

The solution of equations (3.1) and (3.2), denoted by $(\tilde{\beta}, \tilde{H})$, is used as the estimator of (β_0, H_0), where $\tilde{H}$ is a nondecreasing step functions with $H(0) = -\infty$.

Let $\lambda(t)$ be the hazard function, $\dot{\lambda} = d\lambda(t)/dt$. For any u, $t \in [0, \tau]$, define

$$B(\beta_0, t, u) = \exp\left\{ \int_u^t \left\{ \frac{E[Y(t)\dot{\lambda}\{\beta_0' Z + H_0(u)\}]}{E[Y(t)\lambda\{\beta_0' Z + H_0(u)\}]} \right\} dH_0(u) \right\}, \tag{3.3}$$

$$\mu_Z(\beta_0, t) = \frac{E[ZY(t)\lambda\{\beta_0' Z + H_0(X)\} B(t, X)]}{E[Y(t)\lambda\{\beta_0' Z + H_0(t)\}]}, \tag{3.4}$$

$$M_*(\beta_0) = \int_0^{\tau} \{Z - \mu_Z(\beta_0, t)\} dM(\beta_0, t), \tag{3.5}$$

$$A(\beta_0) = E\left[\int_0^{\tau} \{Z - \mu_Z(\beta_0, t)\} Z' Y(t) \dot{\lambda}\{\beta_0' Z + H_0(t)\} dH_0(t) \right], \tag{3.6}$$

$$\Sigma_1(\beta_0) = E\left[\frac{1}{\pi} \left(\int_0^{\tau} \{Z - \mu_Z(\beta_0, t)\} dM(\beta_0, t) \right)^{\otimes 2} \right], \tag{3.7}$$

where M_* is a martingale process with mean $E[M_*] = 0$, A and Σ_1 are assumed to be finite and nonsingular (see Huang and Wang (2010)).

In addition, Huang and Wang (2010) proved that

$$\sqrt{n}(\tilde{\beta} - \beta_0) \to N(0, A^{-1}(\beta_0)\Sigma_1(\beta_0)(A^{-1}(\beta_0))'),$$

in distribution, as $n \to \infty$.

However, π is unknown, that is, more common in practice. Intuitively, to perform the estimation of β and $H(\cdot)$, the non-missingness probability π in (3.1) and (3.2) can be replaced by its nonparametric estimator $\hat{\pi}$. Then the estimating equations become

$$\sum_{i=1}^{n} \frac{V_i}{\hat{\pi}_i} [dN_i(t) - Y_i(t) d\Lambda\{\beta' Z_i + H(t)\}] = 0, \tag{3.8}$$

and

$$\sum_{i=1}^{n} \int_0^{\infty} \frac{V_i}{\hat{\pi}_i} Z_i [dN_i(t) - Y_i(t) d\Lambda\{\beta' Z_i + H(t)\}] = 0. \tag{3.9}$$

Denote the solution of (3.8) and (3.9) as $(\hat{\beta}, \hat{H}(\cdot))$, Huang and Wang (2010) proved that under the regularity conditions given in the Appendix, one has that

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, A^{-1}(\beta_0)\Sigma_2(\beta_0)(A^{-1}(\beta_0))'),$$

in distribution, as $n \rightarrow \infty$, where

$$\Sigma_2(\beta_0) = \Sigma_1(\beta_0) - E\left[\left(\frac{1}{\pi} - 1\right)(E(M_* | W))^{\otimes 2}\right]. \tag{3.10}$$

The procedure to estimate the non-missingness probability π is introduced as follows. Let $W = (W^{(1)}, W^{(2)})$ be the variable on which π is allowed to depend, where $W^{(1)}$ denotes continuous components and $W^{(2)}$ denotes discrete components. A consistent estimator of $\pi(\cdot)$ is defined as

$$\hat{\pi}(w^{(1)}, w^{(2)}) = \frac{\sum_{j=1}^{n} V_j I(W_j^{(2)} = w^{(2)}) K_{h_n}(w^{(1)} - W_j^{(1)})}{\sum_{j=1}^{n} I(W_j^{(2)} = w^{(2)}) K_{h_n}(w^{(1)} - W_j^{(1)})}, \tag{3.11}$$

where $K_{h_n}(\cdot) = K(\cdot/h_n)$, $K(\cdot)$ is a kernel function, h_n is a bandwidth. Two special cases are considered, when W is discrete, $\hat{\pi}(w)$ in equation (3.11) reduces to

$$\hat{\pi}(w) = \frac{\sum_{j=1}^{n} V_j I(W_j = w)}{\sum_{j=1}^{n} I(W_j = w)}, \tag{3.12}$$

while W is continuous, $\hat{\pi}(w)$ in equation (3.11) reduces to

$$\hat{\pi}(w) = \frac{\sum_{j=1}^{n} V_j K_{h_n}(w - W_j)}{\sum_{j=1}^{n} K_{h_n}(w - W_j)}, \tag{3.13}$$

which is Nadaraya-Watson estimator.

Then, to establish the confidence region of β, some definitions are given as follows. Define

$$\hat{B}(\beta,t,u) = \exp\left\{ \int_u^t \left\{ \frac{\sum_{i=1}^{n} V_i Y_i(u) \dot{\lambda}\{\beta' Z_i + \hat{H}(\beta,u)\}/\hat{\pi}_i}{\sum_{i=1}^{n} V_i Y_i(t) \lambda\{\beta' Z_i + \hat{H}(\beta,u)\}/\hat{\pi}_i} \right\} d\hat{H}(\beta,u) \right\},$$

$$\hat{M}_i(\beta,t) = N_i(t) - \int_0^t Y_i(u) d\Lambda \left\{ \beta' Z_i + \hat{H}(\beta,u) \right\},$$

$$\bar{Z}(\beta,t) = \frac{\sum_{i=1}^{n} V_i Z_i Y_i(t) \lambda\{\beta' Z_i + \hat{H}(\beta,t)\} \hat{B}(\beta,t,X_i)/\hat{\pi}_i}{\sum_{i=1}^{n} V_i Y_i(t) \lambda\{\beta' Z_i + \hat{H}(\beta,t)\}/\hat{\pi}_i},$$

and let $M_{**}(W) = E[M_* \,|W]$, we have the estimation equation of $M_{**}(W)$ is

$$\hat{M}_{**,i}(\beta, w_i^{(1)}, w_i^{(2)}) = \frac{\sum_{j=1}^{n} V_j \hat{M}_{*,j}(\beta) I(W_j^{(2)} = w_i^{(2)}) K_{h_n}(w_i^{(1)} - W_j^{(1)})}{\sum_{j=1}^{n} V_j I(W_j^{(2)} = w_i^{(2)}) K_{h_n}(w_i^{(1)} - W_j^{(1)})}, \tag{3.14}$$

where

$$\begin{aligned}
\hat{M}_{*,j}(\beta) &= \int_0^\tau \{Z_j - \bar{Z}(\beta,t)\} d\hat{M}_j(\beta,t) \\
&= \int_0^\tau \{Z_j - \bar{Z}(\beta,t)\} \left[dN_j(t) - Y_j(t) d\Lambda \left\{ \beta' Z_j + \hat{H}(\beta,t) \right\} \right].
\end{aligned}$$

Also, one can define the consistent estimators of A and Σ_2, as $\hat{A}(\beta)$ and $\hat{\Sigma}_2(\beta)$,

$$\hat{A}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{V_i}{\hat{\pi}_i}\int_0^{\tau}\left\{Z_i - \bar{Z}(\beta,t)\right\}Z_i'Y_i(t)\dot{\lambda}\{\beta' Z_i + \hat{H}(\beta,t)\}d\hat{H}(\beta,t),$$

$$\hat{\Sigma}_2(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{V_i}{\hat{\pi}_i^2}\left(\int_0^{\tau}\left\{Z_i - \bar{Z}(\beta,t)\right\}d\hat{M}_i(\beta,t)\right)^{\otimes 2} - \frac{1}{n}\sum_{i=1}^{n}\frac{V_i(1-\hat{\pi}_i)}{\hat{\pi}_i^2}\hat{M}_{**,i}(\beta,w_i^{(1)},w_i^{(2)})^{\otimes 2}.$$

Therefore, the $100(1-\alpha)\%$ normal approximation (NA) based confidence region for β can be established as

$$R_{\alpha}^{NA} = \left\{\beta : n(\hat{\beta}-\beta)'\hat{A}(\hat{\beta})(\hat{\Sigma}_2(\hat{\beta}))^{-1}(\hat{A}(\hat{\beta}))'(\hat{\beta}-\beta) \le \chi_p^2(\alpha)\right\},$$

where $\chi_p^2(\alpha)$ is the upper α-quantile of distribution of χ_p^2.

3.2.2 Empirical likelihood method

Huang and Wang (2010) provided a good way to make inferences for β. However, the methodology sacrificed the computational resource on estimating the variance covariance matrices. In this subsection, motivated by Huang and Wang (2010), we construct empirical likelihood for the transformation model, which avoids estimating the complicated matrices. Define

$$d\hat{M}_{**,i}(\beta,t,w_i^{(1)},w_i^{(2)}) = \frac{\sum_{j=1}^{n}V_j d\hat{M}_{*,j}(\beta,t)I(W_j^{(2)}=w_i^{(2)})K_{h_n}(w_i^{(1)}-W_j^{(1)})}{\sum_{j=1}^{n}V_j I(W_j^{(2)}=w_i^{(2)})K_{h_n}(w_i^{(1)}-W_j^{(1)})},$$

where

$$d\hat{M}_{*,j}(\beta,t) = (Z_i - \bar{Z}(\beta,t))d\hat{M}_j(\beta,t).$$

Motivated by the estimating equation (3.9), for $i = 1, ..., n$, we propose

$$\begin{aligned} U_{ni}(\beta) &= \frac{V_i}{\hat{\pi}_i} \int_0^\tau (Z_i - \bar{Z}(\beta,t))d\hat{M}_i(\beta,t) + \left(1 - \frac{V_i}{\hat{\pi}_i}\right) \int_0^\tau E[(Z_i - \bar{Z}(\beta,t))d\hat{M}_i(\beta,t)\,|W_i] \\ &= \frac{V_i}{\hat{\pi}_i} \int_0^\tau (Z_i - \bar{Z}(\beta,t))d\hat{M}_i(\beta,t) + \left(1 - \frac{V_i}{\hat{\pi}_i}\right) \int_0^\tau E[d\hat{M}_{*,i}(\beta,t)\,|W_i] \\ &= \frac{V_i}{\hat{\pi}_i} \int_0^\tau (Z_i - \bar{Z}(\beta,t))d\hat{M}_i(\beta,t) + \left(1 - \frac{V_i}{\hat{\pi}_i}\right) \int_0^\tau d\hat{M}_{**,i}(\beta,t,w_i^{(1)},w_i^{(2)}). \end{aligned}$$

The empirical likelihood $L_n(\beta)$ at β is defined as

$$L_n(\beta) = \sup\left\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i U_{ni}(\beta) = 0, \sum_{i=1}^n p_i = 1, p_i \geq 0, i = 1, ..., n\right\}.$$

Since under the constrains that $\sum_{i=1}^n p_i = 1$, and $p_i \geq 0$, $\prod_{i=1}^n p_i$ reaches its maximum n^{-n} at $p_i = n^{-1}$. The empirical likelihood ratio at β is defined as

$$R_n(\beta) = \sup\left\{\prod_{i=1}^n np_i : \sum_{i=1}^n p_i U_{ni}(\beta) = 0, \sum_{i=1}^n p_i = 1, p_i \geq 0, i = 1, ..., n\right\}.$$

Therefore, the corresponding empirical log-likelihood ratio, $l_n(\beta) = -2\log R_n(\beta)$, can be expressed as

$$l_n(\beta) = -2\sup\left\{\sum_{i=1}^n \log(np_i) : \sum_{i=1}^n p_i U_{ni}(\beta) = 0, \sum_{i=1}^n p_i = 1, p_i \geq 0, i = 1, ..., n\right\}.$$

By using Lagrange multiplier method, one has

$$l_n(\beta) = 2\sum_{i=1}^n \log\left(1 + (\lambda(\beta))' U_{ni}(\beta)\right),$$

where $\lambda(\beta)$ is a Lagrange multiplier that satisfies the following equation

$$\frac{1}{n}\sum_{i=1}^n \frac{U_{ni}(\beta)}{1 + (\lambda(\beta))' U_{ni}(\beta)} = 0. \tag{3.15}$$

Theorem 3.1. *Let β_0 be the true value of β. Under the conditions (D.1)-(D.9) given in the Appendix, we have that*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0) \rightarrow N(0, \Sigma_2(\beta_0)),$$

in distribution, as $n \longrightarrow \infty$. Moreover, we have that

$$\frac{1}{n}\sum_{i=1}^{n} U_{ni}(\beta_0)U_{ni}(\beta_0)' \rightarrow \Sigma_2(\beta_0),$$

in probability, as $n \longrightarrow \infty$.

By using Theorem 3.1, the limiting distribution of $l_n(\beta_0)$ is given in Theorem 3.2.

Theorem 3.2. *Assume that the same regularity conditions given in Theorem 3.1 hold. Let β_0 be the true value of β, one has that*

$$l_n(\beta_0) \rightarrow \chi^2_p,$$

in distribution, as $n \longrightarrow \infty$, where χ^2_p is a standard chi-squared random variable with p degrees of freedom.

Thus, by Theorem 3.2, the $100(1-\alpha)\%$ empirical likelihood (EL) confidence region for β can be constructed as

$$R^{EL}_{\alpha} = \left\{\beta : l_n(\beta) \leq \chi^2_p(\alpha)\right\},$$

where $\chi^2_p(\alpha)$ is the upper α-quantile of distribution of χ^2_p.

Theorem 3.2 enables us to make global inference about β_0. But sometimes we may be interested in some components of the regression parameters, for instance, β_1, which is a sub-vector of β. Define $\beta_0 = (\beta'_{10}, \beta'_{20})'$. We are only interested in the inference about the q-dimensional sub-vector β_{10}. Then we can propose profile empirical likelihood method to handle nuisance parameters, and construct confidence region for β_1.

Define the profile empirical likelihood ratio as follows:

$$l_n^*(\beta_1) = \inf_{\beta_2} l_n(\beta_1', \beta_2')'.$$

The similar result can be obtained in Theorem 3.3 for the proposed profile log-empirical likelihood ratio $l_n^*(\beta_1)$.

Theorem 3.3. *Assume that the same regularity conditions given in Theorem 3.1 hold. Let β_{10} be the true value of β_1, one has that*

$$l_n^*(\beta_{10}) \rightarrow \chi_q^2,$$

in distribution, as $n \longrightarrow \infty$, where χ_q^2 is a standard chi-squared random variable with q degrees of freedom.

According to Theorem 3.3, the $100(1-\alpha)\%$ profile EL confidence region for β_1 can be established as

$$R_\alpha^{EL*} = \left\{\beta_1 : l_n^*(\beta_1) \leq \chi_q^2(\alpha)\right\},$$

where $\chi_q^2(\alpha)$ is the upper α-quantile of distribution of χ_q^2.

3.3 Simulation studies

A comprehensive simulation study is conducted to investigate the proposed EL method. We adopt the same simulation settings as those in Huang and Wang (2010), and compare the coverage probability of empirical likelihood confidence region for relatively small samples to large samples with normal approximation (NA) confidence region proposed by Huang and Wang (2010). We select $\beta_1 = -0.5$ and $\beta_2 = 0.5$. The transformation function is $H_0(t) = \log(t)$, and the hazard function of ε is generated form $\lambda(t) = \exp(t)/\{1 + r\exp(t)\}$. When $r = 0$ and $r = 1$, the models become the proportional hazards model and the proportional odds model, respectively. In this simulation study, we consider $r = 1$ and $r = 1.5$. Sample sizes n are chosen as 20, 50 and 100 for simulations.

Moreover, we generate missing covariate Z^m and observed covariate Z^c independently. Z^m follows the standard normal distribution, and Z^c is Bernoulli distributed with a parameter of 0.5. We simulate censoring times from three different distributions. For the first data setting, we use a uniform censoring time, which is independent of covariates and with upper limit selected. In other two settings, the censoring times depend on covariates Z^m and Z^c, and it follows uniform distribution from 0 to c, where c is chosen to adjust the censoring rate. Then, non-missingness probabilities in the three simulation settings are associated with different variables. The first setting gives $\pi_1(\delta) = 0.9\delta + 0.4(1-\delta)$. In addition, $\pi_2(X, \delta) = (1 + exp(1.5 - \delta - X))^{-1}$ and $\pi_3(X, \delta, Z^c) = (1 + exp(2 - \delta - X - Z^c))^{-1}$ are considered in the second and third data settings, respectively.

Furthermore, we calculate the estimators β_1 and β_2 for the following different cases: full cohort assuming that all data are available and using all cases, complete-case analysis are using only the available data and selected cases, "true $\pi(\cdot)$" and $\hat{\pi}(\cdot)$ are used. The kernel function $K(\cdot)$ is selected as $K(u) = 3(1-u^2)/4$, $|u| \leq 1$, and the bandwidth h_n is considered to be $h_n = n^{-1/3}$. Then the 95% coverage probabilities (CP) and average lengths (AL) of confidence intervals for the regression parameters are calculated. We repeat all the simulations 1000 times.

Table 3.1 shows the first simulation setting results. The complete-case analysis estimators have the lowest coverage probabilities for both methods, while the estimators based on $\hat{\pi}(\delta, X, Z^c)$ have the highest coverage probabilities. Due to using more information from partially incomplete data, the coverage probabilities for β_1 and β_2 are improved. Furthermore, when $n = 20$, the EL method has better performs than the NA does in terms of coverage probability. For instance, CPs of confidence intervals with estimators based on $\hat{\pi}(\delta)$ are 0.851 and 0.869 for the NA method; while 0.884 and 0.887 for the EL method. When the sample size is 50, the EL method performs slightly better than the NA method in most cases. Moreover, as the sample size increases from 50 to 100, the coverage probabilities become close to 95% nominal levels. Both approaches perform well under large sample sizes.

Table 3.2 displays the results from the second simulation setting. The complete-case analysis still has the lowest coverage probabilities for NA and EL methods. While very similar coverage probabilities are obtained for both methods in most cases, especially for $n = 100$. For instance, CPs for β_1 calculated by EL based on $\hat{\pi}(\delta, X, Z^c)$ is 0.946, and CPs calculated by NA is 0.944. These results indicate that when the non-missingness probability relies on more variables, the coverage probability is getting improved.

In addition, we obtain the similar findings in Table 3.3. For both methods, when the non-missingness probabilities are calculated based on (X, δ), it leads to the lower coverage probabilities than those CPs based on true π and $\hat{\pi}(\delta, X, Z^c)$. We also consider $r = 1.5$ in the fourth simulation setting. Just like $r = 1$, we find the same trends for coverage probabilities in Table 3.4.

3.4 Application to real data

In this section, we analyze the mouse leukemia study (Kalbfleisch and Prentice (1980)) to illustrate the proposed EL method. This study was conducted to investigate if the genetic and viral factors may have impacts in the progress of spontaneous leukemia in mice. The original dataset contains 204 observations with both mortalities caused by thymic or non-thymic leukemia. Other information includes sex, coat color, survival times, "Type" of death (natural or terminated), MHC phenotype, antibody level and so on. Among all factors, two covariates that are the Gpd-1 phenotype and the level of endogenous murine leukemia virus are examined. According to Huang and Wang (2010), we have the MAR assumption in this application.

Following Qi et al. (2005) and Huang and Wang (2010), 29 observations with missing endogenous murine leukemia virus are excluded, and we used 175 mice in our analysis. Moreover, if a virus level $< 10^4$ PFU/ml, the virus level is categories with $Z^c = 0$, otherwise, $Z^c = 1$. We conduct two separate analyses for the death of thymic leukemia and the death of thymic or nonthymic leukemia as the endpoints, respectively. Two different transformation models are considered. The hazard function of ε has the form $\lambda(t) = \exp(t)/\{1 + r\exp(t)\}$ with $r = 1$ and 1.5. To estimate the non-missingness probability, the kernel function is selected as $K(u) = 3(1-u^2)/4$, $|u| \leq 1$, and bandwidth is $h_n = n^{-1/3}$.

First, we consider the survival time of the mice, whose mortality is due to thymic leukemia as the endpoint. Table 3.5 displays the lower bound (LB) and upper bound (UB) of the Gpd-1 and Virus level and the lengths of 95% confidence interval by using the NA and EL methods. By examining Table 3.5, both approaches show that the Gpd-1 genotype and virus level are significantly related to thymic leukemia mortality. Furthermore, all results indicate the negative association between Gpd-1 and death, while the virus level has a positive impact. Also, when the non-missingness probabilities based on all variables, the shortest CI lengths are produced for both methods, indicating that using incomplete cases is helpful to the statistical analysis.

Moreover, we also consider the time to mortality due to thymic or nonthymic leukemia as the failure time. We summarize the results in Table 3.6. The same trend for Gpd-1 genotype is shown, while the interval lengths for virus suggest that the level of endogenous murine leukemia virus does not play a significant role on nonthymic leukemia mortality as it does on thymic leukemia mortality. There is a weaker association between virus level and the death caused by either thymic or nonthymic leukemia. Overall, the interval lengths for regression parameters based on $\hat{\pi}(X, \delta, Z^c)$ are the shortest, and our findings are consistent with the conclusion in Huang and Wang (2010).

3.5 Discussion

Motived by Huang and Wang (2010), we apply empirical likelihood method to make inferences for semiparametric transformation models with missing covariates. The theoretical results provide asymptotic properties, including limiting distribution of the empirical likelihood ratio statistics, which follows the standard chi-squared distribution. Furthermore, the simulation results demonstrate that coverage probability of EL confidence interval is higher than the coverage probability for the NA method with the relatively small sample size. Compared with the traditional normal approximation method, the EL method has less computational cost. There are other topics, such as making inferences for other regression models for survival data with missing covariates, which could be studied in the future.

Table 3.1 Simulation results when the true non-missing probability is $\pi_1(\delta)$.

Approach	NA				EL			
	CP		AL		CP		AL	
	$r=1$, $n=20$, censoring rate=31%, and missing rate=25%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.903	0.911	2.851	2.757	0.914	0.913	2.748	3.011
Complete-case	0.689	0.702	3.569	3.388	0.713	0.712	2.844	2.679
True $\pi_1(\delta)$	0.842	0.864	3.447	3.335	0.881	0.875	2.964	3.217
$\hat{\pi}_1(\delta)$	0.851	0.869	3.114	3.206	0.884	0.887	3.014	3.418
$\hat{\pi}_1(\delta, Z^c)$	0.836	0.844	3.126	3.181	0.893	0.902	3.224	3.257
$\hat{\pi}_1(\delta, X)$	0.852	0.861	3.417	3.069	0.884	0.891	3.175	3.241
$\hat{\pi}_1(\delta, X, Z^c)$	0.873	0.893	3.016	3.157	0.911	0.911	3.003	3.179
	$r=1$, $n=50$, censoring rate=28%, and missing rate=22%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.932	0.939	1.084	1.042	0.927	0.935	1.377	1.284
Complete-case	0.749	0.801	1.256	1.221	0.761	0.775	1.504	1.429
True $\pi_1(\delta)$	0.908	0.927	1.207	1.197	0.914	0.914	1.414	1.337
$\hat{\pi}_1(\delta)$	0.911	0.917	1.183	1.168	0.916	0.917	1.328	1.191
$\hat{\pi}_1(\delta, Z^c)$	0.915	0.919	1.138	1.150	0.922	0.923	1.284	1.297
$\hat{\pi}_1(\delta, X)$	0.917	0.920	1.149	1.159	0.922	0.923	1.379	1.484
$\hat{\pi}_1(\delta, X, Z^c)$	0.924	0.923	1.140	1.138	0.925	0.928	1.390	1.328
	$r=1$, $n=100$, censoring rate=29%, and missing rate=23%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.944	0.945	0.847	0.652	0.948	0.944	1.013	0.947
Complete-case	0.834	0.856	0.896	0.753	0.805	0.843	1.123	0.845
True $\pi_1(\delta)$	0.922	0.930	1.324	0.894	0.935	0.934	1.243	1.179
$\hat{\pi}_1(\delta)$	0.928	0.934	1.175	0.883	0.934	0.939	1.204	1.155
$\hat{\pi}_1(\delta, Z^c)$	0.918	0.927	1.114	1.135	0.930	0.939	1.243	1.211
$\hat{\pi}_1(\delta, X)$	0.934	0.941	0.957	0.765	0.933	0.939	1.147	1.032
$\hat{\pi}_1(\delta, X, Z^c)$	0.937	0.943	0.938	0.764	0.943	0.943	1.154	1.008

Table 3.2 Simulation results when the true non-missing probability is $\pi_2(\delta, X)$.

Approach	NA				EL			
	CP		AL		CP		AL	
	$r=1$, $n=20$, censoring rate=43%, and missing rate=41%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.902	0.916	2.492	2.314	0.912	0.926	2.743	2.829
Complete-case	0.783	0.795	2.892	2.650	0.768	0.810	2.944	2.732
True $\pi_2(\delta, X)$	0.840	0.841	2.690	2.575	0.893	0.896	2.823	2.685
$\hat{\pi}_2(\delta, X)$	0.851	0.868	2.379	2.388	0.899	0.914	2.521	2.503
$\hat{\pi}_2(\delta, X, Z^c)$	0.876	0.887	2.579	2.355	0.910	0.914	2.814	2.499
	$r=1$, $n=50$, censoring rate=42%, and missing rate=45%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.944	0.941	0.973	1.003	0.940	0.940	1.855	1.732
Complete-case	0.819	0.788	1.197	1.246	0.821	0.814	1.723	1.904
True $\pi_2(\delta, X)$	0.912	0.906	1.174	1.205	0.909	0.920	1.624	1.523
$\hat{\pi}_2(\delta, X)$	0.914	0.903	1.087	0.978	0.923	0.918	1.511	1.425
$\hat{\pi}_2(\delta, X, Z^c)$	0.928	0.939	1.004	0.998	0.929	0.934	1.408	1.693
	$r=1$, $n=100$, censoring rate=42%, and missing rate=46%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.951	0.953	0.835	0.558	0.944	0.950	1.033	1.204
Complete-case	0.845	0.877	1.146	1.242	0.792	0.804	1.179	1.084
True $\pi_2(\delta, X)$	0.926	0.941	1.007	1.152	0.930	0.934	1.335	1.317
$\hat{\pi}_2(\delta, X)$	0.935	0.944	1.002	0.731	0.939	0.946	1.256	1.203
$\hat{\pi}_2(\delta, X, Z^c)$	0.944	0.951	0.914	0.824	0.946	0.945	1.179	1.287

Table 3.3 Simulation results when the true non-missing probability is $\pi_3(\delta, X, Z^c)$ and $r = 1$.

Approach	NA				EL			
	CP		AL		CP		AL	
	$r=1$, $n=20$, censoring rate=38%, and missing rate=42%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.843	0.896	1.902	2.004	0.884	0.911	2.411	2.286
Complete-case	0.741	0.767	2.754	2.831	0.773	0.784	2.579	2.603
True $\pi_3(\delta, X, Z^c)$	0.852	0.864	2.459	2.678	0.864	0.872	2.210	2.117
$\hat{\pi}_3(\delta, X)$	0.821	0.849	2.219	2.449	0.843	0.858	2.164	2.335
$\hat{\pi}_3(\delta, X, Z^c)$	0.885	0.890	2.339	2.403	0.874	0.915	2.479	2.415
	$r=1$, $n=50$, censoring rate=40%, and missing rate=44%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.924	0.936	1.134	1.108	0.924	0.934	1.284	1.359
Complete-case	0.771	0.811	1.682	1.427	0.812	0.776	1.738	1.405
True $\pi_3(\delta, X, Z^c)$	0.902	0.914	1.442	1.378	0.914	0.926	1.643	1.512
$\hat{\pi}_3(\delta, X)$	0.919	0.925	1.602	1.388	0.910	0.922	1.764	1.503
$\hat{\pi}_3(\delta, X, Z^c)$	0.920	0.927	1.408	1.221	0.931	0.934	1.507	1.444
	$r=1$, $n=100$, censoring rate=41%, and missing rate=42%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.943	0.944	1.006	0.751	0.946	0.946	1.085	1.179
Complete-case	0.826	0.835	1.402	0.951	0.824	0.847	1.521	1.400
True $\pi_3(\delta, X, Z^c)$	0.937	0.942	1.426	0.999	0.938	0.937	1.254	1.307
$\hat{\pi}_3(\delta, X)$	0.922	0.935	1.448	1.086	0.924	0.939	1.499	1.452
$\hat{\pi}_3(\delta, X, Z^c)$	0.939	0.941	1.226	0.910	0.948	0.948	1.370	1.228

Table 3.4 Simulation results when the true non-missing probability is $\pi_3(\delta, X, Z^c)$ and $r = 1.5$.

Approach	NA				EL			
	CP		AL		CP		AL	
	$r = 1.5$, $n = 20$, censoring rate=41%, and missing rate=42%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.884	0.868	1.539	1.412	0.909	0.914	1.823	1.794
Complete-case	0.710	0.723	2.329	2.092	0.725	0.794	2.011	1.805
True $\pi_3(\delta, X, Z^c)$	0.841	0.863	1.675	1.872	0.852	0.870	1.771	1.813
$\hat{\pi}_3(\delta, X)$	0.803	0.812	1.732	1.844	0.814	0.838	1.728	1.903
$\hat{\pi}_3(\delta, X, Z^c)$	0.864	0.851	1.706	1.641	0.885	0.874	1.804	1.883
	$r = 1.5$, $n = 50$, censoring rate=41%, and missing rate=42%							
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full cohort	0.918	0.931	1.174	1.244	0.923	0.931	1.447	1.358
Complete-case	0.735	0.797	1.785	1.607	0.750	0.804	1.908	1.792
True $\pi_3(\delta, X, Z^c)$	0.884	0.899	1.647	1.826	0.866	0.885	1.504	1.693
$\hat{\pi}_3(\delta, X)$	0.832	0.901	1.589	1.446	0.842	0.904	1.684	1.724
$\hat{\pi}_3(\delta, X, Z^c)$	0.894	0.920	1.504	1.407	0.910	0.934	1.778	1.650

Table 3.5 Interval lengths (LB, UB) of regression parameters for the mouse leukemia data (Thymic leukemia)

Model	Method			
$r = 1$	NA		EL	
	Gpd-1	Virus	Gpd-1	Virus
Complete-case	1.08 (-2.22, -1.14)	1.37 (0.73, 2.09)	1.11 (-2.17, -1.06)	1.33 (0.52, 1.85)
$\hat{\pi}(X)$	1.17 (-2.29, -1.13)	1.28 (0.68, 1.96)	1.23 (-2.26, -1.03)	1.41 (0.74, 2.15)
$\hat{\pi}(X, \delta)$	0.93 (-2.10, -1.18)	1.18 (0.86, 2.04)	0.88 (-2.33, -1.45)	1.25 (0.79, 2.04)
$\hat{\pi}(X, \delta, Z^c)$	0.76 (-1.97, -1.21)	1.16 (1.18, 2.34)	0.82 (-2.01, -1.19)	1.12 (1.18, 2.30)
$r = 1.5$	NA		EL	
	Gpd-1	Virus	Gpd-1	Virus
Complete-case	1.14 (-2.36, -1.22)	1.36 (0.79, 2.15)	1.23 (-2.19, -0.96)	1.47 (0.95, 2.42)
$\hat{\pi}(X)$	1.35 (-2.51, -1.17)	1.40 (0.69, 2.09)	1.41 (-2.30, -0.89)	1.33 (0.84, 2.17)
$\hat{\pi}(X, \delta)$	1.13 (-2.36, -1.24)	1.37 (0.82, 2.18)	1.12 (-2.48, -1.36)	1.50 (0.90, 2.40)
$\hat{\pi}(X, \delta, Z^c)$	0.92 (-2.15, -1.23)	1.28 (1.21, 2.49)	0.94 (-2.04, -1.10)	1.39 (1.14, 2.53)

Table 3.6 Interval lengths (LB, UB) of regression parameters for the mouse leukemia data (Thymic and nonthymic leukemia)

Model	Method			
$r = 1$	NA		EL	
	Gpd-1	Virus	Gpd-1	Virus
Complete-case	1.08 (-2.23, -1.15)	1.34 (0.73, 2.07)	1.08 (-2.10, -1.02)	1.32 (0.42, 1.74)
$\hat{\pi}(X)$	1.08 (-2.44, -1.36)	1.28 (0.67, 1.95)	1.10 (-2.42, -1.32)	1.28 (0.47, 1.75)
$\hat{\pi}(X, \delta)$	1.02 (-2.34, -1.32)	1.22 (0.67, 1.89)	1.09 (-2.27, -1.18)	1.36 (0.73, 2.09)
$\hat{\pi}(X, \delta, Z^c)$	0.90 (-2.22, -1.32)	1.13 (1.18, 2.32)	0.94 (-2.45, -1.51)	1.07 (0.65, 1.72)
$r = 1.5$	NA		EL	
	Gpd-1	Virus	Gpd-1	Virus
Complete-case	1.34 (-2.53, -1.19)	1.36 (0.81, 2.17)	1.35 (-2.39, -1.04)	1.45 (0.51, 1.96)
$\hat{\pi}(X)$	1.24 (-2.73, -1.49)	1.41 (0.70, 2.10)	1.25 (-2.48, -1.21)	1.47 (0.82, 2.29)
$\hat{\pi}(X, \delta)$	1.22 (-2.70, -1.48)	1.36 (0.87, 2.23)	1.22 (-2.31, -1.09)	1.39 (0.80, 2.19)
$\hat{\pi}(X, \delta, Z^c)$	1.18 (-2.72, -1.54)	1.24 (1.24, 2.49)	1.07(-2.50, -1.43)	1.18 (0.76, 1.94)

CHAPTER 4

JACKKNIFE EMPIRICAL LIKELIHOOD INFERENCE FOR THE ACCELERATED FAILURE TIME MODEL

4.1 Background

In survival analysis, covariate effects are considered to be associated with failure time. Researchers proposed the well-known proportional hazards model with right censoring (Cox (1972)) to explore the relationship between survival time of a patient and several explanatory variables. However, the proportional hazards model does not always fit the real data set. In fact, not satisfying the model assumption will result in incorrect regression parameter estimations. The accelerated failure time (AFT) model is a direct generalization of linear model to censored survival data analysis. It relates the logarithm of the failure time linearly to the covariates (Kalbfleisch and Prentice (1980)). As a result of its direct physical interpretation, the accelerated failure time model can be viewed as an alternative to proportional hazards model.

Various tools have been developed to perform nonparametric estimations. Among which, two approaches have drawn special attention. One is the Buckley-James estimator proposed by Buckley and James (1979), which provides an adjustment for censored observations using the Kaplan-Meier weights. Another is the rank-based estimator motivated by Prentice (1978). This rank-based method is heavily used by many researchers, for instance, Tsiatis (1990), Wei et al. (1990), and Ying (1993). An important estimation procedure for the accelerated failure time model is the rank-based estimating equations with a general weight, and a common choice is Gehan-type weight. Fygenson and Ritov (1994) used Gehan weight function that can lead to the monotone estimating equation in the accelerated failure time model, and linear programming for computation is required. Jin et al. (2003) estimated the variance through re-sampling method, and obtained the Gehan estimators via linear programming. Zeng and Lin (2008) proposed a simpler re-sampling strategy that does not require solving estimating equations.

Unfortunately, the re-sampling programming technique is still computationally demanding due to the lack of smoothness of estimating functions, especially, when many covariates and large sample sizes are involved. Furthermore, in order to bypass the computation challenge caused by the non-smooth step estimating functions, several authors further have promoted a useful induced smoothing approach with Gehan weight arising in the accelerated failure time model. This idea proposes to smooth the non-smooth estimating functions by adding a continuous normal noise to the regression coefficients, which leads to continuously differentiable estimating equation that can be dealt with using standard numerical methods, for example, Brown and Wang (2007), Heller (2007), Johnson and Strawderman (2009), Wang and Fu (2011) and Chiou et al. (2014a).

Despite the advantages of smoothing methods, the performances are often limited when used in small to moderate samples due to lack of accuracy of variance estimations. Furthermore, while such procedure can be implemented with relative numerical methods, the computational burden can still be high, especially with large datasets. An empirical likelihood (EL) approach can overcome these present challenges nevertheless.

Thomas and Grunkemeier (1975) were the earliest to propose inverting a nonparametric likelihood ratio test to obtain confidence intervals for the survival probability. Following that, Owen (1988) pioneered the empirical likelihood (EL) method, and this idea became more popular among researchers. By using EL method, researchers can determine the shape of confidence region without estimating the variance. A number of researchers today implement this method, and it has been recognized as a useful tool in statistical sciences. The biggest advantage of EL method is that it can avoid the need for specifying a distribution for the data. Theoretical results have also been established by Owen (1990), which states that the Wilk's theorem (Wilks (1938)) holds under mild conditions. For the high-dimensional case, Chen et al. (2015) established the asymptotic normality of the log empirical likelihood-ratio statistic when the sample size and the dimension of the data are comparable by adding two pseudo-observations to the original data set.

In this chapter, we provide a simpler and more computationally reliable method, jackknife empirical likelihood (JEL) method, for implementing the aforementioned estimators. JEL method is proposed by Jing et al. (2009), and this method constructs confidence region by introducing jackknife pseudo-values into EL methods. When dealing with more complicated computational problems, the JEL method has shown to offer more advantages. JEL method has been widely used ever since, and most recently seen in Zhao et al. (2015b) and Yang et al. (2016). These research studies indicate that the JEL might come in handy in dealing with more general statistics, and also substantially ease the computational burden of covariance matrix estimation. Building on Brown and Wang (2007), we apply JEL approach to the induced smoothed Gehan estimating equation to construct the confidence region for the parameters of interest.

The rest of the chapter is organized as follows. In Section 4.2, an inference procedure by using the jackknife empirical likelihood method is introduced. Simulation studies are then carried out to demonstrate the performance of the proposed JEL method in Section 4.3. In Section 4.4, we move on to present a real data application, followed by further discussions and concluding remarks in Section 4.5. Proofs are provided in the Appendix.

4.2 Main results

4.2.1 Accelerated failure time model and Gehan-type weight function

Let $\{T_i, C_i, X_i\}$, $i = 1, ..., n$, be n independent copies of $\{T, C, X\}$, where T_i and C_i are log-transformed failure time and log-transformed censoring time, X_i is a $p \times 1$-dimensional covariate vector (see Chiou et al. (2014a)). The accelerated failure time model is defined as

$$T_i = {X_i}^T\beta + \varepsilon_i, \ i = 1, ..., n,$$

where β is an $q \times 1$ unknown vector of regression parameters, ε's are the random variables with an unspecified distribution, and we assume that ε's are independent of X. Due to censoring, one can observe $\{Y_i, \Delta_i, X_i\}$, $i = 1, ..., n$, where $Y_i = \min(T_i, C_i)$, and $\Delta_i = I[T_i < C_i]$.

Fygenson and Ritov (1994) proposed a rank based estimating equation with Gehan's weight that is written as

$$\tilde{S}_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i (X_i - X_j) I[e_j(\beta) \geq e_i(\beta)], \tag{4.1}$$

where $e_i(\beta) = Y_i - X_i^T \beta$. It has been proven by Tsiatis (1990) that the solution to equation (4.1) is a consistent estimator of the true parameter β_0 and is asymptotically normal. However, solving the estimating equation is challenging because $\tilde{S}_n(\beta)$ is a discontinuous step function, due to lack of smoothness, estimating the regression parameters is computationally problematic. Moreover, if the derivative does not exist, it is more difficult to obtain the covariance matrix.

Then Brown and Wang (2007) proposed an induced smoothing approach for rank-based inference with Gehan's weight. They replaced the estimating function $\tilde{S}_n(\beta)$ in (4.1) with $E[\tilde{S}_n(\beta + \Gamma_n^{1/2} W)]$, where W is p-dimensional standard normal random vector, and Γ_n is a working covariance matrix of $\hat{\beta}_n$. Chiou et al. (2015) claimed that the forms of Γ_n generally have minimal impact on the bias and standard error (see Johnson and Strawderman (2009) and Chiou et al. (2014a)), and a choice of Γ_n is the identity matrix (see Brown and Wang (2005, 2007)). The induced smoothing approach smooths the discontinuous estimating function in a way that keeps the asymptotic property of the non-smooth estimating function. According to Brown and Wang (2007), the weight-adjusted estimating equation becomes

$$S_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i (X_i - X_j) \Phi \left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}} \right], \tag{4.2}$$

where $r_{ij}^2 = (X_i - X_j)^T (X_i - X_j)/n$, and $\Phi(\cdot)$ is the distribution function of $N(0, 1)$.

Recall that Heller (2007) directly approximated the indicator function with $1 - \Upsilon(u/h)$, where $\Upsilon(\cdot)$ denotes a local distribution function and h is a bandwidth. The resulting estimating equation becomes

$$S_n^*(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_i (X_i - X_j)\Upsilon\left[\frac{e_j(\beta) - e_i(\beta)}{h}\right]. \tag{4.3}$$

Comparing equations (4.2) and (4.3), if taking $\Upsilon(\cdot)$ to be the standard normal distribution $\Phi(\cdot)$, and replacing the bandwidth h with r_{ij}, the equation (4.3) becomes equation (4.2).

However, obtaining efficient and optimal bandwidth h is a very challenging task, which requires intensive computations. One of the potential advantages of utilizing the covariate-dependent bandwidth $r_{ij}^2 = (X_i - X_j)^T(X_i - X_j)/n$ in equation (4.2) is that expression of r_{ij} provides a closed-form of bandwidth. The other advantage is that equation (4.2) uses some information from data set X, and the smoothing parameter respects the scaling structure of the solution sequence.

Furthermore, the solution to equation (4.2) is still consistent to β_0 and has the same asymptotic distribution to equation (4.1). Brown and Wang (2007) suggest that asymptotic covariance is another sandwich formula, which is $D_n(\beta_0)^{-1}B_n(\beta_0)(D_n(\beta_0)^{-1})^T$, where $\mathrm{D_n}(\beta_0) = \mathrm{E}[\partial \mathrm{S_n}(\mathrm{Y};\beta)\}/\partial\beta]_{\beta_0}$, and $B_n(\beta_0) = \mathrm{cov}\{S_n(Y;\beta_0)\}$. It is worth to note that as $n \to \infty$, when $e_j(\beta) > e_i(\beta)$, $\Phi\left[\frac{e_j(\beta)-e_i(\beta)}{r_{ij}}\right] \to 1$; while when $e_j(\beta) < e_i(\beta)$, $\Phi\left[\frac{e_j(\beta)-e_i(\beta)}{r_{ij}}\right] \to 0$. Thus, $S_n(\beta)$ is a proper approximation to $\tilde{S}_n(\beta)$. Johnson and Strawderman (2009) further showed that under regularity conditions, S_n in (4.2) is asymptotically equivalent to $\tilde{S}_n$ in (4.1).

4.2.2 Jackknife empirical likelihood method

In order to avoid complicated covariance matrix estimation, we propose a jackknife empirical likelihood method to make inference for the accelerated failure time model in this subsection. Then we establish the asymptotic chi-square distribution of the empirical log-likelihood ratio and construct the confidence region.

Let $Z_i = (Y_i, \Delta_i, X_i)$. We can rewrite the weight and weight-adjusted estimating equations, given in (4.1) and (4.2) as U-statistic of degree 2 with symmetric kernel functions, respectively. See Appendix for details.

$$\tilde{S}_n(\beta) = \binom{n}{2}\left[\binom{n}{2}^{-1}\sum_{1\le i<j\le n} H(Z_i, Z_j;\beta)\right] \equiv \binom{n}{2} W_n(\beta).$$

The kernel function is defined below

$$H(Z_i, Z_j;\beta) = (X_i - X_j)\left\{\Delta_i I[e_j(\beta) \ge e_i(\beta)] - \Delta_j I[e_i(\beta) \ge e_j(\beta)]\right\}.$$

Also,

$$S_n(\beta) = \binom{n}{2}\left[\binom{n}{2}^{-1}\sum_{1\le i<j\le n} K(Z_i, Z_j;\beta)\right] \equiv \binom{n}{2} U_n(\beta),$$

with the kernel function is defined as follows

$$K(Z_i, Z_j;\beta) = (X_i - X_j)\left\{\Delta_i \Phi\left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}}\right] - \Delta_j \Phi\left[\frac{e_i(\beta) - e_j(\beta)}{r_{ij}}\right]\right\}.$$

The above U-statistic $U_n(\beta)$ is considered as a version of analogue in $S_n(\beta)$. Thus, we can apply JEL to this U-statistic instead of $S_n(\beta)$ to get a smoothed JEL.

Motivated by Jing et al. (2009), we construct the jackknife pseudo-sample for $U_n(\beta)$. Denote the jackknife pseudo-values $\hat{Q}_i(\beta) = nU_n(\beta) - (n-1)U_{n-1}^{(-i)}(\beta)$, $i = 1, ..., n$, where $U_{n-1}^{(-i)}(\beta) = U(Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_n)$ is computed on the $n-1$ samples formed from the original data set by deleting the ith observation. It can be shown that

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \hat{Q}_i(\beta).$$

Furthermore, according to the consistency (Johnson and Strawderman (2009)), we have that as $n \to \infty$, $EU_n(\beta_0) \to 0$.

Moreover, define $U_{n-1}^{(-i)}(\beta)$ as

$$U_{n-1}^{(-i)}(\beta) = \frac{1}{(n-1)(n-2)} \sum_{1\le q<j\le n, q\ne i} K_{n-1}^{(-i)}(Z_q, Z_j; \beta),$$

where $K_{n-1}^{(-i)}(\beta)$ is the kernel function corresponding to $U_{n-1}^{(-i)}(\beta)$ based on $n-1$ samples formed from the original data set by deleting the ith observation. Then, for $i = 1, ..., n$,

$$\begin{aligned}
E\hat{Q}_i(\beta_0) &= nEU_n(\beta_0) - (n-1)EU_{n-1}^{(-i)}(\beta_0) \\
&= \frac{n}{n(n-1)} E\left[\sum_{1\le i<j\le n} K(Z_i, Z_j; \beta_0)\right] \\
&- \frac{n-1}{(n-1)(n-2)} E\left[\sum_{1\le q<j\le n, q\ne i} K_{n-1}^{(-i)}(Z_q, Z_j; \beta_0)\right] \\
&= \frac{1}{n-1} \sum_{1\le i<j\le n} E\left[K(Z_i, Z_j; \beta_0)\right] - \frac{1}{n-2} \sum_{1\le q<j\le n, q\ne i} E\left[K_{n-1}^{(-i)}(Z_q, Z_j; \beta_0)\right] \\
&= \frac{1}{n-1} \times \frac{n(n-1)}{2} E\left[K(Z_1, Z_2; \beta_0)\right] - \frac{1}{n-2} \times \frac{(n-1)(n-2)}{2} E\left[K(Z_1, Z_2; \beta_0)\right] \\
&= \left(\frac{n}{2} - \frac{n-1}{2}\right) E\left[K(Z_1, Z_2; \beta_0)\right] \\
&= \frac{E\left[K(Z_1, Z_2; \beta_0)\right]}{2}.
\end{aligned}$$

Since $E\left[K(Z_1, Z_2; \beta_0)\right] \to 0$, we can derive that $E\hat{Q}_i(\beta_0) \to 0$ as $n \to \infty$.

Then the empirical likelihood at β is defined by

$$L(\beta) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i\hat{Q}_i(\beta) = 0, \sum_{i=1}^{n} p_i = 1, p_i \ge 0\right\}.$$

Hence, the jackknife empirical likelihood ratio at β is as follows

$$R(\beta) = \frac{L(\beta)}{n^{-n}} = \max\left\{\prod_{i=1}^{n} \{np_i\} : \sum_{i=1}^{n} p_i\hat{Q}_i(\beta) = 0, \sum_{i=1}^{n} p_i = 1, p_i \ge 0\right\}.$$

Using Lagrange multipliers, we get the jackknife empirical log-likelihood ratio as follows,

$$l(\beta) = -2\log\{R(\beta)\} = 2\sum_{i=1}^{n}\log\{1+\lambda(\beta)^T\hat{Q}_i(\beta)\}, \tag{4.4}$$

where λ is the solution to the equation

$$f(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Q}_i(\beta)}{1+\lambda(\beta)^T\hat{Q}_i(\beta)} = 0. \tag{4.5}$$

Let β_0 denote the true value of β. We establish the Wilk's theorem as follows.

Theorem 4.1. *Under some mild regularity conditions stated in the Appendix, as $n \longrightarrow \infty$, one has*

$$-2\log R(\beta_0) \xrightarrow{d} \chi_p^2,$$

where χ_p^2 is a standard chi-squared random variable with p degrees of freedom.

The $100(1-\alpha)\%$ JEL confidence region for β can be established as

$$R_\alpha = \left\{\beta : -2\log R(\beta) \leq \chi_p^2(\alpha)\right\},$$

where $\chi_p^2(\alpha)$ is the upper α-quantile of distribution of χ_p^2.

In practice, if we are interested in certain component of the regression parameter, the aforementioned procedure can be used to tackle the nuisance parameter by profiling the empirical likelihood. Denote $\beta = (\beta_1^T, \beta_2^T)^T$, where $\beta_1 \in R^q$ and $\beta_2 \in R^{p-q}$. Similar to Yang and Zhao (2012a) and Zhao and Yang (2012), we define the profile JEL ratio and log-likelihood ratio as follows:

$$R^*(\beta_1) = \max_{\beta_2} R(\beta),$$

and

$$l^*(\beta_1) = -2\log\{R^*(\beta_1)\}.$$

We obtain the standard result as Theorem 4.1 for the proposed profile JEL.

Theorem 4.2. *Under some mild regularity conditions stated in the Appendix, as $n \longrightarrow \infty$, one has*

$$-2 \log R^*(\beta_{10}) \xrightarrow{d} \chi^2_q,$$

where β_{10} is the true value of the parameter of interest β_1.

Thus, the $100(1-\alpha)\%$ JEL confidence region for β_1 can be established as

$$R^*_\alpha = \left\{ \beta_1 : -2 \log R^*(\beta_1) \leq \chi^2_q(\alpha) \right\},$$

where $\chi^2_q(\alpha)$ is the upper α-quantile of distribution of χ^2_q.

4.3 Simulation studies

In this section, we carry out simulation studies to compare the relative performance of the proposed JEL method with normal approximation (NA) procedure proposed by Brown and Wang (2007), which requires variance estimations.

For the comparison of the JEL and NA methods, we adopt the same simulation settings as that of Chiou et al. (2014a). As in Chiou et al. (2014a), we simulate failure times T from the accelerated failure time model

$$\log(T) = 2 + X_1 + X_2 + X_3 + \varepsilon,$$

where X_1 is generated from Bernoulli with rate 0.5, X_2 and X_3 are uncorrelated standard normal variables. ε is sampled from standard normal distribution. We assume censoring rate (CR) as four levels, approximately 15%, 30%, 45% and 60%, representing the different disease status. We choose three sample sizes n = 30, 60 and 100. All the simulations are conducted with 1000 repetitions.

Tables 4.1 to 4.4 display coverage probabilities and average lengths of confidence intervals for the 90% and 95% nominal confidence levels under four different censoring rates 15%, 30%, 45% and 60%, respectively. Abbreviations JEL and ISMB are used to denote JEL and

NA methods, IS indicates the induced smoothing approach and MB is using the multiplier bootstrap for variance calculation. R package *aftgee* developed by Chiou et al. (2014b) is used for the variance computation.

Comparing Table 4.1 to Table 4.4, at the same sample size, the average lengths of both methods increase as the CR increases. The JEL and NA methods have similar performances when the sample sizes are 60 and 100. Especially, when the censoring rate is 15%, the estimated coverage probabilities for JEL and NA are close to nominal levels. Overall, for both methods, when the sample size increases, the average lengths of confidence intervals are getting shorter.

When the sample size is small, $n = 30$, the JEL outperforms the NA method in terms of converge probabilities. We can observe that the JEL method has higher coverage probability than the NA method in most of cases. For instance, when CR = 15%, for β_1, under nominal level 0.95, the coverage probability of the JEL method is 0.931 whereas that of the ISMB is 0.924. Another example, that, is, when CR = 30%, for β_3, under nominal level 0.90, the coverage probability of the JEL method is 0.865, while that of the ISMB is 0.835. This shows that JEL method does have advantage when the sample size is 30.

Tables 4.3 and 4.4 show the performances for both JEL and NA methods under heavy censoring settings (CR = 45% and CR = 60%). When the censoring rate is getting higher, indicating we have less information about the data. Two methods perform similarly when the sample size is 100. However, when the sample size is only 30, the JEL performs sightly better than the NA method.

4.4 Application to real data

In this section, we analyzed two real data sets to illustrate the proposed JEL method. The first data set is Kidney catheter data, and the other one is National Wilm's tumor study data.

The first data set kidney catheter data can be found in the R package *survival*. This data set has been analyzed by McGilchrist and Aisbett (1991) using a log-normal frailty distribution. In this data set, recurrence times of kidney patients using portable dialysis equipment at the point of insertion of the catheter were recorded. When the infection occurs, the catheter is removed, and reinserted after certain time. The time to infection is considered as censored when the catheters are removed for some reasons, that are not infection. The risk variables are age (in years), sex (0 = male, 1 = female), and disease type coded as 0 = GN, 1 = AN, 2 = PKD, 3 = other. Thus, the five regression variables considered are age, sex, and presence/absence of disease types GN, AN, PKD. The data set contains 76 data points, which contain 38 patients, with each having exactly two-insertion information, and the censoring rate is 24%.

We fit the data set using the accelerated failure time model for the recurrence times T, that is,

$$\log(T) = \sum_{i=1}^{5} \beta_i X_i + \varepsilon,$$

where X_1 is the age, X_2 is the sex, X_3 is the disease-GN, X_4 is the disease-AN, and X_5 is the disease-PKD.

The second data set selected to demonstrate the performance of our proposed method is National Wilm's Tumor Study data (D'angio et al. (1989)). This data set can be found in the R package *survival* as well. Wilms tumor is a type of cancer that starts in the kidneys. It is the most common type of kidney cancer in young children. The study was conducted by the National Wilm's Tumor Study Group (NWTSG) and the interest of the study was to assess the relationship between the tumor histology measurement (histol) and the time to tumor relapse (edrel). Depending on the cell type, the tumor histology can be classified into favorable or unfavorable categories. Besides histology measurement, other covariates are patient age (age), disease stages (stage) and study group (study). More detailed introduction was given in Chiou et al. (2014a).

In this data set, there were a total of 4028 subjects, among whom, 571 were cases who experienced the relapse of tumor (rel=1) and 3457 patients did not (rel=0). Thus, the censoring rate is about 86 %. We consider the accelerated failure time model for the time to relapse T, that is,

$$\log(T) = \sum_{i=1}^{4} \beta_i X_i + \varepsilon.$$

The covariates that are taken into account are: X_1 = central histology measurement (1 = favorable, 0 = unfavorable), X_2 = age (measure in year) at diagnosis, X_3 = tumor stage indicators (with 4 being the latest and most severest) and X_4 = a study group indicator (NWTSG-3 and NWTSG-4).

To check the performance of the proposed method, for both application cases, we use R package *aftgee* to estimate β and obtain confidence intervals. The sandwich variance estimation method used is IS-MB. Tables 4.5 and 4.6 report interval estimates based on JEL and NA methods, along with the point estimator (PE) from NA method. LB, UB and Length denote the lower bound, upper bound and length of the 95% confidence interval.

In Table 4.5, we can conclude that the only statistically significant covariate is sex, indicating that gender had a significant effect on the time to infection. This suggests that female patients tend to have longer recurrence times to infection. This is a consistent finding as in McGilchrist and Aisbett (1991).

In Table 4.6, the lower and upper bounds indicate that the coefficients of central histological diagnosis for both methods are negative and significantly different from zero. This suggests that patients with unfavorable central histology measurement tend to have shorter time to tumor relapse.

4.5 Discussion

Based on Brown and Wang (2007), we propose a jackknife empirical likelihood method to make statistical inference for the accelerated failure time model with right censored data. The empirical log-likelihood ratio is proved to have the standard chi-squared distribution and the corresponding confidence intervals are developed. The simulation results suggest that the proposed JEL outperforms the normal approximation method in terms of coverage probability, especially when the sample size is small. Moreover, JEL method performs well when the censoring rate is higher than 45%. Therefore, it is worthwhile to apply the proposed method to heavy censored data sets. The results of two real data applications support our conclusion.

In summary, for the small to moderate sample sizes, when the censoring rate is heavy, we recommend to use the proposed JEL method to make inference for the accelerated failure time model, which improves the existing methods. Another advantage of the proposed JEL method is that it can be easily implemented in a standard software environment. Moreover, JEL methods can be applied to the kernel estimating equation with bandwidth h (see equation (4.3)) as well. Since it is very similar to our proposed estimators, which allows more broad applications than existing estimations in practice.

Table 4.1 Simulation results with the CR = 15%. AL is the average length of confidence intervals of β; CP is the coverage probability.

	JEL (90%)		ISMB (90%)		JEL (95%)		ISMB (95%)	
	AL	CP	AL	CP	AL	CP	AL	CP
$n = 30$								
β_1	0.874	0.879	0.799	0.875	1.041	0.931	0.929	0.924
β_2	0.853	0.880	0.789	0.855	1.016	0.929	0.931	0.914
β_3	0.866	0.879	0.795	0.876	1.032	0.937	0.937	0.929
$n = 60$								
β_1	0.721	0.880	0.532	0.889	0.859	0.941	0.635	0.937
β_2	0.704	0.874	0.533	0.892	0.839	0.942	0.636	0.933
β_3	0.710	0.875	0.538	0.888	0.846	0.940	0.641	0.937
$n = 100$								
β_1	0.428	0.887	0.396	0.904	0.510	0.950	0.488	0.949
β_2	0.427	0.890	0.367	0.887	0.509	0.949	0.489	0.945
β_3	0.428	0.884	0.368	0.888	0.510	0.949	0.488	0.949

Table 4.2 Simulation results with the CR = 30%.

	JEL (90%)		ISMB (90%)		JEL (95%)		ISMB (95%)	
	AL	CP	AL	CP	AL	CP	AL	CP
$n = 30$								
β_1	0.962	0.871	0.931	0.871	1.146	0.917	1.109	0.912
β_2	0.953	0.869	0.938	0.857	1.135	0.914	1.118	0.917
β_3	0.944	0.865	0.929	0.835	1.125	0.912	1.107	0.914
$n = 60$								
β_1	0.811	0.874	0.515	0.886	0.966	0.930	0.614	0.933
β_2	0.814	0.860	0.518	0.861	0.970	0.922	0.617	0.920
β_3	0.828	0.871	0.523	0.879	0.987	0.930	0.623	0.933
$n = 100$								
β_1	0.566	0.890	0.398	0.910	0.674	0.942	0.474	0.943
β_2	0.579	0.893	0.397	0.889	0.510	0.942	0.474	0.946
β_3	0.581	0.892	0.398	0.883	0.692	0.944	0.475	0.938

Table 4.3 Simulation results with the CR = 45%.

	JEL (90%)		ISMB (90%)		JEL (95%)		ISMB (95%)	
	AL	CP	AL	CP	AL	CP	AL	CP
$n = 30$								
β_1	1.104	0.860	0.861	0.862	1.315	0.914	1.026	0.916
β_2	1.102	0.848	0.848	0.839	1.313	0.910	1.012	0.899
β_3	0.997	0.851	0.862	0.845	1.188	0.900	1.025	0.896
$n = 60$								
β_1	0.945	0.866	0.583	0.878	1.126	0.921	0.694	0.920
β_2	0.941	0.866	0.587	0.871	1.121	0.922	0.695	0.924
β_3	0.939	0.862	0.586	0.881	1.119	0.918	0.699	0.929
$n = 100$								
β_1	0.684	0.872	0.443	0.886	0.815	0.927	0.528	0.937
β_2	0.683	0.872	0.444	0.884	0.814	0.929	0.528	0.926
β_3	0.683	0.872	0.437	0.896	0.814	0.929	0.521	0.933

Table 4.4 Simulation results with the CR = 60%.

	JEL (90%)		ISMB (90%)		JEL (95%)		ISMB (95%)	
	AL	CP	AL	CP	AL	CP	AL	CP
$n = 30$								
β_1	1.254	0.830	0.948	0.828	1.494	0.908	1.130	0.893
β_2	1.247	0.823	0.964	0.802	1.486	0.909	1.159	0.882
β_3	1.245	0.814	0.972	0.804	1.483	0.901	1.164	0.868
$n = 60$								
β_1	1.009	0.860	0.664	0.885	1.206	0.910	0.791	0.909
β_2	1.012	0.851	0.662	0.852	1.205	0.893	0.784	0.908
β_3	1.015	0.848	0.673	0.852	1.208	0.890	0.802	0.899
$n = 100$								
β_1	0.720	0.864	0.537	0.885	0.858	0.911	0.640	0.929
β_2	0.715	0.861	0.526	0.852	0.854	0.912	0.637	0.918
β_3	0.717	0.863	0.525	0.852	0.851	0.909	0.626	0.912

Table 4.5 Interval lengths of CIs for the kidney catheter data.

	JEL			ISMB			
	LB	UB	Length	LB	UB	Length	PE
Age	-0.014	0.071	0.085	-0.019	0.023	0.042	0.002
Sex	1.025	2.223	1.198	0.923	2.191	1.267	1.557
disease-GN	-1.200	0.592	1.792	-1.267	0.319	1.586	-0.474
disease-AN	-0.271	1.277	1.548	-1.339	0.145	1.484	-0.597
disease-PKD	-1.042	1.871	2.913	-0.765	1.861	2.626	0.548

Table 4.6 Interval lengths of CIs for the National Wilm's tumor study.

	JEL			ISMB			
	LB	UB	Length	LB	UB	Length	PE
Histol	-3.516	-2.874	0.642	-3.126	-2.592	0.533	-2.859
Age	-0.257	-0.114	0.143	-0.223	-0.113	0.110	-0.168
Stage	-0.612	-0.329	0.283	-0.757	-0.483	0.274	-0.620
Study	-0.429	0.174	0.603	-0.479	0.199	0.678	-0.140

CHAPTER 5

CONCLUSIONS

In the first part of this dissertation, we apply empirical likelihood (EL) method to construct confidence intervals for the regression parameters in the semiparametric transformation models with length-biased sampling. The transformation model plays an essential role in survival analysis. Length-biased data are left-truncated and right censored data under the stationary assumption, which is commonly seen in the context of observational studies. Motivated by Wang and Wang (2015), we introduce the EL inference procedure and prove that the log-likelihood ratio has the asymptotic distribution of χ_p^2. One of the most significant advantages of the empirical likelihood method is that it avoids estimating the complex covariance matrix comparing to normal approximation method. Furthermore, to improve the performances of the empirical likelihood method regarding coverage probability, we propose the adjusted empirical likelihood (AEL), which is initially developed by Chen et al. (2008). In the simulation study, compared to the normal approximation method, our proposed AEL method outperforms in most cases, especially when the censoring rate is relatively heavy.

Next, missing covariates are frequently encountered problems in survey studies. Motivated by the weighted estimating idea proposed by Qi et al. (2005) and Huang and Wang (2010), we employ EL method to make statistical inference for the semiparametric transformation model with missing covariates. Kernel smoothing technique is used to estimate the non-missingness probability. Under mild conditions, we prove that the empirical log-likelihood ratio is asymptotically chi-squared distribution. Numerically, we carry out an extensive simulation study to demonstrate the performance of our proposed method. In contrast to the normal approximation method, our approach shows better performance in the small samples.

The accelerated failure time (AFT) is a direct generalization of a linear model to censored survival data analysis. As a result of its straightforward physical interpretation, one can view the accelerated failure time model as an alternative to the proportional hazards model. We generate pseudo-jackknife sample to develop the jackknife empirical likelihood and propose the JEL method for the inference of the AFT model. Moreover, by using the jackknife pseudo-sample technique for the estimation equation, we prove that the Wilks' theorem for the JEL still holds. Comparing to the traditional empirical likelihood method, the JEL has a significant advantage in saving computational cost. Furthermore, we conduct the simulation studies. The coverage probability and average length of confidence intervals are calculated to support our conclusions.

In addition to the extensive simulation studies, we also provide four real date examples (Dementia disease, Mouse leukemia study, Kidney catheter, National Wilm's tumor study) to illustrate the use of the proposed methods, and demonstrate the comparability and superiority of our methods to some existing approaches. Moreover, we present some discussions on each topic and suggest some further research ideas.

REFERENCES

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory*, pages 1–25. Springer.

Addona, V. and Wolfson, D. B. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime data analysis*, 12(3):267–284.

An, Y. and Zhao, Y. (2017). Jackknife empirical likelihood for the difference of two volumes under ROC surfaces. *Annals of the Institute of Statistical Mathematics*, in press:1–18.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in medicine*, 2(2):273–277.

Bouadoumou, M., Zhao, Y., and Lu, Y. (2015). Jackknife empirical likelihood for the accelerated failure time model with censored data. *Communications in Statistics-Simulation and Computation*, 44(7):1818–1832.

Brown, B. and Wang, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, 92(1):149–158.

Brown, B. and Wang, Y.-G. (2007). Induced smoothing for rank regression with censored survival times. *Statistics in Medicine*, 26(4):828–836.

Buckley, J. (1984). Additive and multiplicative models for relative survival rates. *Biometrics*, 40(1):51–62.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.

Chen, B., Pan, G., Yang, Q., and Zhou, W. (2015). Large dimensional empirical likelihood. *Statistica Sinica*, 25:1659–1677.

Chen, H. Y. and Little, R. J. (1999). Proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, 94(447):896–908.

Chen, J., Variyath, A. M., and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17(2):426–443.

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, 89(3):659–668.

Cheng, S., Wei, L., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82(4):835–845.

Chiou, S., Kang, S., and Yan, J. (2015). Rank-based estimating equations with general weight for accelerated failure time models: an induced smoothing approach. *Statistics in Medicine*, 34(9):1495–1510.

Chiou, S. H., Kang, S., and Yan, J. (2014a). Fast accelerated failure time modeling for case-cohort data. *Statistics and Computing*, 24(4):559–568.

Chiou, S. H., Kang, S., Yan, J., et al. (2014b). Fitting accelerated failure time models in routine survival analysis with R package aftgee. *Journal of Statistical Software*, 61(11):1–23.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society. Series A (General)*, 148(2):82–117.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Cuzick, J. (1988). Rank regression. *The Annals of Statistics*, 16(4):1369–1389.

Dabrowska, D. M. and Doksum, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics*, 15(1):1–23.

D'angio, G. J., Breslow, N., Beckwith, J. B., Evans, A., Baum, E., Delorimier, A., Fernbach, D., Hrabovsky, E., Jones, B., Kelalis, P., et al. (1989). Treatment of wilms' tumor. results of the third national wilms' tumor study. *Cancer*, 64(2):349–360.

Fleming, T. R. and Harrington, D. P. (2011). *Counting Processes and Survival Analysis*, volume 169. John Wiley & Sons, New York.

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

Heller, G. (2007). Smoothed rank regression with censored data. *Journal of the American Statistical Association*, 102(478):552–559.

Huang, B. and Wang, Q. (2010). Semiparametric analysis based on weighted estimating equations for transformation models with missing covariates. *Journal of Multivariate Analysis*, 101(9):2078–2090.

Huang, H. and Zhao, Y. (2018). Empirical likelihood for the bivariate survival function under univariate censoring. *Journal of Statistical Planning and Inference*, 194:32–46.

Huffer, F. and McKeague, I. (1991). Weighted least-squares estimation for aalen's additive risk model. technical report. Technical report, Stanford Univ., CA (United States). Dept. of Statistics.

Jin, Z., Lin, D., Wei, L., and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, 90(2):341–353.

Jing, B.-Y., Yuan, J., and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*, 104(487):1224–1232.

Johnson, L. M. and Strawderman, R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. *Biometrika*, 96(3):577–590.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kleinbaum, D. G. (1998). Survival analysis, a self-learning text. *Biometrical Journal*, 40(1):107–108.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica*, 4:199–218.

Li, Z., Xu, J., and Zhou, W. (2016). On nonsmooth estimating functions via jackknife empirical likelihood. *Scandinavian Journal of Statistics*, 43(1):49–69.

Lin, D. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*, 88(424):1341–1349.

Lin, H.-L., Li, Z., Wang, D., and Zhao, Y. (2017). Jackknife empirical likelihood for the error variance in linear models. *Journal of Nonparametric Statistics*, 29(2):151–166.

Liu, X. and Zhao, Y. (2012). Semi-empirical likelihood inference for the ROC curve with missing data. *Journal of Statistical Planning and Inference*, 142(12):3123–3133.

Lu, W. and Liang, Y. (2006). Empirical likelihood inference for linear transformation models. *Journal of Multivariate Analysis*, 97(7):1586–1599.

McGilchrist, C. and Aisbett, C. (1991). Regression with frailty in survival analysis. *Biometrics*, 47(2):461–466.

Murphy, S., Rossini, A., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92(439):968–976.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. B. (2001). *Empirical Likelihood*. CRC Press, London.

Pettitt, A. (1984). Proportional odds models for survival data and estimates using ranks. *Applied Statistics*, 33(2):169–175.

Pierce, D. A. and Preston, D. L. (1984). Hazard function modelling for dose-response analysis of cancer incidence in the a-bomb survivor data. Technical report.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, 65(1):167–179.

Qi, L., Wang, C., and Prentice, R. L. (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, 100(472):1250–1263.

Qi, L., Zhang, X., Sun, Y., Wang, L., and Zhao, Y. (2018). Weighted estimating equations for additive hazards models with missing covariates. *Annals of the Institute of Statistical Mathematics*, in press.

Qin, J. (2017). Examples and basic theories for length biased sampling problems. In *Biased Sampling, Over-identified Parameter Problems and Beyond*. Springer.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22:300–325.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Shen, Y., Ning, J., and Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association*, 104(487):1192–1202.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70(352):865–871.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, 9(1):93–108.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, 18(1):354–372.

Wang, C. and Chen, H. Y. (2001). Augmented inverse probability weighted estimator for Cox missing covariate regression. *Biometrics*, 57(2):414–419.

Wang, D., Tian, L., and Zhao, Y. (2017). Smoothed empirical likelihood for the Youden index. *Computational Statistics & Data Analysis*, 115:1–10.

Wang, D. and Zhao, Y. (2016). Jackknife empirical likelihood for comparing two Gini indices. *Canadian Journal of Statistics*, 44(1):102–119.

Wang, D., Zhao, Y., and Gilmore, D. W. (2016). Jackknife empirical likelihood confidence interval for the Gini index. *Statistics & Probability Letters*, 110:289–295.

Wang, Q.-H. and Jing, B.-Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. *Annals of the Institute of Statistical Mathematics*, 53(3):517–527.

Wang, S., Cui, H., and Li, R. (2013). Empirical likelihood inference for semi-parametric estimating equations. *Science China Mathematics*, 56:1247–1262.

Wang, X. and Wang, Q. (2015). Estimation for semiparametric transformation models with length-biased sampling. *Journal of Statistical Planning and Inference*, 156:80 – 89.

Wang, Y.-G. and Fu, L. (2011). Rank regression for accelerated failure time model with clustered and censored data. *Computational Statistics & Data Analysis*, 55(7):2334–2343.

Wei, L.-J., Ying, Z., and Lin, D. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika*, 77(4):845–851.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.

Yang, H., Liu, S., and Zhao, Y. (2016). Jackknife empirical likelihood for linear transformation models with right censoring. *Annals of the Institute of Statistical Mathematics*, 68(5):1095–1109.

Yang, H., Lu, K., and Zhao, Y. (2017). A nonparametric approach for partial areas under ROC curves and ordinal dominance curves. *Statistica Sinica*, 27:357–371.

Yang, H. and Zhao, Y. (2012a). New empirical likelihood inference for linear transformation models. *Journal of Statistical Planning and Inference*, 142:1659–1668.

Yang, H. and Zhao, Y. (2012b). Smoothed empirical likelihood for ROC curves with censored data. *Journal of Multivariate Analysis*, 109:254–263.

Yang, H. and Zhao, Y. (2013). Smoothed jackknife empirical likelihood inference for the difference of ROC curves. *Journal of Multivariate Analysis*, 115:270–284.

Yang, H. and Zhao, Y. (2015). Smoothed jackknife empirical likelihood inference for ROC curves with missing data. *Journal of Multivariate Analysis*, 140:123–138.

Yang, H. and Zhao, Y. (2017). Smoothed jackknife empirical likelihood for the difference of two quantiles. *Annals of the Institute of Statistical Mathematics*, 69(5):1059–1073.

Yang, H. and Zhao, Y. (2018). Smoothed jackknife empirical likelihood for the one-sample difference of quantiles. *Computational Statistics & Data Analysis*, 120:58–69.

Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression. *The Annals of Statistics*, 40(2):1102–1131.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, 21(1):76–99.

Yu, W., Sun, Y., and Zheng, M. (2011). Empirical likelihood method for linear transformation models. *Annals of the Institute of Statistical Mathematics*, 63(2):331–346.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640.

Zeng, D. and Lin, D. (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostatistics*, 9(2):355–363.

Zhang, Z. and Zhao, Y. (2013). Empirical likelihood for linear transformation models with interval-censored failure time data. *Journal of Multivariate Analysis*, 116:398–409.

Zhao, M., Zhao, Y., and McKeague, I. W. (2015a). Empirical likelihood inference for the odds ratio of two survival functions under right censoring. *Statistics & Probability Letters*, 107:304–312.

Zhao, Y. (2010). Semiparametric inference for transformation models via empirical likelihood. *Journal of Multivariate Analysis*, 101(8):1846–1858.

Zhao, Y. and Chen, F. (2008). Empirical likelihood inference for censored median regression model via nonparametric kernel estimation. *Journal of Multivariate Analysis*, 99(2):215–231.

Zhao, Y. and Jinnah, A. (2011). Inference for Coxs regression models via adjusted empirical likelihood. *Computational Statistics*, 27:1–12.

Zhao, Y., Meng, X., and Yang, H. (2015b). Jackknife empirical likelihood inference for the mean absolute deviation. *Computational Statistics & Data Analysis*, 91:92–101.

Zhao, Y. and Yang, S. (2012). Empirical likelihood confidence intervals for regression parameters of the survival rate. *Journal of Nonparametric Statistics*, 24(1):59–70.

Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika*, 92(2):492–498.

Zhou, M. and Li, G. (2008). Empirical likelihood analysis of the Buckley–James estimator. *Journal of Multivariate Analysis*, 99(4):649–664.

APPENDICES

Appendix A

Proof of theorems for Chapter 2

In this Appendix, we provide the proofs of Theorems 2.1, 2.2, 2.3 and 2.4.

Proof of Theorem 2.1.

In order to ensure the central limit theorem for counting process martingales, certain regularity conditions, which can be found in Fleming and Harrington (2011), need to hold. On the other hand, as the same conditions as in Wang and Wang (2015) are stated.

(D.1) $\lambda(\cdot)$ is positive and $\dot{\lambda}(\cdot)$ is bounded and continuous on $(-\infty, B)$, where B is a finite constant.

(D.2) For some constant $C > 0$, $P(\|Z\| < C) = 1$.

(D.3) $H_0(\cdot)$ has continuous and positive derivatives on $[0, \tau]$.

(D.4) Σ^* and Σ_* are assumed to be finite and nondegenerate.

(D.5) $E\Big[\int_0^\infty ZY(t)r(t,\tilde{T},\delta)d\Lambda_\varepsilon\{\beta_0' Z + H_0(t)\}\Big]^2 < \infty$.

Notice that

$$\sum_{i=1}^n d\hat{M}_i(\beta_0, t) = \sum_{i=1}^n \Big[dN_i(t) - Y_i(t)\hat{r}(t, \tilde{T}_i, \delta_i)d\Lambda_\varepsilon\{\beta_0' Z_i + \hat{H}(\beta_0, t)\}\Big] = 0.$$

In addition, following Zhao and Yang (2012), and recall that

$$d\hat{\Lambda}_C = \sum_{i=1}^n dI(V_i \wedge C_i \leq t, \delta = 0) \Big/ \sum_{i=1}^n I(V_i \wedge C_i \geq t)$$

satisfies

$$\sum_{i=1}^n d\hat{M}_{C_i}(t) = \sum_{i=1}^n \Big\{dI(V_i \wedge C_i \leq t, \delta = 0) - I(V_i \wedge C_i \geq t)d\hat{\Lambda}_C(t)\Big\}$$

$$= \sum_{i=1}^{n} \left\{ dI(V_i \wedge C_i \leq t, \delta = 0) - I(V_i \wedge C_i \geq t) \times \frac{\sum_{i=1}^{n} dI(V_i \wedge C_i \leq t, \delta = 0)}{\sum_{i=1}^{n} I(V_i \wedge C_i \geq t)} \right\}$$

$$= \sum_{i=1}^{n} I(V_i \wedge C_i \leq t, \delta = 0) - \sum_{i=1}^{n} I(V_i \wedge C_i \geq t) \times \frac{\sum_{i=1}^{n} dI(V_i \wedge C_i \leq t, \delta = 0)}{\sum_{i=1}^{n} I(V_i \wedge C_i \geq t)}$$

$$= 0.$$

It is easy to show that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \left\{ (Z_i - \hat{z}(\beta_0, t))\, d\hat{M}_i(\beta_0, t) + \hat{a}(\beta_0, t) d\hat{M}_{C_i}(t) \right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} Z_i d\hat{M}_i(\beta_0, t) - \frac{1}{\sqrt{n}} \int_0^{\tau} \hat{z}(\beta_0, t) \sum_{i=1}^{n} d\hat{M}_i(\beta_0, t) \\
&+ \frac{1}{\sqrt{n}} \int_0^{\tau} \hat{a}(\beta_0, t) \sum_{i=1}^{n} d\hat{M}_{Ci}(t) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} Z_i \left[dN_i(t) - Y_i(t) \hat{r}(t, \tilde{T}_i, \delta_i) d\Lambda_{\varepsilon}\{\beta_0' Z_i + \hat{H}(\beta_0, t)\} \right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i(\beta_0) + o_p(1).
\end{aligned}$$

The last equation directly comes from the Appendix of Wang and Wang (2015). Therefore, according to the conclusion of proofs in Step 3 in the Appendix of Wang and Wang (2015), one has that when $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta_0) \xrightarrow{d} N(0, \Sigma^*(\beta_0)).$$

It can also be shown that for $i = 1, ..., n$,

$$\begin{aligned}
W_{ni}(\beta_0) &= W_i(\beta_0) + \int_0^{\tau} [z(t) - \hat{z}(\beta_0, t)] dN_i \\
&+ Z_i \left(\int_0^{\tau} Y_i(t) \hat{r}(t, \tilde{T}_i, \delta_i) d\Lambda_{\varepsilon}\{\beta_0' Z_i + \hat{H}(\beta_0, t)\} - \int_0^{\tau} Y_i(t) r(t, \tilde{T}_i, \delta_i) d\Lambda_{\varepsilon}\{\beta_0' Z_i + H_0(t)\} \right)
\end{aligned}$$

$$+\left(\int_0^\tau \hat{z}(\beta_0,t)Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)d\Lambda_\varepsilon\{\beta_0'Z_i+\hat{H}(\beta_0,t)\} - \int_0^\tau z(t)Y_i(t)r(t,\tilde{T}_i,\delta_i)d\Lambda_\varepsilon\{\beta_0'Z_i+H_0(t)\}\right)$$
$$+\int_0^\tau (\hat{a}(\beta_0,t)-a(t))d\left(I(V_i\wedge C_i\le t,\delta_i=0)\right)$$
$$+\left(\int_0^\tau \hat{a}(\beta_0,t)I(V_i\wedge C_i\ge t)d\hat{\Lambda}_C(t) - \int_0^\tau a(t)I(V_i\wedge C_i\ge t)d\Lambda_C(t)\right)$$
$$= W_i(\beta_0)+r_{i1}+r_{i2}+r_{i3}+r_{i4}+r_{i5}.$$

First, let us prove the following result holds,

$$\left\|\int_0^\tau Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)d\Lambda_\varepsilon\{\beta_0'Z_i+\hat{H}(\beta_0,t)\} - \int_0^\tau Y_i(t)r(t,\tilde{T}_i,\delta_i)d\Lambda_\varepsilon\{\beta_0'Z_i+H_0(t)\}\right\| = o_p(1). \tag{A.1}$$

By the consistency of $\hat{r}(t,\tilde{T}_i,\delta_i)$ to $r(t,\tilde{T}_i,\delta_i)$ and (A.9)

$$\hat{H}(\beta_0,t)-H_0(t) = \frac{1}{n}\sum_{j=1}^n\int_0^t \frac{B(s,t)}{B_2(s)}dM_j(s)+o_p(n^{-1/2})$$

in the Appendix of Wang and Wang (2015), one can have that

$$\int_0^\tau Y_i(t)\hat{r}(t,\tilde{T}_i,\delta_i)d\Lambda_\varepsilon\{\beta_0'Z_i+\hat{H}(\beta_0,t)\} - \int_0^\tau Y_i(t)r(t,\tilde{T}_i,\delta_i)d\Lambda_\varepsilon\{\beta_0'Z_i+H_0(t)\}$$
$$=(1+o_p(1))\int_0^\tau Y_i(t)r(t,\tilde{T}_i,\delta_i)d[\Lambda_\varepsilon\{\beta_0'Z_i+\hat{H}(\beta_0,t)\}-\Lambda_\varepsilon\{\beta_0'Z_i+H_0(t)\}]$$
$$=(1+o_p(1))\int_0^\tau Y_i(t)r(t,\tilde{T}_i,\delta_i)d\left[\Lambda_\varepsilon\{\beta_0'Z_i+H_0(t)\}(\hat{H}(\beta_0,t)-H_0(t))\right]$$
$$=(1+o_p(1))\int_0^\tau Y_i(t)r(t,\tilde{T}_i,\delta_i)d\left[\Lambda_\varepsilon\{\beta_0'Z_i+H_0(t)\}\left(\frac{1}{n}\sum_{j=1}^n\int_0^t\frac{B(u,t)}{B_2(u)}dM_j(u)+o_p(n^{-1/2})\right)\right]$$
$$=(1+o_p(1))\int_0^\tau\left\{Y_i(t)r(t,\tilde{T}_i,\delta_i)\left[\Lambda_\varepsilon\{\beta'{}_0Z_i+H_0(t)\}\left(\frac{1}{n}\sum_{j=1}^n\frac{B(t,t)}{B_2(t)}dM_j(t)\right)\right.\right.$$
$$\left.\left.+\left(\frac{1}{n}\sum_{j=1}^n\int_0^t\frac{B(u,t)}{B_2(u)}dM_j(u)+o_p(n^{-1/2})\right)d\Lambda_\varepsilon\{\beta'{}_0Z_i+H_0(t)\}\right]\right\}.$$

By condition (D.5), $\left|\int_0^\tau Y(t)r(t,\tilde{T}_i,\delta)d\Lambda_\varepsilon\{\beta'{}_0Z+H_0(t)\}\right|$ has a finite second moment. As $B(u,t)/B_2(t)$ is bounded and $M_j(t)$ $(j=1,...,n)$ is a martingale, the above equation will be $o_p(1)$. Thus, (A.1) is valid.

It follows the uniform consistency of Kaplan-Meier and Nelson-Aalen estimators $\hat{S}_C$, $\hat{\Lambda}_C$ and $\hat{\omega}(t, \tilde{T}_i, \delta_i)$, recalling the definition of $z(t)$ and $a(t)$, one can obtain the uniform consistency of $\hat{z}(\beta_0, t)$ and $\hat{a}(\beta_0, t)$, that are,

$$\sup_{0 \le t \le \tau} |\hat{z}(\beta_0, t) - z(t)| \xrightarrow{p} 0, \tag{A.2}$$

and

$$\sup_{0 \le t \le \tau} |\hat{a}(\beta_0, t) - a(t)| \xrightarrow{p} 0. \tag{A.3}$$

Given condition (D.2), (A.1), (A.2) and (A.3), $\|r_{i1}\| = o_p(1)$, $\|r_{i2}\| = o_p(1)$, $\|r_{i3}\| = o_p(1)$ and $\|r_{i4}\| = o_p(1)$ hold.

Furthermore, we have that

$$\begin{aligned}
r_{i5} &= \int_0^\tau \hat{a}(\beta_0, t) I(V_i \wedge C_i \ge t) d\hat{\Lambda}_C(t) - \int_0^\tau a(t) I(V_i \wedge C_i \ge t) d\Lambda_C(t) \\
&= \int_0^\tau \hat{a}(\beta_0, t) I(V_i \wedge C_i \ge t) d\hat{\Lambda}_C(t) - \int_0^\tau \hat{a}(\beta_0, t) I(V_i \wedge C_i \ge t) d\Lambda_C(t) \\
&+ \int_0^\tau \hat{a}(\beta_0, t) I(V_i \wedge C_i \ge t) d\Lambda_C(t) - \int_0^\tau a(t) I(V_i \wedge C_i \ge t) d\Lambda_C(t) \\
&= \int_0^\tau \hat{a}(\beta_0, t) I(V_i \wedge C_i \ge t) d(\hat{\Lambda}_C(t) - \Lambda_C(t)) \\
&+ \int_0^\tau (\hat{a}(\beta_0, t) - a(t)) d\left(I(V_i \wedge C_i \le t, \delta_i = 0)\right) \\
&= (1 + o_p(1)) \int_0^\tau a(t) I(V_i \wedge C_i \ge t) d(\hat{\Lambda}_C(t) - \Lambda_C(t)) \\
&+ \int_0^\tau (\hat{a}(\beta_0, t) - a(t)) d\left(I(V_i \wedge C_i \le t, \delta_i = 0)\right) \\
&= (1 + o_p(1)) \int_0^\tau \frac{a(t) I(V_i \wedge C_i \ge t) \sum_{j=1}^n dM_{Cj}(t)}{\sum_{j=1}^n I(V_j \wedge C_j \ge t)} + o_p(1).
\end{aligned}$$

For each i, since $a(t) I(V_i \wedge C_i \ge t) \Big/ \sum_{j=1}^n I(V_j \wedge C_j \ge t)$ is predicable and finite, $M_{Cj}(t)$ ($j =$

$1, ..., n)$ is a martingale. The following term

$$\int_0^\tau a(t)I(V_i \wedge C_i \geq t)\sum_{j=1}^n dM_{Cj}(t) \Big/ \sum_{j=1}^n I(V_j \wedge C_j \geq t)$$

is a martingale integral (Andersen et al. (2012)), which converges to 0 in probability. It results in that $\|r_{i5}\| = o_p(1)$.

As a result, for $i = 1, 2, ..., n$,

$$W_{ni}(\beta_0) = W_i(\beta_0) + o_p(1). \tag{A.4}$$

Subsequently, for any $k \in R^p$, the following decompositions is provided,

$$\begin{aligned} & k'\left(\frac{1}{n}\sum_{i=1}^n W_{ni}(\beta_0)(W_{ni}(\beta_0))' - \frac{1}{n}\sum_{i=1}^n W_i(\beta_0)(W_i(\beta_0))'\right)k \\ & = \frac{1}{n}\sum_{i=1}^n \left[k'\left(W_{ni}(\beta_0) - W_i(\beta_0)\right)\right]^2 + \frac{2}{n}\sum_{i=1}^n \left(k'W_i(\beta_0)\right) \\ & \times \left[k'(W_{ni}(\beta_0) - W_i(\beta_0)\right]. \end{aligned}$$

By (A.4), it can be shown that the two parts of right-hand side of the above equation are both $o_p(1)$. Then one has

$$\frac{1}{n}\sum_{i=1}^n W_{ni}(\beta_0)(W_{ni}(\beta_0))' = \frac{1}{n}\sum_{i=1}^n W_i(\beta_0)(W_i(\beta_0))' + o_p(1).$$

According to the law of large numbers, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n W_i(\beta_0)(W_i(\beta_0))' \xrightarrow{p} E\left[W_i(\beta_0)(W_i(\beta_0))'\right] = \Sigma^*(\beta_0).$$

Furthermore, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n W_{ni}(\beta_0)(W_{ni}(\beta_0))' \xrightarrow{p} \Sigma^*(\beta_0).$$

□

Proof of Theorem 2.2.

Similar to Yu et al. (2011), we need to show that (1): $\max_{1\le i\le n}\|W_{ni}(\beta_0)\| = o_p(n^{1/2})$ and (2): $\theta = O_p(n^{-1/2})$, where θ is the solution of equation (2.9).

Recall that $W_{ni}(\beta_0) = W_i(\beta_0) + o_p(1)$ $(i = 1, 2, ..., n)$ is proven, hence, we only need to show that $\max_{1\le i\le n}\|W_i(\beta_0)\| = o_p(n^{1/2})$. Having that $W_i(\beta_0)$ are i.i.d. random variables, and $E\left[W_i(\beta_0)(W_i(\beta_0))'\right] = \Sigma^*(\beta_0) < \infty$, which indicates that $W_i(\beta_0)$ has finite second moment, by Lemma 11.2 of Owen (2001), one has that (1): $\max_{1\le i\le n}\|W_{ni}(\beta_0)\| = o_p(n^{1/2})$ is valid.

Next, similar arguments in the proofs of the Theorem 3.2 of Owen (2001) and Lu and Liang (2006) are followed. Let $g(\theta) = \left(\sum_{i=1}^{n} W_{ni}(\beta_0)/(1+\theta' W_{ni}(\beta_0))\right)\Big/ n$, $\theta = \rho\eta$, where $\rho > 0$ and $\|\eta\| = 1$, we obtain that

$$\begin{aligned}
0 &= \|g(\theta)\| \\
&= \|g(\rho\eta)\| \\
&\ge |\eta' g(\rho\eta)| \\
&\ge \frac{1}{n}\left|\eta' \sum_{i=1}^{n} \frac{W_{ni}(\beta_0)}{1+\theta' W_{ni}(\beta_0)}\right| \\
&= \frac{1}{n}\left|\eta' \left\{\sum_{i=1}^{n} W_{ni}(\beta_0) - \rho \sum_{i=1}^{n} \frac{W_{ni}(\beta_0)\eta' W_{ni}(\beta_0)}{1+\rho\eta' W_{ni}(\beta_0)}\right\}\right| \\
&\ge \frac{\rho}{n}\eta' \sum_{i=1}^{n} \frac{W_{ni}(\beta_0)(W_{ni}(\beta_0))'}{1+\rho\eta' W_{ni}(\beta_0)}\eta - \frac{1}{n}\left|\sum_{i=1}^{n} W_{ni}(\beta_0)\right| \\
&\ge \frac{\rho\eta' Q_n(\beta_0)\eta}{1+\rho W_n(\beta_0)} - \frac{1}{n}\left|\sum_{i=1}^{n} W_{ni}(\beta_0)\right|,
\end{aligned}$$

where $Q_n(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} W_{ni}(\beta_0)(W_{ni}(\beta_0))'$, and $W_n(\beta_0) = \max_{1\le i\le n}\|W_{ni}(\beta_0)\| = o_p(n^{1/2})$.

By the law of large numbers, one has $Q_n(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} W_i(\beta_0)(W_i(\beta_0))' + o_p(1)$. Therefore, as $n \to \infty$, $\lim_{n\to\infty} Q_n(\beta_0) = E\left[W_i(\beta_0)(W_i(\beta_0))'\right] = \Sigma^*(\beta_0)$. In addition, based on Theorem 2.1,

$$\frac{1}{n}\left|\sum_{i=1}^{n} W_{ni}(\beta_0)\right| = O_p(n^{-1/2})$$

is valid. Thus, one can prove that (2): $\theta = O_p(n^{-1/2})$ is also true.

From the results (1) and (2) above, it can be shown that

$$\theta = \left(\frac{1}{n}\sum_{i=1}^{n} W_{ni}(\beta_0)(W_{ni}(\beta_0))'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} W_{ni}(\beta_0)\right) + o_p(n^{-1/2}).$$

Then, by Taylor expansion, we can derive that

$$\begin{aligned} l_n(\beta_0) &= 2\sum_{i=1}^{n}\theta' W_{ni}(\beta_0) - \sum_{i=1}^{n}\theta' W_{ni}(\beta_0)(W_{ni}(\beta_0))'\theta + o_p(1) \\ &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right)'\left(\frac{1}{n}\sum_{i=1}^{n} W_{ni}(\beta_0)(W_{ni}(\beta_0))'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right) + o_p(1). \end{aligned}$$

Combining the Slutsky Lemma and Theorem 2.1, we complete the proof of Theorem 2.2. □

Proof of Theorem 2.3.

The proof is to follow the arguments in Yu et al. (2011). Note that $\beta_0 = (\beta_{10}', \beta_{20}')'$, and the corresponding $Z = (Z_1', Z_2')'$. Define

$$\tilde{\Sigma}_*(\beta_0) = E\left[\int_0^\tau \{Z - z(t)\}Z_2' Y(t) r(t, \tilde{T}, \delta)\dot{\lambda}_\varepsilon\{\beta_0' Z + H_0(t)\} dH_0(t)\right].$$

Since Σ_* is positive definite, $\tilde{\Sigma}_*$ is of rank $p - q$. Denote $\hat{\beta}_2(\beta_{10}) = \arg\inf_{\beta_2} l_n(\beta_{10}', \beta_2')'$. Let

$$\Pi = \tilde{\Sigma}_*(\beta_0)'\Sigma^*(\beta_0)^{-1}\tilde{\Sigma}_*(\beta_0).$$

By similar arguments to Qin and Lawless (1994) and Yu et al. (2011), we can show that

$$\sqrt{n}(\hat{\beta}_2 - \beta_{20}) = -\Pi(\beta_0)^{-1}\tilde{\Sigma}_*(\beta_0)'\Sigma^*(\beta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0) + o_p(1),$$

and the Lagrange multiplier θ_2 satisfies that

$$\sqrt{n}\theta_2 = \left(I - \Sigma^*(\beta_0)^{-1}\tilde{\Sigma}_*(\beta_0)\Pi(\beta_0)^{-1}\tilde{\Sigma}_*(\beta_0)'\right)\Sigma^*(\beta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0) + o_p(1).$$

Thus, by Taylor's expansion, we have that

$$\begin{aligned} l_n^*(\beta_{10}) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right)' \left(\Sigma^*(\beta_0)^{-1} - \Sigma^*(\beta_0)^{-1}\tilde{\Sigma}_*(\beta_0)(\tilde{\Sigma}_*(\beta_0)'\Sigma^*(\beta_0)^{-1}\tilde{\Sigma}_*(\beta_0))^{-1}\tilde{\Sigma}_*(\beta_0)'\Sigma^*(\beta_0)^{-1}\right) \\ &\quad \times \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right) + o_p(1) \\ &= \left(\Sigma^*(\beta_0)^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right)' S \left(\Sigma^*(\beta_0)^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right) + o_p(1), \end{aligned}$$

where

$$S = I - \Sigma^*(\beta_0)^{-1/2}\tilde{\Sigma}_*(\beta_0)(\tilde{\Sigma}_*(\beta_0)'\Sigma^*(\beta_0)^{-1}\tilde{\Sigma}_*(\beta_0))^{-1}\tilde{\Sigma}_*(\beta_0)'\Sigma^*(\beta_0)^{-1/2}$$

is a symmetric and idempotent matrix with trace q. By Theorem 2.1,

$$\Sigma^*(\beta_0)^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0) \xrightarrow{d} N(0, I).$$

We proved Theorem 2.3. □

Proof of Theorem 2.4.

As Chen et al. (2008) did, we can obtain Theorem 2.4 from Theorem 2.2. Similarly, we need to show that $\theta^{ad} = O_p(n^{-1/2})$. Define

$$v(\theta^{ad}) = \frac{1}{n+1}\sum_{i=1}^{n+1} \frac{W_{ni}^{ad}(\beta_0)}{1 + (\theta^{ad})' W_{ni}^{ad}(\beta_0)},$$

and $\theta^{ad} = \rho_1\eta_1$, where $\rho_1 > 0$ and $\|\eta_1\| = 1$. Using the result of Theorem 2.2 and equation (2.10), one has

$$0 = \left\|v(\theta^{ad})\right\|$$

$$
\begin{aligned}
&= \|v(\rho_1\eta_1)\| \\
&\geq |\eta_1' v(\rho_1\eta_1)| \\
&\geq \frac{1}{n}\left|{\eta'}_1 \sum_{i=1}^{n+1} \frac{W_{ni}^{ad}(\beta_0)}{1+(\theta^{ad})' W_{ni}^{ad}(\beta_0)}\right| \\
&= \frac{1}{n}\left|\eta_1' \left\{\sum_{i=1}^{n+1} W_{ni}^{ad}(\beta_0) - \rho_1 \sum_{i=1}^{n+1} \frac{W_{ni}^{ad}(\beta_0)\eta_1' W_{ni}^{ad}(\beta_0)}{1+\rho_1\eta_1' W_{ni}^{ad}(\beta_0)}\right\}\right| \\
&\geq \frac{\rho_1}{n(1+\rho_1 W_n^{ad}(\beta_0))}\sum_{i=1}^{n}\left(\eta_1' W_{ni}^{ad}(\beta_0)\right)^2 - \frac{1}{n}\left|\sum_{i=1}^{n} W_{ni}^{ad}(\beta_0)\right|(1-a_n/n) \\
&= \frac{\rho_1}{n(1+\rho_1 W_n^{ad}(\beta_0))}\sum_{i=1}^{n}\left(\eta_1' W_{ni}^{ad}(\beta_0)\right)^2 - \frac{1}{n}\left|\sum_{i=1}^{n} W_{ni}^{ad}(\beta_0)\right| + O_p(n^{-2/3}a_n),
\end{aligned}
$$

where $W_n^{ad}(\beta_0) = \max_{1\leq i\leq n}\left|W_{ni}^{ad}(\beta_0)\right|$. Following the same steps in the proof of Theorem 2.2 and Chen et al. (2008), one has $\theta^{ad} = O_p(n^{-1/2})$ as $a_n = o_p(n)$.

Moreover, one has

$$
\theta^{ad} = \left(\frac{1}{n}\sum_{i=1}^{n} W_{ni}^{ad}(\beta_0)(W_{ni}^{ad}(\beta_0))'\right)^{-1}\sum_{i=1}^{n} W_{ni}^{ad}(\beta_0) + o_p(n^{-1/2}).
$$

The adjusted EL ratio is as follows.

$$
\begin{aligned}
l_n^{ad}(\beta_0) &= -2\log R^{ad}(\beta_0) \\
&= 2\sum_{i=1}^{n+1}\log(1+(\theta^{ad})' W_{ni}^{ad}(\beta_0)) \\
&= 2\sum_{i=1}^{n+1}\{(\theta^{ad})' W_{ni}^{ad}(\beta_0) - (\theta^{ad})' W_{ni}^{ad}(\beta_0)(W_{ni}^{ad}(\beta_0))'\theta^{ad}/2\} + o_p(1).
\end{aligned}
$$

Substituting the expansion of θ^{ad}, one can get that

$$
l_n^{ad}(\beta_0) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}^{ad}(\beta_0)\right)'\left(\frac{1}{n}\sum_{i=1}^{n} W_{ni}^{ad}(\beta_0)(W_{ni}^{ad}(\beta_0))'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}^{ad}(\beta_0)\right) + o_p(1).
$$

Thus, $l_n^{ad}(\beta_0)$ converges to χ_p^2 distribution. We proved Theorem 2.4. □

Appendix B

Proof of theorems for Chapter 3

In this Appendix, we prove Theorems 3.1, 3.2 and 3.3 presented in Section 3.2.

Proof of Theorem 3.1.

We assume that the same conditions as in Huang and Wang (2010).

(D.1) $P(Y(\tau) = 1) > 0$.

(D.2) For some constant $C > 0$, $P(\|Z\| < C) = 1$.

(D.3) A, Σ_1 and Σ_2 are assumed to be positive definite.

(D.4) (i) The selection probability $\pi(w)$ has r continuous and bounded partial derivatives with respect to the continuous components of W almost surely;

(ii) $\inf_w(\pi(w)) > 0$.

(D.5) (i) The probability function $f(w)$ of W has r continuous and bounded partial derivatives with respect to the continuous components of W almost surely;

(ii) $0 < \inf_w f(w) \leq \sup_w f(w) < \infty$.

(D.6) The kernel function $K(\cdot)$ is a bounded kernel function with bounded support, and $K(\cdot)$ is a kernel of order $r(> d)$, where d is the number of elements in $W^{(1)}$.

(D.7) $nh_n^{2d} \to \infty$ and $nh_n^{2r} \to \infty$ as $n \to \infty$.

(D.8) The conditional expectation $E[M_* |W]$ has r continuous and bounded partial derivatives with respect to the continuous components of W almost surely.

(D.9) $P(X \geq t)$ is continuous for $t \in [0, \tau]$.

Here, we have the following result.

$$\sum_{i=1}^{n} \frac{V_i}{\hat{\pi}_i} d\hat{M}_i(\beta_0, t) = \sum_{i=1}^{n} \frac{V_i}{\hat{\pi}_i} \left[dN_i(t) - Y_i(t) d\Lambda\{\beta_0' Z_i + \hat{H}(\beta_0, t)\} \right] = 0.$$

Mimicking Qi et al. (2005), to prove Theorem 3.1, in addition to the regularity conditions

(D.1)-(D.9), the following regularity conditions are needed:

(c.1) $\sup_{t\in[0,\tau]} \left\|\bar{M}_{Edn}(t)\right\| \to 0$ in probability as $n \to \infty$, where

$$\bar{M}_{Edn}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\pi_i}\right) \int_0^t \Big[E[d\hat{M}_i(u)\,|W_i] - E[dM_i(u)\,|W_i]\Big].$$

(c.2) $\sup_{t\in[0,\tau]} \left\|\bar{M}_{Eqn}(t)\right\| \to 0$ in probability as $n \to \infty$, where

$$\bar{M}_{Eqn}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\hat{\pi}_i}\right) \int_0^t \Big[E[d\hat{M}_i(u)\,|W_i] - E[dM_i(u)\,|W_i]\Big].$$

(c.3) $\bar{M}_{En}(t)$ converges to a mean-zero Gaussian process with continuous sample paths, where

$$\bar{M}_{En}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i(\hat{\pi}_i - \pi_i)}{\pi_i^2} E[M_i(t)\,|W_i].$$

From Lemma A.3 in Huang and Wang (2010), we can easily obtain that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\pi_i}\right) E[M_{*,i}\,|W_i] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i}{\hat{\pi}_i} M_{*,i} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i}{\pi_i} M_{*,i} + o_p(1).$$

Combining with Lemma A.1 in Huang and Wang (2010), we have that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\pi_i}\right) E[M_{*,i}\,|W_i] = o_p(1). \tag{B.1}$$

Furthermore, by Taylor expansion of $1/\hat{\pi}_i$ about $1/\pi_i$, one has that

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\hat{\pi}_i}\right) E[M_{*,i}\,|W_i] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\pi_i}\right) E[M_{*,i}\,|W_i] - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i(\hat{\pi}_i - \pi_i)}{\pi_i^2} E[M_{*,i}\,|W_i] + o_p(1). \end{aligned} \tag{B.2}$$

Then by the condition (c.3), equation (B.2) can be written as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\hat{\pi}_i}\right) E[M_{*,i}\,|W_i] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{V_i}{\pi_i}\right) E[M_{*,i}\,|W_i] + o_p(1).$$

According to the condition (c.2) and equation (B.1), we can obtain that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{V_i}{\hat{\pi}_i}\right)E[\hat{M}_{*,i}\,|W_i] \\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{V_i}{\hat{\pi}_i}\right)E[M_{*,i}\,|W_i]+o_p(1) \\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{V_i}{\pi_i}\right)E[M_{*,i}\,|W_i]+o_p(1) \\
&=o_p(1). \qquad\qquad (B.3)
\end{aligned}$$

It is clear that one has that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_{ni}(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left\{\frac{V_i}{\hat{\pi}_i}(Z_i-\bar{Z}(\beta_0,t))d\hat{M}_i(\beta_0,t)+\left(1-\frac{V_i}{\hat{\pi}_i}\right)d\hat{M}_{**,i}(\beta_0,t,w_i^{(1)},w_i^{(2)})\right\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\frac{V_i}{\hat{\pi}_i}(Z_i-\bar{Z}(\beta_0,t))d\hat{M}_i(\beta_0,t)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{V_i}{\hat{\pi}_i}\right)E[\hat{M}_{*,i}\,|W_i] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\frac{V_i}{\hat{\pi}_i}Z_i d\hat{M}_i(\beta_0,t)-\int_0^{\tau}\bar{Z}(\beta_0,t)\sum_{i=1}^{n}\frac{V_i}{\hat{\pi}_i}d\hat{M}_i(\beta_0,t) \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{V_i}{\hat{\pi}_i}\right)E[\hat{M}_{*,i}\,|W_i] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\beta_0)+o_p(1).
\end{aligned}$$

By proofs in the Appendix of Huang and Wang (2010), one has that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\beta_0)\rightarrow N(0,\Sigma_2(\beta_0)),$$

in distribution, as $n\rightarrow\infty$.

Moreover, we can have that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_{ni}(\beta_0)\rightarrow N(0,\Sigma_2(\beta_0)),$$

in distribution, as $n\rightarrow\infty$.

Furthermore, one need to prove that

$$U_{ni}(\beta_0) = U_i(\beta_0) + o_p(1). \tag{B.4}$$

Write

$$\begin{aligned}
U_{ni}(\beta_0) &= U_i(\beta_0) + \frac{V_i}{\hat{\pi}_i}\int_0^\tau Z_i\left(d\hat{M}_i(\beta_0,t) - dM_i(\beta_0,t)\right)\\
&+ \frac{V_i}{\hat{\pi}_i}\int_0^\tau \left(\mu_Z dM_i(\beta_0,t) - \bar{Z}(\beta_0,t)d\hat{M}_i(\beta_0,t)\right)\\
&+ \left(1-\frac{V_i}{\hat{\pi}_i}\right)\int_0^\tau \left(d\hat{M}_{**,i}(\beta_0,t,w_i^{(1)},w_i^{(2)}) - dM_{**,i}(\beta_0,t,w_i^{(1)},w_i^{(2)})\right)\\
&=: U_i(\beta_0) + r_{i1} + r_{i2} + r_{i3},
\end{aligned}$$

where

$$\begin{aligned}
r_{i1} &= \frac{V_i}{\hat{\pi}_i}Z_i\left(\int_0^\tau Y_i(t)d\Lambda\{\beta_0'Z_i + \hat{H}(\beta_0,t)\} - \int_0^\tau Y_i(t)d\Lambda\{\beta_0'Z_i + H_0(t)\}\right)\\
r_{i2} &= \frac{V_i}{\hat{\pi}_i}\int_0^\tau \left(\mu_Z dM_i(\beta_0,t) - \bar{Z}(\beta_0,t)d\hat{M}_i(\beta_0,t)\right)\\
\mathrm{r}_{i3} &= \left(1-\frac{V_i}{\hat{\pi}_i}\right)\int_0^\tau \left(d\hat{M}_{**,i}(\beta_0,t,w_i^{(1)},w_i^{(2)}) - dM_{**,i}(\beta_0,t,w_i^{(1)},w_i^{(2)})\right).
\end{aligned}$$

It suffices to show that

$$\left\|\int_0^\tau Y_i(t)d\Lambda\{\beta_0'Z_i + \hat{H}(\beta_0,t)\} - \int_0^\tau Y_i(t)d\Lambda\{\beta_0'Z_i + H_0(t)\}\right\| = o_p(1). \tag{B.5}$$

Some arguments similar to the proofs of Theorems 2.1 and 2.2 in the Appendix of Huang and Wang (2010) can be used here. Recall $B(\beta_0, t, u)$ in equation (3.3). Let $a > 0$, define

$$\lambda^*\{H_0(t)\} = B(\beta_0, t, a),$$

$$B_1(\beta_0, t) = \int_a^t E[Y(t)\dot{\lambda}\{\beta_0'Z + H_0(u)\}]dH_0(u),$$

$$B_2(\beta_0, t) = E[Y(t)\lambda\{\beta_0' Z + H_0(u)\}].$$

Like Huang and Wang (2010), it is easy to prove that

$$B(\beta_0, t, u) = \lambda^*\{H_0(t)\}/\lambda^*\{H_0(u)\},$$

and

$$d\lambda^*\{H_0(t)\} = [\lambda^*\{H_0(t)\}/B_2(\beta_0, t)]dB_1(\beta_0, t).$$

Since

$$\begin{aligned}\sum_{i=1}^n dM_i(\beta_0, t) &= \sum_{i=1}^n dN_i(t) - \sum_{i=1}^n Y_i(t) d\Lambda\{\beta_0' Z_i + H_0(t)\} \\ &= \sum_{i=1}^n Y_i(t) d\Lambda\{\beta_0' Z_i + \hat{H}(\beta_0, t)\} - \sum_{i=1}^n Y_i(t) d\Lambda\{\beta'_0 Z_i + H_0(t)\} \\ &= (1 + o_p(1)) \sum_{i=1}^n Y_i(t) d\left(\lambda\{\beta_0' Z_i + H_0(t)\}\left[\hat{H}(\beta_0, t) - H_0(t)\right]\right),\end{aligned}$$

similar to the equation (A.9) in Wang and Wang (2015), one has that

$$\hat{H}(\beta_0, t) - H_0(t) = \frac{1}{n}\sum_{j=1}^n \int_0^t \frac{B(\beta_0, u, t)}{B_2(\beta_0, u)} dM_j(\beta_0, u) + o_p(n^{-1/2}). \quad \text{(B.6)}$$

Then by equation (B.6), one has that

$$\begin{aligned}&\int_0^\tau Y_i(t) d\Lambda\{\beta_0' Z_i + \hat{H}(\beta_0, t)\} - \int_0^\tau Y_i(t) d\Lambda\{\beta_0' Z_i + H_0(t)\} \\ =& \int_0^\tau Y_i(t) d[\Lambda\{\beta_0' Z_i + \hat{H}(\beta_0, t)\} - \Lambda\{\beta_0' Z_i + H_0(t)\}] \\ =& \int_0^\tau Y_i(t) d\left[\Lambda\{\beta_0' Z_i + H_0(t)\}(\hat{H}(\beta_0, t) - H_0(t))\right] \\ =& \int_0^\tau Y_i(t) d\left[\Lambda\{\beta_0' Z_i + H_0(t)\}\left(\frac{1}{n}\sum_{j=1}^n \int_0^t \frac{B(\beta_0, u, t)}{B_2(\beta_0, u)} dM_j(\beta_0, u) + o_p(n^{-1/2})\right)\right] \\ =& \int_0^\tau \left\{Y_i(t)\left[\Lambda\{\beta_0' Z_i + H_0(t)\}\left(\frac{1}{n}\sum_{j=1}^n \frac{B(\beta_0, t, t)}{B_2(\beta_0, t)} dM_j(\beta_0, t)\right)\right.\right.\end{aligned}$$

$$+\left(\frac{1}{n}\sum_{j=1}^{n}\int_0^t \frac{B(\beta_0,u,t)}{B_2(\beta_0,u)}dM_j(\beta_0,u)+o_p(n^{-1/2})\right)d\Lambda\{\beta_0'Z_i+H_0(t)\}\Big]\Big\}$$
$$=o_p(1).$$

Thus, (B.5) is valid.

Then since

$$\begin{aligned}
r_{i2} &= \frac{V_i}{\hat{\pi}_i}\int_0^\tau \Big(\mu_Z dM_i(\beta_0,t)-\bar{Z}(\beta_0,t)d\hat{M}_i(\beta_0,t)\Big)\\
&= \frac{V_i}{\hat{\pi}_i}\Big[\int_0^\tau (\mu_Z-\bar{Z}(\beta_0,t))dN_i\\
&- \Big(\int_0^\tau \mu_Z Y_i(t)d\Lambda\{\beta_0'Z_i+H_0(t)\}-\int_0^\tau \bar{Z}(\beta_0,t)Y_i(t)d\Lambda\{\beta_0'Z_i+\hat{H}(\beta_0,t)\}\Big)\Big]\\
&= \frac{V_i}{\hat{\pi}_i}\Big[\int_0^\tau (\mu_Z-\bar{Z}(\beta_0,t))dN_i\\
&- \int_0^\tau \mu_Z Y_i(t)d\Lambda\{\beta_0'Z_i+H_0(t)\}+\int_0^\tau \bar{Z}(\beta_0,t)Y_i(t)d\Lambda\{\beta_0'Z_i+H_0(t)\}\\
&+ \int_0^\tau \bar{Z}(\beta_0,t)Y_i(t)d\Lambda\{\beta_0'Z_i+\hat{H}(\beta_0,t)\}-\int_0^\tau \bar{Z}(\beta_0,t)Y_i(t)d\Lambda\{\beta_0'Z_i+H_0(t)\}\Big]\\
&= \frac{V_i}{\hat{\pi}_i}\Big[\int_0^\tau (\mu_Z-\bar{Z}(\beta_0,t))dN_i\\
&- \int_0^\tau \big(\mu_Z-\bar{Z}(\beta_0,t)\big)Y_i(t)d\Lambda\{\beta_0'Z_i+H_0(t)\}\\
&+ \int_0^\tau \bar{Z}(\beta_0,t)\Big[Y_i(t)d\Lambda\{\beta_0'Z_i+\hat{H}(\beta_0,t)-Y_i(t)d\Lambda\{\beta_0'Z_i+H_0(t)\}\Big]\Big],
\end{aligned}$$

this together with the uniform consistency of $\bar{Z}(\beta_0,t)$ to μ_Z, that is,

$$\sup_{0\le t\le\tau}\left|\bar{Z}(\beta_0,t)-\mu_Z\right|\to 0, \tag{B.7}$$

in probability, as $n\to\infty$, we can obtain that $\|r_{i2}\|=o_p(1)$. Moreover, from condition (D.2), equations (B.5) and (B.7), $\|r_{i1}\|=o_p(1)$ can be proven.

Furthermore, we have that

$$\mathrm{r}_{i3}=\left(1-\frac{V_i}{\hat{\pi}_i}\right)\int_0^\tau \Big(d\hat{M}_{**,i}(\beta_0,t,w_i^{(1)},w_i^{(2)})-dM_{**,i}(\beta_0,t,w_i^{(1)},w_i^{(2)})\Big)$$

$$= \left(1 - \frac{V_i}{\hat{\pi}_i}\right)\left(\int_0^\tau E[(Z_i - \bar{Z}(\beta_0, t))d\hat{M}_i(\beta_0, t)\,|W_i] - \int_0^\tau E[(Z_i - \mu_Z(\beta_0, t))dM_i(\beta_0, t)\,|W_i]\right)$$
$$= \left(1 - \frac{V_i}{\hat{\pi}_i}\right)\left(E\left[\int_0^\tau (Z_i - \bar{Z}(\beta_0, t))d\hat{M}_i(\beta_0, t)\,|W_i\right] - E\left[\int_0^\tau (Z_i - \mu_Z(\beta_0, t))dM_i(\beta_0, t)\,|W_i\right]\right).$$

Define

$$\begin{aligned} R &= \int_0^\tau (Z_i - \bar{Z}(\beta_0, t))d\hat{M}_i(\beta_0, t) - (Z_i - \mu_Z(\beta_0, t))dM_i(\beta_0, t) \\ &= \int_0^\tau (\mu_Z(\beta_0, t) - \bar{Z}(\beta_0, t))dN_i \\ &- \left(\int_0^\tau \mu_Z Y_i(t) d\Lambda\{\beta_0' Z_i + H_0(t)\} - \int_0^\tau \bar{Z}(\beta_0, t) Y_i(t) d\Lambda\{\beta_0' Z_i + \hat{H}(\beta_0, t)\}\right). \end{aligned}$$

Similar arguments lead to $R = o_p(1)$. In addition, we can obtain that

$$\int_0^\tau (Z_i - \bar{Z}(\beta_0, t))d\hat{M}_i(\beta_0, t) \to \int_0^\tau (Z_i - \mu_Z(\beta_0, t))dM_i(\beta_0, t)$$

in probability. Moreover, $\int_0^\tau (Z_i - \mu_Z(\beta_0, t))dM_i(\beta_0, t)$ and $\int_0^\tau (Z_i - \bar{Z}(\beta_0, t))d\hat{M}_i(\beta_0, t)$ are bounded, one has that

$$E\left(\int_0^\tau (Z_i - \bar{Z}(\beta_0, t))d\hat{M}_i(\beta_0, t)\,|W_i\right) \to E\left(\int_0^\tau (Z_i - \mu_Z(\beta_0, t))dM_i(\beta_0, t)\,|W_i\right),$$

in probability. Therefore, $\|r_{i3}\| = o_p(1)$. The proof of equation (B.4) is completed.
Then we conduct the following decompositions,

$$\begin{aligned} & c'\left(\frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0)(U_{ni}(\beta_0))' - \frac{1}{n}\sum_{i=1}^n U_i(\beta_0)(U_i(\beta_0))'\right)c \\ & = \frac{1}{n}\sum_{i=1}^n \left[c'\left(U_{ni}(\beta_0) - U_i(\beta_0)\right)\right]^2 + \frac{2}{n}\sum_{i=1}^n (c' U_i(\beta_0)) \\ & \times \left[c'(U_{ni}(\beta_0) - U_i(\beta_0)\right], \end{aligned}$$

where c is some constant.

From equation (B.4), it can be easily shown that the two parts of right-hand side of the above equation are both $o_p(1)$. Hence

$$\frac{1}{n}\sum_{i=1}^{n} U_{ni}(\beta_0)(U_{ni}(\beta_0))' = \frac{1}{n}\sum_{i=1}^{n} U_i(\beta_0)(U_i(\beta_0))' + o_p(1).$$

By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} U_i(\beta_0)(U_i(\beta_0))' \to E\left[U_i(\beta_0)(U_i(\beta_0))'\right] = \Sigma_2(\beta_0),$$

in probability, as $n \to \infty$.

Thus,

$$\frac{1}{n}\sum_{i=1}^{n} U_{ni}(\beta_0)(U_{ni}(\beta_0))' \to \Sigma_2(\beta_0),$$

in probability, as $n \to \infty$. Finally, we proved Theorem 3.1. □

Proof of Theorem 3.2.

We need to show that

$$\max_{1\le i\le n} \|U_{ni}(\beta_0)\| = o_p(n^{1/2}), \tag{B.8}$$

and

$$\lambda = O_p(n^{-1/2}), \tag{B.9}$$

where the Lagrange multiplier λ is the solution of equation (3.15).

Since $U_i(\beta_0)$ are i.i.d. random variables, and $E\left[U_i(\beta_0)(U_i(\beta_0))'\right] = \Sigma_2(\beta_0) < \infty$, $U_i(\beta_0)$ has finite second moment. Moreover, for $i = 1, 2, ..., n$, $U_{ni}(\beta_0) = U_i(\beta_0) + o_p(1)$ is proven. It is sufficient to show that $\max_{1\le i\le n} \|U_i(\beta_0)\| = o_p(n^{1/2})$. It follows by Lemma 11.2 of Owen (2001), we have equation (B.8) is true.

Then, following the similar steps in the proofs of the Theorem 3.2 of Owen (2001) and Lu and Liang (2006), we can prove equation (B.9) is valid as well.

Let $f(\lambda) = \left(\sum_{i=1}^{n} U_{ni}(\beta_0)/(1+\lambda' U_{ni}(\beta_0))\right)\Big/ n$, $\lambda = \rho\theta$, where $\rho > 0$ and $\|\theta\| = 1$, we obtain that

$$\begin{aligned}0 &= \|f(\lambda)\| \\ &= \|f(\rho\theta)\| \\ &\geq |\theta' f(\rho\theta)| \\ &\geq \frac{1}{n}\left|\theta' \sum_{i=1}^{n} \frac{U_{ni}(\beta_0)}{1+\lambda' U_{ni}(\beta_0)}\right| \\ &= \frac{1}{n}\left|\theta' \left\{\sum_{i=1}^{n} U_{ni}(\beta_0) - \rho \sum_{i=1}^{n} \frac{U_{ni}(\beta_0)\theta' U_{ni}(\beta_0)}{1+\rho\theta' U_{ni}(\beta_0)}\right\}\right| \\ &\geq \frac{\rho}{n}\theta' \sum_{i=1}^{n} \frac{U_{ni}(\beta_0)(U_{ni}(\beta_0))'}{1+\rho\theta' U_{ni}(\beta_0)}\theta - \frac{1}{n}\left|\sum_{i=1}^{n} U_{ni}(\beta_0)\right| \\ &\geq \frac{\rho\theta' S_n(\beta_0)\theta}{1+\rho U_n(\beta_0)} - \frac{1}{n}\left|\sum_{i=1}^{n} U_{ni}(\beta_0)\right|,\end{aligned}$$

where $S_n(\beta_0) = \left(\sum_{i=1}^{n} U_{ni}(\beta_0)(U_{ni}(\beta_0))'\right) \Big/ n$, and $U_n(\beta_0) = \max_{1\leq i\leq n} \|U_{ni}(\beta_0)\| = o_p(n^{1/2})$.

By the law of large numbers, one has that

$$S_n(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} U_i(\beta_0)(U_i(\beta_0))' + o_p(1).$$

Therefore, as $n \to \infty$,

$$\lim_{n\to\infty} S_n(\beta_0) = E\left[U_i(\beta_0)(U_i(\beta_0))'\right] = \Sigma_2(\beta_0).$$

Combining with Theorem 3.1, we have that

$$\frac{1}{n}\left|\sum_{i=1}^{n} U_{ni}(\beta_0)\right| = O_p(n^{-1/2}).$$

We complete that proof of equation (B.9).

Using the results (B.8) and (B.9), we can obtain that

$$\lambda = \left(\frac{1}{n} \sum_{i=1}^{n} U_{ni}(\beta_0)(U_{ni}(\beta_0))' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} U_{ni}(\beta_0) \right) + o_p(n^{-1/2}).$$

It follows from the Taylor expansion, Slutsky lemma and Theorem 3.1,

$$\begin{aligned} l_n(\beta_0) &= 2 \sum_{i=1}^{n} \lambda' U_{ni}(\beta_0) - \sum_{i=1}^{n} \lambda' U_{ni}(\beta_0)(U_{ni}(\beta_0))' \lambda + o_p(1) \\ &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{ni}(\beta_0) \right)' \left(\frac{1}{n} \sum_{i=1}^{n} U_{ni}(\beta_0)(U_{ni}(\beta_0))' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{ni}(\beta_0) \right) + o_p(1). \end{aligned}$$

We complete the proof of Theorem 3.2. □

Proof of Theorem 3.3.

Following the arguments as in Yu et al. (2011), we prove the theorem. Corresponding to $(\beta_{10}', \beta_{20}')'$, we denote $Z = (Z_1', Z_2')'$. Define

$$A^*(\beta_0) = E\left[\int_0^\tau \{Z - \mu_Z(\beta_0, t)\} Z_2{}' Y(t) \dot{\lambda}\{\beta_0' Z + H_0(t)\} dH_0(t) \right].$$

Since A is positive definite, A^* is of rank $p - q$. Let

$$\hat{\beta}_2 = \arg \inf_{\beta_2} l_n(\beta_{10}, \beta_2),$$

similar to Qin and Lawless (1994), we can show that

$$\sqrt{n}(\hat{\beta}_2 - \beta_{20}) = -(A^{*\prime} \Sigma_2^{-1} A^*)^{-1} A^{*\prime} \Sigma_2^{-1} n^{-1/2} \sum_{i=1}^{n} U_{ni}(\beta_0) + o_p(1),$$

$$\sqrt{n}\lambda_2 = \left(I - \Sigma_2^{-1} A^* (A^{*\prime} \Sigma_2^{-1} A^*)^{-1} A^{*\prime} \right) \Sigma_2^{-1} n^{-1/2} \sum_{i=1}^{n} U_{ni}(\beta_0) + o_p(1),$$

where λ_2 is the corresponding Lagrange multiplier.

Hence, by Taylor's expansion, we have that

$$\begin{aligned} l_n^*(\beta_{10}) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0)\right)' \left(\Sigma_2^{-1} - \Sigma_2^{-1}A^*\left(A^{*'}\Sigma_2^{-1}A^*\right)^{-1}A^{*'}\Sigma_2^{-1}\right) \\ &\times \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0)\right) + o_p(1) \\ &= \left(\Sigma_2^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0)\right)' \Gamma \left(\Sigma_2^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0)\right) + o_p(1), \end{aligned}$$

where

$$\Gamma = I - \Sigma_2^{-1/2}A^*(A^{*'}\Sigma_2^{-1}A^*)^{-1}A^{*'}\Sigma_2^{-1/2}$$

is a symmetric and idempotent matrix with trace q. By Theorem 3.1,

$$\Sigma_2^{-1/2}(\beta_0)\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0) \to N(0, I),$$

in distribution, as $n \to \infty$. The proof of Theorem 3.3 is completed. □

Appendix C

Proof of theorems for Chapter 4

To derive the asymptotic properties of $l(\beta_0)$ and $l^*(\beta_{10})$, we assume some regularity conditions hold.

(D.1) X is bounded, that is, $P(\|X\| \leq M) = 1$ for some $0 < M < \infty$.

(D.2) The conditional distribution $F_{e_1(\beta)|X_1}(t)$ of $e_1(\beta) = Y_1 - \beta^T X_1$ given X_1 is twice continuously differentiable in t for all X.

(D.3) For any X, the conditional density function $F'_{e_1(\beta)|X_1}(t) = f_{e_1(\beta)|X_1}(t) > 0$ for t in a neighborhood of 0.

Firstly, we re-express the smoothed rank estimating function $S_n(\beta)$ in (4.2) as a U-statistic with a symmetric kernel function.

$$\begin{aligned}
S_n(\beta) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_i (X_i - X_j)\Phi\left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}}\right] \\
&= \sum_{1\leq i<j\leq n} \Delta_i (X_i - X_j)\Phi\left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}}\right] + \sum_{1\leq j<i\leq n} \Delta_i (X_i - X_j)\Phi\left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}}\right] \\
&= \sum_{1\leq i<j\leq n} \Delta_i (X_i - X_j)\Phi\left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}}\right] + \sum_{1\leq i<j\leq n} \Delta_j (X_j - X_i)\Phi\left[\frac{e_i(\beta) - e_j(\beta)}{r_{ji}}\right],
\end{aligned}$$

because of $r_{ij}^2 = (X_i - X_j)^T (X_i - X_j)/n$, one has that $r_{ij} = r_{ji}$,

$$\begin{aligned}
S_n(\beta) &= \sum_{1\leq i<j\leq n} (X_i - X_j)\left\{\Delta_i \Phi\left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}}\right] - \Delta_j \Phi\left[\frac{e_i(\beta) - e_j(\beta)}{r_{ij}}\right]\right\} \\
&= \begin{pmatrix} n \\ 2 \end{pmatrix}\left[\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{1\leq i<j\leq n} K(Z_i, Z_j; \beta)\right] \\
&\equiv \begin{pmatrix} n \\ 2 \end{pmatrix} S_n^*(\beta),
\end{aligned}$$

where $S_n^*(\beta)$ is a U-statistic of degree 2

$$S_n^*(\beta) = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{1 \le i < j \le n} K(Z_i, Z_j; \beta) \equiv U_n(\beta),$$

with the kernel function

$$K(Z_i, Z_j; \beta) = (X_i - X_j) \left\{ \Delta_i \Phi \left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}} \right] - \Delta_j \Phi \left[\frac{e_i(\beta) - e_j(\beta)}{r_{ij}} \right] \right\}.$$

Similarly, we can also derive $\tilde{S}_n(\beta)$ in (4.1) as a U-statistic with a symmetric kernel function, that is,

$$\begin{aligned}
\tilde{S}_n(\beta) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i (X_i - X_j) I[e_j(\beta) \ge e_i(\beta)], \\
&= \sum_{1 \le i < j \le n} \Delta_i (X_i - X_j) I[e_j(\beta) \ge e_i(\beta)] + \sum_{1 \le j < i \le n} \Delta_i (X_i - X_j) I[e_j(\beta) \ge e_i(\beta)] \\
&= \sum_{1 \le i < j \le n} \Delta_i (X_i - X_j) I[e_j(\beta) \ge e_i(\beta)] + \sum_{1 \le i < j \le n} \Delta_j (X_j - X_i) I[e_i(\beta) \ge e_j(\beta)] \\
&= \sum_{1 \le i < j \le n} (X_i - X_j) \left\{ \Delta_i I[e_j(\beta) \ge e_i(\beta)] - \Delta_j I[e_i(\beta) \ge e_j(\beta)] \right\} \\
&= \begin{pmatrix} n \\ 2 \end{pmatrix} \left[\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{1 \le i < j \le n} H(Z_i, Z_j; \beta) \right] \\
&\equiv \begin{pmatrix} n \\ 2 \end{pmatrix} W_n(\beta),
\end{aligned}$$

with the kernel function

$$H(Z_i, Z_j; \beta) = (X_i - X_j) \left\{ \Delta_i I[e_j(\beta) \ge e_i(\beta)] - \Delta_j I[e_i(\beta) \ge e_j(\beta)] \right\}.$$

Fygenson and Ritov (1994) pointed out that when evaluated at $\beta = \beta_0$, $W_n(\beta_0)$ is asymptotically normal and has expectation zero. Furthermore, by (A.7) in the Appendix of Johnson and Strawderman (2009), we can have the asymptotically equivalence of $U_n(\beta_0)$ to $W_n(\beta_0)$, $\sqrt{n}\,\|U_n(\beta_0) - W_n(\beta_0)\| \xrightarrow{p} 0$, that is,

$$U_n(\beta_0) = W_n(\beta_0) + o_p(n^{-1/2}). \tag{C.1}$$

Then $EU_n(\beta_0) = EW_n(\beta_0) + E\left[o_p(n^{-1/2})\right]$. Hence, we can have $EU_n(\beta_0) \to 0$ as $n \to \infty$. Before proving Theorem 4.1, we first list some notations. Define

$$\begin{cases} \hat{V}_i(\beta) = nW_n(\beta) - (n-1)W_{n-1}^{(-i)}(\beta),\ i = 1, ..., n, \\ W_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i(\beta), \\ G(\beta) = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i(\beta)\hat{V}_i^T(\beta), \\ G^*(\beta) = \frac{1}{n}\sum_{i=1}^{n}\hat{Q}_i(\beta)\hat{Q}_i^T(\beta), \\ \phi(z,\beta) = (\phi_1(z,\beta), ..., \phi_p(z,\beta))^T = EH(z, Z_1; \beta), \\ \psi(x,y,\beta) = H(x,y;\beta) - \phi(x,\beta) - \phi(y,\beta), \\ g(z,\beta) = (g_1(z,\beta), ..., g_p(z,\beta))^T = 2\phi(z,\beta), \\ \sigma_l^2(\beta) = Var(\phi_l(Z_1,\beta)), l = 1, ..., p, \\ \sigma_{st}^2(\beta) = Cov(\phi_s(Z_1,\beta), \phi_t(Z_1,\beta)), s,t = 1, ..., p, \\ \Sigma_{p\times p}^{(\beta)} : \text{the asymptotic variance} - \text{covariance matrix of} \sqrt{n}W_n(\beta), \\ \text{with elements}\, 4\sigma_{st}(\beta),\ s,\ t = 1, ..., p. \end{cases}$$

Under conditions (D.1)-(D.3), following Jing et al. (2009) and Li et al. (2016), we will prove Lemmas A.1 to A.5.

Lemma A.1. Under conditions (D.1)-(D.3), as $n \to \infty$, we have

$$\sqrt{n}U_n(\beta_0) \xrightarrow{d} N(0, \Sigma_{p\times p}^{(\beta_0)}).$$

Proof. From Li et al. (2016), we can conclude that $\sqrt{n}W_n(\beta_0)$ tends to have a normal distribution with mean 0 and covariance $\Sigma_{p\times p}^{(\beta_0)}$. Then as $n \to \infty$, by (C.1), we can derive that

$$
\begin{aligned}
E[\sqrt{n}U_n(\beta_0)] &= E[\sqrt{n}(W_n(\beta_0) + o_p(n^{-1/2}))] \\
&= E[\sqrt{n}W_n(\beta_0) + o_p(1)] = E[\sqrt{n}W_n(\beta_0)] + o_p(1) \to 0,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{cov}(\sqrt{n}U_n(\beta_0)) = \mathrm{cov}(\sqrt{n}(W_n(\beta_0) + o_p(n^{-1/2}))) = \mathrm{cov}(\sqrt{n}W_n(\beta_0) + o_p(1)) \\
&= \mathrm{cov}(\sqrt{n}W_n(\beta_0)) + 2\mathrm{cov}(\sqrt{n}W_n(\beta_0), o_p(1)) + \mathrm{cov}(o_p(1)) \\
&= \mathrm{cov}(\sqrt{n}W_n(\beta_0)) + 2\left[E(\sqrt{n}W_n(\beta_0) \times o_p(1)) - E(\sqrt{n}W_n(\beta_0)) \times E(o_p(1))\right] + o_p(1) \\
&= \Sigma_{p\times p}^{(\beta_0)} + o_p(1) \\
&\to \Sigma_{p\times p}^{(\beta_0)}.
\end{aligned}
$$

Thus, Lemma A.1 holds. □

Lemma A.2. Under conditions (D.1)-(D.3), with probability tending to one as $n \to \infty$, the zero vector is contained in the interior of the convex hull of $\left\{\hat{Q}_1(\beta_0), ..., \hat{Q}_n(\beta_0)\right\}$. *Proof.* To get the representation of $\hat{Q}_i(\beta_0)$, by the Hoeffding decomposition, from Li et al. (2016), we have that

$$
W_{n,l}(\beta_0) = \frac{2}{n}\sum_{i=1}^{n}\phi_l(Z_i, \beta_0) + \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i<j}^{n}\psi_l(Z_i, Z_j, \beta_0). \tag{C.2}
$$

Combining (C.1) with (C.2), we can derive that

$$
U_{n,l}(\beta_0) = \frac{2}{n}\sum_{i=1}^{n}\phi_l(Z_i, \beta_0) + \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i<j}^{n}\psi_l(Z_i, Z_j, \beta_0) + o_p(n^{-1/2}).
$$

By some calculations, one has that

$$\begin{aligned}\hat{Q}_{i,l}(\beta_0) &= 2\phi_l(Z_i,\beta_0) + \frac{2}{n-1}\sum_{j=1,j\neq i}^{n}\psi_l(Z_i,Z_j,\beta_0) \\ &- \begin{pmatrix} n-1 \\ 2 \end{pmatrix}^{-1} \sum_{i_1<i_2,i_1\neq i,i_2\neq i}^{n} \psi_l(Z_{i1},Z_{i2},\beta_0) + o_p(n^{-1/2}) \\ &:= g_l(Z_i,\beta_0) + r_{ni,l}(\beta_0) + o_p(n^{-1/2}).\end{aligned}$$

Note Li et al. (2016) proved that

$$\begin{aligned}&Er_{ni,l}^2(\beta_0) \leq Cn^{-1}E\psi_l^2(Z_i,Z_j,\beta_0) + Cn^{-2}E\psi_l^2(Z_i,Z_j,\beta_0) \\ &\to 0, \end{aligned} \tag{C.3}$$

where C is some generic constant. From (C.3), it is clear that

$$r_{ni,l}(\beta_0) = O_p(n^{-1/2}) \to 0.$$

Thus,

$$\hat{Q}_{i,l}(\beta_0) \xrightarrow{p} g_l(Z_i,\beta_0), i = 1,....n, l = 1,...,p. \tag{C.4}$$

Similar to Li et al. (2016), we have the following conclusions,

$$E \max_{1\leq i\leq n, 1\leq l\leq p}\left|\frac{2}{n-1}\sum_{j=1,j\neq i}^{n}\psi_l(Z_i,Z_j,\beta_0)\right|^4 = O(n^{-1}),$$

and

$$E \max_{1\leq i\leq n, 1\leq l\leq p}\left|\begin{pmatrix} n-1 \\ 2 \end{pmatrix}^{-1}\sum_{i_1<i_2,i_1\neq i,i_2\neq i,}^{n}\psi_l(Z_{i_1},Z_{i_2},\beta_0)\right|^2 = O(n^{-1}).$$

For the random vector $g(X_1, \beta_0)$, any arbitrary $C > 0$, one has that

$$\inf_{\omega \in \Omega} P(g(X_1, \beta_0)^T \omega \geq C) > 0,$$

and we claim that there exists a positive constant C_0 such that

$$a = \inf_{\omega \in \Omega} P(g(X_1, \beta_0)^T \omega \geq C_0) > 0. \tag{C.5}$$

Because the variance covariance matrix of $g(X_1, \beta_0)$ is $\Sigma_{p\times p}^{(\beta_0)}$, the function $g(X_1, \beta_0)$ is non-degenerate. Then following the proof of Lemma A.2 in Owen (1990), we can prove that there exists a unit vector ω_0 such that

$$P(g(X_1, \beta_0)^T \omega_0 \geq C) = 0.$$

Because of the arbitrariness of C, we conclude that

$$P(g(X_1, \beta_0)^T \omega_0 > 0) = 0,$$

which also indicates that

$$P(g(X_1, \beta_0)^T \omega_0 < 0) = 0.$$

However, the assumption $Eg(X_1, \beta_0)^T \omega_0 = 0$ leads to that $g(X_1, \beta_0)^T \omega_0 = 0$ a.s. This is a contradiction to the condition that $\Sigma_{p\times p}^{(\beta_0)}$ is positive definite. Thus, (C.5) is correct.

Following Li et al. (2016), we have that as $n \to \infty$,

$$\sup_{\omega \in \Omega} \left| P(g(X_1, \beta_0)^T \omega \geq C_0) - \frac{1}{n} \sum_{i=1}^{n} I\left(g(X_i, \beta_0)^T \omega \geq C_0\right) \right| \to 0 \ a.s.$$

Based on (C.4), it can be shown that as $n \to \infty$,

$$P\left(\inf_{\omega \in \Omega} \frac{1}{n} \sum_{i=1}^{n} I\left(\hat{Q}_i^T(\beta_0)\omega > C_0/2\right) > a/2 \right) \to 1,$$

where a is defined in (C.5).

We complete the proof of Lemma A.2. $\square$

Lemma A.3. Under conditions (D.1)-(D.3), we have $G^*(\beta_0) = \Sigma_{p\times p}^{(\beta_0)} + o(1)$, a.s.

Proof. Combining Lemma A.1 in Li et al. (2016) and strong law of large numbers for U-statistics, we get $W_n(\beta_0) = o(1)\ a.s.$ For $l = 1, ..., r$, let $\sigma_{H,l}^2(\beta_0) = Var(H_l(Z_1, Z_2; \beta_0))$. Since $E[H_l^2(Z_1, Z_2; \beta_0)] < \infty$, $\sigma_{H,l}^2(\beta_0) < \infty$. As a result,

$$\begin{aligned} G(\beta_0) &= \frac{1}{n}\sum_{i=1}^{n} \hat{V}_i(\beta_0)\hat{V}_i^T(\beta_0) \\ &= \frac{1}{n}\sum_{i=1}^{n} \Big[\hat{V}_i(\beta_0) - W_n(\beta_0) + W_n(\beta_0)\Big]\Big[\hat{V}_i(\beta_0) - W_n(\beta_0) + W_n(\beta_0)\Big]^T \\ &= \frac{1}{n}\sum_{i=1}^{n} \Big[\hat{V}_i(\beta_0) - W_n(\beta_0)\Big]\Big[\hat{V}_i(\beta_0) - W_n(\beta_0)\Big]^T + W_n(\beta_0)W_n^T(\beta_0) \\ &= \frac{1}{n}\sum_{i=1}^{n} \Big[nW_n(\beta_0) - (n-1)W_{n-1}^{(-i)}(\beta_0) - W_n(\beta_0)\Big]\Big[nW_n(\beta_0) - (n-1)W_{n-1}^{(-i)}(\beta_0) - W_n(\beta_0)\Big]^T \\ &+ W_n(\beta_0)W_n^T(\beta_0) \\ &= \frac{(n-1)^2}{n}\sum_{i=1}^{n} \Big[W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)\Big]\Big[W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)\Big]^T + o(1)\ a.s. \end{aligned} \tag{C.6}$$

From Lemma A.3 in Li et al. (2016), we have that

$$G(\beta_0) = \Sigma_{p\times p}^{(\beta_0)} + o(1)\ a.s.$$

Also, since

$$\begin{aligned} W_n(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} \hat{V}_n(\beta_0) &= \frac{1}{n}\sum_{i=1}^{n} \Big[nW_n(\beta_0) - (n-1)W_{n-1}^{(-i)}(\beta_0)\Big] \\ &= nW_n(\beta_0) - \frac{n-1}{n}\sum_{i=1}^{n} W_{n-1}^{(-i)}(\beta_0), \end{aligned} \tag{C.7}$$

which leads to

$$\sum_{i=1}^{n} W_{n-1}^{(-i)}(\beta_0) = nW_n(\beta_0). \tag{C.8}$$

Furthermore,

$$\begin{aligned}
G^*(\beta_0) &= \frac{1}{n}\sum_{i=1}^{n} \hat{Q}_i(\beta_0)\hat{Q}_i^T(\beta_0) \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[\hat{Q}_i(\beta_0) - U_n(\beta_0) + U_n(\beta_0)\right]\left[\hat{Q}_i(\beta_0) - U_n(\beta_0) + U_n(\beta_0)\right]^T \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[\hat{Q}_i(\beta_0) - U_n(\beta_0)\right]\left[\hat{Q}_i(\beta_0) - U_n(\beta_0)\right]^T + U_n(\beta_0)U_n^T(\beta_0). \qquad \text{(C.9)}
\end{aligned}$$

Note that the first term $\sum_{i=1}^{n} \hat{Q}_i(\beta_0)\hat{Q}_i^T(\beta_0)/n$ in equation (C.9) can be proved as $\Sigma_{p\times p}^{(\beta_0)} + o(1)$ $a.s.$

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \left[\hat{Q}_i(\beta_0) - U_n(\beta_0)\right]\left[\hat{Q}_i(\beta_0) - U_n(\beta_0)\right]^T \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[nU_n(\beta_0) - (n-1)U_{n-1}^{(-i)}(\beta_0) - U_n(\beta_0)\right]\left[nU_n(\beta_0) - (n-1)U_{n-1}^{(-i)}(\beta_0) - U_n(\beta_0)\right]^T \\
&= \frac{(n-1)^2}{n}\sum_{i=1}^{n} \left[U_n(\beta_0) - U_{n-1}^{(-i)}(\beta_0)\right]\left[U_n(\beta_0) - U_{n-1}^{(-i)}(\beta_0)\right]^T \\
&= \frac{(n-1)^2}{n}\sum_{i=1}^{n} \left[(W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)) + o(n^{-1/2}))\right]\left[(W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)) + o(n^{-1/2}))\right]^T \\
&= \frac{(n-1)^2}{n}\sum_{i=1}^{n} \left[W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)\right]\left[W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)\right]^T + o(1) \\
&+ \frac{2(n-1)^2}{n}\sum_{i=1}^{n} o(n^{-1/2})\left[W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)\right] \\
&= \frac{(n-1)^2}{n}\sum_{i=1}^{n} \left[W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)\right]\left[W_n(\beta_0) - W_{n-1}^{(-i)}(\beta_0)\right]^T + o(1) \\
&+ \frac{2(n-1)^2}{n} o(n^{-1/2})\left(nW_n(\beta_0) - \sum_{i=1}^{n} W_{n-1}^{(-i)}(\beta_0)\right) \\
&= G(\beta_0) + o(1)
\end{aligned}$$

$$= \Sigma_{p\times p}^{(\beta_0)} + o(1)\ a.s.$$

Also, based on the strong law of large number for U-statistics, we can have $U_n(\beta_0) = o(1)$ a.s. Therefore, $G^*(\beta_0) = \Sigma_{p\times p}^{(\beta_0)} + o(1)\ a.s.$ □

Lemma A.4. Let $A_n = \max_{1\le i\neq j\le n} \|K(Z_1, Z_2; \beta_0)\|$. Under the condition (D.1), we have $A_n = o(n^{1/2})$ a.s.

Proof. By a chaining argument, it suffices to prove that $2^{-n/2} \times \max_{1\le j\le 2^n} \|K(Z_j, Z_{2^n}; \beta_0)\| \to 0$ a.s. For each $\varepsilon > 0$, we have

$$\begin{aligned}
&\sum_{n=1}^{\infty} P\left\{\max_{1\le j\le 2^n} \|K(Z_j, Z_{2^n}; \beta_0)\| \ge \varepsilon 2^{n/2}\right\} \\
&\le \sum_{n=1}^{\infty} 2^n P\left\{\|K(Z_j, Z_{2^n}; \beta_0)\| \ge \varepsilon 2^{n/2}\right\} \\
&= \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} 2^n P\left\{2^{(m+1)/2} > \varepsilon^{-1}\|K(Z_j, Z_{2^n}; \beta_0)\| \ge 2^{m/2}\right\} \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{m} 2^n P\left\{2^{(m+1)/2} > \varepsilon^{-1}\|K(Z_j, Z_{2^n}; \beta_0)\| \ge 2^{m/2}\right\} \\
&\le \sum_{m=1}^{\infty} 2^{m+1} P\left\{2^{(m+1)/2} > \varepsilon^{-1}\|K(Z_j, Z_{2^n}; \beta_0)\| \ge 2^{m/2}\right\} \\
&\le 2\varepsilon^{-2} E[\|K(Z_1, Z_2; \beta_0)\|^2] \\
&\le 2\varepsilon^{-2} \times 16 E[\left\|XX^T\right\|^2] \\
&\le 32\varepsilon^{-2} M^2 \\
&< \infty.
\end{aligned}$$

Then by Borel-Cantelli Lemma, we have $2^{-n/2} \times \max_{1\le j\le 2^n} \|K(Z_j, Z_{2^n}; \beta_0)\| \to 0$ a.s. Thus, $A_n = o(n^{1/2})$ a.s. □

Lemma A.5. Let $B_n = \max_{1\le i\le n} \left\|\hat{Q}_i(\beta_0)\right\|$. Under conditions (D.1)-(D.3), $B_n = o(n^{1/2})$ and $n^{-1}\sum_{i=1}^{n} \left\|\hat{Q}_i(\beta_0)\right\|^3 = o(n^{1/2})$.

Proof. We can check that

$$\begin{aligned} U_n(\beta_0) &= \frac{1}{n(n-1)} \sum_{l=1}^{n} \sum_{j=1, j\neq l}^{n} K(Z_l, Z_j; \beta_0) \\ &= \frac{2}{n(n-1)} \sum_{j=1, j\neq i}^{n} K(Z_i, Z_j; \beta_0) + \frac{n-2}{n} U_{n-1}^{(-i)}(\beta_0). \end{aligned}$$

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

$$\begin{aligned} \left\| \hat{Q}_i(\beta_0) \right\| &= \left\| \frac{2}{n-1} \sum_{j=1, j\neq i}^{n} K(Z_i, Z_j; \beta_0) - U_{n-1}^{(-i)}(\beta_0) \right\| \\ &\leq 3 \max_{1\leq i\neq j\leq n} |K(Z_i, Z_j; \beta_0)| \\ &= 3A_n. \end{aligned} \tag{C.10}$$

Combining (C.10) and the result of Lemma A.4, that is, $A_n = o(n^{1/2})$ a.s. Thus,

$$B_n = \max_{1\leq i\leq n} \left\| \hat{Q}_i(\beta_0) \right\| = o(n^{1/2}),$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \left\| \hat{Q}_i(\beta_0) \right\|^3 \leq B_n \times \frac{1}{n} \sum_{i=1}^{n} \left\| \hat{Q}_i(\beta_0) \right\|^2 = o(n^{1/2}).$$

□

Proof of Theorem 4.1

Following Owen (2001) and Lu and Liang (2006), we let $\lambda = \rho\theta$, where $\rho \geq 0$ and $\|\theta\| = 1$. According to (4.5), we obtain that

$$\begin{aligned} 0 &= \|f(\lambda)\| \\ &= \|f(\rho\theta)\| \\ &\geq \left|\theta^T f(\rho\theta)\right| \\ &= \frac{1}{n} \left| \theta^T \left\{ \sum_{i=1}^{n} \hat{Q}_i(\beta_0) - \rho \sum_{i=1}^{n} \frac{\hat{Q}_i(\beta_0)\theta^T \hat{Q}_i(\beta_0)}{1 + \rho\theta^T \hat{Q}_i(\beta_0)} \right\} \right| \end{aligned}$$

$$\geq \frac{\rho}{n}\theta^T \sum_{i=1}^{n} \frac{\hat{Q}_i(\beta_0)\hat{Q}_i^T(\beta_0)}{1+\rho\theta^T\hat{Q}_i(\beta_0)}\theta - \frac{1}{n}\left|\sum_{j=1}^{p} e_j^T \sum_{i=1}^{n} \hat{Q}_i(\beta_0)\right|$$
$$\geq \frac{\rho\theta^T G^*(\beta_0)\theta}{1+\rho B_n} - \frac{1}{n}\left|\sum_{j=1}^{p} e_j^T \sum_{i=1}^{n} \hat{Q}_i(\beta_0)\right|,$$

where e_j is the unit vector on the jth coordinate direction.

By the central limit theorem, the second term is $O_p(n^{-1/2})$. We also have $G^*(\beta_0) = \Sigma_{p\times p}^{(\beta_0)} + o(1)$ a.s. from Lemma A.3. Thus, one has

$$\|\lambda\| = \rho = O_p(n^{-1/2}). \tag{C.11}$$

Denote $\eta_i = \lambda^T \hat{Q}_i(\beta_0)$, from Lemma A.5 and (C.11), it can be proved that

$$\max_{1\leq i\leq n} |\eta_i| = O_p(n^{-1/2})o_p(n^{1/2}) = o_p(1),$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{Q}(\beta_0)\frac{\eta_i^2}{1+\eta_i}\right\| = o_p(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{1/2}).$$

Note that

$$0 = f(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\hat{Q}(\beta_0)(1-\eta_i+\frac{\eta_i^2}{1+\eta_i})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\hat{Q}(\beta_0) - \frac{1}{n}\sum_{i=1}^{n}\hat{Q}(\beta_0)\eta_i + \frac{1}{n}\sum_{i=1}^{n}\hat{Q}(\beta_0)\frac{\eta_i^2}{1+\eta_i}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\hat{Q}(\beta_0) - G^*(\beta_0)\lambda + \frac{1}{n}\sum_{i=1}^{n}\hat{Q}(\beta_0)\frac{\eta_i^2}{1+\eta_i}.$$

Therefore, we can write

$$\lambda = (G^*(\beta_0))^{-1}U_n(\beta_0) + \gamma,$$

where $\|\gamma\| = o_p(n^{-1/2})$.

Then, by Taylor expansion, we can obtain that

$$l(\beta_0) = 2\sum_{i=1}^{n}\log(1+\eta_i) = 2\sum_{i=1}^{n}\eta_i - \sum_{i=1}^{n}\eta_i^2 + o_p(1)$$
$$= 2n\lambda^T U_n(\beta_0) - n\lambda^T G^*(\beta_0)\lambda + o_p(1)$$
$$= nU_n^T(\beta_0)(G^*(\beta_0))^{-1}U_n(\beta_0) - n\gamma^T G^*(\beta_0)\gamma + o_p(1). \qquad \text{(C.12)}$$

In (C.12), the first term can be shown converging to the chi-squared distribution, which is $nU_n^T(\beta_0)(G^*(\beta_0))^{-1}U_n(\beta_0) \xrightarrow{d} \chi_p^2$. Moreover, for the second term, we can obtain that $n\gamma^T G^*(\beta_0)\gamma = no_p(n^{-1/2})O_p(1)o_p(n^{-1/2}) = o_p(1)$. Therefore, $-2\log R(\beta_0) \xrightarrow{d} \chi_p^2$. □

Proof of Theorem 4.2

We follow the similar arguments in Yu et al. (2011) and Yang and Zhao (2012a). Corresponding to $\beta_0 = (\beta_{10}^T, \beta_{20}^T)^T$, we denote $Z = (Z_1^T, Z_2^T)^T$. Recall that under some suitable regularity conditions, $\sqrt{n}(\hat{\beta} - \beta_0)$ was shown to be asymptotically normally distributed with mean zero and variance-covariance matrix $D_n(\beta_0)^{-1}B_n(\beta_0)(D_n(\beta_0)^{-1})^T$.

Define

$$\bar{D}(\beta_0) = \lim_{n\to\infty} E[\partial S_n/\partial\beta_2]_{\beta_0}.$$

Since D is positive definite, $\bar{D}$ is of rank $p-q$. Denote

$$\hat{\beta}_2 = \arg\inf_{\beta_2} l\left[(\beta_{10}^T, \beta_2^T)^T\right].$$

Similar to Qin and Lawless (1994), we can show that

$$\sqrt{n}(\hat{\beta}_2 - \beta_{20}) = -(\bar{D}(\beta_0)^T(\Sigma_{p\times p}^{(\beta_0)})^{-1}\bar{D}(\beta_0))^{-1}\bar{D}(\beta_0)^T(\Sigma_{p\times p}^{(\beta_0)})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Q}_i(\beta_0) + o_p(1),$$

and

$$\sqrt{n}\lambda_2 = \left(I - \left(\Sigma_{p\times p}^{(\beta_0)}\right)^{-1}\bar{D}(\beta_0)\left(\bar{D}(\beta_0)^T\left(\Sigma_{p\times p}^{(\beta_0)}\right)^{-1}\bar{D}(\beta_0)\right)^{-1}\bar{D}(\beta_0)^T\right)(\Sigma_{p\times p}^{(\beta_0)})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Q}_i(\beta_0) + o_p(1),$$

where λ_2 is the corresponding Lagrange multiplier.

Recall that

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \hat{Q}_i(\beta).$$

Hence, by Taylor's expansion, one has that

$$\begin{aligned}
l^*(\beta_{10}) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Q}_i(\beta_0)\right)^T \left((\Sigma_{p\times p}^{(\beta_0)})^{-1} - (\Sigma_{p\times p}^{(\beta_0)})^{-1}\bar{D}(\beta_0)(\bar{D}(\beta_0)^T(\Sigma_{p\times p}^{(\beta_0)})^{-1}\bar{D}(\beta_0))^{-1}\bar{D}(\beta_0)^T(\Sigma_{p\times p}^{(\beta_0)})^{-1}\right) \\
&\quad \times \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Q}_i(\beta_0)\right) + o_p(1) \\
&= \left((\Sigma_{p\times p}^{(\beta_0)})^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Q}_i(\beta_0)\right)^T \Psi \left((\Sigma_{p\times p}^{(\beta_0)})^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{Q}_i(\beta_0)\right) + o_p(1) \\
&= \left((\Sigma_{p\times p}^{(\beta_0)})^{-1/2}\sqrt{n}U_n(\beta_0)\right)^T \Psi \left((\Sigma_{p\times p}^{(\beta_0)})^{-1/2}\sqrt{n}U_n(\beta_0)\right) + o_p(1),
\end{aligned}$$

where

$$\Psi = I - (\Sigma_{p\times p}^{(\beta_0)})^{-1/2}\bar{D}(\beta_0)(\bar{D}(\beta_0)^T(\Sigma_{p\times p}^{(\beta_0)})^{-1}\bar{D}(\beta_0))^{-1}\bar{D}(\beta_0)^T(\Sigma_{p\times p}^{(\beta_0)})^{-1/2}.$$

Note that Ψ is a symmetric and idempotent matrix with trace q. By Lemma A.1,

$$(\Sigma_{p\times p}^{(\beta_0)})^{-1/2}\sqrt{n}U_n(\beta_0) \xrightarrow{d} N(0, I_{p\times p}).$$

Then we have that

$$-2\log R^*(\beta_{10}) \xrightarrow{d} \chi_q^2.$$

The proof of Theorem 4.2 is completed. □