12-16-2020

Novel Empirical Likelihood Inference Procedures for Zero-Inflated and Right Censored Data and Their Applications

Faysal I. Satter
Georgia State University

NOVEL EMPIRICAL LIKELIHOOD INFERENCE PROCEDURES FOR ZERO-INFLATED AND RIGHT CENSORED DATA AND THEIR APPLICATIONS

by

FAYSAL SATTER

Under the Direction of Yichuan Zhao, PhD

ABSTRACT

The empirical likelihood method is a reliable data analysis tool in all statistical areas for its nonparametric features with parametric likelihood benefits. Because of the versatility of this method, we investigate its performance under survival and non-survival data structures. Zero-inflated data may arise in many areas where there are many zero values, and the non-zero values are often highly positively skewed. Confidence intervals based on a normal approximation for such zero-inflated data may have low coverage probabilities. We study empirical likelihood (EL) based inference techniques to construct a nonparametric confidence interval for the mean of a zero-inflated population, the mean difference of two zero-inflated skewed populations, and the quantile difference of a zero-inflated population.

We also apply the empirical likelihood method in two different kinds of survival data. First, we consider panel count data. In panel count data, each study subject can only be observed at discrete time points rather than continuously. The total number of events between the two observation times are known, but the exact time of events is unknown. Furthermore, the observation times can be different among subjects and carry important information about the underlying recurrent process. The second dataset comes from cohort study data. Collecting covariate information on all study subjects makes cohort studies very expensive. One way to reduce the cost while keeping sufficient covariate information is to use a case-cohort study design. We consider case-cohort data to make inferences about the regression parameters of semiparametric transformation models. For both datasets, an empirical likelihood ratio is formulated, and the Wilks' theorem is established.

Extensive simulation studies are carried out to assess all the methods mentioned earlier in various data settings. We compare the performance in terms of coverage probabilities and average lengths by NA and EL methods' confidence intervals. The applicability of the methods is also illustrated by real datasets.

INDEX WORDS: Empirical likelihood, Jackknife empirical likelihood, Confidence interval, Wilks' Theorem, Zero-inflation, Quantile, Panel count data, Case-cohort study design

NOVEL EMPIRICAL LIKELIHOOD INFERENCE PROCEDURES FOR ZERO-INFLATED AND RIGHT CENSORED DATA AND THEIR APPLICATIONS

by

FAYSAL SATTER

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

2020

NOVEL EMPIRICAL LIKELIHOOD INFERENCE PROCEDURES FOR ZERO-INFLATED AND RIGHT CENSORED DATA AND THEIR APPLICATIONS

by

FAYSAL SATTER

Committee Chair: Yichuan Zhao

Committee: Jing Zhang
Yichen Cheng
Jun Kong

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2020

DEDICATION

This dissertation is dedicated to my family, and friends.

ACKNOWLEDGEMENTS

This dissertation work would not have been possible without the support of many people.

I want to express my utmost gratitude to my advisor, Dr. Yichuan Zhao, who nurtured and guided me throughout the doctoral program. I am fortunate enough to have Dr. Zhao as my advisor from the beginning of my Ph.D. study. The strong motivation, aspiration to learn, and dedication to research, I got all of them from him. Whenever I needed any help or guidance, may it be academic or personal, Dr. Zhao was there. I would not continue this Ph.D. journey without his selfless support and advice.

I would also like to express my sincere gratitude to the committee members Dr. Jing Zhang, Dr. Yichen Cheng, and Dr. Jun Kong, for accepting my invitation to be a committee member. I want to thank them for their thoughtful and constructive comments and suggestions. These suggestions helped me to conduct my research and write the dissertation in the right direction.

I want to thank the department chair, Dr. Guantao Chen, and the associate chair Dr. Alexandra Smirnova for their academic and teaching assistantship support. These supports are crucial and pivotal for this doctoral study.

I want to express my appreciation to my fellow students, Jameson Stillwell, Xue Yu, Kangni Alemdjrodo, Husneara Rahman, Edem Defor, Bing Liu, Shuenn Siang Ng, Alan Dills for any personal and academic help or discussion, and made me feel at home. The open-up conservation with them on any topics helped me relieve stress and stay focused on my study.

I want to thank my parents for believing in me and supporting me emotionally throughout my whole life. Sending the only son abroad for higher studies was hard for them, and yet they made that sacrifice to achieve my dream. My wife understands the focus and dedication needed for the Ph.D. study, and she is super considerate and supportive. I want to thank my sister, brother-in-law, and other family members for always being by my side.

I want to thank the “Touch the Earth”, Recreational Services’ outdoor recreation program at GSU. The weekend outdoor activities are an integral part of my PhD life.

Finally, I want to thank everyone for helping me finish this long doctoral journey, one way or another. This is about achieving a long-sought dream, and you all are part of it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- AEL - Adjusted empirical likelihood
- AJEL - Adjusted jackknife empirical likelihood
- AL - Average length
- BCML - Bias-corrected maximum likelihood
- CCI - Case-cohort study with the same number of cases and controls
- CCII - Case-cohort study with twice as many controls as cases
- CDF - Cumulative distribution function
- CI - Confidence interval
- CP - Coverage probability
- DNA - Deoxyribonucleic acid
- ECDF - Empirical cumulative distribution function
- EL - Empirical likelihood
- ELR - Empirical likelihood ratio
- FH - Favourable histology
- FULL - Full cohort study
- i.i.d. - Independent and identically distributed
- JEL - Jackknife empirical likelihood
- NA - Normal approximation

- TEL - Transformed empirical likelihood
- TJEL - Transformed jackknife empirical likelihood
- TAJEL - Transformed adjusted jackknife empirical likelihood
- UH - Unfavorable histology
- VACURG - Veterans administration cooperative urological research group

CHAPTER 1

INTRODUCTION

The idea of nonparametric likelihood is first sketched by Thomas and Grunkemeier (1975) in survival data analysis, where they proposed a likelihood ratio method to construct a confidence interval for survival probabilities for right censored data. Owen (1988) and Owen (1990) extended this methodology in more general settings and introduced "empirical likelihood" method to construct a confidence interval for the univariate mean with theoretical justification. This method effectively combines the benefits of the parametric likelihood approach and the reliability of the nonparametric method. Some notable advantages of empirical likelihood (EL) are: (1) Unlike parametric likelihood-based methods, the EL technique does not assume any known family of distribution, yet it produces more efficient and powerful estimation and tests. (2) EL confidence region is range-preserving and transformation invariant (3) The shape and orientation of the confidence region are determined entirely by the data. (4) EL regions are Bartlett correctable, which increases the coverage accuracy. (5) The variance or variance-covariance matrix is not needed to construct a confidence interval. (6) Unlike normal approximation and the bootstrap method, EL does not require a pivotal quantity to construct a confidence interval. For a comprehensive review, see Owen (2001). Because of these attractive features, EL methods have been applied in a wide range of research areas. In this dissertation, we study the performance and applicability of empirical likelihood with three different data structures: zero-inflated data, panel count data, and case-cohort data.

In many situations, we may observe many zeros in the data. These zero values are valid observations and an indispensable part of the data. These excess zeros make the data zero-inflated. And often, zero-inflated data is positively skewed. Panel count data is a special kind of survival data, where study subjects are only observed at discrete time points during the study period. As a result, one can only know the total number of events between two observation time points instead of the events' actual time. The subject becomes censored if there are no events at the end of the study period. Also, the observation history may offer information about the recurrent event process. Cohort studies are designed to evaluate the association between exposure and disease. This design is very costly and time-consuming because all the study subjects are followed over the whole study period. A case-cohort study is designed to reduce the cost with the same level of efficiency. Covariate information is collected only from a subsample of the full cohort and all cases. This reduces the need to get covariate information from all study subjects.

1.1 Empirical Likelihood

In this section, we will review the formulation of EL. The cumulative distribution function (CDF) of a random variable X is defined as

$$F(x) = \Pr(X \leq x), -\infty < x < \infty.$$

The corresponding empirical cumulative distribution function (ECDF) of independent random samples $X_1, \ldots, X_n$ is

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x), -\infty < x < \infty.$$

Denote $p_i = \Pr(X = X_i) = F(X_i) - F(X_i-)$, where $F(x-) = \Pr(X < x)$. The empirical likelihood function is

$$L(F) = \prod_{i=1}^{n} p_i.$$

It can be easily shown that $F_n(\cdot)$ maximizes the likelihood function $L(F)$. Owen (1988) defines the nonparametric empirical likelihood ratio (ELR) as

$$R(F) = \frac{L(F)}{L(F_n)},$$

with the constraints $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$, where p_i is a probability vector. For example, we can define the empirical likelihood for a population mean μ as

$$L(\mu) = \max\left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i X_i = \mu, \sum_{i=1}^n p_i = 1, p_i \geq 0. \right\}.$$

Since $L(\mu)$ is attains its maximum at n^{-1} under the constraint $\sum_{i=1}^n p_i = 1$, the empirical likelihood ratio at μ is

$$R(\mu) = \max\left\{ \prod_{i=1}^n np_i : \sum_{i=1}^n p_i X_i = \mu, \sum_{i=1}^n p_i = 1, p_i \geq 0. \right\}. \tag{1.1}$$

This is a maximization problem with two constraints. The Lagrange multiplier method can be used to solve this optimization problem. Let λ be the Lagrange multiplier, which can be computed by solving the equation

$$\sum_{i=1}^n \frac{X_i - \mu}{1 + \lambda'(X_i - \mu)} = 0.$$

Plugging λ into eqn. (1.1), and then taking the log yields the empirical log-likelihood ratio

$$\log R(\mu) = -\sum_{i=1}^n \log[1 + \lambda(X_i - \mu)].$$

Let μ_0 be the true value of μ. Owen (1990) proved the nonparametric version of Wilks' theorem for $\log R(\mu_0)$, i.e., $-2\log R(\mu_0)$ converges to a limiting chi-squared distribution with one degree of freedom. Using the Wilks' theorem, the $100(1-\alpha)\%$ confidence interval for μ is given by

$$I(\alpha) = \{\mu : -2\log R(\mu_0) \leq \chi_1^2(\alpha)\}.$$

Despite having all the benefits, the EL method may suffer from some drawbacks, such as computational complexity, under-coverage problem, and convex-hull problem. Many researchers proposed several versions of empirical likelihood to overcome these issues. We highlight a few of them below.

1.1.1 Jackknife empirical likelihood

In this section, we go over jackknife empirical likelihood (JEL). Jing et al. (2009) pointed out that EL method is computationally very difficult for non-linear functionals, such as U-statistics. They proposed a variant of EL method, called jackknife empirical likelihood method, by combining the EL and the jackknife pseudo-values (Quenouille (1956)). The idea is to calculate jackknife pseudo-values from the parameter of interest and then apply EL method to those pseudo-values. Since this is simply a sample mean based EL application, it is easy to apply and more capable of handling complex non-linear constraints. JEL enjoys all the EL benefits and Jing et al. (2009) showed that the Wilks' theorem still holds for JEL method.

Let θ be the parameter of interest. For n independent random variables $X_1, \ldots, X_n$, define a consistent estimator of θ as the following U-statistics,

$$Y_n = T(X_1, \ldots, X_n).$$

Let $Y_{n-1}^{(-i)}$ be the statistic from $(n-1)$ variables after deleting i-th observation from the data. Then the jackknife pseudo-values are defined as

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

The average of these pseudo-values is the jackknife estimator of θ

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

Shi (1984) showed that under mild condition, the pseudo-values are asymptotically independent. The next step is to apply EL method to these pseudo-values.

Let $p = (p_1, \ldots, p_n)$ be a probability vector. Then $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$ for $1 \leq i \leq n$.

Let $\theta_p = \sum_{i=1}^{n} p_i E\widehat{V}_i$. Then the empirical likelihood, evaluated at θ, is

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

Then the jackknife empirical likelihood ratio at θ is

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

Using the Lagrange multiplier method, we obtain the jackknife empirical log-likelihood ratio as follows,

$$\log R(\theta) = -\sum_{i=1}^{n} \log[1 + \lambda(\widehat{V}_i - \theta_p)],$$

where λ satisfies the equation

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

Let θ_0 be the true value of θ. Jing et al. (2009) proved the nonparametric Wilks' theorem for $\log R(\theta_0)$, i.e.,

$$-2\log R(\theta_0) \xrightarrow{\mathfrak{D}} \chi^2_1.$$

1.1.2 Adjusted empirical likelihood

We sketch the procedure of adjusted empirical likelihood (AEL) in reference to constructing a empirical likelihood confidence interval for population mean described in Section 1.1. one pivotal step in EL method is to find the Lagrange multiplier by solving the eqn.

$$\sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda'(X_i - \mu)} = 0.$$

If zero is not in the interior of the convex hull of $(X_i - \mu)$, the solution does not exists and it creates a convex-hull problem for solving the equation. In some situations, empirical likelihood can suffer from this convex-hull problem, and the overall performance can be worse. To address this issue, Chen et al. (2008) proposed the adjusted empirical likelihood. The key idea is to add one more artificial data point to the existing dataset to make sure that zero is in the interior of the convex hull of $(X_i - \mu)$.

We denote $g_i(\mu) = (X_i - \mu)$. The extra artificial value is obtained by

$$g_{n+1}(\mu) = -\frac{a_n}{n}\sum_{i=1}^{n} g_i(\mu),$$

where a_n is a positive constant. A general recommendation by Chen et al. (2008) is to use $a_n = \max(1, \log(\mathrm{n})/2)$, but other values can be appropriate in different situations. With the new value, the adjusted empirical likelihood ratio at μ is

$$R^a(\mu) = \max\left\{\prod_{i=1}^{n+1}(n+1)p_i : \sum_{i=1}^{n+1} p_i X_i = \mu, \sum_{i=1}^{n+1} p_i = 1, p_i \geq 0\right\}.$$

We can optimize the AEL ratio using Lagrange multiplier and Chen et al. (2008) showed that the Wilks' theorem also holds for the AEL, i.e.,

$$-2\log R^a(\mu_0) \xrightarrow{\mathfrak{D}} \chi_1^2.$$

This simple adjusted improve the coverage probabilities of the empirical likelihood, especially when the sample size is small.

1.1.3 Transformed empirical likelihood

Now, we review the transformed empirical likelihood (TEL). TEL is proposed by Jing et al. (2017) to improve the coverage probability of EL method without adding any complexity or theoretical justification. This is done by a simple transformation of the original empirical likelihood. Jing et al. (2017) suggested a set of four properties the transformation should possess in order to be considered as a good candidate for the transformation function. Using these properties and the extended empirical likelihood (proposed by Tsao and Wu (2013)) as starting point, a transformed empirical likelihood method is developed.

Let $l(\theta)$ be the original empirical log-likelihood ratio for the parameter θ. Many functions can satisfy the four properties to be considered as good transformation function. The truncated quadratic transformation $g_t(l(\theta); \gamma)$ of $l(\theta)$ is a good option, which is defined as

$$g^t(l(\theta); \gamma) = l(\theta) \times \max[1 - l(\theta)/n, 1 - \gamma],$$

where $\gamma \in [0, 1]$ is a constant, and n is the sample size. The default choice for γ is 0.5. Then, the corresponding transformed empirical log-likelihood ratio is

$$\begin{aligned} l^t(\theta) &= g^t[l(\theta); \gamma = 0.5)] \\ &= l(\theta) * \max[1 - l(\theta)/n, 0.5], \end{aligned}$$

which can be written as

$$l^t(\theta) = \begin{cases} l(\theta)[1 - l(\theta)/n] & \text{if} \quad l(\theta) \leq n/2 \\ l(\theta)/2 & \text{if} \quad l(\theta) > n/2. \end{cases}$$

TEL has the same asymptotic properties as EL, i.e., Wilks' theorem holds for the TEL such that [cf. Jing et al. (2017)]

$$-2\log l^t(\theta_0) \xrightarrow{\mathfrak{D}} \chi_1^2.$$

The remaining dissertation is organized as follows. In Chapter 2, we illustrate a novel empirical likelihood and adjusted empirical likelihood methods for the mean of a zero-inflated population. In Chapter 3, we discuss jackknife empirical likelihood and adjusted jackknife empirical likelihood methods for the mean difference of two zero-inflated populations. We study the jackknife empirical likelihood methods for the quantile difference of a zero-inflated data in Chapter 4. We propose empirical likelihood confidence intervals method for panel count data in Chapter 5. Case-cohort data is analyzed by empirical likelihood and adjusted empirical likelihood methods in Chapter 6. Summary and some concluding remarks are included in Chapter 7.

CHAPTER 2

EMPIRICAL LIKELIHOOD FOR THE MEAN OF A ZERO-INFLATED POPULATION

2.1 Background

In recent years, ecology, public health, medicine, and environmental science often produce data, which are highly positively skewed and contain significant proportion of zero values. Zero-inflated data refer to data with a heterogeneous distribution, which has attracted more attention for the psychology and health care research. These zero-inflated and positive skewed populations are contrary to the simple homogeneous population and unlikely to follow the normal distribution. As a result, the existing methods, which do not consider this special kind of data setting may give low coverage probabilities.

Parametric and empirical likelihood methods are used to find a confidence interval for the population containing many zero values. Welsh et al. (1996) presented parametric conditional models that incorporate excess zeros with an application to the mean abundance of rare animals. Dobbie and Welsh (2001) proposed a method of modeling correlated zero-inflated data. Fletcher et al. (2005) used a combination of ordinary and logistic regression for skewed data with many zeros. Min and Agresti (2002) proposed a model for nonnegative observations clumping at zero. Application of zero-inflated Poisson and binomial regression is given by Hall (2000) and Lambert (1992). An example of Hurdle models for count data with extra zeros is given by Hu et al. (2011).

Empirical likelihood (EL) method was proposed by Owen (1988, 1990), which has the advantage of having parametric likelihood benefits without having any distributional assumptions. The shape of the confidence region depends on the data, and it is range respective and transformation invariant. Furthermore, Diciccio et al. (1991) showed that this method is Bartlett correctable which is later extended by Corcoran (1998) by introducing two broader classes of nonparametric discrepancy measures. Hall and La Scala (1990) gave few methodologies and algorithms of the empirical likelihood. A comprehensive overview of empirical likelihood can be found in Owen (2001). Also adding artificial observations to the original data is discussed in the literature by many authors under different settings. For example, to address the convex hull problem in the high-dimensional data, Chen et al. (2015) proposed the balanced augmented EL method, which added two pseudo-observations to the original data. This simple technique puts the zero vector in the convex hull, which produces good performance for the high-dimensional case. Another use of two extra pseudo-values to the data is illustrated in Cheng et al. (2018). They focused on the convex hull problem in the jackknife empirical likelihood (JEL) (Jing et al. (2009)) inference with the small sample size. The simulation studies indicated that this modification gave the better performance than the JEL when zero may not be in the convex hull.

Because of the advantages of empirical likelihood over parametric methods, it has been using extensively in all statistical areas. Chen and Qin (2003) proposed nonparametric empirical likelihood method to find confidence intervals for data with zero observations. Chen et al. (2003) also proposed a similar method without separating the zero and nonzero values. Kang et al. (2010) gave an application of empirical and parametric likelihood interval estimation of population having many zero observations. Zhou and Zhou (2005) proposed empirical likelihood for the mean difference of two skewed populations with many zero values. Pailden and Chen (2013) also proposed similar empirical likelihood for two zero-inflated populations. But the empirical likelihood for this data setting is very complicated and computationally intensified.

In this chapter, we propose an empirical likelihood confidence interval for a mean of the skewed population with additional zero values. We generate artificial values using the data structure and apply an empirical likelihood method to those values. We also propose an adjusted empirical likelihood (AEL) method. The adjusted empirical likelihood method was proposed by Chen et al. (2008), which enabled us to adjust the empirical likelihood function by adding one more data point. And this adjustment will produce better coverage probability than the empirical likelihood counterpart. Liu and Chen (2010) adjusted the empirical likelihood with a specific level to attain high level precision even for small sample or high dimension estimating function. The advantage of our proposed EL and AEL methods is that they are easy to implement and has simpler interpretation than existing ones.

This chapter is organized as follows. In Section 2.2, empirical likelihood and adjusted empirical likelihood method are proposed. We conduct simulation studies to evaluate the proposed methods in Section 2.3. Then we apply this method to a real life data set in Section 2.4. Some concluding remarks are given in Section 2.5. All proofs are given in the Appendix A.

2.2 Main Results

2.2.1 Zero-inflated population

we consider an unknown skewed population that contains a significant proportion of zero values. Let $X_1, X_2, \ldots, X_n$ be non-negative i.i.d. random samples from the skewed population X with $E(X) = \mu$. Let $1 > \delta = P(X = 0) > 0$ be the probability of having zero values. For convenience, we assume that n_1 represents the number of positive values in X_i's. Denote these values as $x_1, x_2, \ldots, x_{n_1}$. Then, one has $n_0 = n - n_1$ number of zero values in the sample. It can be shown that

$$\mu = E(X) = E(X|X > 0)P(X > 0) = (1 - \delta)E(X|X > 0).$$

The nonparametric estimator T_{n_1} of μ can be written as:

$$T_{n_1} = \frac{\sum_{i=1}^{n_1} x_i}{n} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} \cdot \frac{n_1}{n}.$$

We denote:

$$\hat{V}_i = \frac{n_1 x_i}{n}, \quad i = 1, 2, \ldots, n_1.$$

As $n \to \infty$ we have that $n_1/n \to 1-\delta$ and

$$T_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} \hat{V}_i = \frac{n_1}{n}\frac{\sum_{i=1}^{n_1} x_i}{n_1} \to (1-\delta)E(X|X>0) = \mu.$$

2.2.2 Empirical likelihood procedure

We now apply the empirical likelihood method on these values to make inference for μ. The empirical likelihood ratio at the parameter μ is given by

$$R(\mu) = \sup\left\{\prod_{i=1}^{n_1}(np_i) : \sum_{i=1}^{n_1} p_i = 1, \sum_{i=1}^{n_1} p_i(\hat{V}_i - \mu) = 0, p_i \geq 0, i = 1, \ldots, n_1\right\}.$$

By using the Lagrange multiplier method, the empirical log-likelihood ratio is

$$\log R(\mu) = -\sum_{i=1}^{n_1} \log[1 + \lambda(\hat{V}_i - \mu)].$$

where λ satisfies the following equation

$$\frac{1}{n_1}\sum_{i=1}^{n_1} \frac{\hat{V}_i - \mu}{1 + \lambda(\hat{V}_i - \mu)} = 0. \tag{2.1}$$

We establish the asymptotic chi-squared distribution of the empirical log-likelihood ratio statistic and construct the EL and AEL confidence intervals.

Theorem 2.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. non-negative observations of X. Assume $0 < \delta < 1$ and $E|X|^2 < \infty$. Let μ_0 be the true value of $\mu = EX = (1-\delta)E(X|X>0)$. Then*

$$-2logR(\mu_0) \to \chi_1^2 \textit{ in distribution as } n \to \infty.$$

Based on Theorem, 2.1, we can construct the asymptotic $100(1-\alpha)\%$ empirical likelihood confidence interval for μ_0 as

$$I_1(\alpha) = \{\mu : -2\log R(\mu) \leq \chi_1^2(\alpha)\},$$

where $\chi_1^2(\alpha)$ is the $(1-\alpha)$ quantile of a chi-square distribution with 1 degree of freedom.

2.2.3 Adjusted empirical likelihood procedure

For the adjusted empirical likelihood, we add one more data point to the artificial values and apply an empirical likelihood method. Let $g_i = \hat{V}_i - \mu, \quad i = 1, 2, \ldots, n_1$, and $\bar{g} = n_1^{-1} \sum_{i=1}^{n_1} g_i$ for any fixed μ. Following Chen et al. (2008), for some $a_{n_1} > 0$, we define

$$g_{n_1+1} = -a_{n_1}\bar{g} = -\frac{a_{n_1}}{n_1}\sum_{i=1}^{n_1} g_i.$$

where $a_{n_1} = \max(1, \log(n_1)/2)$ as suggested by Chen et al. (2008). We apply EL method to these $n_1 + 1$ values. Then the adjusted empirical log-likelihood ratio of μ is defined as

$$R^\star(\mu) = \sup\left\{\sum_{i=1}^{n_1+1} log\left\{(n_1+1)p_i\right\} : \sum_{i=1}^{n_1+1} p_i = 1, \sum_{i=1}^{n_1+1} p_i(g_i) = 0, p_i \geq 0, \right.$$
$$\left. i = 1, \ldots, n_1 + 1\right\}.$$

Using Lagrange multiplier method, we obtain the adjusted empirical log-likelihood ratio as follows,

$$\log R^\star(\mu) = -\sum_{i=1}^{n_1+1} \log[1 + \lambda g_i],$$

where λ satisfies

$$\frac{1}{n_1+1}\sum_{i=1}^{n_1+1} \frac{g_i}{1+\lambda g_i} = 0.$$

Theorem 2.2. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. non-negative observations of X. Suppose $0 < \delta < 1$ and $E|X|^2 < \infty$. Let μ_0 be the true value of $\mu = EX = (1-\delta)E(X|X>0)$. As $n \to \infty$,*

$$-2logR^\star(\mu_0) \to \chi_1^2 \text{ in distribution.}$$

From Theorem 2.2, the asymptotic $100(1-\alpha)\%$ adjusted empirical likelihood (AEL) confidence interval for μ_0 is as follows,

$$I_2(\alpha) = \{\mu : -2\log R^{\star}(\mu) \leq \chi_1^2(\alpha)\}.$$

2.3 Simulation Study

For the simulation, zero values are generated from binomial distribution with proportion 0.1, 0.2 and 0.5. And the non-zero observations are generated from log-normal, chi-squared and exponential distributions. All simulations are repeated 10000 times.

We summarize the coverage probabilities and average lengths for the empirical likelihood (EL), AEL, EL (Chen and Qin (2003)), and NA CIs at a 95% nominal confidence level. We generate the log-normal distribution with parameters 0 and 1 in the first setting. For the log-normal distribution, bias-corrected maximum likelihood (BCML) confidence intervals (Zhou and Tu (2000)) are included for comparison. For the second setting, a data set is generated from the exponential distribution with parameter 1. Chi-square distribution with parameter 1 is used to generate data for the third setting.

From the simulation studies, we find that the coverage probabilities of the our proposed empirical likelihood (EL) confidence interval are better than that of based on normal approximation, maximum likelihood and empirical likelihood (see Chen and Qin (2003)) in most cases. From Tables 2.1, 2.2 and 2.3, our adjusted empirical likelihood confidence interval has even better coverage probability than other ones. The average length of the adjusted empirical likelihood confidence interval is longer than ones of other confidence intervals. And the length gets shorter as the sample size increases.

2.4 Real Data Analysis

To illustrate our method, we use the data from a pilot study aimed to examine the accuracy of the sexual behavior report of adolescence females. We are interested in the mean of contemporaneous daily diary report. The data set has been taken from Tang et al. (2012). The data is zero-inflated. Out of 47 observations, there are 13 zero observations and the non-zero observations are right skewed (the skewness is 1.770). The mean and the standard deviation of diary group are 9.106 and 12.356, respectively.

A point estimate of the mean μ is 9.106. The EL confidence interval is (5.308, 13.929), and the adjusted empirical likelihood confidence interval is (5.415, 14.246). The corresponding confidence interval based on normal approximation is (5.570, 12.642) and the confidence interval with the empirical likelihood (Chen and Qin (2003)) is (5.401, 13.526). Since our simulation study shows that our adjusted empirical likelihood has better coverage than other methods. In our example, one may want to use (5.415, 14.246) as the confidence interval for the mean of contemporaneous daily diary report. One might also consider the EL confidence interval (5.308, 13.929) because of its shorter length than AEL confidence interval even if it may have less coverage probability than AEL as our simulation study shows.

2.5 Conclusion

Contemporary research studies collect information on an array of measurements with the significant amount of zeros. More recently, there has been a huge influx of zero-inflated observations in biology and health care. This has motivated the broad use of statistical procedures for zero-inflated data in practice. In this chapter, we proposed EL and adjusted EL methods to obtain the confidence interval for the mean of a skewed population where the population contains significant number of zero observations. These methods are compared with the normal approximation method and EL method proposed by Chen and Qin (2003). The simulation studies show that the EL and AEL interval estimates have more accurate coverage probability than that of based on normal approximation and empirical likelihood methods proposed by Chen and Qin (2003). Future studies may involve considering other more advanced EL methods, such as the extended empirical likelihood, the balanced augmented EL and we extend the new method to the mean difference of two samples, etc.

Table (2.1) Coverage probabilities (average lengths) with nominal level 0.95 when the skewed data follow the log-normal distribution.

n	δ	NA	BCML	EL (Chen and Qin (2003))	EL	AEL
	.1	.867 (1.593)	.881 (1.629)	.879 (1.648)	.889 (1.724)	.889 (1.772)
20	.2	.836 (1.552)	.882 (1.604)	.851 (1.795)	.891 (1.807)	.895 (1.865)
	.5	.850 (1.581)	.853 (1.605)	.854 (1.590)	.851 (1.602)	.881 (1.694)
	.1	.889 (1.056)	.912 (1.128)	.890 (1.146)	.892 (1.179)	.922 (1.257)
50	.2	.895 (1.034)	.934 (1.109)	.925 (1.301)	.943 (1.265)	.948 (1.358)
	.5	.901 (1.021)	.936 (1.002)	.914 (1.215)	.925 (1.198)	.945 (1.308)
	.1	.906 (0.782)	.944 (0.785)	.933 (0.870)	.937 (0.880)	.943 (0.906)
100	.2	.906 (0.751)	.944 (0.765)	.939 (0.899)	.945 (0.922)	.950 (0.952)
	.5	.915 (0.645)	.943 (0.666)	.941 (0.892)	.947 (0.872)	.952 (0.932)

NOTE:
NA: normal approximation
BCML: bias-corrected maximum likelihood
EL: empirical likelihood
AEL: adjusted empirical likelihood

Table (2.2) Coverage probabilities (average lengths) with nominal level 0.95 when the skewed data follow the exponential distribution.

n	δ	NA	EL (Chen and Qin (2003))	EL	AEL
	.1	.890 (0.817)	.904 (0.840)	.917 (0.860)	.924 (0.972)
20	.2	.907 (0.832)	.921 (0.932)	.924 (0.918)	.935 (0.982)
	.5	.921 (0.822)	.920 (0.889)	.937 (0.957)	.940 (0.984)
	.1	.927 (0.523)	.930 (0.570)	945 (0.573)	.947 (0.593)
50	.2	.928 (0.519)	.933 (0.520)	.945 (0.544)	.945 (0.585)
	.5	.932 (0.520)	.936 (0.545)	.947 (0.561)	.948 (0.566)
	.1	.930 (0.388)	.933 (0.387)	.943 (0.418)	.946 (0.430)
100	.2	.933 (0.381)	.940 (0.418)	.944 (0.440)	.950 (0.455)
	.5	.937 (0.383)	.942 (0.410)	.950 (0.450)	.956 (0.460)

NOTE:
NA: normal approximation
BCML: bias-corrected maximum likelihood
EL: empirical likelihood
AEL: adjusted empirical likelihood

Table (2.3) Coverage probabilities (average lengths) with nominal level 0.95 when the skewed data follow the chi-squared distribution.

n	δ	NA	EL (Chen and Qin (2003))	EL	AEL
	.1	.821 (0.944)	.822 (0.962)	.875 (1.080)	.907 (1.202)
20	.2	.829 (0.905)	.836 (0.998)	.858 (1.004)	.893 (1.133)
	.5	.843 (0.897)	.900 (0.902)	.937 (0.952)	.938 (0.992)
	.1	.912 (0.736)	.917 (0.787)	.919 (0.835)	.930 (0.870)
50	.2	.910 (0.713)	.912 (0.720)	.921 (0.791)	.936 (0.828)
	.5	.901 (0.593)	.908 (0.603)	.910 (0.620)	.916 (0.708)
	.1	.932 (0.430)	.933 (0.421)	.942 (0.466)	.946 (0.487)
100	.2	.929 (0.411)	.940 (0.431)	.945 (0.442)	.952 (0.460)
	.5	.918 (0.427)	.920 (0.444)	.945 (0.446)	.954 (0.458)

NOTE:
NA: normal approximation
BCML: bias-corrected maximum likelihood
EL: empirical likelihood
AEL: adjusted empirical likelihood

CHAPTER 3

JACKKNIFE EMPIRICAL LIKELIHOOD INFERENCE FOR THE MEAN DIFFERENCE OF TWO ZERO-INFLATED SKEWED POPULATIONS

3.1 Background

Many disciplines, such as environmental studies, ecology, biology, biometrics, epidemiology, insurance, meteorology, manufacturing, etc., have the potential to generate datasets containing many zero values, and non-zero values are highly positively skewed. These zeros are valid response outcomes, and therefore should not be ignored. For example, in estimating the mean diagnostic cost (testing charges) of a hospital, it is essential to use the diagnostic cost of all the patients. However, many patients may not have done any diagnostic tests during the period of interest, which will result in many zero test charges. Another example is the rainfall observation records, where there could be no rain for some days.

Various parametric/semiparametric methods have been developed to deal with zero-inflated datasets with the assumption of the population distribution. Lachenbruch (1976) assumed some parametric families, including exponential, log-normal distribution, for positive values and compared two groups using the so-called "two-part" test. Duan et al. (1983) used a two-part model under the assumption of the log-normal distribution for positive values. Zhou and Tu (1999), Zhou and Tu (2000) also employed a two-part model, which is a combination of binomial and log-normal distributions. However, using these methods depends on the justification of the assumptions about the distribution of the population.

Empirical likelihood method, first introduced by Owen (1988, 1990), is a nonparametric method for small samples with superior performance. Advantages of empirical likelihood (EL) include data-driven confidence regions, transformation invariance, and many more. Also, Diciccio et al. (1991) showed that the Bartlett correction improves the coverage rate from $O(n^{-1})$ to $O(n^{-2})$ for the sample size n, which is a notable improvement compared with classic EL. Zhou and Zhou (2005) proposed the empirical likelihood method to find the confidence interval for the mean difference of two zero-inflated skewed populations. Pailden and Chen (2013) also developed a similar empirical likelihood method for two zero-inflated populations.

Nevertheless, the empirical likelihood method involving non-linear estimation equations is complicated and computationally intensive. Jing et al. (2009) proposed jackknife empirical likelihood (JEL) method to overcome this computational difficulty. This method essentially converts the statistic of interest into a sample mean using jackknife pseudo-values. It then applies empirical likelihood procedures on the mean of those pseudo-values (see Jing et al. (2009)). This JEL method makes the estimation problem much simpler and computationally efficient. For the inference of only a part of parameters, Li et al. (2011) implemented JEL in the profile empirical likelihood.

Furthermore, Chen et al. (2008) proposed an adjusted empirical likelihood (AEL) method to remove the zero convex hull problem in computing the profile likelihood function. They added one more data point to the original dataset, and it improved the coverage probability substantially. Liu and Yu (2010) studied the Bartlett correction on two-sample adjusted empirical likelihood. More work was done by Liu and Chen (2010), who considered adjusted empirical likelihood for higher order precision. Zhao et al. (2015) proposed adjusted jackknife empirical likelihood for the mean absolute deviation. Chen and Ning (2016) later combined JEL and AEL methods and proposed an adjusted jackknife empirical likelihood (AJEL) for one-sample and two-sample U-statistics. Another example of using artificial data points is Chen et al. (2015). They considered the high-dimensional data setting when the sample size and the data dimension are comparable. In that setting, they dealt with the convex hull problem by adding two more data points to the data set. They showed the asymptotic normality of the empirical log-likelihood ratio statistic. More recently, Cheng et al. (2018) proposed balanced augmented jackknife empirical likelihood for two-sample U-statistics, where they added two artificial data points to the pseudo-values produced by the jackknife resampling process.

The aim of this chapter is to find a better inference method for the mean difference of two independent zero-inflated skewed populations. The empirical likelihood method proposed by Zhou and Zhou (2005) involves finding solutions to complicated non-linear equations. Zhou and Zhou (2005) treated $\delta_1 = P(X = 0) > 0$ and $\delta_2 = P(Y = 0) > 0$, which are the probabilities of having zero values in the two populations, as the nuisance parameters and resulting Wilks' statistic was maximized over δ_1 and δ_2 (see the details in Section 2). This EL formulation added more computational costs. To alleviate this complexity, we propose the JEL method, which makes the estimation process simpler and computationally more efficient. The idea is to first estimate δ_1 and δ_2 consistently. Then, we use jackknife pseudo-values to construct EL ratio for the mean difference. Motivated by Wang and Zhao (2016) and Alemdjrodo and Zhao (2019), we also propose AJEL method, which improves the estimation performance in terms of coverage accuracy.

The rest of the chapter is organized as follows. Jackknife empirical likelihood and adjusted jackknife empirical likelihood methods for the mean difference are proposed in Section 3.2. In Section 3.3, we carry out extensive simulation studies. We apply two real-life datasets to illustrate the proposed methodology in Section 3.4. Section 3.5 includes the conclusion and some remarks. All proofs are provided in the Appendix B.

3.2 Main Results

3.2.1 Model setup

In this section, we extend the EL approach for the mean of a zero-inflated population developed in Chapter 2 and propose jackknife empirical likelihood and adjusted jackknife empirical likelihood for the mean difference of two independent zero-inflated skewed populations.

Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be non-negative random samples from two independent skewed populations X, Y with means $E(X) = \mu_x$ and $E(Y) = \mu_y$, respectively. We are interested in constructing a confidence interval for the mean difference,

$\theta = \mu_x - \mu_y$. We assume both distributions have a significant amount of zero observations. Let $\delta_1 = P(X = 0) > 0$ and $\delta_2 = P(Y = 0) > 0$ be the probability of having zero values of these two populations. Suppose that there are m_1 positive values in X_i's, and denote these values as $x_1, x_2, \ldots, x_{m_1}$. Then, there are $m_0 = m - m_1$ zero values in the first sample. Similarly, there are n_1 positive values in Y_i's, and denote them as $y_1, y_2, \ldots, y_{n_1}$. Thus, there are $n_0 = n - n_1$ zero values in the second sample. We can write μ_x and μ_y as

$$\mu_x = E(X|X > 0)P(X > 0) = (1 - \delta_1)E(X|X > 0),$$
$$\mu_y = E(Y|Y > 0)P(Y > 0) = (1 - \delta_2)E(Y|Y > 0).$$

3.2.2 Jackknife empirical likelihood procedure

We write the estimator of μ_x as $\bar{X} = m^{-1}\sum_{i=1}^{m_1} x_i = m_1^{-1}\sum_{i=1}^{m_1} x_i \cdot m_1/m$. Similarly the estimator of μ_y is $\bar{Y} = n_1^{-1}\sum_{j=1}^{n_1} y_j \cdot n_1/n$. Let $t = m_1 + n_1$. Denote $\hat{\delta}_1 = m_0/m$, and $\hat{\delta}_2 = n_0/n$. A consistent estimator of the parameter θ is (cf. Satter and Zhao (2020) and the proof of Lemma 3.1 in the Appendix B)

$$\begin{aligned}
T_t = \bar{X} - \bar{Y} &= \frac{\sum_{i=1}^{m_1} x_i}{m_1}\frac{m_1}{m} - \frac{\sum_{j=1}^{n_1} y_j}{n_1}\frac{n_1}{n} \\
&= (1 - \hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1 - \hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&= \frac{1}{m_1}\frac{1}{n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}\left(\frac{m_1 x_i}{m} - \frac{n_1 y_j}{n}\right) \\
&= \frac{1}{m_1}\frac{1}{n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}((1 - \hat{\delta}_1)x_i - (1 - \hat{\delta}_2)y_j) \\
&= \frac{1}{m_1}\frac{1}{n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1} h(x_i, y_j) \\
&= T(x_1, \ldots, x_{m_1}, y_1, \ldots, y_{n_1}),
\end{aligned}$$

where $h(x_i, y_j) = (1 - \hat{\delta}_1)x_i - (1 - \hat{\delta}_2)y_j, i = 1, 2, \ldots, m_1, j = 1, 2, \ldots, n_1$ is a kernel function with

$$Eh(X_i, Y_j) = E((1 - \hat{\delta}_1)x_i - (1 - \hat{\delta}_2)y_j)$$

$$
\begin{aligned}
&= E(E((1-\hat{\delta}_1)x_i - (1-\hat{\delta}_2)y_j)|m_1, n_1) \\
&= E((1-\hat{\delta}_1)E(x_i|x_i > 0) - (1-\hat{\delta}_2)E(y_j|y_j > 0)) \\
&= E(1-\hat{\delta}_1)E(x_i|x_i > 0) - E(1-\hat{\delta}_2)E(y_j|y_j > 0) \\
&= E\left(\frac{m_1}{m}\right)E(x_i|x_i > 0) - E\left(\frac{n_1}{n}\right)E(y_j|y_j > 0) \\
&= \left(\frac{E(1(x_1 > 0) + \cdots + 1(x_m > 0))}{m}\right)E(x_i|x_i > 0) \\
&\quad -\left(\frac{E(1(y_1 > 0) + \cdots + 1(y_n > 0))}{n}\right)E(y_j|y_j > 0) \\
&= P(X > 0)E(X|X > 0) - P(Y > 0)E(Y|Y > 0) \\
&= \mu_x - \mu_y \\
&= \theta.
\end{aligned}
$$

Note that the consistent estimate T_t is not a standard U-statistic, since the kernel function is not fixed. Therefore, the JEL for standard U-statistic by Jing et al. (2009) needs a modification. Using the estimator T_t, we can generate t jackknife pseudo-values

$$\hat{V}_k = tT_t - (t-1)T_{t-1}^{(-k)}, k = 1, 2, \ldots, t,$$

where $T_{t-1}^{(-k)}$ is based on the $(t-1)$ samples after deleting the ith observation from the original sample dataset.

Specifically, for $k = 1, \ldots, m_1$,

$$
\begin{aligned}
\hat{V}_k &= (m_1+n_1)\left[(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}\right] \\
&\quad -(m_1+n_1-1)\left[(1-\hat{\delta}_1)\frac{\sum_{i\neq k, i=1}^{m_1} x_i}{m_1-1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}\right] \\
&= (1-\hat{\delta}_1)\left[(m_1+n_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (m_1+n_1-1)\frac{\sum_{i\neq k, i=1}^{m_1} x_i}{m_1-1}\right] \\
&\quad -(1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1},
\end{aligned}
\tag{3.1}
$$

and for $k = m_1 + 1, \ldots, t$,

$$\begin{aligned} \hat{V}_k &= (m_1+n_1)\left[(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}\right] \\ &\quad -(m_1+n_1-1)\left[(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j\neq k-m_1, j=1}^{n_1} y_j}{n_1-1}\right] \\ &= (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} \\ &\quad -(1-\hat{\delta}_2)\left[(m_1+n_1)\frac{\sum_{j=1}^{n_1} y_j}{n_1} - (m_1+n_1-1)\frac{\sum_{j\neq k-m_1, j=1}^{n_1} y_j}{n_1-1}\right]. \end{aligned} \tag{3.2}$$

The eqns. (3.1) and (3.2) are useful to the proofs of lemmas in the Appendix B.

It can be shown that

$$T_t = \frac{1}{t}\sum_{i=1}^{t} \hat{V}_i.$$

Next, we apply the JEL to these pseudo-values. We can define the jackknife empirical likelihood ratio at θ as

$$R(\theta) = \sup\left\{\prod_{i=1}^{t}(tp_i) : \sum_{i=1}^{t} p_i = 1, \sum_{i=1}^{t} p_i(\hat{V}_i - \theta) = 0, p_i \geq 0, i = 1, \ldots, t\right\}.$$

Using the Lagrange multiplier method, the jackknife empirical log-likelihood ratio is

$$\log R(\theta) = -\sum_{i=1}^{t} \log[1 + \mu(\theta)(\hat{V}_i - \theta)],$$

where $\mu(\theta)$ satisfies the following equation

$$\frac{1}{t}\sum_{i=1}^{t} \frac{\hat{V}_i - \theta}{1 + \mu(\theta)(\hat{V}_i - \theta)} = 0.$$

Define

$$h_0(x, y) = (1-\delta_1)x - (1-\delta_2)y,$$

$$h_{01}(x) = Eh_0(x, Y) = (1-\delta_1)x - (1-\delta_2)E(Y|Y>0),\ \ \sigma_1^2 = \text{var}(h_{01}(X)),$$

$$h_{02}(y) = Eh_0(X, y) = (1-\delta_1)E(X|X>0) - (1-\delta_2)y,\ \sigma_2^2 = \text{var}(h_{02}(Y)).$$

We now establish the Wilks' theorem, i.e., an asymptotic chi-squared distribution of the log-likelihood ratio statistics of the proposed method, and show how to construct confidence intervals for the mean difference θ.

Theorem 3.1. *Assume that* $0 < \delta_1 < 1, 0 < \delta_2 < 1$. *Also let* $E{h_0}^2(X, Y) < \infty, \sigma_1^2 > 0, \sigma_2^2 > 0$, *and* $m/n \to c, 0 < c < \infty$ *as* $m, n \to \infty$. *Then for the true value* θ_0 *of the mean difference* θ,

$$-2logR(\theta_0) \xrightarrow{\mathfrak{D}} \chi_1^2 \text{ as } t \to \infty.$$

Based on Theorem 3.1, the asymptotic $100(1-\alpha)\%$ jackknife empirical likelihood confidence interval for θ_0 is given by

$$I_1(\alpha) = \{\theta : -2\text{log}R(\theta) \le \chi_1^2(\alpha)\},$$

where $\chi_1^2(\alpha)$ is the $(1-\alpha)$ quantile of a chi-square distribution with one degree of freedom.

3.2.3 Adjusted jackknife empirical likelihood procedure

Motivated by Chen et al. (2008), we add one more data point to the jackknife pseudo-values. We employ the adjusted jackknife empirical method by applying the empirical likelihood method to all of these values. Define $g_i = g_i(\theta) = \hat{V}_i - \theta, \quad i = 1, 2, \ldots, t$, and $\bar{g}_t = \bar{g}_t(\theta) = t^{-1}\sum_{i=1}^t g_i(\theta)$ for any fixed θ. For some $a_t > 0$, the additional observation is defined as

$$g_{t+1}(\theta) = -a_t\bar{g}_t(\theta) = -\frac{a_t}{t}\sum_{i=1}^t g_i(\theta).$$

Chen et al. (2008) suggested $a_t = \max(1, \log(t)/2)$. Applying empirical likelihood method to these $t+1$ values, we define the adjusted jackknife empirical log-likelihood ratio at θ as follows [cf. Lin et al. (2017)],

$$R^{\star}(\theta) = \sup\left\{\sum_{i=1}^{t+1} \log\{(t+1)p_i)\} : \sum_{i=1}^{t+1} p_i = 1, \sum_{i=1}^{t+1} p_i g_i(\theta) = 0, p_i \geq 0, \right.$$

$$\left. i = 1, 2, \ldots, t+1 \right\}.$$

The adjusted jackknife empirical log-likelihood ratio is

$$\log R^{\star}(\theta) = -\sum_{i=1}^{t+1} \log[1 + \lambda(\theta) g_i(\theta)],$$

where the Lagrange multiplier $\lambda(\theta)$ satisfies the following equation

$$\frac{1}{t+1} \sum_{i=1}^{t+1} \frac{g_i(\theta)}{1 + \lambda(\theta) g_i(\theta)} = 0. \tag{3.3}$$

Now the Wilks' theorem for the AJEL method as follows:

Theorem 3.2. *Assume that* $0 < \delta_1 < 1, 0 < \delta_2 < 1$, $Eh_0{}^2(X,Y) < \infty, \sigma_1^2 > 0, \sigma_2^2 > 0$, $m/n \to c, 0 < c < \infty$ *as* $m, n \to \infty$. *Also let* θ_0 *be the the true value of the mean difference* θ. *Then as* $t \to \infty$,

$$-2 log R^{\star}(\theta_0) \to \chi_1^2 \textit{ in distribution.}$$

Based on Theorem 3.2, we obtain the asymptotic $100(1-\alpha)\%$ AJEL confidence interval for θ_0 as follows,

$$I_2(\alpha) = \{\theta : -2\log R^{\star}(\theta) \leq \chi_1^2(\alpha)\}.$$

3.3 Simulation Study

Simulation studies are carried out to assess the performance of the proposed JEL methods. We compare the jackknife empirical likelihood and adjusted jackknife empirical likelihood confidence intervals with the confidence intervals by the normal approximation (NA) and the EL method (see Zhou and Zhou (2005)) in terms of coverage accuracy and average length of the confidence intervals.

The data from exponential, chi-squared and Poisson distributions are generated for the simulation purposes. We use the binomial distribution to generate zero values with proportion 0.1, 0.2, 0.3 and 0.4. For Table 3.1, the non-zero observations for the first and second samples are generated from an exponential distribution with rate 2 and rate 5, respectively. For Table 3.2, non-zero observations for the first and the second samples are generated from a chi-squared distribution with df = 1 and df = 4, respectively. And the Poisson distributions with mean 1 and mean 2 are used to generate non-zero observations for the first and the second samples, respectively, for Table 3.3. All simulation results are based on 10000 repetitions.

Simulation studies show that the coverage probabilities of the 95% jackknife empirical likelihood (JEL) confidence intervals are better than those of NA and EL confidence intervals proposed by Zhou and Zhou (2005). The AJEL confidence intervals have even better coverage probabilities than JEL confidence intervals, especially for smaller sample sizes, although the AJEL and JEL have longer average length than EL and NA confidence intervals in most cases.

3.4 Application to Real Data

Morrison-Beedy et al. (2011) conducted a large randomized controlled trial to assess the effectiveness of HIV-prevention intervention among urban adolescent females. Adolescent girls from western New York and others (who were interested in that program) were selected for the trial and sexual behavior data were collected. One usual way of collecting these data is retrospective recall method, in which the participants recall their sexual behavior over the study period. But because of possible cognitive (e.g., recollection) biases of this method, a contemporaneous daily diary method can be used, where the participants record their daily sexual activities in a diary. A pilot study was conducted before the primary longitudinal research, and the sexual behavior of adolescence females was collected by contemporaneous daily diary and retrospective recall methods. The dataset became zero-inflated due to no sexual behavior being reported for many females during the relevant time period. We are interested in testing

$$H_0 : \text{mean difference of the sexual behaviors of these two groups} = 0.$$

Each group has 47 observations, of which 13 are zeros giving $\hat{\delta}_1$ and $\hat{\delta}_2$ for both groups are 0.277. The skewnesses for the diary and the recall group are 1.770 and 3.200, respectively. The mean and the standard deviation of diary group are 9.106 and 12.356, respectively. And for the recall group, they are 13.808 and 23.559, respectively. The 95% confidence intervals and the corresponding interval lengths are calculated using the four methods and summarized in Table 3.4.

All the confidence intervals contain zero, meaning no significant difference between the two groups. The normal approximation confidence interval has the shortest interval length, and the adjusted JEL has the longest interval length. Following the conclusion of our simulation study, JEL and AJEL have more accurate coverage accuracy than the other two methods, although JEL and AJEL have longer interval lengths.

The second dataset is collected from Neuh*ä*user (2012), which is originally taken from Siegmund et al. (2004). This is a study of methylation pattern measurement on cancer cells. Hypermethylation can serve as a biomarker of cancer, i.e., the presence of methylation may indicate the possible cancer cell. The absence or partial presence of methylation gives negative results, which are considered as zero values. Substantial presence of methylation gives positive values, which are often highly skewed.

The dataset has the methylation test results from two groups: small lung cancer cells and non-small lung cancer cells. The small lung cancer cell group has 41 observations, out of which 25 are negative/zero values giving $\hat{\delta}_1 = 0.610$. The second group has 46 observations, and 16 of them are negative/zeros giving $\hat{\delta}_2 = 0.348$. The skewnesses for these groups are 5.089 and 1.681, respectively. The mean and standard deviation for the first group are 5.339 and 20.169, respectively. And those for the second groups are 19.641 and 31.892, respectively.

We are interested in testing

$$H_0 : \text{methylation mean difference of small and non-small lung cancer cells } = 0.$$

All four inference methods are applied to this data set. The confidence intervals and the corresponding interval lengths are given in Table 3.5.

The 95% confidence intervals by all methods for the methylation pattern dataset do not contain zero, which shows there is a significant difference between the two groups. The normal approximation confidence interval has shortest length; whereas, the adjusted jackknife empirical likelihood has the longest length. However, as our simulation studies suggest, we can rely more on jackknife empirical likelihood and adjusted jackknife empirical likelihood confidence intervals because of their better coverage probabilities.

3.5 Conclusion

Zero-inflated skewness is a natural pattern in many studies. Because of the essence of the zero values, it is important to incorporate them into the statistical analysis. Several parametric and semi-parametric methods have been developed to deal with zero-inflated skewed data under the assumption of the population distribution. This assumption may restrict the wide use of those techniques. One feasible alternative nonparametric method is empirical likelihood proposed by Zhou and Zhou (2005), but it can be computationally intensive. Considering these difficulties, we proposed jackknife empirical likelihood and adjusted jackknife empirical likelihood methods. Our JEL methods are computationally simple, and simulation studies demonstrate that they have better coverage probabilities than those of normal approximation and empirical likelihood confidence intervals.

The innovative contribution of this chapter is the new formulation of JEL in this setting, which is beyond the standard U-statistic. The derivation of the Wilks' theorem for the JEL is more challenging than the standard setting like U-statistic. The researcher can apply the techniques developed in this chapter to other more complicated statistical problems and makes EL inference with better performance for the small sample size. Two real-life datasets analyses show the applicability of these methods in practical situations. Future studies of the proposed methods may involve a further calibration such as the Bartlett correction. In the future, we use these methods to compare the means of $k \geq 3$ zero-inflated populations. We will also apply mean empirical likelihood proposed by Liang et al. (2019) to the mean difference and construct JEL confidence intervals to achieve better performance.

Table (3.1) Coverage probabilities (average lengths) with confidence level 0.95 when the skewed data follow the exponential distribution.

m	n	Method	$\delta_1 = 0.1$ $\delta_2 = 0.1$	$\delta_1 = 0.2$ $\delta_2 = 0.2$	$\delta_1 = 0.3$ $\delta_2 = 0.3$	$\delta_1 = 0.4$ $\delta_2 = 0.4$	$\delta_1 = 0.2$ $\delta_2 = 0.3$
20	20	AJEL	0.928 (0.484)	0.913 (0.455)	0.904 (0.429)	0.879 (0.396)	0.914 (0.453)
		JEL	0.915 (0.457)	0.896 (0.427)	0.881 (0.400)	0.854 (0.366)	0.896 (0.424)
		EL	0.879 (0.402)	0.886 (0.417)	0.872 (0.366)	0.860 (0.352)	0.881 (0.396)
		NA	0.876 (0.382)	0.873 (0.375)	0.876 (0.367)	0.861 (0.351)	0.873 (0.375)
50	50	AJEL	0.937 (0.298)	0.929 (0.282)	0.910 (0.266)	0.890 (0.246)	0.920 (0.279)
		JEL	0.930 (0.290)	0.921 (0.274)	0.900 (0.257)	0.877 (0.237)	0.911 (0.271)
		EL	0.904 (0.254)	0.893 (0.248)	0.886 (0.241)	0.880 (0.215)	0.894 (0.262)
		NA	0.886 (0.246)	0.890 (0.242)	0.885 (0.236)	0.877 (0.226)	0.883 (0.241)
70	70	AJEL	0.936 (0.250)	0.927 (0.236)	0.919 (0.222)	0.898 (0.207)	0.922 (0.234)
		JEL	0.931 (0.245)	0.921 (0.231)	0.912 (0.216)	0.888 (0.201)	0.916 (0.229)
		EL	0.909 (0.227)	0.895 (0.221)	0.902 (0.207)	0.887 (0.180)	0.899 (0.212)
		NA	0.889 (0.209)	0.892 (0.205)	0.894 (0.200)	0.884 (0.192)	0.891 (0.204)
20	30	AJEL	0.922 (0.474)	0.918 (0.449)	0.903 (0.423)	0.885 (0.389)	0.915 (0.444)
		JEL	0.911 (0.453)	0.903 (0.426)	0.891 (0.399)	0.863 (0.364)	0.901 (0.420)
		EL	0.897 (0.398)	0.883 (0.405)	0.889 (0.396)	0.860 (0.354)	0.878 (0.389)
		NA	0.860 (0.373)	0.870 (0.368)	0.868 (0.359)	0.858 (0.342)	0.868 (0.366)

NOTE:
NA: normal approximation
EL: empirical likelihood
AEL: adjusted empirical likelihood
AJEL: adjusted jackknife empirical likelihood

Table (3.2) Coverage probabilities (average lengths) with confidence level 0.95 when the skewed data follow the chi-squared distribution.

m	n	Method	$\delta_1 = 0.1$ $\delta_2 = 0.1$	$\delta_1 = 0.2$ $\delta_2 = 0.2$	$\delta_1 = 0.3$ $\delta_2 = 0.3$	$\delta_1 = 0.4$ $\delta_2 = 0.4$	$\delta_1 = 0.2$ $\delta_2 = 0.3$
20	20	AJEL	0.928 (2.873)	0.906 (2.722)	0.880 (2.549)	0.867 (2.365)	0.891 (2.259)
		JEL	0.909 (2.714)	0.885 (2.556)	0.888 (2.373)	0.857 (2.181)	0.889 (2.428)
		EL	0.884 (2.412)	0.882 (2.267)	0.872 (2.354)	0.879 (2.302)	0.889 (2.401)
		NA	0.883 (2.343)	0.882 (2.374)	0.874 (2.352)	0.885 (2.296)	0.888 (2.375)
50	50	AJEL	0.930 (1.754)	0.914 (1.669)	0.892 (1.568)	0.888 (1.454)	0.899 (1.584)
		JEL	0.924 (1.708)	0.915 (1.621)	0.890 (1.517)	0.878 (1.404)	0.889 (1.536)
		EL	0.890 (1.624)	0.899 (1.613)	0.891 (1.511)	0.880 (1.419)	0.884 (1.520)
		NA	0.893 (1.495)	0.895 (1.517)	0.890 (1.507)	0.891 (1.470)	0.891 (1.512)
70	70	AJEL	0.929 (1.470)	0.914 (1.391)	0.898 (1.304)	0.888 (1.215)	0.891 (1.323)
		JEL	0.922 (1.437)	0.907 (1.360)	0.889 (1.272)	0.876 (1.181)	0.890 (1.292)
		EL	0.889 (1.321)	0.876 (1.278)	0.882 (1.275)	0.871 (1.161)	0.889 (1.291)
		NA	0.888 (1.266)	0.898 (1.282)	0.887 (1.272)	0.890 (1.243)	0.891 (1.280)
20	30	AJEL	0.933 (2.433)	0.911 (2.305)	0.892 (2.170)	0.878 (2.010)	0.886 (2.207)
		JEL	0.919 (2.321)	0.896 (2.188)	0.891 (2.102)	0.864 (1.880)	0.884 (2.087)
		EL	0.899 (2.201)	0.890 (2.102)	0.887 (2.091)	0.873 (1.901)	0.879 (2.001)
		NA	0.897 (2.007)	0.894 (2.024)	0.889 (2.014)	0.887 (1.957)	0.883 (2.026)

NOTE:
NA: normal approximation
EL: empirical likelihood
AEL: adjusted empirical likelihood
AJEL: adjusted jackknife empirical likelihood

Table (3.3) Coverage probabilities (average lengths) with confidence level 0.95 when the skewed data follow the Poisson distribution.

m	n	Method	$\delta_1 = 0.1$ $\delta_2 = 0.1$	$\delta_1 = 0.2$ $\delta_2 = 0.2$	$\delta_1 = 0.3$ $\delta_2 = 0.3$	$\delta_1 = 0.4$ $\delta_2 = 0.4$	$\delta_1 = 0.2$ $\delta_2 = 0.3$
20	20	AJEL	0.940 (1.551)	0.921 (1.476)	0.906 (1.388)	0.885 (1.297)	0.906 (1.413)
		JEL	0.924 (1.465)	0.902 (1.385)	0.878 (1.293)	0.853 (1.196)	0.887 (1.322)
		EL	0.893 (1.302)	0.876 (1.289)	0.866 (1.214)	0.852 (1.148)	0.870 (1.264)
		NA	0.877 (1.237)	0.872 (1.228)	0.864 (1.199)	0.855 (1.155)	0.863 (1.208)
50	50	AJEL	0.937 (0.947)	0.915 (0.897)	0.902 (0.841)	0.884 (0.784)	0.904 (0.861)
		JEL	0.932 (0.922)	0.906 (0.871)	0.891 (0.814)	0.872 (0.755)	0.894 (0.835)
		EL	0.903 (0.883)	0.883 (0.819)	0.879 (0.780)	0.870 (0.739)	0.886 (0.803)
		NA	0.894 (0.801)	0.879 (0.793)	0.876 (0.774)	0.873 (0.747)	0.878 (0.784)
70	70	AJEL	0.936 (0.794)	0.917 (0.749)	0.899 (0.705)	0.881 (0.655)	0.898 (0.719)
		JEL	0.929 (0.778)	0.910 (0.733)	0.892 (0.688)	0.871 (0.637)	0.890 (0.702)
		EL	0.911 (0.753)	0.889 (0.701)	0.884 (0.660)	0.869 (0.628)	0.879 (0.678)
		NA	0.889 (0.681)	0.884 (0.672)	0.880 (0.658)	0.871 (0.633)	0.872 (0.664)
20	30	AJEL	0.940 (1.350)	0.921 (1.280)	0.908 (1.204)	0.885 (1.123)	0.912 (1.240)
		JEL	0.927 (1.288)	0.906 (1.215)	0.888 (1.136)	0.863 (1.050)	0.894 (1.172)
		EL	0.907 (1.173)	0.879 (1.103)	0.870 (1.061)	0.860 (1.001)	0.877 (1.143)
		NA	0.881 (1.085)	0.874 (1.072)	0.867 (1.045)	0.861 (1.006)	0.869 (1.061)

NOTE:
NA: normal approximation
EL: empirical likelihood
AEL: adjusted empirical likelihood
AJEL: adjusted jackknife empirical likelihood

Table (3.4) Confidence intervals and lengths with 95% confidence level for the sexual health study data.

Method	Lower Limit	Upper Limit	Length
AJEL	-14.357	1.658	16.015
JEL	-14.043	1.448	15.491
EL	-14.202	1.798	16.000
NA	-11.096	1.692	12.788

NOTE:
NA: normal approximation
EL: empirical likelihood
AEL: adjusted empirical likelihood
AJEL: adjusted jackknife empirical likelihood

Table (3.5) Confidence intervals and lengths with 95% confidence level for the methylation pattern measurement data.

Method	Lower Limit	Upper Limit	Length
AJEL	-25.452	-2.277	23.175
JEL	-24.948	-2.817	22.131
EL	-25.102	-3.141	21.961
NA	-23.676	-4.929	18.747

NOTE:
NA: normal approximation
EL: empirical likelihood
AEL: adjusted empirical likelihood
AJEL: adjusted jackknife empirical likelihood

CHAPTER 4

JACKKNIFE EMPIRICAL LIKELIHOOD FOR THE QUANTILE DIFFERENCE OF A ZERO-INFLATED POPULATION

4.1 Background

The inference problem that we are addressing in this chapter has two key features: one-sample quantile difference and zero-inflated data. A quantile is a useful and essential statistical measure that gives a robust and meaningful summary statistics of the location of the data. On the other hand, the one-sample quantile difference provides a descriptive measure of the spread of the data, especially if the distribution of the population is asymmetric. One popular quantile difference is the interquartile range, which captures the middle 50% of the data - enclosed by first and third quartiles. In risk management, quantile difference is used as a value-at-risk measure for an investment risk. The quantile difference is also used in risk analysis, reliability analysis, lifetime data analysis, etc. The inference problem of quantile difference becomes challenging if the data is zero-inflated.

Zero-inflated data naturally arises in many fields, such as epidemiology, astronomy, ecology, biology, insurance, meteorology, engineering, psychology, etc. A significant proportion in such a dataset are zeros, and non-zero values are often highly positively skewed. For example, one may observe zero-inflation in health insurance data, where there may be no claims for many insurers. These zero values are “true zeros” and an integral part of the data. Also, there might be some very high insurance claims. In addictive behavior research, there might be many non-substance use reports and some high drug abuse cases. Since the zero values are correct and valid, any inference without using these zero values might give a misleading result.

Owen (1988, 1990) developed the empirical likelihood (EL) method for the population mean. The empirical likelihood is a nonparametric technique, i.e., no parametric assumption about the distribution is needed. Yet, it preserves some critical parametric properties. For example, the empirical log-likelihood ratio for the population mean follows the standard chi-squared distribution. The shape and orientation of the confidence intervals are determined by the data only. Also, the confidence intervals are transformation invariant and range preserving. Qin and Lawless (1994) combined EL with an estimating equation. Owen (2001) has a more general overview of the EL method. However, the empirical likelihood method loses its efficiency when there are some nonlinear constraints involved. Solving these nonlinear constraints makes the EL method very computationally costly. To alleviate this computation burden, Jing et al. (2009) proposed the jackknife empirical likelihood (JEL) method. The JEL method avoids solving complex nonlinear constraints by transforming the statistics of interest into the mean of jackknife pseudo-values. the simplicity and computation efficiency, the JEL method has been used in many statistics fields.

There exist many parametric and nonparametric inference methods for zero-inflated data. A two-part model is introduced by Duan et al. (1983) under the lognormal distribution. Zhou and Tu (2000) assumed a lognormal distribution for non-zero values to find study diagnostic test charge data. Hall (2000) did a case study with zero-inflated Poisson and binomial regression. Hasan and Krishnamoorthy (2018) considered the parametric confidence intervals technique for the mean and percentile for zero-inflated lognormal data. Since using the parametric method requires the justification of the distribution assumption, the non-parametric EL method is also used for zero-inflated data. See Chen and Qin (2003), Chen et al. (2003), Zhou and Zhou (2005), Pailden and Chen (2013), Satter and Zhao (2020), Satter and Zhao (2021), among others.

Empirical likelihood performance can be affected by two factors: convex hull problem and under-coverage issue. The adjusted empirical likelihood (AEL) method is first proposed by Chen et al. (2008) to address the EL convex hull problem by adding one extra pseudo-value. This is very easy to implement and gives a better result than the EL method. Emerson and Owen (2009) and Liu and Chen (2010) also tried to tackle the convex hull problem by adding artificial data points. The under-coverage issue can be mitigated by Bartlett correction (Diciccio et al. (1991), Chen and Cui (2007)). However, Barlett correction is a complex computational operation. Jing et al. (2017) investigate another way to improve coverage probability by transforming the empirical likelihood ratio by a simple transformation. Stewart and Ning (2020) studied several versions of EL to construct confidence intervals for the mean of a zero-inflated population.

The sample quantile is not an efficient estimator [cf. Falk (1984)]. To improve the efficiency, Yang (1985) gave a nonparametric smooth kernel estimator and established the asymptotic normality. Sheather and Marron (1990) proposed the kernel quantile estimators to improve the efficiency. Chen and Hall (1993) first proposed a smoothed empirical likelihood (EL) method for the quantile inference. To avoid solving nonlinear equations, Adimari (1996) presented a simpler version of an EL method for quantiles, which Zhou and Jing (2003a) improve with kernel density function. Zhou and Jing (2003b) later extended the smooth EL for the inference of the quantile difference. As the EL method can be challenging for nonlinear constraints, Yang and Zhao (2018) developed the JEL method with a kernel smooth estimator of quantile difference. Quantile difference for two independent populations is also proposed by Yang and Zhao (2016). To the best of our knowledge, there is no inference procedure for the quantile difference of zero-inflated data.

In this chapter, we propose the JEL method to construct a confidence interval for the quantile difference of zero-inflated data. To further improve the performance, we also present the adjusted jackknife empirical likelihood (AJEL), transformed jackknife empirical likelihood (TJEL), and transformed the adjusted jackknife empirical likelihood (TAJEL) method.

The rest of the chapter is organized as follows. Jackknife empirical likelihood, adjusted jackknife empirical likelihood, transformed jackknife empirical likelihood, and adjusted transformed jackknife empirical likelihood methods for the quantile difference of one sample are proposed in Section 4.2. Simulation studies are performed in Section 4.3. We apply the proposed methods to a real dataset to illustrate the application to a real-world problem in Section 4.4. Conclusion and some related remarks are added in Section 4.5. Finally, proofs and corresponding lemmas are provided in the Appendix C.

4.2 Main Results

4.2.1 Jackknife empirical likelihood method

Let $X_1, X_2, \ldots, X_n$ be a random sample of X from a zero -inflated skewed population. Let $\delta = P(X = 0) > 0$ be the probability of having zero values of the population. Suppose that there are n_1 positive values in X_i's, and denote these values as $x_1, x_2, \ldots, x_{n_1}$. Then, there are $n_0 = n - n_1$ zero values in the sample. We denote $\hat{\delta} = n_0/n$.

Let the positive values $x_1, x_2, \ldots, x_{n_1}$ come from a distribution with distribution function $F(x)$. For any $p \in (0, 1)$ and $y \in \Re$, we can define p-th quantile function as $F^{-1}(p) = \inf\{y : F(y) \geq p\}$. For zero-inflated data, the distribution function is defined as

$$G(x;\delta) = \begin{cases} \delta & \text{if} \quad x = 0 \\ \delta + (1-\delta)F(x) & \text{if} \quad x > 0. \end{cases} \tag{4.1}$$

The p-th quantile, denoted by q_p, can be obtained by the relation $G(q_p; \delta) = p$. Then from equation (4.1), we get

$$q_p = \begin{cases} 0 & \text{if} \quad p < \delta \\ F^{-1}\left(\frac{p-\delta}{1-\delta}\right) & \text{if} \quad p > \delta. \end{cases} \tag{4.2}$$

The focus of this chapter is to construct a confidence interval for

$$\begin{aligned} \theta(s,t) &= G^{-1}(t;\delta) - G^{-1}(s;\delta) \\ &= F^{-1}\left(\frac{t-\delta}{1-\delta}\right) - F^{-1}\left(\frac{s-\delta}{1-\delta}\right), \end{aligned} \tag{4.3}$$

where $\delta < s < t < 1$. Define the empirical estimator of F as $F_{n_1}(x) = 1/n_1 \sum_{i=1}^{n_1} I(x_i \leq x)$.

By using the empirical estimator of F^{-1} and δ, we can estimate $\theta(s,t)$ as

$$\hat{\theta} = \hat{\theta}(s,t) = F_{n_1}^{-1}\left(\frac{t-\hat{\delta}}{1-\hat{\delta}}\right) - F_{n_1}^{-1}\left(\frac{s-\hat{\delta}}{1-\hat{\delta}}\right),$$

where $F_{n_1}^{-1}(\cdot)$ is the estimator of F^{-1}.

Let w be a symmetric density function and $K(x)$ be the smooth distribution function with $K(x) = \int_{u\le x} w(u)du$. For a positive bandwidth h, we consider the smoothed estimating equation for θ as

$$T_{n_1}(\theta; s,t) = \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i-\theta)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}}.$$

After deleting j-th observation from n_1 positive values, we have

$$T_{n_1}^{-j}(\theta; s,t) = \frac{1}{n_1-1}\sum_{i=1, i\neq j}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}^{-j}(x_i-\theta)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}}, \quad 1\le j\le n_1,$$

where

$$F_{n_1}^{-j}(y) = \frac{1}{n_1-1}\sum_{1\le i\le n_1, i\neq j} I(x_i\le y), i=1,\ldots,n_1.$$

We can define the jackknife pseudo-values as

$$\hat{V}_j(\theta; s,t) = n_1 T_{n_1}(\theta; s,t) - (n_1-1)T_{n_1}^{-j}(\theta; s,t), j=1,\ldots,n_1.$$

The jackknife empirical likelihood ratio at the parameter θ is given by

$$R(\theta; s,t) = \sup\left\{\prod_{i=1}^{n_1} n_1 p_i : \sum_{i=1}^{n_1} p_i = 1, \sum_{i=1}^{n_1} p_i \hat{V}_i(\theta; s,t) = 0, p_i\ge 0, i=1,\ldots,n_1\right\}. \quad (4.4)$$

By using the Lagrange multiplier method, the jackknife empirical log-likelihood ratio

$$l(\theta; s,t) \quad = \quad \log R(\theta; s,t)$$

$$= \quad -\sum_{i=1}^{n_1} \log[1+\lambda(\theta;s,t)\hat{V}_i(\theta;s,t)],$$

where $\lambda(\theta; s,t)$ satisfies the following equation

$$\frac{1}{n_1}\sum_{i=1}^{n_1} \frac{\hat{V}_i(\theta;s,t)}{1+\lambda(\theta;s,t)\hat{V}_i(\theta;s,t)} = 0.$$

We establish Wilks' theorem for the jackknife empirical log-likelihood ratio as follows.

Theorem 4.1. *Assume that* $0 < \delta < 1$ *and the regularity conditions* $C1 - C4$ *in the Appendix C hold. Then for the true value* θ_0 *of the quantile difference* θ*, as* $n \to \infty$

$$-2l(\theta_0; s,t) \xrightarrow{\mathfrak{D}} \chi_1^2.$$

Based on Theorem 4.1, the asymptotic $100(1-\alpha)\%$ jackknife empirical likelihood confidence interval for θ_0 is given by

$$I_1(\alpha) = \{\theta : -2l(\theta; s,t) \leq \chi_1^2(\alpha)\},$$

where $\chi_1^2(\alpha)$ is the $(1-\alpha)$ quantile of a chi-square distribution with one degree of freedom.

4.2.2 Adjusted jackknife empirical likelihood method

To deal with the zero convex hull problem encountered in the empirical likelihood, Chen et al. (2008) proposed the adjusted empirical likelihood (AEL) method. This AEL method involves adding one more data point to $\hat{V}_i(\theta; s,t)$. Let $u_i(\theta; s,t) = \hat{V}i(\theta; s,t), i = 1, 2, \ldots, n_1$. Then, the extra data point is generated by

$$u_{n_1+1}(\theta; s,t) = -a_{n_1}\sum_{i=1}^{n_1} \frac{u_i(\theta;s,t)}{n_1},$$

where a_{n_1} is a positive constant. The general recommendation is to choose a_{n_1} as $\max(1, \log(n_1)/2)$. Now with the $n_1 + 1$ data points, the adjusted jackknife empirical likelihood (AJEL) ratio is defined as follows,

$$R^a(\theta;s,t) = \sup\left\{\prod_{i=1}^{n_1+1}(n_1+1)p_i : \sum_{i=1}^{n_1+1} p_i = 1, \sum_{i=1}^{n_1} p_i u_i(\theta;s,t) = 0, p_i \geq 0, \right.$$
$$\left. i = 1,\ldots,n_1+1\right\}. \quad (4.5)$$

By using the Lagrange multiplier method, we have the adjusted jackknife empirical log-likelihood ratio at θ as

$$\begin{aligned} l^a(\theta;s,t) &= \log R^a(\theta;s,t) \\ &= -\sum_{i=1}^{n_1+1}\log[1+\lambda^a(\theta;s,t)u_i(\theta;s,t)], \end{aligned}$$

where $\lambda^a(\theta;s,t)$ satisfies the following equation

$$\frac{1}{n_1+1}\sum_{i=1}^{n_1+1}\frac{u_i(\theta;s,t)}{1+\lambda^a(\theta;s,t)u_i(\theta;s,t)} = 0.$$

We establish Wilks' theorem for the adjusted jackknife empirical log-likelihood ratio as follows.

Theorem 4.2. *Assume that* $0 < \delta < 1$ *and the regularity conditions* $C1-C4$ *in the Appendix C hold. Then for the true value* θ_0 *of the quantile difference* θ*, as* $n \to \infty$

$$-2l^a(\theta_0;s,t) \xrightarrow{\mathfrak{D}} \chi_1^2.$$

Based on Theorem 4.2, the asymptotic $100(1-\alpha)\%$ jackknife empirical likelihood confidence interval for θ_0 is given by

$$I_2(\alpha) = \{\theta : -2l^a(\theta;s,t) \leq \chi_1^2(\alpha)\}.$$

4.2.3 Transformed jackknife empirical likelihood method

Instead of dealing with the under-coverage problem in empirical likelihood either by increasing the rate of convergence through the Bartlett correction or by fixing convex hull constraint problem, Jing et al. (2017) proposed a simple transformation based EL method. Here we apply transformed empirical likelihood method (TEL) to the jackknife pseudo-values and develop transformed jackknife empirical likelihood (TJEL) method. Define

$$w(l(\theta; s, t, \gamma)) = l(\theta; s, t) * \max(1 - l(\theta; s, t)/n_1, 1 - \gamma),$$

where $\gamma \in [0, 1]$ is a constant and $w(l(\theta; s, t))$ is a truncated quadratic transformation of $l(\theta; s, t)$. Following the recommendation by Jing et al. (2017), we set $\gamma = 0.5$. Then the transformed jackknife empirical log-likelihood ratio at the parameter θ is

$$\begin{aligned} l^{tr}(\theta; s, t) &= w(l(\theta; s, t, \gamma = 0.5)) \\ &= l(\theta; s, t) * \max(1 - l(\theta; s, t)/n_1, 0.5), \end{aligned}$$

which can be written explicitly as

$$l^{tr}(\theta; s, t) = \begin{cases} l(\theta; s, t)[1 - l(\theta; s, t)/n_1] & \text{if} \quad l(\theta; s, t) \leq n_1/2 \\ l(\theta; s, t)/2 & \text{if} \quad l(\theta; s, t) > n_1/2. \end{cases}$$

The TJEL has the same asymptotic properties as JEL [cf. Jing et al. (2017)]. Therefore, the Wilks' theorem for the transformed adjusted empirical log-likelihood ratio is as follows.

Theorem 4.3. *Assume that* $0 < \delta < 1$ *and the regularity conditions* $C1-C4$ *in the Appendix C hold. Then for the true value* θ_0 *of the quantile difference* θ*, as* $n \to \infty$

$$-2l^{tr}(\theta_0; s, t) \xrightarrow{\mathfrak{D}} \chi^2_1.$$

Based on Theorem 4.3, the asymptotic $100(1-\alpha)\%$ transformed jackknife empirical likelihood confidence interval for θ_0 is given by

$$I_3(\alpha) = \{\theta : -2l^{tr}(\theta; s, t) \leq \chi_1^2(\alpha)\}.$$

4.2.4 Transformed adjusted jackknife empirical likelihood method

Here we combine the TEL and AEL method and apply this to the jackknife pseudo-values to develop transformed adjusted jackknife empirical likelihood (TAJEL) method.

Recall that we obtain the adjusted jackknife empirical log-likelihood ratio $l^a(\theta; s, t)$. Now, for a constant $\gamma \in [0, 1]$, define

$$w^{ta}(l^a(\theta; s, t, \gamma)) = l^a(\theta; s, t) * \max(1 - l^a(\theta; s, t)/(n_1 + 1), 1 - \gamma),$$

where $w^{ta}(l^a(\theta; s, t, \gamma))$ is a truncated quadratic transformation of $l^a(\theta; s, t)$. Here, we choose $\gamma = 0.5$. Thus, the TAJEL ratio is

$$\begin{aligned} l^{ta}(\theta; s, t) &= w(l^a(\theta; s, t, \gamma = 0.5)) \\ &= l^a(\theta; s, t) * \max(1 - l^a(\theta; s, t)/(n_1 + 1), 0.5), \end{aligned}$$

which can be written explicitly as

$$l^{ta}(\theta; s, t) = \begin{cases} l^a(\theta; s, t)[1 - l^a(\theta; s, t)/(n_1 + 1)] & \text{if} \quad l^a(\theta; s, t) \leq (n_1 + 1)/2 \\ l^a(\theta; s, t)/2 & \text{if} \quad l^a(\theta; s, t) > (n_1 + 1)/2. \end{cases}$$

The Wilks' theorem for the transformed adjusted empirical log-likelihood ratio is as follows.

Theorem 4.4. *Assume that $0 < \delta < 1$ and the regularity conditions $C1 - C4$ in the Appendix C hold. Then for the true value θ_0 of the quantile difference θ, as $n \to \infty$*

$$-2l^{ta}(\theta_0; s, t) \xrightarrow{\mathfrak{D}} \chi_1^2.$$

Based on Theorem 4.4, the asymptotic $100(1 - \alpha)\%$ transformed adjusted jackknife empirical likelihood confidence interval for θ_0 is given by

$$I_4(\alpha) = \{\theta : -2l^{ta}(\theta; s, t) \leq \chi_1^2(\alpha)\}.$$

4.3 Simulation Study

Three simulation studies are carried out to assess the performance of the proposed JEL methods. We compare the coverage probabilities and average lengths obtained by the proposed methods under various simulation settings. We choose three different proportions of zero values $0.1, 0.2$, and 0.3, and the non-zero values are generated from a chi-square distribution with one degree of freedom, exponential distribution with mean 2, and log-normal distribution with zero mean and one unit standard deviation. Three tables are constructed to present the simulation result for each distribution. We consider random samples of sizes $30, 50$, and 70. The coverage probabilities and average lengths of the confidence intervals are calculated with 2000 repetitions.

We use the biweight kernel function for all simulation studies, which is defined as

$$w(u) = \begin{cases} \frac{15}{16}(1-u^2)^2 & \text{if } |u| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Bandwidth selection is an important aspect of using a kernel function. Here, we implemented the cross-validation method from Bowman and Azzalini (1997) to find an optimal bandwidth for the kernel function. R package ‘sm’ (V2.2-5.6; Bowman and Azzalini (2018)) is used for bandwidth selection, and all the simulations are performed in R (R Core Team (2020)).

Following Zhou and Jing (2003a), and using eqns. (4.1) - (4.3), we can write

$$\sqrt{n_1}(\hat{\theta}(s,t) - \theta_0(s,t)) \xrightarrow{\mathfrak{D}} N(0, \sigma^2(\theta_0, s, t)),$$

where

$$
\begin{aligned}
\sigma^2(\theta_0, s, t) &= \frac{\frac{s-\delta}{1-\delta}\left(1-\frac{s-\delta}{1-\delta}\right)}{(1-\delta)^2 f^2\left\{F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right\}} - \frac{2\left(\frac{s-\delta}{1-\delta}\right)\left(1-\frac{t-\delta}{1-\delta}\right)}{(1-\delta)^2 f\left\{F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right\} f\left\{F^{-1}\left(\frac{t-\delta}{1-\delta}\right)\right\}} \\
&\quad + \frac{\frac{t-\delta}{1-\delta}\left(1-\frac{t-\delta}{1-\delta}\right)}{(1-\delta)^2 f^2\left\{F^{-1}\left(\frac{t-\delta}{1-\delta}\right)\right\}},
\end{aligned}
$$

where f is the density function of F. Note that the asymptotic distribution of $\hat{\theta}(s,t)$ is not rigorously derived. This is remained as future research.

An estimate of $\sigma^2(\theta, s, t))$, denoted by $\hat{\sigma}^2(\hat{\theta}, s, t))$ is obtained by

$$
\begin{aligned}
\hat{\sigma}^2(\hat{\theta}, s, t) &= \frac{\frac{s-\hat{\delta}}{1-\hat{\delta}}\left(1-\frac{s-\hat{\delta}}{1-\hat{\delta}}\right)}{(1-\hat{\delta})^2 \hat{f}^2\left\{F_{n_1}^{-1}\left(\frac{s-\hat{\delta}}{1-\hat{\delta}}\right)\right\}} - \frac{2\left(\frac{s-\hat{\delta}}{1-\hat{\delta}}\right)\left(1-\frac{t-\hat{\delta}}{1-\hat{\delta}}\right)}{(1-\hat{\delta})^2 \hat{f}\left\{F_{n_1}^{-1}\left(\frac{s-\hat{\delta}}{1-\hat{\delta}}\right)\right\} \hat{f}\left\{F_{n_1}^{-1}\left(\frac{t-\hat{\delta}}{1-\hat{\delta}}\right)\right\}} \\
&\quad + \frac{\frac{t-\hat{\delta}}{1-\hat{\delta}}\left(1-\frac{t-\hat{\delta}}{1-\hat{\delta}}\right)}{(1-\hat{\delta})^2 \hat{f}^2\left\{F_{n_1}^{-1}\left(\frac{t-\hat{\delta}}{1-\hat{\delta}}\right)\right\}},
\end{aligned}
$$

where $\hat{f}$ is the estimate of f by the kernel smoothing function. Then the asymptotic $100(1-\alpha)\%$ confidence interval for θ_0 is obtained by

$$
(\hat{\theta} \pm z_{\alpha/2}\ \hat{\sigma}(\hat{\theta}, s, t)),
$$

where $z_{\alpha/2}$ is the upper $\alpha/2$-quantile of $N(0,1)$.

Table 4.1 tabulates the first simulation study where non-zero values are generated from a chi-square distribution with one degree of freedom. Five methods, namely, normal approximation (NA), jackknife empirical likelihood (JEL), adjusted jackknife empirical likelihood (AJEL), transformed jackknife empirical likelihood (TJEL), and transformed adjusted jackknife empirical likelihood (TAJEL), are implemented. The simulation results are compared in terms of coverage probabilities and the average length of the 95% confidence intervals. It is evident from the simulation study that the NA method has an under-coverage problem, and the JEL methods give better coverage probability, especially for small sample sizes. As the sample size increases, the coverage probabilities are get closer to the nominal level. By adding one more data point to deal with the convex hull problem, the AJEL method performs better than the JEL method in terms of coverage probability. TJEL method, which uses a transformation of the log-likelihood ratio to deal with the under-coverage problem, gives the similar performance as the AJEL method does. Finally, TAJEL provides the largest coverage probabilities among all methods, but sometimes it may incur over-coverage issues. In general, the average lengths for JEL and AJEL methods are longer than the average length for NA method. The average length for TJEL is somewhat close to the AJEL method. And TAJEL gives the largest average length. As the sample size increases, the average lengths consistently get shorter. There are over-coverage issue when proportion of zeros are high and the quantile difference is small. In such cases, TJEL or TAJEL perform better. Another finding from the simulation study is that if the smaller quantile is close to the proportion of zeros in the data, the results might not be stable.

Table 4.2 and Table 4.3 give simulation results for exponential and log-normal distributions, respectively. The results for both tables are similar to the results in Table 4.1. The only exception is when $\delta = 0.3$, and $s = 0.50$ and $t = 0.70$. In this setting, the NA method gives better coverage rate than the JEL method for the exponential distribution. The TJEL method gives the best result in such a case. On the other hand, the NA method shows over-coverage issue for the log-normal distribution. TJEL and TAJEL provide better results in such a simulation setting for the log-normal distribution.

4.4 Real Data Analysis

Siegmund et al. (2004) studied the methylation pattern measurement on cancer cells. Methylation in DNA happens when the methyl group is added to the DNA. Methylation regulates gene expression and can often turn a healthy cell into a cancer cell. Identifying hypermethylation is a vital diagnostic tool, and early detection of cancer is very crucial for intervention. The diagnosis of methylation gives real numbers $\{x|x \in \Re\}$. When the diagnosis detects no or partial presence of methylation, it provides negative results. These negative test results generally mean undetectable methylation. Therefore, they are considered as zero.

Often there are many zeros in the methylation diagnosis, and some hypermethylated cell can make the data positively skewed.

Neuh$\ddot{a}$user (2012) has a non-small lung cancer cell methylation test result. There are 46 observations in the dataset, and the proportion of zero is $\hat{\delta} = 0.348$. The mean and standard deviation of the data is 19.641 and 31.892, respectively. The skewness is 1.681, i.e., the data is slightly positively skewed.

We applied NA and the proposed methods to this dataset. The 95% confidence intervals and the corresponding confidence lengths are presented in Table 4.4. The result shows the confidence intervals for the quantile difference at various quantiles by the above-mentioned five methods. NA method has the smallest confidence length and TAJEL provides the longest length. This result is expected as our simulation studies suggested. But since the jackknife empirical likelihood methods has better coverage probabilities, we recommend the proposed methods despite the longer confidence length.

4.5 Conclusion

Zero-inflated data is a relatively common phenomenon. Because of the structure of the data, special statistical technique is required for any inferential analysis. On the other hand, the quantile difference is an essential statistical measure. In this chapter, we proposed the jackknife empirical likelihood method to construct a confidence interval for the quantile difference on zero-inflated data. We introduced a smoothed estimating equation and obtain jackknife pseudo-values. We set up a jackknife empirical log-likelihood ratio and establish Wilks' theorem. To improve the accuracy, we also implement adjusted jackknife empirical likelihood, transformed jackknife empirical likelihood, and transformed adjusted jackknife empirical likelihood. Simulation studies confirm the benefits of using our methods. We also applied DNA methylation data to show the applicability of our proposed methods.

There are more research opportunities in this research area. Future research may involve the inference for the quantile difference of two independent or two correlated zero-inflated populations.

Table (4.1) Coverage probabilities (average lengths) of 95% confidence interval with the non-zero data following chi-square distribution.

δ	s	t	n	NA	JEL	AJEL	TJEL	TAJEL
0.1	0.20	0.90	30	0.804 (2.658)	0.898 (3.748)	0.915 (3.864)	0.913 (3.847)	0.935 (3.970)
			50	0.826 (2.294)	0.879 (3.018)	0.892 (3.137)	0.892 (3.115)	0.898 (3.192)
			70	0.864 (2.062)	0.902 (2.579)	0.909 (2.656)	0.908 (2.647)	0.914 (2.699)
	0.25	0.75	30	0.866 (1.571)	0.889 (1.873)	0.906 (2.035)	0.923 (2.364)	0.946 (2.448)
			50	0.892 (1.251)	0.921 (1.421)	0.932 (1.475)	0.932 (1.481)	0.938 (1.524)
			70	0.898 (1.078)	0.946 (1.138)	0.952 (1.169)	0.952 (1.169)	0.955 (1.200)
	0.35	0.65	30	0.879 (0.984)	0.933 (1.148)	0.946 (1.204)	0.949 (1.217)	0.958 (1.292)
			50	0.890 (0.783)	0.936 (0.819)	0.944 (0.855)	0.944 (0.860)	0.958 (0.933)
			70	0.898 (0.661)	0.937 (0.678)	0.947 (0.712)	0.946 (0.711)	0.953 (0.740)
0.2	0.30	0.90	30	0.837 (2.900)	0.914 (3.806)	0.938 (3.934)	0.938 (3.932)	0.958 (4.048)
			50	0.858 (2.537)	0.906 (2.969)	0.918 (3.050)	0.916 (3.048)	0.928 (3.120)
			70	0.882 (2.250)	0.910 (2.548)	0.924 (2.637)	0.922 (2.625)	0.928 (2.683)
	0.40	0.75	30	0.880 (1.595)	0.924 (1.643)	0.943 (1.736)	0.949 (1.979)	0.970 (2.203)
			50	0.907 (1.253)	0.913 (1.164)	0.935 (1.313)	0.937 (1.323)	0.950 (1.370)
			70	0.920 (1.098)	0.928 (0.988)	0.939 (1.058)	0.939 (1.058)	0.944 (1.091)
	0.45	0.70	30	0.896 (1.233)	0.924 (1.246)	0.934 (1.304)	0.940 (1.346)	0.960 (1.600)
			50	0.908 (0.949)	0.921 (0.885)	0.934 (0.931)	0.936 (0.950)	0.950 (1.022)
			70	0.922 (0.819)	0.924 (0.743)	0.934 (0.782)	0.934 (0.783)	0.942 (0.820)
0.3	0.40	0.90	30	0.884 (3.277)	0.914 (3.687)	0.942 (3.812)	0.944 (3.821)	0.964 (3.956)
			50	0.894 (2.857)	0.890 (2.902)	0.906 (3.015)	0.906 (3.041)	0.915 (3.114)
			70	0.919 (2.540)	0.900 (2.488)	0.918 (2.600)	0.915 (2.580)	0.928 (2.663)
	0.45	0.85	30	0.889 (2.942)	0.906 (2.565)	0.931 (2.768)	0.948 (3.692)	0.966 (4.183)
			50	0.920 (2.262)	0.902 (1.894)	0.916 (2.045)	0.928 (2.350)	0.948 (2.466)
			70	0.942 (1.902)	0.928 (1.604)	0.938 (1.663)	0.938 (1.664)	0.958 (1.906)
	0.50	0.70	30	0.938 (1.245)	0.904 (1.153)	0.914 (1.254)	0.925 (1.299)	0.948 (1.588)
			50	0.949 (0.957)	0.919 (0.797)	0.932 (0.836)	0.935 (0.847)	0.952 (0.936)
			70	0.944 (0.809)	0.930 (0.661)	0.940 (0.689)	0.940 (0.690)	0.956 (0.736)

NOTE:
NA: normal approximation
JEL: jackknife empirical likelihood
AJEL: jackknife adjusted empirical likelihood
TJEL: transformed jackknife empirical likelihood
TAJEL:transformed adjusted jackknife empirical likelihood

Table (4.2) Coverage probabilities (average lengths) of 95% confidence interval with the non-zero data following exponential distribution.

δ	s	t	n	NA	JEL	AJEL	TJEL	TAJEL
0.1	0.20	0.90	30	0.830 (3.575)	0.895 (4.637)	0.918 (4.818)	0.917 (4.809)	0.938 (4.967)
			50	0.842 (3.011)	0.907 (3.611)	0.916 (3.714)	0.915 (3.694)	0.927 (3.801)
			70	0.864 (2.659)	0.924 (3.111)	0.930 (3.194)	0.929 (3.191)	0.934 (3.247)
	0.25	0.75	30	0.890 (2.414)	0.914 (2.503)	0.928 (2.691)	0.934 (3.010)	0.955 (3.224)
			50	0.922 (1.895)	0.932 (1.904)	0.944 (1.995)	0.944 (2.003)	0.952 (2.084)
			70	0.920 (1.607)	0.936 (1.546)	0.945 (1.596)	0.944 (1.596)	0.952 (1.645)
	0.35	0.65	30	0.922 (1.775)	0.920 (1.681)	0.938 (1.801)	0.943 (1.833)	0.956 (1.954)
			50	0.938 (1.383)	0.942 (1.254)	0.950 (1.312)	0.950 (1.317)	0.964 (1.400)
			70	0.937 (1.172)	0.947 (1.048)	0.955 (1.089)	0.955 (1.087)	0.962 (1.129)
0.2	0.30	0.90	30	0.858 (3.943)	0.898 (4.579)	0.920 (4.769)	0.921 (4.777)	0.945 (4.946)
			50	0.878 (3.327)	0.887 (3.601)	0.902 (3.761)	0.903 (3.757)	0.914 (3.863)
			70	0.890 (2.980)	0.912 (3.059)	0.920 (3.168)	0.919 (3.155)	0.924 (3.240)
	0.40	0.75	30	0.928 (2.576)	0.920 (2.278)	0.940 (2.447)	0.946 (2.658)	0.965 (3.033)
			50	0.938 (2.041)	0.932 (1.708)	0.946 (1.863)	0.946 (1.881)	0.956 (1.966)
			70	0.945 (1.728)	0.945 (1.424)	0.954 (1.502)	0.954 (1.503)	0.960 (1.558)
	0.45	0.70	30	0.942 (2.185)	0.920 (1.802)	0.939 (1.918)	0.946 (1.986)	0.958 (2.262)
			50	0.948 (1.694)	0.932 (1.361)	0.944 (1.429)	0.946 (1.442)	0.962 (1.540)
			70	0.962 (1.418)	0.946 (1.126)	0.952 (1.175)	0.952 (1.175)	0.960 (1.229)
0.3	0.40	0.90	30	0.880 (4.345)	0.908 (4.59)	0.934 (4.788)	0.939 (4.817)	0.964 (5.032)
			50	0.902 (3.744)	0.899 (3.527)	0.912 (3.645)	0.912 (3.682)	0.925 (3.795)
			70	0.929 (3.361)	0.914 (3.030)	0.926 (3.143)	0.924 (3.130)	0.934 (3.235)
	0.45	0.85	30	0.920 (3.974)	0.896 (3.324)	0.920 (3.551)	0.939 (4.454)	0.958 (5.277)
			50	0.942 (3.197)	0.912 (2.515)	0.928 (2.661)	0.934 (2.905)	0.952 (3.148)
			70	0.952 (2.717)	0.922 (2.132)	0.938 (2.220)	0.936 (2.229)	0.948 (2.438)
	0.50	0.70	30	0.972 (2.361)	0.906 (1.695)	0.924 (1.844)	0.940 (1.925)	0.960 (2.254)
			50	0.964 (1.804)	0.934 (1.252)	0.944 (1.322)	0.947 (1.336)	0.957 (1.442)
			70	0.973 (1.526)	0.938 (1.047)	0.948 (1.093)	0.949 (1.095)	0.960 (1.151)

NOTE:
NA: normal approximation
JEL: jackknife empirical likelihood
AJEL: jackknife adjusted empirical likelihood
TJEL: transformed jackknife empirical likelihood
TAJEL:transformed adjusted jackknife empirical likelihood

Table (4.3) Coverage probabilities (average lengths) of 95% confidence interval with the non-zero data following lognormal distribution.

δ	s	t	n	NA	JEL	AJEL	TJEL	TAJEL
0.1	0.20	0.90	30	0.794 (3.320)	0.883 (5.782)	0.905 (5.974)	0.904 (5.943)	0.924 (6.104)
			50	0.826 (2.902)	0.893 (4.42)	0.902 (4.553)	0.900 (4.497)	0.910 (4.610)
			70	0.844 (2.578)	0.899 (3.722)	0.904 (3.816)	0.903 (3.817)	0.908 (3.877)
	0.25	0.75	30	0.888 (1.774)	0.898 (2.211)	0.912 (2.437)	0.926 (2.831)	0.950 (2.968)
			50	0.908 (1.396)	0.924 (1.594)	0.938 (1.663)	0.939 (1.675)	0.948 (1.735)
			70	0.927 (1.214)	0.925 (1.268)	0.935 (1.309)	0.936 (1.308)	0.945 (1.349)
	0.35	0.65	30	0.919 (1.214)	0.930 (1.246)	0.940 (1.327)	0.945 (1.347)	0.958 (1.436)
			50	0.924 (0.919)	0.939 (0.887)	0.954 (0.929)	0.954 (0.934)	0.964 (1.003)
			70	0.942 (0.784)	0.934 (0.747)	0.943 (0.778)	0.943 (0.777)	0.949 (0.809)
0.2	0.30	0.90	30	0.829 (3.630)	0.892 (5.776)	0.914 (5.954)	0.915 (5.952)	0.942 (6.097)
			50	0.852 (3.172)	0.888 (4.346)	0.906 (4.468)	0.905 (4.442)	0.916 (4.554)
			70	0.880 (2.764)	0.906 (3.549)	0.916 (3.636)	0.914 (3.620)	0.920 (3.712)
	0.40	0.75	30	0.923 (1.823)	0.922 (1.869)	0.940 (1.993)	0.947 (2.254)	0.966 (2.605)
			50	0.936 (1.431)	0.936 (1.312)	0.951 (1.462)	0.952 (1.476)	0.960 (1.534)
			70	0.948 (1.227)	0.944 (1.086)	0.953 (1.159)	0.953 (1.160)	0.960 (1.201)
	0.45	0.70	30	0.940 (1.463)	0.922 (1.384)	0.940 (1.468)	0.949 (1.514)	0.960 (1.792)
			50	0.951 (1.124)	0.923 (0.971)	0.936 (1.024)	0.938 (1.038)	0.950 (1.123)
			70	0.959 (0.952)	0.942 (0.810)	0.947 (0.847)	0.946 (0.847)	0.956 (0.889)
0.3	0.40	0.90	30	0.866 (4.056)	0.914 (5.760)	0.944 (5.915)	0.944 (5.929)	0.962 (6.084)
			50	0.900 (3.479)	0.894 (4.092)	0.910 (4.260)	0.909 (4.269)	0.922 (4.410)
			70	0.918 (3.019)	0.906 (3.395)	0.918 (3.500)	0.915 (3.487)	0.926 (3.577)
	0.45	0.85	30	0.904 (3.332)	0.909 (3.335)	0.929 (3.614)	0.938 (5.239)	0.963 (6.181)
			50	0.934 (2.587)	0.920 (2.329)	0.933 (2.465)	0.935 (2.873)	0.950 (3.053)
			70	0.944 (2.201)	0.920 (1.885)	0.928 (1.969)	0.927 (1.975)	0.941 (2.260)
	0.50	0.70	30	0.970 (1.615)	0.895 (1.261)	0.916 (1.373)	0.926 (1.429)	0.947 (1.751)
			50	0.979 (1.189)	0.929 (0.877)	0.936 (0.923)	0.938 (0.934)	0.946 (1.025)
			70	0.978 (1.004)	0.927 (0.718)	0.940 (0.750)	0.943 (0.752)	0.954 (0.796)

NOTE:
NA: normal approximation
JEL: jackknife empirical likelihood
AJEL: jackknife adjusted empirical likelihood
TJEL: transformed jackknife empirical likelihood
TAJEL:transformed adjusted jackknife empirical likelihood

Table (4.4) Upper limit and lower limit of 95% confidence interval (CI) and length for the real data with normal approximation (NA), jackknife empirical likelihood (JEL), adjusted jackknife empirical likelihood (AJEL), transformed jackknife empirical likelihood (TJEL), and transformed adjusted jackknife empirical likelihood (TAJEL).

s	t		NA	JEL	AJEL	TJEL	TAJEL
		Lower limit	58.099	69.780	69.770	69.770	69.770
0.40	0.95	Upper limit	96.508	135.530	135.530	135.530	135.530
		Length	38.409	65.750	65.760	65.760	65.760
		Lower limit	57.457	70.320	69.490	69.490	69.490
0.45	0.95	Upper limit	96.65	135.320	135.320	135.320	135.320
		Length	39.193	65.000	65.830	65.830	65.830
		Lower limit	57.003	70.040	70.040	70.040	69.790
0.50	0.95	Upper limit	96.555	135.140	135.140	135.140	135.140
		Length	39.552	65.100	65.100	65.100	65.350
		Lower limit	56.602	64.630	64.630	64.630	63.800
0.55	0.95	Upper limit	88.155	130.560	130.560	134.710	134.710
		Length	31.553	65.930	65.930	70.080	70.910

NOTE:
NA: normal approximation
JEL: jackknife empirical likelihood
AJEL: jackknife adjusted empirical likelihood
TJEL: transformed jackknife empirical likelihood
TAJEL: transformed adjusted jackknife empirical likelihood

CHAPTER 5

EMPIRICAL LIKELIHOOD INFERENCE FOR THE PANEL COUNT DATA WITH INFORMATIVE OBSERVATION PROCESS

5.1 Background

In long-term event-history or longitudinal studies, it is desirable to follow the study subjects continuously over time so that all the recurrent events of each subject can be monitored in real time and the exact times of the events are observed. Unfortunately, sometimes it may not be feasible or even unrealistic to keep continuous track of those study subjects. Instead they are only observed at discrete time points within the study period and only the number of events occurred between the two time points are known. Such interval-censored data are commonly known as panel count data (see Sun and Zhao (2013)). Such data may occur in medical follow-up, demography, epidemiology, psychology, tumorigenicity and reliability studies, among others. For example, in a cancer study, patients are often scheduled to come to the hospital to be screened by the medical professionals at finite discrete time points, e.g., three months, six months etc. This is a feasible approach for a follow up study from a practical point of view as patients may visit early or late, or even miss a scheduled visit. At each visit, only the number of new events, e.g., new cancer cell, is recorded since the last screening procedure. Since there is an observation gap between two subsequent screening procedures, the exact timing of the new cancer cell formation is unknown. Subjects are only observed intermittently, and observations may also be right-censored.

In panel count data, number of observations differ among study subjects. Also, observation times can vary from subject to subject. These characteristics lead to two key processes that control panel count data: the recurrent event process and the observation process. The number of events occurring between two time points can be viewed as the realizations of the underlying recurrent event process and the observation times for each subject is the result of observation process. The relationship between these processes plays a vital role in analyzing the panel count data.

Kalbfleisch and Lawless (1985) formulated algorithms for maximum likelihood estimation under a continuous-time Markov model. Estimation of the mean function of the underlying point processes that govern the panel count data was discussed by Sun and Kalbfleisch (1995), which was later extended by Wellner and Zhang (2000). For treatment comparison of panel count data, several nonparametric methods were proposed. For example, Thall and Lachin (1988) adopted nonparametric procedure to compare the recurrence rates of two treatment groups, Sun and Fang (2003) proposed nonparametric test for the comparison of underlying recurrent point processes. Zhang (2006) and Balakrishnan and Zhao (2009) devised nonparametric procedures for k ($k \geq 2$) comparisons. Regression analysis of panel count data was also explored. Among others, Sun and Wei (2000) proposed semiparametric regression methods based on some estimation equations, whereas Zhang (2002) proposed semiparametric maximum pseudolikelihood estimation for regression parameters. He et al. (2008) studied regression analysis of multivariate panel count data.

However, in various situations, the two processes may not be independent and the observation process may carry some important information about the underlying recurrent event process. E.g., more severely ill subjects are monitored more closely, next follow-up visit could be chosen based on the current disease status. Or, a subject decides to go to the clinic when they are in poor conditions. Considering such dependency, Huang et al. (2006), Sun et al. (2007), He et al. (2009) considered some frailty models to make the connection between the observation and event processes, and the regression parameters were then estimated by solving some estimating equations. Li et al. (2010) proposed a more flexible estimation method via estimating equation approach. Zhao et al. (2013) gave a similar model, but also included a terminal event in their model. Buzkova (2010) used a joint modeling approach after predicting the observation times by some time-varying factors such as outcome at the last visit or cumulative exposure. A more general and robust estimation procedure was proposed by Zhao et al. (2013). Zhang and Zhao (2013) adopt robust joint modeling approach for the estimation of multivariate panel count data with informative observation process. More recently, Fang et al. (2017) presented a joint model using two latent variables for panel count data with time-dependent covariates and informative observation process.

With informative observation process, previous analyses of panel count data involve semiparametric transformation model or joint modeling. In this chapter, we employ non-parametric empirical likelihood procedure for the inference about the parameters of the model. We adopted the model representation by Li et al. (2010). Our simulation studies reveal that when the sample size is small, the normal approximation method suffers from under-coverage issue. We try to solve this issue by the new nonparametric procedure.

Owen (1988, 1990, 1991) introduced empirical likelihood (EL) as a nonparametric alternative to the classical statistical methods. Since then EL was extensively used in many statistical areas because of its advantages over other methods. Advantages include transformation invariance, range preservation, shape and orientation determination of the confidence region by the data itself, and Bartlett correctability Diciccio et al. (1991), Chen and Cui (2006)). See Owen (2001) for a more comprehensive review.Hall and La Scala (1990) gave some fundamental algorithms of empirical likelihood. Another benefit of EL method is that unlike traditional normal approximation (NA) method, it does not require estimating covariance matrix to construct confidence regions. Instead an implicit studentization is carried out internally. Also many papers showed better performance of EL methods in comparison to NA methods for the small sample size through their numerical results. These attractive features encourage many researchers to extend the EL methods to the survival and longitudinal data analysis. Yu et al. (2011) proposed EL for linear transformation model for right censored survival data.Liu et al. (2014) examined EL for the additive hazard model with current status data, i.e., one observation time for each study subject. Zhang and Zhao (2013) formulated EL methods for linear transformation models for interval-censored failure time data. Wang et al. (2010) presented two generalized EL methods for analyzing longitudinal data: element-wise and subject wise empirical likelihood methods. Hu and Lin (2012) analyzed longitudinal data with within-subject correlation by empirical likelihood. Dauxois et al. (2016) applied EL for mean functions of recurrent events with competing risks under random censorship and with a terminal event. Longitudinal neuroimaging data are, for example, analyzed using two-stage empirical likelihood method by Shi et al. (2011).

The rest of this chapter is organized as follows. Section 5.2 illustrates the notations and assumptions, which are needed to present the model. An overview of existing NA approaches, followed by EL inference procedure with the asymptotic distributions of the empirical likelihood ratio test statistic is also given in Section 5.2. In Section 5.3, some simulation studies are presented to compare the empirical likelihood method with the normal approximation based method. Section 5.4 includes the application of the proposed method to a real dataset. Finally, some concluding remarks are given in Section 5.4. The proofs of Theorems are provided in the Appendix D.

5.2 Main Results

5.2.1 Model setup

We will review the normal approximation method developed by Li et al. (2010) for the completeness. We use the similar notation as Li et al. (2010) did. Consider a recurrent event study involving n independent study subjects. For subject i, let $Y_i(t)$ be the underlying recurrent event process that represents the cumulative number of event occurrences before or at time t and $O_i(t)$ denote the observation process that counts the total observations before or at time t. Suppose that $Y_i(t)$ is observed only at the potential time points $0 < T_{i,1} < \cdots < T_{i,M_i} \leq \tau$, where M_i is the total number of possible observation point for subject i and τ denotes the maximum study duration. These observation times are assumed to be generated from the observation process $O_i(t) = \sum_{j=1}^{M_i} I(T_{i,j} \leq t),\ i = 1, \cdots, n$. Define $O_i^*(t) = O_i\{\min(t, C_i)\}$, where C_i is the follow-up or censoring time for subject $i, i = 1, \cdots, n$. The censoring time $C_i \leq \tau$ is either the last observation time or the administrative study end time. Thus, $O_i^*(t)$ represents the observed/actual observation process that jumps only at observation time. In other words, one can observe panel count data on the $Y_i(t)$'s at time t only when $O_i^*(t)$ jumps. Therefore, the total number of actual observation for subject i, $m_i^* = O_i(C_i) = O_i^*(C_i)$. Additionally, a vector of covariates, possibly time-dependent, is also available from a left-continuous covariate process $Z_i(t)$ for each subject i. Since $O_i^*(t)$ and $Z_i(t)$ are only known and observable as long as individual i is under study and not censored, we have $0 \leq t \leq C_i; i = 1, \cdots, n$. Combining all of them, one can observe the dataset:

$$\{O_i^*(t), Z_i(t), Y_i(T_{i,1}), \cdots, Y_i(T_{i,M_i}); 0 \leq t, T_{i,M_i} \leq C_i, i = 1, \cdots, n, K = 1, \cdots, M\},$$

i.e., we only have panel count data on $Y_i(t)$'s.

Define $\mathcal{F}_{it} = \{O_i(s), 0 \leq s < t\}$ as the history or filtration of the observation process O_i up to time $t-$, $i = 1, \cdots, n$. Assume that C_i is independent of $\{O_i(t), Y_i(t)\}$ conditional on $Z_i(t)$. Let $dO_i(t)$ denote the indicator of whether the ith subject to be observed in $[t, t+dt)$. Following Li et al. (2010) and Zhao et al. (2013), the observation process $O_i(t)$ follows the proportional rate model [cf. Cook and Lawless (2007)]

$$E\{dO_i(t)|Z_i(t)\} = e^{\gamma Z_i(t)}\lambda_0(t)dt, \tag{5.1}$$

where γ is a vector of unknown parameters, $Z_i(t)$ is the covariate process at time t, and $\lambda_0(\cdot)$ is an unspecified baseline rate function.

Model (1) can be expressed as [cf. Lin et al. (2000)]

$$E\{dO_i(t)|Z_i(t)\} = e^{\gamma Z_i(t)}d\Lambda_0(t), \tag{5.2}$$

where $\Lambda_0(\cdot)$ is an unknown function. Here, model (5.1) implies model (5.2) with $d\Lambda_0(t) = \lambda_0(t)$. Model (5.2) is more attractive because of it's flexibility in allowing various dependent structures among recurrent events and applicability to any recurrent event counting process.

Considering the effects of covariates on the recurrent event process $Y_i(t)$, we model the conditional mean function of $Y_i(t)$ given $Z_i(t)$ and $\mathcal{F}_{it}$ as

$$E\{Y_i(t)|Z_i(t), \mathcal{F}_{it}\} = g\{\mu_0(t)e^{\beta_1' Z_i(t)+\beta_2' H(\mathcal{F}_{it})}\}, \tag{5.3}$$

where $g(\cdot)$ is a known twice continuously differentiable and strictly increasing function, $\mu_0(t)$ denotes an unknown arbitrary function of t, β_1 and β_2 are vectors of unknown regression parameters, and $H(\cdot)$ is a vector of known functions of $\mathcal{F}_{it}$. It is also assumed that given $Z_i(t)$ and $\mathcal{F}_{it}$, $O_i(t)$ and $Y_i(t)$ are independent. The model as given in equation (5.3) is chosen for its apparent flexibility. Please see Li et al. (2010) for details.

Let β_{10}, β_{20} and γ_0 denote the true values of β_1, β_2 and γ. Also let $X_i(t) = (Z_i(t)', H(\mathcal{F}_{it})')'$, $\beta = (\beta_1', \beta_2')'$, $\beta_0 = (\beta_{10}', \beta_{20}')'$ for easy presentation.

5.2.2 Semiparametric transformation model

We go over the normal approximation (NA) method in detail proposed by Li et al. (2010). The NA method is the foundation of our EL method. Define

$$M_i(t;\beta_1,\beta_2,\gamma) = \int_0^t Y_i(u)\Delta_i(u)dO_i(u) - \int_0^t g\{\mu_0(u)e^{\beta_1' Z_i(u)+\beta_2' H(\mathcal{F}_{iu})}\}\Delta_i(u)e^{\gamma' Z_i(u)}d\Lambda_0(u),$$

where $\Delta_i(t) = I(C_i \geq t)$ is the at risk indicator for subject i, $i = 1, \ldots, n$. Then under models (5.1) and (5.3), it can be easily shown that $E\{M_i(t;\beta_{10},\beta_{20},\gamma_0)\} = 0$. Then $M_i(t;\beta_{10},\beta_{20},\gamma_0)$'s are zero-mean stochastic processes.

Li et al. (2010) proposed the following generalized estimating equations.

$$\sum_{i=1}^n \Big[Y_i(t)\Delta_i(t)dO_i(t) - g\{\mu_0(t)e^{\beta' X_i(t)}\}\Delta_i(t)e^{\gamma' Z_i(t)}d\Lambda_0(t)\Big] = 0, \quad 0 \leq t \leq \tau, \tag{5.4}$$

$$U(\beta;\gamma) = \sum_{i=1}^n \int_0^\tau W(t)X_i(t)\left[Y_i(t)\Delta_i(t)dO_i(t) - g\{\mu_0(t)e^{\beta' X_i(t)}\}\Delta_i(t)e^{\gamma' Z_i(t)}d\Lambda_0(t)\right], \tag{5.5}$$

to estimate $\mu_0(t)$ and β, respectively, where $W(t)$ is a possibly data-dependent weight function. For fixed γ, let $\hat{\mu}_0(t)$ and $\hat{\beta} = (\hat{\beta}_1{}', \hat{\beta}_2{}')'$ be the estimates of $\mu_0(t)$ and β_0 obtained by solving equations (5.4) and (5.5), respectively.

Let the limit of $S^{(j)}(t;\gamma), E_X(t;\beta,\gamma)$, and $R(t;\beta,\gamma)$ be denoted by $s^{(j)}(t;\gamma), e_x(t;\beta,\gamma)$ and $r(t;\beta,\gamma)$, respectively,

where, same as in Li et al. (2010)

$$E_X(t;\beta,\gamma) = \frac{\sum_{i=1}^n \Delta_i(t)X_i(t)\dot{g}\{\mu_0(t)e^{\beta' X_i(t)}\}e^{\beta' X_i(t)+\gamma' Z_i(t)}}{\sum_{i=1}^n \Delta_i(t)\dot{g}\{\mu_0(t)e^{\beta' X_i(t)}\}e^{\beta' X_i(t)+\gamma' Z_i(t)}},$$

$$R(t;\beta,\gamma) = n^{-1}\sum_{i=1}^n \{X_i(t) - E_X(t;\beta,\gamma)\}\Delta_i(t)g\{\mu_0(t)e^{\beta' X_i(t)}\}e^{\gamma' Z_i(t)}.$$

$$S^{(j)}(t;\gamma) = n^{-1}\sum_{i=1}^n \Delta_i(t)Z_i(t)^j e^{\gamma' Z_i(t)}, j = 1, 2.$$

Also let $\bar{z}(t;\gamma) = s^{(1)}(t;\gamma)/s^{(0)}(t;\gamma)$. Define

$$M_i(t;\beta,\gamma) = \int_0^t Y_i(u)\Delta_i(u)dO_i(u) - \int_0^t g\{\mu_0(u)e^{\beta' X_i(u)}\}\Delta_i(u)e^{\gamma' Z_i(u)}d\Lambda_0(u),$$

$$M_i^*(t;\gamma) = \int_0^t \Delta_i(u)dO_i(u) - \int_0^t \Delta_i(u)e^{\gamma' Z_i(u)}d\Lambda_0(u),$$

$$\begin{aligned} P(\beta,\gamma) \quad = \quad & E\Big[\int_0^\tau w(t)\Delta_i(t)g\{\mu_0(t)e^{\beta' X_i(t)}\}e^{\gamma' Z_i(t)}\{X_i(t) \\ & -e_x(t;\beta,\gamma)\}\{Z_i(t) - \bar{z}(t;\gamma)\}' d\Lambda_0(t)\Big], \end{aligned}$$

$$D(\gamma) = E\left[\int_0^\tau \{Z_i(t) - \bar{z}(t;\gamma)\}^{\otimes 2}\Delta_i(t)e^{\gamma' Z_i(u)}d\Lambda_0(t)\right].$$

Further denote

$$A(\beta;\gamma) = E\left[\int_0^\tau w(t)\Delta_i(t)\dot{g}\{\mu_0(t)e^{\beta' X_i(t)}\}\{X_i(t) - e_x(t;\beta,\gamma)\}^{\otimes 2}e^{\beta' X_i(t)+\gamma' Z_i(t)}\mu_0(t)d\Lambda_0(t)\right],$$

$$\Sigma(\beta;\gamma) \quad = \quad E\Big[\int_0^\tau w(t)\{X_i(t) - e_x(t;\beta,\gamma)\}dM_i(t;\beta,\gamma) - \int_0^\tau \frac{w(t)r(t;\beta,\gamma)}{s^{(0)}(t;\gamma)}dM_i^*(t;\gamma)$$

$$-P(\beta,\gamma)[D(\gamma)]^{-1}\int_0^{\tau}\{Z_i(t)-\bar{z}(t;\gamma)\}dM_i^*(t;\gamma)\Bigg]^{\otimes 2}.$$

Now, we want to establish the asymptotic properties of $\hat{\beta}$. Under the regularity conditions (D1)-(D5) in the Appendix D, Li et al. (2010) proved that

$$\sqrt{n}(\hat{\beta}-\beta_0)\xrightarrow{\mathfrak{D}} N(0,[A(\beta_0;\gamma_0)]^{-1}\Sigma(\beta_0;\gamma_0)[A(\beta_0;\gamma_0)]^{-1}).$$

At first, γ and $\Lambda_0(t)$ are unknown and need to be estimated. This estimation can be done based on the recurrent event data on the $O_i(t)$'s [cf. Andersen et al. (1993) and Cook and Lawless (2007)]. Then a consistent estimator of γ, denoted by $\hat{\gamma}$ can be estimated by solving the following estimating equation

$$V(\gamma)=\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t)-\bar{Z}(t;\gamma)\}\Delta_i(t)dO_i(t)=0,$$

where $\bar{Z}(t;\gamma)=S^{(1)}(t;\gamma)/S^{(0)}(t;\gamma)$, and $\Lambda_0(t)$ can be estimated by

$$\hat{\Lambda}_0(t;\hat{\gamma})=\sum_{i=1}^{n}\int_0^{t}\frac{\Delta_i(u)dO_i(u)}{nS^{(0)}(u;\hat{\gamma}))}.$$

Then the estimating equation to estimate $\mu_0(t;\beta,\hat{\gamma})$ becomes

$$\sum_{i=1}^{n}\Big[Y_i(t)\Delta_i(t)dO_i(t)-g\{\mu_0(t;\beta,\hat{\gamma})e^{\beta' X_i(t)}\}\Delta_i(t)e^{\hat{\gamma}' Z_i(t)}d\hat{\Lambda}_0(t;\hat{\gamma})\Big]=0,\quad 0\le t\le\tau,\quad (5.6)$$

And after replacing $\mu_0(t;\beta,\hat{\gamma})$ by it's estimate $\hat{\mu}_0(t;\beta,\hat{\gamma})$, the estimating equation $U(\beta;\hat{\gamma})$ becomes

$$U(\beta;\hat{\gamma})=\sum_{i=1}^{n}\int_0^{\tau}W(t)X_i(t)\left[Y_i(t)\Delta_i(t)dO_i(t)-g\{\hat{\mu}_0(t;\beta,\hat{\gamma})e^{\beta' X_i(t)}\}\Delta_i(t)e^{\hat{\gamma}' Z_i(t)}d\hat{\Lambda}_0(t;\hat{\gamma})\right],\quad (5.7)$$

Define

$$\hat{M}_i(t;\beta,\hat{\gamma}) = \int_0^t Y_i(u)\Delta_i(u)dO_i(u) - \int_0^t g\{\hat{\mu}_0(u;\beta,\hat{\gamma})e^{\beta' X_i(u)}\}\Delta_i(u)e^{\hat{\gamma}' Z_i(u)}d\hat{\Lambda}_0(u;\hat{\gamma})\,,$$

$$\hat{M}_i^*(t;\hat{\gamma}) = \int_0^t \Delta_i(u)dO_i(u) - \int_0^t \Delta_i(u)e^{\hat{\gamma}' Z_i(u)}d\hat{\Lambda}_0(u;\hat{\gamma})\,,$$

$$\hat{E}_X(t;\beta,\hat{\gamma}) = \frac{\sum_{i=1}^n \Delta_i(t)X_i(t)\dot{g}\{\hat{\mu}_0(t;\beta,\hat{\gamma})e^{\beta' X_i(t)}\}e^{\beta' X_i(t)+\hat{\gamma}' Z_i(t)}}{\sum_{i=1}^n \Delta_i(t)\dot{g}\{\hat{\mu}_0(t;\beta,\hat{\gamma})e^{\beta' X_i(t)}\}e^{\beta' X_i(t)+\hat{\gamma}' Z_i(t)}}\,,$$

$$\hat{R}(t;\beta,\hat{\gamma}) = n^{-1}\sum_{i=1}^n \{X_i(t) - \hat{E}_X(t;\beta,\hat{\gamma})\}\Delta_i(t)g\{\hat{\mu}_0(t;\beta,\hat{\gamma})e^{\beta' X_i(t)}\}e^{\hat{\gamma}' Z_i(t)}\,,$$

$$\hat{D} = n^{-1}\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}^{\otimes 2}\Delta_i(t)dO_i(t)\,,$$

and

$$\begin{aligned}\hat{P}(\beta,\hat{\gamma}) \;=\; & n^{-1}\sum_{i=1}^n \int_0^\tau W(t)\Delta_i(t)g\{\hat{\mu}_0(t;\beta,\hat{\gamma})e^{\beta' X_i(t)}\}e^{\hat{\gamma}' Z_i(t)}\{X_i(t)\hat{E}_X(t;\beta.\hat{\gamma})\} \\ & \{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}' d\hat{\Lambda}_0(u;\hat{\gamma})\,.\end{aligned}$$

In the above, $\dot{g}(t) = dg(t)/dt$, and $\upsilon^{\otimes 2} = \upsilon\upsilon'$ for a vector υ.

Then as $n \to \infty$, $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically approximated by a normal distribution with mean zero and a covariance matrix that can be consistently estimated by $(\hat{A}(\hat{\beta},\hat{\gamma}))^{-1}\hat{\Sigma}(\hat{\beta},\hat{\gamma})(\hat{A}(\hat{\beta},\hat{\gamma}))^{-1}$ [cf. Li et al. (2010)],

where

$$\begin{aligned}\hat{A}(\beta,\hat{\gamma}) \;=\; & n^{-1}\sum_{i=1}^n \int_0^\tau W(t)\Delta_i(t)\dot{g}\{\hat{\mu}_0(t;\beta,\hat{\gamma})e^{\beta' X_i(t)}\}\{X_i(t) - \hat{E}_X(t;\beta,\hat{\gamma})\}^{\otimes 2} \\ & e^{\beta' X_i(t)+\hat{\gamma}' Z_i(t)}\hat{\mu}_0(t;\beta,\hat{\gamma})d\hat{\Lambda}_0(t;\hat{\gamma})\end{aligned}$$

and

$$\hat{\Sigma}(\beta,\hat{\gamma}) \;=\; n^{-1}\sum_{i=1}^n \Bigg[\int_0^\tau W(t)\{X_i(t) - \hat{E}_X(t;\beta,\hat{\gamma})\}d\hat{M}_i(t;\beta,\hat{\gamma})$$

$$-\int_0^\tau \frac{W(t)\hat{R}(t,\beta,\hat{\gamma})}{S^{(0)}(t;\hat{\gamma})}d\hat{M}_i^*(t;\hat{\gamma})$$
$$-\hat{P}(\beta,\hat{\gamma})\hat{D}^{-1}\int_0^\tau \{Z_i(t)-\bar{Z}(t;\hat{\gamma})\}d\hat{M}_i^*(t;\hat{\gamma})\Bigg]^{\otimes 2}.$$

Li et al. (2010) showed overall satisfactory performance of the normal approximation method in their simulation analysis for the sample size 100 and 300. This method, however, requires rigorous variance computations. Also, there may exist an under-coverage issue when the sample size is small as shown by our simulation studies. To deal with these issues, we propose an empirical likelihood method in the next section.

5.2.3 Empirical likelihood

Motivated by Owen (1988, 1990) and Qin and Lawless (1994), we apply empirical likelihood method to find the confidence intervals of the parameters. Define

$$U_{ni}(\beta;\hat{\gamma}) = \int_0^\tau W(t)\{X_i(t)-\hat{E}_X(t;\beta,\hat{\gamma})\}d\hat{M}_i(t;\beta,\hat{\gamma}) - \int_0^\tau \frac{W(t)\hat{R}(t;\beta,\hat{\gamma})}{S^{(0)}(t,\hat{\gamma})}d\hat{M}_i^*(t;\hat{\gamma}) -$$

$$\hat{P}(\beta,\hat{\gamma})\hat{D}^{-1}\int_0^\tau \{Z_i(t)-\bar{Z}(t;\hat{\gamma})\}d\hat{M}_i^*(t;\hat{\gamma}), i=1,2,\ldots,n,$$

where $\hat{E}_X(t;\beta,\hat{\gamma})$, $\hat{R}(t;\beta,\hat{\gamma})$, $\hat{M}_i(t;\beta,\hat{\gamma})$, $\hat{M}_i^*(t;\hat{\gamma})$, $\hat{P}(\beta,\hat{\gamma})$ and $\hat{D}$ are defined as in Section 5.2.2.

The idea of empirical likelihood is to consider the data observations as if they are from a fixed and unknown distribution F. To model F by a multinomial distribution concentrated on the observations, with p_i as the probability mass at the ith observation. The empirical likelihood is $L(F)=\prod_{i=1}^n p_i$ and the empirical likelihood ratio is of the form $L(F)/L(F_n)$, where F_n is the empirical distribution dividing the probability equally among n observations.

Let $p=(p_1,p_2,\cdots,p_n)$ be a probability vector, i.e., $\sum_{i=1}^n p_i = 1$, and $p_i \geq 0$ for all i. The empirical likelihood at true parameter value β_0 is defined as:

$$L(\beta_0) = \sup\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i U_{ni}(\beta_0;\hat{\gamma}) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1\}.$$

A unique maximum exists for $L(\beta_0)$ for a given β_0, provided that 0 in inside the convex hull of the points $\{U_{n_i}\}, i = 1, 2, \ldots, n$. Also, $\prod_{i=1}^n$ attains its maximum at $p_i = 1/n$ under the restrictions $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0, i = 1, 2, \ldots, n$. Then, the empirical likelihood ratio at true parameter value β_0 is defined by

$$R(\beta_0) = \sup\{\prod_{i=1}^n np_i : \sum_{i=1}^n p_i U_{ni}(\beta_0; \hat{\gamma}) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1\}.$$

Using the Lagrange multiplier method, $R(\beta_0)$ is maximized when

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda' U_{ni}(\beta; \hat{\gamma})},$$

where $\lambda = (\lambda_1, \cdots, \lambda_p)'$ is the solution to

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

Then the empirical log-likelihood ratio at β is

$$\begin{aligned} l(\beta) &= -2 \log R(\beta) \\ &= 2 \sum_{i=1}^n \log\{1 + \lambda' U_{ni}(\beta; \hat{\gamma})\}. \end{aligned}$$

We establish the Wilks' theorem for the empirical log-likelihood ratio as follows.

Theorem 5.1. *Under the regularity conditions stated in the Appendix D, $l(\beta_0)$ converges in distribution to χ_p^2 as $n \to \infty$, where χ_p^2 is a chi-square distribution with p degrees of freedom.*

Based on Theorem 5.1, we can construct the asymptotic $100(1-\alpha)\%$ empirical likelihood confidence region for β_0 as

$$I_1(\alpha) = \{\beta : l(\beta) \leq \chi_p^2(\alpha)\},$$

where $\chi_p^2(\alpha)$ is the $(1 - \alpha)$ quantile of χ_p^2.

In practice, one may be more interested in the inference of a single component of the parameter β. Subramanian (2007), Yu et al. (2011) and Zhao (2011) proposed the profile empirical likelihood method for this purpose. The idea is to profile out nuisance parameters from the full EL, then construct confidence interval for the actual components that one is interested in.

Define $\beta_0 = ((\beta_0^{(1)})', (\beta_0^{(2)})')'$. Suppose we want to construct EL confidence regions for a q-dimensional ($q < p$) subvector $\beta_0^{(1)}$. This can be done by profiling out $\beta_0^{(2)}$ from the full EL. Then the profile EL ratio at $\beta^{(1)}$ is

$$l_{\text{profile}}(\beta^{(1)}) = \min_{\beta^{(2)}} l((\beta_0^{(1)})', (\beta_0^{(2)})').$$

We establish Wilks' theorem for $l_{\text{profile}}(\beta^{(1)})$ as follows:

Theorem 5.2. *Under the regularity conditions stated in the Appendix D, the profile EL statistic $l_{profile}(\beta_0^{(1)})$ converges in distribution to χ_q^2 as $n \to \infty$, where χ_q^2 is a chi-square distribution with q degrees of freedom.*

Based on Theorem 5.2, we can construct the asymptotic $100(1-\alpha)\%$ empirical likelihood confidence region for $\beta_0^{(1)}$ as

$$I_2(\alpha) = \{\beta^{(1)} : l_{\text{profile}}(\beta_0^{(1)}) \leq \chi_q^2(\alpha)\},$$

where $\chi_q^2(\alpha)$ is the $(1-\alpha)$ quantile of χ_q^2.

5.3 Simulation Study

To evaluate the performance of the proposed empirical likelihood method, an extensive simulation study is conducted. First, the covariate $Z_i(t)$ and follow-up time C_i were generated from a Bernoulli ($p = 0.5$) and a U ($\tau/2$, τ) distribution, respectively.

We assumed that $O_i(t)$ is a Poisson process with $\lambda_0(t) = c/\tau$, where c is a constant. Then given Z_i, the number of observations, m_i^*, follows the Poisson$\left(\text{mean} = \frac{cC_ie^{\gamma Z_i}}{\tau}\right)$ and the observation times $(T_{i,1}, ..., T_{i,m_i^*})$ were obtained from the order statistics of a random sample of size m_i^* from $\mathrm{U}(0, C_i)$. For the generation of panel count data $Y_i(T_{i,j})$, we assumed that

$$Y_i(T_{i,j}) = Y_i^*(T_{i,1}) + Y_i^*(T_{i,2} - T_{i,1}) + \cdots + Y_i^*(T_{i,j} - T_{i,j-1}),$$

with $T_{i,0} = 0$. Also given $Z_i(t)$ and $\mathcal{F}_{it}$, we assumed that $Y_i^*(T_{i,1})$ and $Y_i^*(T_{i,j} - T_{i,j-1})$ followed Poisson distribution with the mean functions

$$g\{\mu_0(T_{i,1})e^{\beta_1 Z_i + \beta_2 H(\mathcal{F}_{iT_{i,1}})}\}$$

and

$$g\{\mu_0(T_{i,j})e^{\beta_1 Z_i + \beta_2 H(\mathcal{F}_{iT_{i,j}})}\} - g\{\mu_0(T_{i,j-1})e^{\beta_1 Z_i + \beta_2 H(\mathcal{F}_{it_{i,j-1}})}\},$$

respectively, $i = 1, \cdots, n, j = 1, \cdots, m_i^*$.

For all situations, we took $W(t) = 1, c = 5, \gamma = 1$, and $H(\mathcal{F}_{it}) = O_i(t-)$. We choose 4 different sets for the true value of $\beta = (\beta_1, \beta_2)$: $(0.1, 0.1), (0.3, 0), (0, 0.1)$, $(0.3, 0.1)$ and considered samples sized $n = 30, 50, 70$. The simulation results in terms of coverage probabilities (CP) and average lengths (AL) of the confidence intervals for $\beta = (\beta_1, \beta_2)$ with $\{g(t) = t, \mu_0(t) = t\}$, $\{g(t) = \log(t), \mu_0(t) = t\}$ and $\{g(t) = t^2, \mu_0(t) = \exp(t)\}$ were calculated, and shown in Table 5.1, Table 5.2 and Table 5.3, respectively. Also two different settings for $\mu_0(t)$ are considered. All simulation results were based on 2,000 repetitions.

From Table 5.1, we see that in almost all cases, the empirical likelihood method outperforms the normal approximation method in terms of coverage probabilities of confidence intervals. The result is noteworthy specially when the sample size is small. The empirical likelihood approach remains stable across different parameter settings. As the sample size increases, the coverage probabilities get closer to the nominal level of 0.95. It is also noticeable that we achieved the better coverage probabilities for empirical likelihood method at the cost of a bigger average length. We obtain similar results in Table 5.2 and Table 5.3.

The results are similar for normal approximation and empirical likelihood with bigger sample sizes. Also the empirical likelihood method is computationally expensive. And the computation would become very time-consuming for the high-dimensional data. Considering these, one may prefer empirical likelihood method when the sample size is relatively small.

5.4 Real Data Analysis

The bladder cancer dataset is a well-known and well-analyzed example of panel count data. This dataset was originated from the cancer follow-up study conducted by Veterans Administration Cooperative Urological Research Group (VACURG) of USA, and was presented in Andrews and Herzberg (1985). In the beginning, 118 patients with superficial bladder tumors were selected for the study. The number of tumors and the size of the largest tumor were recorded. After the tumors were transurethrally removed, the patients were randomly assigned to one of the three treatment groups: Placebo, thiotepa and pyridoxine. The patients were then scheduled to visit the medical centers at pre-determined intervals. However, patients failed to visit the medical centers as scheduled (reasons might include personal, family, health issues). As a result, the visiting times and censoring times differ from patient to patient. At each visit, the visiting times and the number of new bladder tumors since the last visit were recorded. Unfortunately, the exact times of the new tumor occurrences were unknown. The new tumors were transurethrally removed and the patients were followed again until the patient died or the study ended. As two patients had no follow-up visits, they were discarded from the data. The main aim of the study was to investigate the treatment effect on the rate of tumor occurrences.

The study gives us panel count data, which consists of the observation or visit times, number of new tumors between two visits, and three covariates: type of treatment, the number of initial tumors and the size of the largest tumor. As studies showed that pyridoxine was ineffective in reducing the tumor recurrence, we focus our analysis on two other groups: placebo and thiotepa. These two groups have 47 and 38 patients, respectively. The average number of observations for placebo and thiotepa groups are 8.66 and 15.30, respectively. The average number of new tumors found for the placebo and thiotepa groups are 0.70 and 0.23, respectively. Also, the average follow up times for these two groups are 16.8 and 17.4

months, respectively. The complete dataset of these 85 patients can be found in Sun and Zhao (2013).

For the analysis, define covariate $Z' = (Z_1', Z_2', Z_3')'$, where Z_1 is the treatment indicator with $Z_1 = 1$ if the patients are in the thiotepa group and $Z_1 = 0$ if the patients are in the placebo group, Z_2 is the size of the largest tumor and Z_3 is the number of initial tumors. Then β_1, β_2 and β_3 represent the effects of the treatment, the size of the largest tumor, and the number of initial tumors, respectively. In addition, β_4 is the effect of the observation or visit process on the tumor recurrence process.

We apply NA method and our proposed EL method to the above panel count data. We assume that total number of patients' visit may carry some information about the recurrence of bladder tumors, i.e., $H(\mathcal{F}_{it}) = O_i(t-)$ and we choose $W(t) = 1$.

Table 5.4 and Table 5.5 show the result with the link function $g(t) = t$ and $g(t) = \log(t)$, respectively. Each table gives the 95% confidence intervals with the length of the confidence intervals for the normal approximation and empirical likelihood method. The results in Table 5.4 and Table 5.5 are also presented in Figure 5.1 and Figure 5.2, respectively.

The results in Table 5.4 and Table 5.5 suggest that the thiotepa treatment has a significant effect in reducing the recurrence rate of bladder tumor. Also the initial number of tumors has positive significant relation with the recurrence of the tumor, while the largest tumor size has no significant effect. It is also evident from Table 5.4, which considers the link function $g(t) = t$, that the recurrence rate of the tumors is significantly related to the observation process, i.e., the observation does carry important information about the underlying recurrent process. However, with the link function $g(t) = \log(t)$, Table 5.5 gives us a different result for the significance of observation history. In Table 5.5, the normal approximation confidence interval gives a significant positive effect, whereas the empirical likelihood confidence interval includes zero, affirming no significance effect of observation history on the recurrent event. This arises the question of selecting the best link function, $g(t)$. Li et al. (2010) provided a goodness-of-the-fit test that evaluates the model (5.2) for a given g. The results from this goodness-of-the-fit suggest that it is more reasonable to use the link function $g(t) = t$ than $g(t) = \log(t)$. It is also noteworthy that in almost all cases, the empirical likelihood gives us confidence intervals with longer length than the normal approximation confidence intervals. But our simulation study shows that empirical likelihood performs better than the normal approximation does in terms of coverage probability at the cost of longer average length. One may rely more on the empirical likelihood confidence intervals for the small data set.

5.5 Conclusion

We propose an empirical likelihood procedure for panel count data with informative observation times. Motivated by the flexible semiparametric transformation model proposed by Li et al. (2010), the empirical log-likelihood ratio is developed. We proved the Wilks' theorem of this log-likelihood ratio, i.e., the empirical log-likelihood ratio follows limiting chi-square distribution. Using this asymptotic distribution, one can construct a confidence region of the parameters. To get the inference for a subset of parameters, e.g., to get a confidence interval of a parameter, we also formulate a profile empirical likelihood method by profiling nuisance parameters from the full EL.

One advantage of the proposed methods is that they perform better in terms of coverage probabilities. We performed numerical studies under various simulation settings. We found that EL methods have better coverage probabilities than the NA methods, especially when the sample size is small. Also, unlike NA based methods, EL confidence intervals can be constructed without computing complex variance-covariance matrix. However, our proposed method involves solving a non-linear equation, which can be computation-intensive. Other EL methods, such as jackknife empirical likelihood, can be investigated, which can avoid or reduce the computation burden.

A critical aspect of the semiparametric transformation model is to select a robust link function. The same issue also exists for the empirical likelihood. We rely on the same goodness-of-fit test to choose the right candidate for the link function. Further research can be done to find an optimal link function.

Different researchers model panel count data with informative observation times differently. Consequently, the analysis procedures are different. Also, other factors, such as terminal events, follow-up processes, etc., can be included in the new model. Future research may include these extensions into the empirical likelihood procedure.

Table (5.1) Coverage probabilities (average lengths) for normal approximation (NA) and empirical likelihood (EL) with nominal level 0.95, $g(t) = t$ and $\mu_0(t) = t$.

	$\tau = 1$				$\tau = 5$			
	β_1		β_2		β_1		β_1	
(β_1, β_2)	NA	EL	NA	EL	NA	EL	NA	EL
				$n = 30$				
(0.1, 0.1)	0.888 (1.996)	0.894 (2.071)	0.836 (0.236)	0.857 (0.259)	0.885 (1.061)	0.889 (1.092)	0.836 (0.153)	0.869 (0.168)
(0.3, 0)	0.896 (2.444)	0.895 (2.538)	0.850 (0.325)	0.877 (0.353)	0.900 (1.199)	0.902 (1.233)	0.859 (0.180)	0.893 (0.196)
(0, 0.1)	0.901 (2.077)	0.907 (2.156)	0.851 (0.248)	0.887 (0.273)	0.889 (1.076)	0.892 (1.110)	0.831 (0.156)	0.872 (1.172)
(0.3, 0.1)	0.886 (1.916)	0.891 (1.986)	0.847 (0.226)	0.879 (0.248)	0.854 (1.043)	0.860 (1.073)	0.820 (0.149)	0.864 (0.164)
				$n = 50$				
(0.1, 0.1)	0.908 (1.578)	0.911 (1.614)	0.868 (0.188)	0.899 (0.202)	0.922 (0.858)	0.922 (0.877)	0.862 (0.121)	0.890 (0.133)
(0.3, 0)	0.906 (1.927)	0.911 (1.970)	0.882 (0.260)	0.909 (0.276)	0.928 (0.950)	0.931 (0.971)	0.894 (0.145)	0.922 (0.157)
(0, 0.1)	0.919 (1.618)	0.924 (1.653)	0.869 (0.193)	0.905 (0.209)	0.898 (0.860)	0.906 (0.881)	0.851 (0.121)	0.890 (0.133)
(0.3, 0.1)	0.908 (1.513)	0.913 (1.545)	0.853 (0.176)	0.894 (0.190)	0.916 (0.840)	0.915 (0.860)	0.838 (0.119)	0.885 (0.131)
				$n = 70$				
(0.1, 0.1)	0.914 (1.353)	0.916 (1.376)	0.880 (0.162)	0.912 (0.174)	0.907 (0.728)	0.911 (0.745)	0.869 (0.104)	0.908 (0.115)
(0.3, 0)	0.922 (1.654)	0.922 (1.682)	0.902 (0.227)	0.921 (0.238)	0.936 (0.816)	0.935 (0.833)	0.913 (0.126)	0.933 (0.136)
(0, 0.1)	0.920 (1.378)	0.919 (1.402)	0.889 (0.165)	0.913 (0.177)	0.919 (0.745)	0.924 (0.762)	0.886 (0.106)	0.919 (0.116)
(0.3, 0.1)	0.926 (1.300)	0.924 (1.322)	0.878 (0.150)	0.916 (0.163)	0.913 (0.720)	0.918 (0.935)	0.878 (0.104)	0.909 (0.114)

Table (5.2) Coverage probabilities (average lengths) for normal approximation (NA) and empirical likelihood (EL) with nominal level 0.95, $g(t) = log(t)$ and $\mu_0(t) = t$.

	$\tau = 1$				$\tau = 5$			
	β_1		β_2		β_1		β_1	
(β_1, β_2)	NA	EL	NA	EL	NA	EL	NA	EL
				$n = 30$				
(0.1, 0.1)	0.904	0.910	0.844	0.866	0.905	0.927	0.853	0.888
	(1.205)	(1.433)	(0.282)	(0.290)	(1.765)	(1.891)	(0.393)	(0.388)
(0.3, 0)	0.929	0.938	0.869	0.879	0.906	0.909	0.861	0.873
	(1.023)	(1.302)	(0.208)	(0.228)	(1.634)	(1.786)	(0.332)	(0.343)
(0, 0.1)	0.907	0.909	0.845	0.852	0.915	0.917	0.846	0.858
	(1.134)	(1.369)	(0.267)	(0.275)	(1.685)	(1.816)	(0.380)	(0.380)
(0.3, 0.1)	0.906	0.915	0.839	0.845	0.912	0.915	0.842	0.856
	(1.290)	(1.480)	(0.303)	(0.308)	(1.815)	(1.959)	(0.414)	(0.411)
				$n = 50$				
(0.1, 0.1)	0.916	0.921	0.879	0.890	0.926	0.928	0.871	0.877
	(1.954)	(1.311)	(0.234)	(0.253)	(1.424)	(1.662)	(0.328)	(0.337)
(0.3, 0)	0.934	0.940	0.896	0.912	0.921	0.925	0.902	0.910
	(0.803)	(1.196)	(0.166)	(0.213)	(1.298)	(1.581)	(0.274)	(0.298)
(0, 0.1)	0.921	0.925	0.861	0.882	0.923	0.929	0.879	0.889
	(0.913)	(1.309)	(0.226)	(0.251)	(1.388)	(1.643)	(1.323)	(0.334)
(0.3, 0.1)	0.921	0.925	0.866	0.871	0.929	0.930	0.886	0.896
	(1.048)	(1.426)	(0.256)	(0.276)	(1.507)	(1.757)	(0.344)	(0.351)
				$n = 70$				
(0.1, 0.1)	0.937	0.943	0.885	0.901	0.930	0.935	0.893	0.895
	(0.817)	(1.150)	(0.207)	(0.247)	(1.210)	(1.556)	(0.286)	(0.304)
(0.3, 0)	0.933	0.949	0.905	0.921	0.937	0.941	0.898	0.911
	(0.681)	(1.047)	(0.146)	(0.207)	(1.117)	(1.494)	(0.237)	(0.266)
(0, 0.1)	0.929	0.933	0.883	0.899	0.924	0.931	0.904	0.910
	(0.816)	(1.153)	(0.209)	(0.248)	(1.200)	(1.568)	(0.287)	(0.306)
(0.3, 0.1)	0.924	0.931	0.880	0.897	0.921	0.927	0.905	0.906
	(0.910)	(1.387)	(0.225)	(0.259)	(1.281)	(1.607)	(0.302)	(0.316)

Table (5.3) Coverage probabilities (average lengths) for normal approximation (NA) and empirical likelihood (EL) with nominal level 0.95, $g(t) = t^2$ and $\mu_0(t) = \exp(t)$.

	$\tau = 1$				$\tau = 2$			
	β_1		β_2		β_1		β_1	
(β_1, β_2)	NA	EL	NA	EL	NA	EL	NA	EL
				$n = 30$				
(0.1, 0.1)	0.836 (0.494)	0.847 (0.533)	0.826 (0.080)	0.866 (0.105)	0.818 (0.596)	0.840 (0.657)	0.801 (0.088)	0.840 (0.116)
(0.3, 0)	0.893 (0.503)	0.898 (0.525)	0.840 (0.078)	0.865 (0.084)	0.858 (0.462)	0.863 (0.480)	0.840 (0.073)	0.852 (0.078)
(0, 0.1)	0.830 (0.490)	0.845 (0.531)	0.817 (0.078)	0.868 (0.103)	0.817 (0.597)	0.825 (0.661)	0.817 (0.089)	0.943 (0.116)
(0.3, 0.1)	0.867 (0.527)	0.884 (0.581)	0.832 (0.082)	0.871 (0.109)	0.829 (0.663)	0.863 (0.747)	0.803 (0.090)	0.844 (0.121)
				$n = 50$				
(0.1, 0.1)	0.860 (0.399)	0.868 (0.422)	0.836 (0.063)	0.884 (0.084)	0.842 (0.507)	0.855 (0.545)	0.818 (0.072)	0.860 (0.095)
(0.3, 0)	0.917 (0.403)	0.921 (0.414)	0.875 (0.064)	0.891 (0.066)	0.887 (0.375)	0.899 (0.385)	0.857 (0.058)	0.875 (0.060)
(0, 0.1)	0.867 (0.400)	0.873 (0.425)	0.845 (0.063)	0.874 (0.083)	0.835 (0.494)	0.843 (0.531)	0.808 (0.070)	0.860 (0.092)
(0.3, 0.1)	0.873 (0.432)	0.895 (0.466)	0.838 (0.063)	0.886 (0.085)	0.861 (0.556)	0.879 (0.604)	0.808 (0.072)	0.858 (0.098))
				$n = 70$				
(0.1, 0.1)	0.869 (0.351)	0.877 (0.366)	0.833 (0.054)	0.891 (0.071)	0.856 (0.453)	0.876 (0.483)	0.828 (0.061)	0.900 (0.088)
(0.3, 0)	0.924 (0.346)	0.931 (0.354)	0.903 (0.055)	0.912 (0.056)	0.912 (0.327)	0.924 (0.341)	0.883 (0.050)	0.928 (0.058)
(0, 0.1)	0.850 (0.346)	0.870 (0.369)	0.826 (0.054)	0.915 (0.076)	0.821 (0.436)	0.849 (0.469)	0.810 (0.061)	0.898 (0.087)
(0.3, 0.1)	0.892 (0.381)	0.915 (0.414)	0.840 (0.055)	0.922 (0.080)	0.882 (0.494)	0.904 (0.536)	0.829 (0.063)	0.905 (0.091)

Table (5.4) Confidence interval (CI) and length for normal approximation (NA) and empirical likelihood (EL) with nominal level 0.95 with $g(t) = t$.

		NA		EL	
	Estimate (SE)	95% CI	Length	95% CI	Length
β_1	-1.742 (0.380)	(-2.487, -0.997)	1.490	(-2.497, -0.979)	1.518
β_2	-0.101 (0.108)	(-0.313, 0.111)	0.424	(-0.342, 0.103)	0.445
β_3	0.286 (0.066)	(0.157, 0.415)	0.258	(0.161, 0.436)	0.275
β_4	0.053 (0.024)	(0.006, 0.100)	0.094	(0.002, 0.095)	0.093

Table (5.5) Confidence interval (CI) and length for normal approximation (NA) and empirical likelihood (EL) with nominal level 0.95 with $g(t) = \log(t)$.

		NA		EL	
	Estimate (SE)	95% CI	Length	95% CI	Length
β_1	-0.816 (0.372)	(-1.545, -0.087)	1.458	(-2.274, -0.720)	1.554
β_2	0.021 (0.096)	(-0.167, 0.209)	0.376	(-0.320, 0.090)	0.409
β_3	0.212 (0.063)	(0.089, 0.335)	0.246	(0.152, 0.544)	0.392
β_4	0.047 (0.023)	(0.002, 0.092)	0.090	(-0.017, 0.112)	0.129

Figure 5.1: Plot of confidence interval for normal approximation (NA) and empirical likelihood (EL) with nominal level 0.95 with $g(t) = t$.

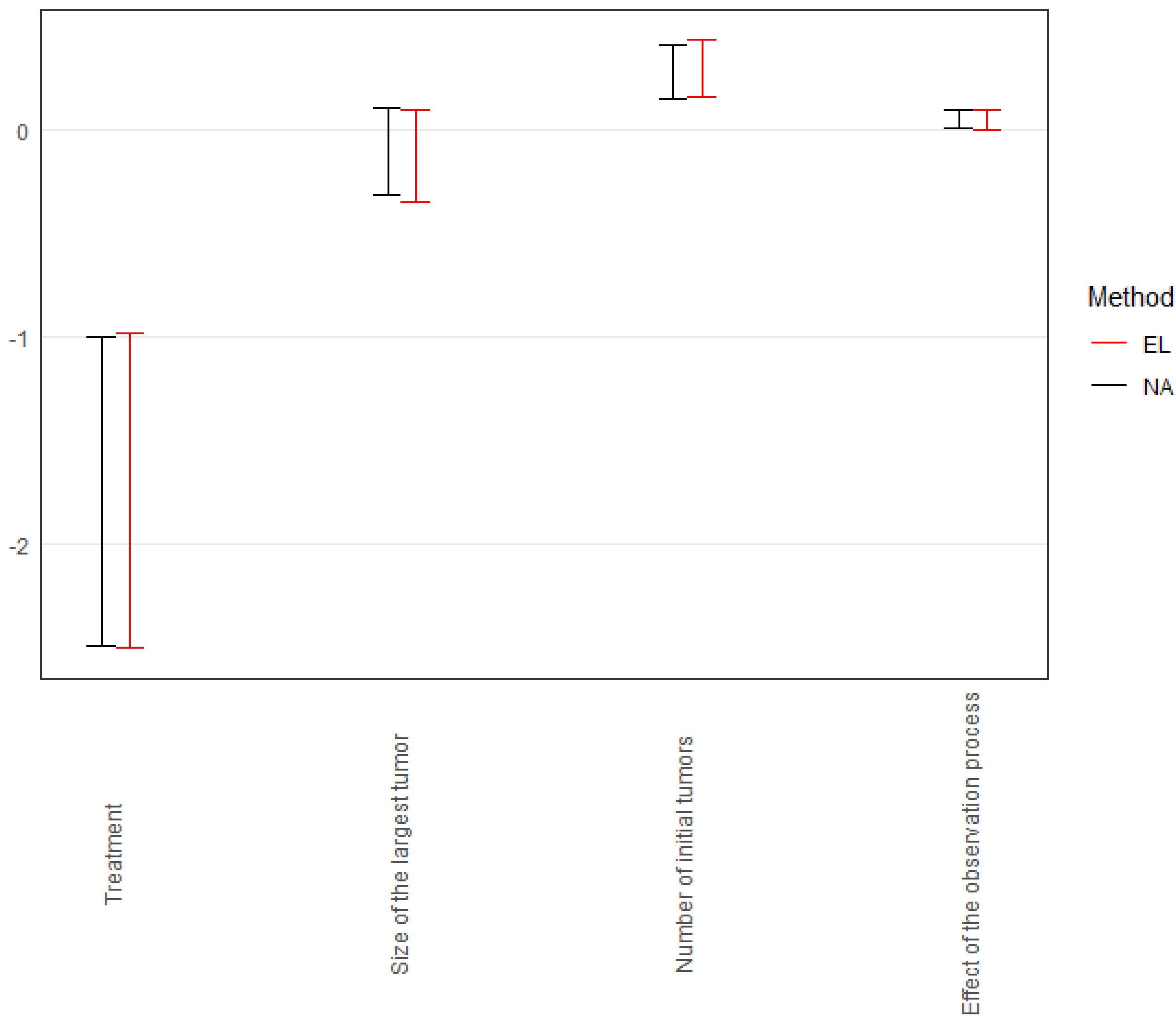
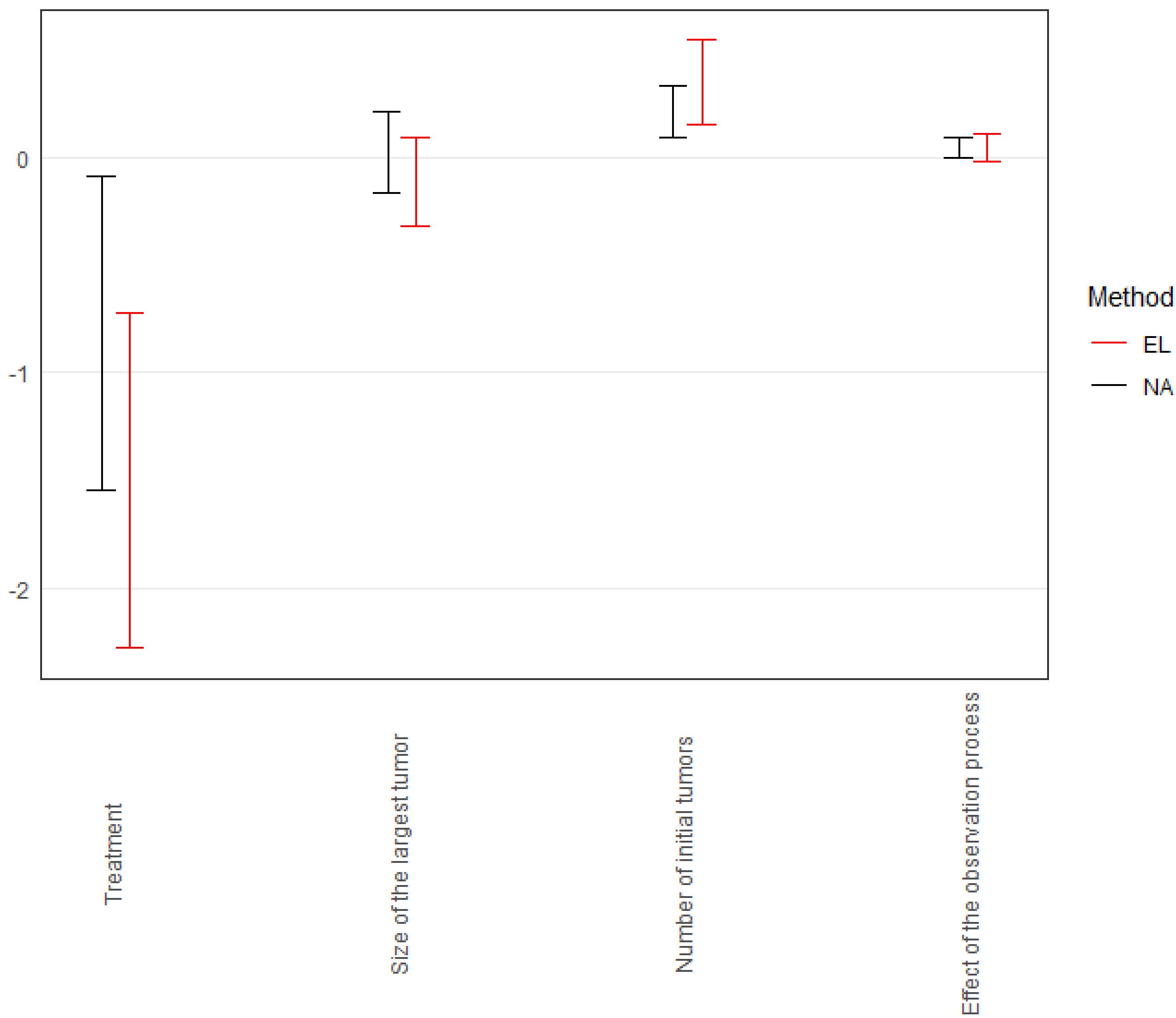


Figure 5.2: Plot of confidence interval for normal approximation (NA) and empirical likelihood (EL) with nominal level 0.95 with $g(t) = \log(t)$.



CHAPTER 6

EMPIRICAL LIKELIHOOD INFERENCE FOR THE TRANSFORMATION MODEL WITH THE CASE-COHORT STUDY

6.1 Background

Case-cohort study, proposed by Prentice (1986), is designed to use the information from a cohort study in a more cost-efficient and flexible way. Under this study design, a random subsample is selected from the full cohort. Covariate information is collected only from the chosen subsample and any cases, i.e., any subject who experiences the event of interest. This design efficiently uses covariate information from the redundant full cohort. It also reduces the selection and information bias as all cases and controls are sampled from the same population, and the investigators are blind to the case status. Other advantages include studying multiple outcomes using the same subcohort, and measuring risk at any time up to a given time.

Prentice (1986) design was based on the Cox model (Cox (1972)) under the proportional hazards model assumption. He adopted a pseudo-likelihood approach to estimate the relative risk regression parameters. Self and Prentice (1988) further justified the method by giving asymptotic distribution theory and efficiency results using martingale and finite population convergence results. Lin and Ying (1993) extended the pseudo-likelihood approach in more general missing covariate data under the Cox Model, which includes the case-cohort design as a special case. These estimators are later improved by Chen and Lo (1999), which gives more asymptotic efficiency. Chen (2001b) proposed a unified approach to generalize the case-cohort study that can include nested case-cohort, case-cohort, and classical case-control designs. As the assumptions for proportional hazards models may not be valid in some situations, there were other studies that incorporate different survival models. Kulich and Lin (2000) proposed an estimation procedure for the additive hazards regression model for the case-cohort design. The proportional odds model under a case-cohort study is investigated by Chen (2001b). Nan et al. (2006) and Kong and Cai (2009) developed a semiparametric accelerated failure time model for the case-cohort study. Recently, Sun et al. (2016) discussed a class of general additive-multiplicative hazard models which includes the Cox model and additive hazard model as special cases.

A more general semiparametric linear transformation model has also been used for case-cohort studies. The semiparametric transformation model is given by

$$H(T) = -\beta' Z + \epsilon, \tag{6.1}$$

where H is an unknown monotone increasing function, T is the failure time, β is a p dimensional unknown vector of the parameter of interest, Z is a vector of covariates, and ϵ is a random variable with a known continuous distribution. ϵ is assumed to be independent of both Z and the censoring time C. The well-known proportional hazards model and the proportional odds model are special cases of the model (6.1). Let $\Lambda(t)$ be the cumulative hazard function for ϵ, i.e., $\mathrm{pr}(\epsilon > t) = \exp\{-\Lambda(t)\}$ with $\Lambda(-\infty) = 0$ and $t \in (-\infty, \infty)$. Model (6.1) becomes a proportional hazards model if $\Lambda(t) = \exp(t)$, i.e., ϵ follows an extreme value distribution, and proportional odds model if $\Lambda(t) = \log(1 + \exp(t))$, i.e., ϵ follows a logistic distribution.

Several inference procedures exist for the semiparametric transformation model under the covariate independent censoring assumption. Cheng et al. (1995) and Fine et al. (1998) proposed inference procedures based on rank-based estimating equations, which is later extended by Kong et al. (2004) for the classical case-cohort design. Also, a conditional profile likelihood-based method is proposed for a modified version of the case-cohort study in which the censoring times of all the censored subjects in the cohort are observed (cf. Chen (2001a)). But the assumption of censoring variable and covariate may not be appropriate in many situations. Chen et al. (2002) proposed martingale based estimating equations, which can be applied with a covariate-censoring dependency assumption. Lu and Tsiatis

(2006) adopted this martingale-based inference method for the case-cohort design. On the other hand,Chen and Zucker (2009) proposed approximate profile likelihood and pseudo-partial likelihood methods for full-cohort data and then extended those to accommodate the case-cohort design.

In this chapter, we apply an empirical likelihood (EL) method to make inference about the regression parameters of the case-cohort design based on martingale representation by Lu and Tsiatis (2006). The empirical likelihood method was first introduced by Owen (1988, 1990) as a nonparametric approach to constructing confidence regions for the parameter of interest. It is a nonparametric likelihood approach, meaning that it enjoys the benefit of the likelihood method without any parametric distributional assumptions. Also, this method does not require complex variance computation to construct confidence intervals. Other key advantages of EL include transformation invariance, Bartlett correctability (Diciccio et al. (1991), data determining the shape of the confidence region, etc. Interested readers are referred to the comprehensive review by Owen (2001) and Chen and Van Keilegom (2009). Because of all the appealing and useful properties, the empirical likelihood method has been used in many statistical analysis areas. A class of functionals of survival functions was studied by Wang and Jing (2001). Li and Wang (2003) proposed EL for linear regression for right-censored data using a synthetic data approach. Zhou (2016) explored some important survival analysis areas using empirical likelihood. Lu and Liang (2006) developed the empirical likelihood method for censored survival data under the linear transformation models. They showed that the limiting distribution of the empirical likelihood ratio is a weighted sum of the standard chi-squared distributions. Yu et al. (2011) later modified that EL algorithm in which the limiting distribution of the empirical likelihood ratio follows standard chi-squared distribution. Other research on the empirical likelihood for linear transformation models includes Shen (2012), Zhang and Zhao (2013), Yang et al. (2017). So far, there is no work for the linear transformation model for the case-cohort study design.

The rest of the chapter is organized as follows. We set up the model and define the counting processes in Section 6.2.1. Section 6.2.2 gives a review of the existing normal approximation method based on inference techniques by the weighted estimation equation. In Sections 6.2.3 and 6.2.4, we develop the empirical likelihood and the adjusted empirical likelihood methods, respectively, for a vector of regression parameters, along with the profile empirical likelihood method for a subset of regression parameters. Extensive simulation studies are carried out in Section 6.3. A real dataset is used to apply our method in Section 6.4. Lastly, a short discussion is given in Section 6.5. All the technical details and proofs are provided in the Appendix E.

6.2 Main Results

6.2.1 Model setup

We use similar notations as in Lu and Tsiatis (2006). Let V and C represent the time to failure from the onset and the censoring time, respectively. For n subjects, one can observe the complete data as n i.i.d. random vectors $T_i, \delta_i, Z_i, i = 1, 2, \ldots, n$, where $T_i = \min(V_i, C_i)$ and $\delta_i = I(V_i \le C_i)$. Using the usual counting process notations, for $i = 1, 2, \ldots, n$, we can define the following

$$\begin{aligned} N_i(t) &= I(T_i \le t, \delta_i = 1) \\ Y_i(t) &= I(T_i \ge t), \end{aligned}$$

and a martingale process (Anderson et al., 1993)

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

where (β_0, H_0) are the true value of (β, H).

In the classical case-cohort study design (see Prentice (1986)), a random sample of size n_s, named subcohort, is selected from the full cohort without replacement. Let ξ_i =I(subject i is in the subcohort) be the subcohort indicator. The covariate information Z_i is collected for all the subjects in the subcohort, i.e., $\xi_i = 1$, and all remaining cases, i.e., $\xi_i = 0$ and

$\delta_i = 1$. We assume that the censoring time C is independent of T, given Z. Then, one can observe $[T_i, \delta_i, \xi_i, \{\delta_i + (1-\delta_i)\xi_i\}Z_i]$ for $i = 1, 2, \ldots, n$. Note that ξ_i is independent of $(T_i, \delta_i, Z_i), i = 1, 2, \ldots, n$, but the ξ_i's, $i = 1, 2, \ldots, n$ are dependent because of the sampling without replacement.

Let the probabilities for the subjects to be selected in the subsample be equal and denoted by $\tilde{p}, \tilde{p} \in (0, 1)$. Because of sampling without replacement, we can write $p = n_s/n$. Here as in Lu and Tsiatis (2006), we assume that $\tilde{p}$ converges to a positive constant p as $n_s, n \to \infty$. Let the limit of the $\tilde{p}$ be denoted by p.

6.2.2 Normal approximation method

A common method to utilize the available information from the full-cohort data is to use the inverse selection probabilities to define a weight for each subject in the full-cohort. In this section, we review the normal approximation method proposed by Lu and Tsiatis (2006). We adopt the similar notations as Lu and Tsiatis (2006) did. As in Kong et al. (2004), Lu and Tsiatis (2006), we define a weight $\rho_i = \delta_i + (1-\delta_i)\xi_i/\tilde{p}$ for each subject in the full cohort by the inverse selection probabilities. Here ρ_i takes

$$\rho_i = \begin{cases} 1 & \text{if } \delta_i = 1; \\ 1/\tilde{p} & \text{if } \delta_i = 0, \xi_i = 1\ ; \\ 0 & \text{if } \delta_i = 0, \xi_i = 0. \end{cases}$$

Using the results from Robins et al. (1994), we can replace $\tilde{p}$ by its empirical estimator p and we have weights $\pi_i = \delta_i + (1-\delta_i)\xi_i/p$ for each subject $i, i = 1, 2, \ldots, n$, that allow us to use information from the full-cohort data.

Lu and Tsiatis (2006) proposed the following estimating equations

$$\sum_{i=1}^{n} \pi_i[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \beta' Z_i\}] = 0, (0 \leq t \leq \tau), H(0) = -\infty, \tag{6.2}$$

$$U(\beta, H) = \sum_{i=1}^{n} \int_0^{\infty} Z_i \pi_i [dN_i(t) - Y_i(t) d\Lambda\{H(t) + \beta' Z_i\}] = 0 \qquad (6.3)$$

to estimate H and β, respectively. Here $\tau = \inf\{t : \text{pr}(V_i > t) = 0\}$. Equations (6.2) and (6.3) can be solved iteratively using a similar algorithm by Chen et al. (2002). Let $(\hat{\beta}, \hat{H})$ be the solution of (6.2) and (6.3). Note that $\hat{H}$ is estimated as a nondecreasing step function with jumps only at the observed failure times. Denote $\dot{\lambda}(t) = d/dt(\lambda)$. For any $t, s \in (0, \tau]$, define

$$B_1(\beta, t) = E[\dot{\lambda}\{H_0(t) + \beta' Z_1\} Y_1(t)],$$

$$B_2(\beta, t) = E[\lambda\{H_0(t) + \beta' Z_1\} Y_1(t)],$$

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

$$\mu_Z(\beta, t) = \frac{E[Z_1 \lambda\{H_0(T_1) + \beta' Z_1\} Y_1(t) B(\beta, t, T_1)]}{E[\lambda\{H_0(t) + \beta' Z_1\} Y_1(t)]},$$

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

$$A(\beta) = \int_0^{\tau} E[\{Z_1 - \mu_Z(\beta, t)\} Z_1' \dot{\lambda}\{H_0(t) + \beta' Z_1\} Y_1(t)] dH_0(t),$$

$$\Sigma(\beta) = E\left(\{\delta_i + (1 - \delta_i)/\tilde{p}\} \left[\int_0^{\tau} \{Z_1 - \mu_Z(\beta, t)\} dM_1(\beta, t) \right]^{\otimes 2} \right)$$
$$- \frac{1 - \tilde{p}}{\tilde{p}} \left(E\left[\int_0^{\tau} \{Z_1 - \mu_Z(\beta, t)\} dM_1(\beta, t) \right] \right)^{\otimes 2}.$$

Then under the regularity conditions given in the Appendix E, Lu and Tsiatis (2006) showed that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathfrak{D}} N(0, (A(\beta_0))^{-1} \Sigma(\beta_0) ((A(\beta_0))^{-1})'),$$

where $v^{\otimes 2} = vv'$, for any real vector v.

The consistent estimators of A and Σ, denoted by $\hat{A}(\hat{\beta})$ and $\hat{\Sigma}(\hat{\beta})$, respectively, can be obtained by

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

$$\begin{aligned}\hat{\Sigma}(\hat{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\pi_i^2\bigg[\int_0^{\tau}\{Z_i - \bar{Z}(\hat{\beta},t)\}d\hat{M}_i(\hat{\beta},t)\bigg]^{\otimes 2} \\ &\quad -\frac{1-p}{p}\bigg[\frac{1}{n}\sum_{i=1}^{n}\delta_i\int_0^{\tau}\{Z_i - \bar{Z}(\hat{\beta},t)\}d\hat{M}_i(\hat{\beta},t)\bigg]^{\otimes 2},\end{aligned}$$

respectively,

where for any $s, t \in (0, \tau]$

$$\bar{Z}(\beta,t) = \frac{\sum_{i=1}^{n}\pi_i Z_i\lambda\{\hat{H}(\beta,T_i) + \beta'Z_i\}Y_i(t)\hat{B}(\beta,t,T_i)}{\sum_{i=1}^{n}\pi_i\lambda\{\hat{H}(\beta,T_i) + \beta'Z_i\}Y_i(t)},$$

$$\hat{B}(\beta,t,u) = \exp\bigg[\int_u^t \big\{\hat{B_2}^{-1}(\beta,u)\hat{B}_1(\beta,u)\big\}d\hat{H}(\beta,u)\bigg],$$

$$\hat{B}_1(\beta,t) = n^{-1}\sum_{i=1}^{n}\pi_i\dot{\lambda}\{\hat{H}(\beta,t) + \beta'Z_i\}Y_i(t),$$

$$\hat{B}_2(\beta,t) = n^{-1}\sum_{i=1}^{n}\pi_i\lambda\{\hat{H}(\beta,t) + \beta'Z_i\}Y_i(t),$$

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

Then, the $100(1-\alpha)\%$ normal approximation based confidence region for β developed by Lu and Tsiatis (2006) can be constructed as

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

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

6.2.3 Empirical likelihood method

Motivated by the estimating equation (6.3), we propose

$$U_{ni}(\beta) = \int_0^t \pi_i(Z_i - \bar{Z}(\beta, t))d\hat{M}_i(\beta, t), i = 1, 2, \ldots, n,$$

where $\bar{Z}(\beta, t)$ and $d\hat{M}_i(t)$ are defined in the previous section.

For a probability vector w_i, the empirical likelihood ratio, evaluated at β is defined as:

$$R(\beta) = \sup\left\{ \prod_{i=1}^{n} nw_i : \sum_{i=1}^{n} w_i U_{ni}(\beta) = 0, \sum_{i=1}^{n} w_i = 1, w_i \geq 0, i = 1, 2, \ldots, n \right\}.$$

The empirical log-likelihood ratio at β is defined as

$$\begin{aligned} l(\beta) &= -2logR(\beta) \\ &= 2\sum_{i=1}^{n} \log\{1 + (\theta(\beta))' U_{ni}(\beta)\}, \end{aligned}$$

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

$$\sum_{i=1}^{n} \frac{U_{ni}(\beta)}{1 + (\theta(\beta))' U_{ni}(\beta)} = 0. \tag{6.4}$$

Thus, the Wilks' theorem for the empirical log-likelihood ratio is given in Theorem 6.1.

Theorem 6.1. *Let β_0 be the true value of β. Under the regularity conditions given in the Appendix E, as $n \to \infty$,*

$$l(\beta_0) \xrightarrow{\mathfrak{D}} \chi_p^2,$$

where χ_p^2 is a standard chi-square distribution with p degrees of freedom.

Based on Theorem 6.1, we can construct the asymptotic $100(1-\alpha)\%$ empirical likelihood confidence region for β_0 as

$$R^{EL}(\alpha) = \{\beta : l(\beta) \leq \chi_p^2(\alpha)\},$$

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

In practice, only a part of the parameters may be of interest. Then one can use profile empirical likelihood (Subramanian (2007), Yu et al. (2011)) by profiling out nuisance parameters from the full EL. Define $\beta_0 = ((\beta_0^{(1)})', (\beta_0^{(2)})')'$. Suppose we want to construct EL confidence regions for a q-dimensional ($q < p$) subvector $\beta_0^{(1)}$. This can be done by profiling out $\beta^{(2)}$ from the full EL. Then the profile EL ratio at $\beta^{(1)}$ is

$$l_{\text{profile}}(\beta^{(1)}) = \min_{\beta^{(2)}} l((\beta^{(1)})', (\beta^{(2)})').$$

We establish Wilks' theorem for $l_{\text{profile}}(\beta^{(1)})$ as follows:

Theorem 6.2. *Under the regularity conditions given in the Appendix E, as $n \to \infty$,*

$$l_{profile}(\beta_0^{(1)}) \xrightarrow{\mathfrak{D}} \chi_q^2,$$

where χ_q^2 is a standard chi-square distribution with q degrees of freedom.

Based on Theorem 6.2, we can construct the asymptotic $100(1-\alpha)\%$ empirical likelihood confidence region for $\beta_0^{(1)}$ as

$$R_{\text{profile}}^{EL}(\alpha) = \{\beta^{(1)} : l_{\text{profile}}(\beta^{(1)}) \leq \chi_q^2(\alpha)\},$$

where $\chi_q^2(\alpha)$ is the $(1-\alpha)$ quantile of χ_q^2.

6.2.4 Adjusted empirical likelihood method

The EL method requires us to find the Lagrange multiplier $\theta(\beta)$ by solving equation (6.4). This may create the convex hull problem or the empty set problem if 0 is not in the convex hull of $\{U_{ni}(\beta), i = 1, 2, \ldots, n\}$. To improve the coverage probability of the empirical likelihood methods, Chen et al. (2008) proposed the adjusted empirical likelihood (AEL) method.

In the AEL method, one artificial data point is added into the data set and we apply the EL method on the new data set. Define the artificial point as

$$U_{n(n+1)}(\beta) = -\frac{a_n}{n}\sum_{i=1}^{n} U_{ni}(\beta),$$

for some positive constant a_n. The general recommendation for a_n is to take $a_n = \max(1, \log(n)/2)$. With the artificial data point, we have the new data set $U_{ni}^{ad}(\beta), i = 1, 2, \ldots, (n+1)$.

Applying EL method on these $(n+1)$ data points, we have the adjusted empirical likelihood ratio function, evaluate at β as

$$R^{ad}(\beta) = \sup\Big\{\prod_{i=1}^{n+1}(n+1)w_i : \sum_{i=1}^{n+1} w_i U_{ni}^{ad}(\beta) = 0, \sum_{i=1}^{n+1} w_i = 1, w_i \geq 0, i = 1, 2, \ldots, n+1\Big\}.$$

Therefore, the adjusted empirical log-likelihood ratio at β is

$$\begin{aligned} l^{ad}(\beta) &= -2logR^{ad}(\beta) \\ &= 2\sum_{i=1}^{n+1} \log\{1 + (\theta^{ad}(\beta))' U_{ni}^{ad}(\beta)\}, \end{aligned}$$

where $\theta^{ad}(\beta)$ is the solution to

$$\sum_{i=1}^{n+1} \frac{U_{ni}^{ad}(\beta)}{1 + (\theta^{ad}(\beta))' U_{ni}^{ad}(\beta)} = 0. \tag{6.5}$$

Theorem 6.3. *Under the regularity conditions given in the Appendix E, as $n \to \infty$,*

$$l^{ad}(\beta_0) \xrightarrow{\mathfrak{D}} \chi_p^2,$$

where χ_p^2 is a standard chi-square distribution with p degrees of freedom.

Based on Theorem 6.3, we can construct the asymptotic $100(1-\alpha)\%$ adjusted empirical likelihood confidence region for β_0 as

$$R^{AEL}(\alpha) = \{\beta : l^{ad}(\beta) \leq \chi^2_p(\alpha)\},$$

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

Similar to Section 6.2.3, we derive the Wilks' Theorem for the profile AEL ratio in Thoerem 6.4.

Theorem 6.4. *Under the regularity conditions given in the Appendix E, as $n \to \infty$,*

$$l^{ad}_{profile}(\beta_0^{(1)}) \xrightarrow{\mathfrak{D}} \chi^2_q,$$

where χ^2_q is a standard chi-square distribution with q degrees of freedom.

Based on Theorem 6.4, we can construct the asymptotic $100(1-\alpha)\%$ adjusted empirical likelihood confidence region for $\beta_0^{(1)}$ as

$$R^{AEL}_{\text{profile}}(\alpha) = \{\beta^{(1)} : l^{ad}_{\text{profile}}(\beta^{(1)}) \leq \chi^2_q(\alpha)\},$$

where $\chi^2_q(\alpha)$ is the $(1-\alpha)$ quantile of χ^2_q.

6.3 Simulation Study

Extensive simulation studies are carried out to assess the finite sample performance of the proposed methods and compare the results with the performance of the existing NA method in Lu and Tsiatis (2006). As in Dabrowska and Doksum (1988), we assume that the hazard function of the error term ϵ is of the form $\lambda(t) = \exp(t)/\{1 + \gamma \exp(t)\}$, with $\gamma = 0, 1, 2$. Note that $\lambda(t)$ with $\gamma = 0$ and 1 corresponds to the proportional hazards and the proportional odds model, respectively. The transformation function $H(t)$ is chosen to be $\log(t)$, $\log(e^t - 1)$ and $\log(0.5e^{2t} - 0.5)$ for $\gamma = 0, 1$, and 2, respectively. Two covariates are generated independently as $Z_1 \sim U(0, 1)$ and $Z_2 \sim \text{Ber}(0.5)$.

For the first set of simulations, the regression parameters are set to be $\beta = (\beta_1, \beta_2)' = (1, 1)'$. The censoring scheme is chosen to be covariate independent. The censoring times are generated from $U(0, c)$, where c is chosen such that the expected proportion of censoring rate is 70% and 85%. Sample sizes 100 and 200 are considered. Two case-cohort study designs are considered. In one setting, a case-cohort study is designed to have the same number of cases and controls (denoted by CCI setting) and in the other setting, twice the number of controls as cases are taken (denoted by CCII setting). A full cohort study, where all the cases and controls are selected, is also considered (denoted by FULL setting). All simulation results are based on 2500 repetitions.

Table 6.1 shows the results from the first simulation studies. The results are shown in terms of coverage probabilities of the 95% confidence intervals and the corresponding average length in the parenthesis. For the proportional hazard model, i.e., when $\gamma = 0$, the normal approximation method has under coverage. The empirical likelihood method improves the coverage probabilities, which can further be improved by the adjusted empirical likelihood method. However, the average lengths of the confidence intervals also increase at the same time. As the sample size increases, the coverage probabilities approach the nominal level with smaller average lengths. When $\gamma = 1$ or $\gamma = 2$, the simulation results are different as, in general, NA method exhibits over coverage probabilities. Applying EL method in such cases gives under-coverage probabilities, but AEL method provides coverage probabilities that are closer to the nominal level.

The second set of simulation studies consider covariate dependent scheme. Here the censoring times are generated from $\beta_c Z_1/5 + U[0, c]$, where $\beta_c = 0.1$ or 0.3, and c is chosen such that the expected proportion of censoring rate is 75%. Two sets of regression parameters, $\beta = (\beta_1, \beta_2)' = (1, 1)'$ and $\beta = (\beta_1, \beta_2)' = (0, 0)'$ are chosen. Only the first case-cohort study design (CCI) is used for this simulation. All other simulation settings are same as the first simulation settings. The results are summarized in Table 6.2. When $\gamma = 0$, the empirical likelihood and the adjusted empirical likelihood methods improve the coverage probabilities. When $\gamma = 1$ or $\gamma = 2$, EL and AEL methods give coverage probabilities closer to 95% nominal level. In general, our proposed methods perform better even with dependent covariate structure.

6.4 Real Data Analysis

To illustrate the practical usage, we apply our proposed methods to the Wilms' Tumor Study (D'Angio et al. (1989); Green et al. (1998)) dataset. Wilms' tumor or nephroblastoma is a rare renal cancer that starts in the kidney precursor cells. Children under 5 are most susceptible to this type of cancer. Wilms' tumor patient with favorable histology (FH) is easier to treat than patients with unfavorable histology (UH). This histological classification of tumor is an important factor that affects the survival and tumor recurrence. Other factors include the cancer spread stage (I-IV) and tumor diameter. There are total $3,915$ subjects in the full cohort and cancer recurrence rate among them is 17%. In our analysis, we treat cancer recurrence as cases and consider 4 indicator variables I(state II), I(stage III), I(stage IV), I(Central UH), and one continuous variable tumor diameter as covariates in our analysis. In this analysis, we want to evaluate the significance of these covariates on tumor recurrence.

The assessment of histology of the tumor was done in two phases: first by pathologists at different local sites and then by experts in a centralized location. The central reevaluation of tumor histology is considered "accurate", but it is an expensive and timely process. The case-cohort sampling scheme can be very useful and cost-effective as we only need to have central pathological histology for cases and for the subsample instead of all patients. In our analysis, we implement "CCI" sampling scheme, i.e., a subsample of size 669 is randomly selected. The central pathological histology and other covariate information are available for all cases and subsample. Same as in simulation studies, we consider three regression models with $\gamma = 0, 1$, and 2. We applied EL and AEL methods to find 95% confidence intervals for the parameters. Then, we compare the results with the confidence intervals by normal approximation proposed by Lu and Tsiatis (2006).

The confidence intervals and their corresponding confidence lengths obtained by aforementioned three models are presented in Table 6.3 and Figure 6.1. From all the three models, covariates Stages II, stage III, stage IV and UH are found to have significant positive association with the tumor recurrence. The other covariate, Diameter, has no significant effect.

6.5 Conclusion

In this chapter, empirical likelihood and adjusted empirical likelihood methods are proposed to make inferences for the semiparametric transformation model for the case-cohort study. A case-cohort study is a very useful technique to save time and cost by efficiently using the covariates without losing the key information. Using the martingale representation, we developed log-likelihood ratio test statistics for both empirical likelihood and adjusted empirical likelihood methods. The limiting distributions of these test statistics follow a standard chi-square distribution, which enables our methods to conveniently make inference about the model parameters by avoiding the variance calculation. Our simulation results depict the usefulness of the proposed methods. Also, we implement our methods to Wilms's tumor dataset to show the practical usage in real life.

Table (6.1) Coverage probabilities (average lengths) for normal approximation (NA), empirical likelihood (EL), adjusted empirical likelihood (AEL) with nominal level 0.95 and covariate independent censoring.

			CR=70%			CR=85%		
			NA	EL	AEL	NA	EL	AEL
$n = 100$								
CCI	$\gamma = 0$	β_1	0.913 (3.484)	0.921 (3.713)	0.928 (3.803)	0.881 (5.977)	0.894 (6.555)	0.903 (6.735)
		β_2	0.924 (2.088)	0.930 (2.260)	0.931 (2.313)	0.913 (3.482)	0.919 (3.965)	0.926 (4.071)
	$\gamma = 1$	β_1	0.937 (4.676)	0.929 (4.693)	0.938 (4.793)	0.930 (7.823)	0.930 (7.282)	0.936 (7.475)
		β_2	0.943 (2.649)	0.944 (2.693)	0.950 (2.920)	0.934 (4.223)	0.928 (4.294)	0.937 (4.413)
	$\gamma = 2$	β_1	0.968 (5.783)	0.934 (5.436)	0.938 (5.573)	0.938 (9.955)	0.930 (8.098)	0.941 (8.293)
		β_2	0.962 (3.200)	0.938 (3.107)	0.942 (3.187)	0.940 (5.472)	0.935 (4.721)	0.944 (4.866)
CCII	$\gamma = 0$	β_1	0.928 (2.925)	0.931 (3.032)	0.939 (3.115)	0.927 (4.928)	0.930 (5.383)	0.936 (5.550)
		β_2	0.930 (1.767)	0.936 (1.841)	0.942 (1.892)	0.925 (2.917)	0.926 (3.322)	0.932 (3.416)
	$\gamma = 1$	β_1	0.961 (3.728)	0.942 (3.654)	0.946 (3.745)	0.963 (6.077)	0.932 (5.886)	0.938 (6.064)
		β_2	0.956 (2.124)	0.941 (2.124)	0.944 (2.171)	0.962 (3.713)	0.938 (3.505)	0.944 (3.614)
	$\gamma = 2$	β_1	0.971 (4.679)	0.942 (4.438)	0.947 (4.559)	0.975 (7.280)	0.933 (6.548)	0.938 (6.744)
		β_2	0.960 (2.595)	0.943 (2.538)	0.949 (2.606)	0.972 (4.847)	0.930 (3.817)	0.937 (3.936)
FULL	$\gamma = 0$	β_1	0.940 (2.557)	0.941 (2.640)	0.946 (2.712)	0.939 (3.687)	0.941 (3.888)	0.945 (4.012)
		β_2	0.958 (1.575)	0.947 (1.638)	0.954 (1.684)	0.946 (2.342)	0.945 (2.606)	0.947 (2.703)
	$\gamma = 1$	β_1	0.959 (3.356)	0.946 (3.302)	0.948 (3.388)	0.963 (4.419)	0.937 (4.374)	0.941 (4.508)
		β_2	0.959 (1.934)	0.945 (1.939)	0.952 (1.990)	0.968 (2.686)	0.945 (2.838)	0.951 (2.932)
	$\gamma = 2$	β_1	0.966 (4.094)	0.938 (3.899)	0.942 (4.004)	0.972 (5.130)	0.935 (4.859)	0.940 (5.006)
		β_2	0.971 (2.298)	0.952 (2.254)	0.955 (2.314)	0.975 (3.182)	0.944 (3.016)	0.949 (3.113)
$n = 200$								
CCI	$\gamma = 0$	β_1	0.936 (2.457)	0.938 (2.563)	0.942 (2.610)	0.906 (4.003)	0.913 (4.207)	0.918 (4.269)
		β_2	0.942 (1.465)	0.942 (1.554)	0.945 (1.584)	0.932 (2.346)	0.934 (2.503)	0.939 (2.545)
	$\gamma = 1$	β_1	0.954 (3.177)	0.942 (3.146)	0.945 (3.189)	0.941 (4.697)	0.939 (4.790)	0.943 (4.847)
		β_2	0.952 (1.811)	0.948 (1.816)	0.950 (1.843)	0.947 (2.657)	0.952 (2.853)	0.954 (2.897)
	$\gamma = 2$	β_1	0.955 (3.833)	0.943 (3.731)	0.947 (3.784)	0.957 (5.371)	0.944 (5.227)	0.948 (5.307)
		β_2	0.958 (2.162)	0.946 (2.141)	0.949 (2.167)	0.959 (2.972)	0.945 (2.936)	0.947 (2.979)
CCII	$\gamma = 0$	β_1	0.931 (2.033)	0.934 (2.082)	0.939 (2.111)	0.941 (3.283)	0.942 (3.395)	0.944 (3.445)
		β_2	0.949 (1.223)	0.949 (1.256)	0.951 (1.272)	0.942 (1.956)	0.943 (2.043)	0.945 (2.074)
	$\gamma = 1$	β_1	0.955 (2.593)	0.946 (2.583)	0.953 (2.619)	0.961 (3.688)	0.944 (3.693)	0.947 (3.747)
		β_2	0.953 (1.487)	0.951 (1.501)	0.954 (1.523)	0.957 (2.120)	0.943 (2.151)	0.947 (2.184)
	$\gamma = 2$	β_1	0.962 (3.114)	0.952 (3.048)	0.956 (3.092)	0.967 (4.299)	0.949 (4.223)	0.952 (4.288)
		β_2	0.955 (1.761)	0.949 (1.747)	0.951 (1.772)	0.969 (2.406)	0.949 (2.394)	0.950 (2.430)
FULL	$\gamma = 0$	β_1	0.945 (1.799)	0.947 (1.836)	0.949 (1.862)	0.947 (2.568)	0.948 (2.644)	0.952 (2.684)
		β_2	0.952 (1.100)	0.948 (1.126)	0.950 (1.143)	0.955 (1.615)	0.949 (1.688)	0.950 (1.714)
	$\gamma = 1$	β_1	0.951 (2.285)	0.945 (2.270)	0.949 (2.303)	0.961 (2.947)	0.952 (2.967)	0.953 (3.007)
		β_2	0.955 (1.328)	0.942 (1.332)	0.946 (1.351)	0.959 (1.771)	0.957 (1.804)	0.958 (1.831)
	$\gamma = 2$	β_1	0.957 (2.757)	0.950 (2.699)	0.954 (2.738)	0.965 (3.321)	0.948 (3.278)	0.950 (3.327)
		β_2	0.962 (1.571)	0.954 (1.560)	0.957 (1.582)	0.965 (1.937)	0.944 (1.930)	0.947 (1.970)

Table (6.2) Coverage probabilities (average lengths) for normal approximation (NA), empirical likelihood (EL), adjusted empirical likelihood (AEL) with nominal level 0.95 and covariate dependent censoring.

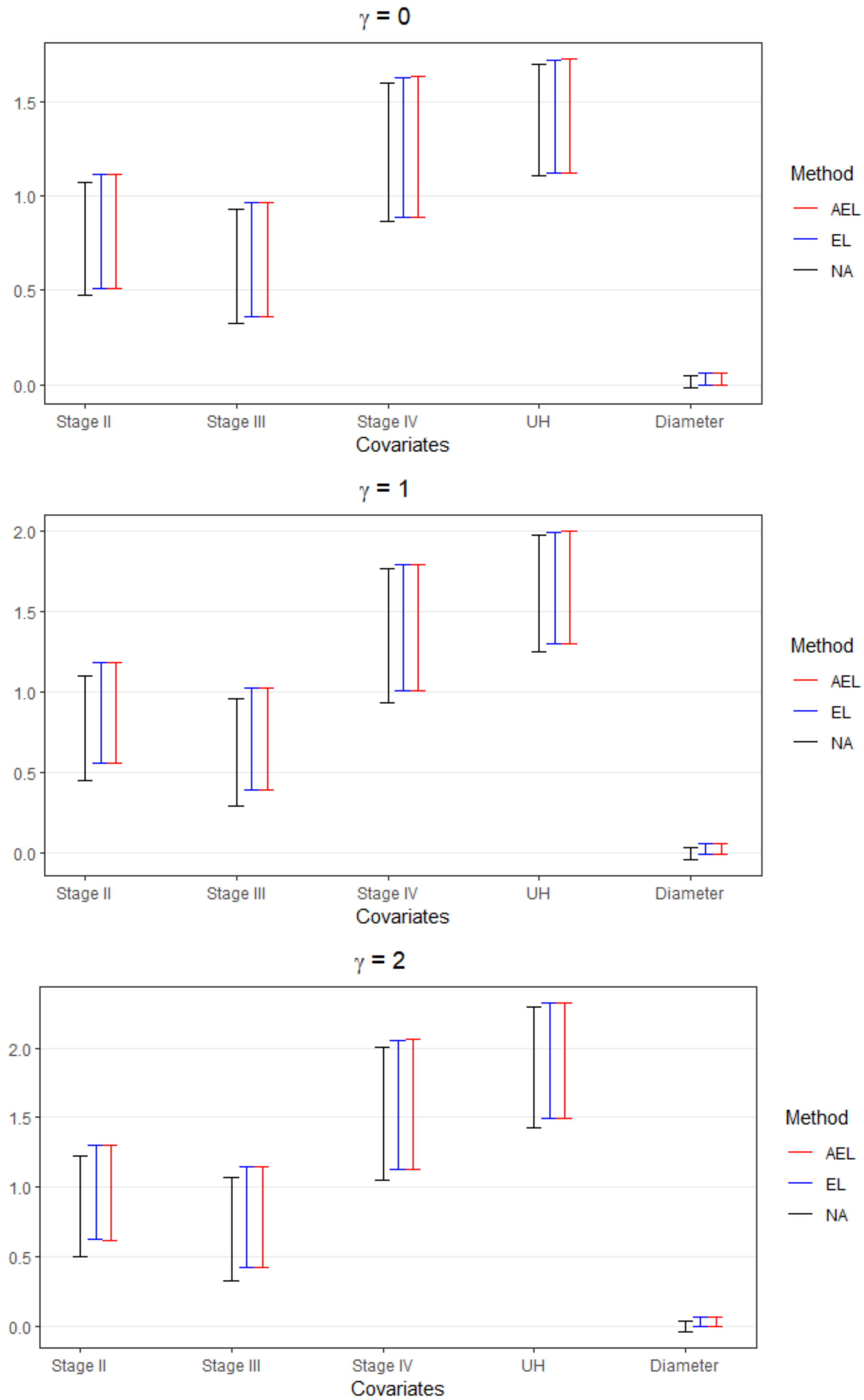
			$\beta_c = 0.1$			$\beta_c = 0.3$		
			NA	EL	AEL	NA	EL	AEL
$n = 100$								
(1, 1)	$\gamma = 0$	β_1	0.909 (3.939)	0.912 (4.138)	0.919 (4.251)	0.920 (4.041)	0.922 (4.254)	0.929 (4.372)
		β_2	0.926 (2.343)	0.928 (2.454)	0.934 (2.521)	0.920 (2.342)	0.922 (2.342)	0.926 (2.519)
	$\gamma = 1$	β_1	0.940 (5.108)	0.934 (4.973)	0.941 (5.111)	0.956 (5.176)	0.934 (5.029)	0.941 (5.169)
		β_2	0.945 (2.881)	0.941 (2.868)	0.948 (2.945)	0.953 (2.849)	0.944 (2.837)	0.947 (2.912)
	$\gamma = 2$	β_1	0.957 (6.240)	0.937 (5.827)	0.944 (5.990)	0.967 (6.252)	0.933 (5.823)	0.939 (5.988)
		β_2	0.958 (3.408)	0.935 (3.296)	0.945 (3.385)	0.960 (3.363)	0.943 (3.255)	0.947 (3.343)
(0, 0)	$\gamma = 0$	β_1	0.917 (3.955)	0.926 (4.061)	0.932 (4.171)	0.916 (3.955)	0.927 (4.060)	0.933 (4.171)
		β_2	0.926 (2.256)	0.929 (2.299)	0.934 (2.359)	0.924 (2.247)	0.930 (2.288)	0.936 (2.348)
	$\gamma = 1$	β_1	0.937 (4.950)	0.930 (4.822)	0.937 (4.953)	0.940 (4.925)	0.935 (4.789)	0.941 (4.920)
		β_2	0.932 (2.751)	0.931 (2.728)	0.937 (2.799)	0.929 (2.730)	0.925 (2.705)	0.931 (2.775)
	$\gamma = 2$	β_1	0.958 (6.038)	0.938 (5.677)	0.943 (5.834)	0.956 (6.004)	0.940 (5.636)	0.945 (5.792)
		β_2	0.945 (3.285)	0.938 (3.202)	0.943 (3.285)	0.957 (3.243)	0.939 (3.157)	0.943 (3.239)
$n = 200$								
(1, 1)	$\gamma = 0$	β_1	0.934 (2.779)	0.940 (2.852)	0.942 (2.893)	0.931 (2.826)	0.934 (2.902)	0.936 (2.942)
		β_2	0.940 (1.640)	0.943 (1.692)	0.945 (1.715)	0.933 (1.632)	0.935 (1.680)	0.939 (1.702)
	$\gamma = 1$	β_1	0.956 (3.440)	0.951 (3.403)	0.952 (3.452)	0.959 (4.109)	0.947 (3.972)	0.950 (4.047)
		β_2	0.944 (1.955)	0.943 (1.958)	0.945 (1.987)	0.940 (1.935)	0.941 (1.942)	0.943 (1.968)
	$\gamma = 2$	β_1	0.954 (4.069)	0.943 (3.948)	0.947 (4.005)	0.961 (4.030)	0.943 (3.908)	0.946 (3.964)
		β_2	0.952 (2.276)	0.946 (2.248)	0.949 (2.284)	0.956 (2.241)	0.951 (2.214)	0.953 (2.248)
(0, 0)	$\gamma = 0$	β_1	0.943 (2.743)	0.944 (2.764)	0.949 (2.803)	0.928 (2.745)	0.929 (2.768)	0.933 (2.807)
		β_2	0.946 (1.575)	0.949 (1.584)	0.951 (1.606)	0.947 (1.567)	0.949 (1.578)	0.952 (1.599)
	$\gamma = 1$	β_1	0.957 (3.337)	0.942 (3.289)	0.945 (3.336)	0.960 (3.322)	0.950 (3.271)	0.953 (3.318)
		β_2	0.939 (1.887)	0.936 (1.876)	0.941 (1.902)	0.944 (1.872)	0.943 (1.861)	0.946 (1.887)
	$\gamma = 2$	β_1	0.955 (3.970)	0.940 (3.855)	0.945 (3.910)	0.960 (3.926)	0.943 (3.809)	0.947 (3.864)
		β_2	0.956 (2.224)	0.949 (2.196)	0.953 (2.226)	0.954 (2.193)	0.945 (2.165)	0.948 (2.194)

Table (6.3) The 95% confidence intervals (CI) and their corresponding lengths by normal approximation (NA), empirical likelihood (EL), adjusted empirical likelihood (AEL) methods for the tumor study dataset.

	NA		EL		AEL	
	CI	Length	CI	Length	CI	Length
			$\gamma = 0$			
Stage II	(0.472, 1.070)	0.598	(0.511, 1.112)	0.601	(0.509, 1.111)	0.602
Stage III	(0.325, 0.931)	0.606	(0.357, 0.966)	0.609	(0.357, 0.967)	0.610
Stage IV	(0.863, 1.599)	0.736	(0.884, 1.625)	0.741	(0.886, 1.630)	0.744
UH	(1.107, 1.695)	0.588	(1.117, 1.714)	0.597	(1.120, 1.722)	0.602
Diameter	(-0.015, 0.048)	0.063	(-0.003, 0.062)	0.065	(-0.003, 0.063)	0.066
			$\gamma = 1$			
Stage II	(0.451, 1.102)	0.651	(0.562, 1.183)	0.621	(0.563, 1.186)	0.623
Stage III	(0.295, 0.961)	0.666	(0.393, 1.027)	0.634	(0.392, 1.028)	0.636
Stage IV	(0.932, 1.771)	0.839	(1.011, 1.793)	0.782	(1.011, 1.795)	0.784
UH	(1.247, 1.971)	0.724	(1.298, 1.993)	0.695	(1.297, 1.997)	0.700
Diameter	(-0.036, 0.035)	0.071	(-0.005, 0.061)	0.066	(-0.005, 0.061)	0.066
			$\gamma = 2$			
Stage II	(0.504, 1.228)	0.724	(0.621, 1.297)	0.676	(0.619, 1.297)	0.678
Stage III	(0.328, 1.070)	0.742	(0.420, 1.142)	0.722	(0.420, 1.145)	0.725
Stage IV	(1.047, 2.004)	0.957	(1.127, 2.058)	0.931	(1.126, 2.060)	0.934
UH	(1.427, 2.299)	0.872	(1.497, 2.320)	0.823	(1.497, 2.321)	0.824
Diameter	(-0.038, 0.041)	0.079	(-0.006, 0.067)	0.073	(-0.006, 0.069)	0.075

NOTE:
NA: normal approximation
EL: empirical likelihood
AEL: adjusted empirical likelihood

Figure 6.1: Plot of 95% confidence intervals by normal approximation (NA), empirical likelihood (EL), adjusted empirical likelihood (AEL) methods for the tumor study dataset.



CHAPTER 7

CONCLUSIONS

We apply empirical likelihood methods in three different survival and non-survival data to construct confidence intervals in this dissertation. Zero-inflated data is more common in many scientific and social study areas. Because of the true nature of the zeros, these zeros should be included in any statistical analysis. But at the same time, using normal approximation based method on zero-inflated data incur under-coverage issue. The empirical likelihood method is a nonparametric alternative in such cases. In empirical likelihood, no parametric assumption about the distribution is required, yet it enjoys all the parametric likelihood benefits. But the empirical likelihood method becomes computationally intensive if it involves solving a non-linear estimating equation.

We seek to alleviate the computation burden by applying the empirical likelihood method to zero-inflated data. We proposed a novel EL method to obtain the confidence interval for the mean of a zero-inflated skewed population. Our methods are computationally easier than other alternatives such as normal approximation or the traditional EL method. Moreover, the simulation study shows that the proposed methods give better coverage probabilities than the existing methods. To further improve coverage probability, we propose the adjusted empirical likelihood by adding one more data point to the original data. Simulation studies are carried out, as well as one real dataset is used to show the application of our proposed methods. Another popular technique to deal with empirical likelihoods' complex computation is the jackknife empirical likelihood. We apply the jackknife empirical likelihood and adjusted jackknife empirical likelihood method for the inference of the mean difference of two independent zero-inflated skewed populations. Our JEL methods are computationally simple, and simulation studies confirm better coverage probabilities than those of normal approximation and empirical likelihood confidence intervals proposed by Zhou and Zhou (2005). We apply two real-life datasets to illustrate the proposed techniques. We also apply jackknife empirical likelihood and its variants to construct confidence intervals for the quantile difference of a zero-inflated population. We show that our proposed methods provide better accuracy through the simulation study. A practical use is also illustrated by a real dataset.

We also investigate the performance of empirical likelihood in survival data. A panel count data is a special type of survival data, where each study subject is observed only at discrete time points. Also, the observation times itself can be very informative about the underlying recurrent process. We apply the empirical likelihood method on the panel count data. Numerical studies reveal better performance regarding coverage probability with empirical likelihood method than NA based method. Bladder tumor data is used to illustrate the actual usage of the proposed method with real-world data. Finally, we propose EL and AEL methods for case-cohort data. Case-cohort is a cost-effective technique to analyze cohort data. Under this study design, only cases and a random subsample of the study subjects provide the covariate information. The simulation study shows the proposed EL method has better performance than other existing NA based method. We demonstrate the proposed methods by analyzing the Wilms' tumor dataset.

REFERENCES

Adimari, G. (1998). An empirical likelihood statistic for quantiles. *Journal of Statistical Computation and Simulation*, 60(1), 85-95.

Alemdjrodo, K. and Zhao, Y. (2019). Reduce the computation in jackknife empirical likelihood for comparing two correlated Gini indices. *Journal of Nonparametric Statistics*, 31(4): 849-866.

An, Y. and Zhao, Y. (2018). Jackknife empirical likelihood for the difference of two volumes under ROC surfaces. *Annals of the Institute of Statistical Mathematics*, 70: 789-806.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). Statistical Models Based on Counting Processes. *New York: Springer-Verlag.*

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.

Andrews, D. F. and Herzberg, A. M. (1985). Data: A Collection of Problems from Many Fields for the Student and Research Worker. *New York: Springer-Verlag.*

Balakrishnan, N. and Zhao, X. (2009). New multi-sample nonparametric tests for panel count data. *The Annals of Statistics*, 37(3): 1112-1149.

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.

Bowman, A. W. and Azzalini, A. (1997). Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations. *Oxford University Press, Oxford.*

Bowman, A. W. and Azzalini, A. (2018). R package 'sm': nonparametric smoothing methods (version 2.2-5.6). URL http://www.stats.gla.ac.uk/∼adrian/sm.

Buzkova, P. (2010). Panel count data regression with informative observation times. *The International Journal of Biostatistics*, 6(1): Article 30.

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

Chen, H. Y. (2001a). Fitting semiparametric transformation regression models to data from a modified case-cohort design. *Biometrika*, 88: 255–268.

Chen, H. Y. (2001b). Weighted semiparametric likelihood method for fitting a proportional odds regression model to data from the case-cohort design. *Journal of the American Statistical Association*, 96(456): 1446–1457.

Chen, J., Chen, S-Y. and Rao, J. N. K. (2003). Empirical likelihood confidence intervals for the mean of a population containing many zero values. *Canadian Journal of Statistics*, 31(1): 53-68.

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.

Chen, K. and Lo, S.H. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika*, 86(4): 755–764.

Chen, S. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, 21(3), 1166-1181.

Chen, S. X. and Cui, H. J. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance paramters. *Bimetrika*, 16: 1101-1115.

Chen S. X. and Cui, H. J. (2007). On the second-order properties of empirical likelihood with moment restrictions. *Journal of Econometrics*, 141(2): 492-516.

Chen, S. X. and Qin J. (2003). Empirical likelihood-based confidence intervals for data with possible zero observations. *Statistics and Probability Letters 65*, 29-37.

Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression (with discussion), *Test*, 18: 415–447.

Chen, Y-H. and Zucker, D. M. (2009). Case-cohort analysis with semiparametric transformation models. *Journal of Statistical Planning and Inference*, 139: 3706-3717.

Chen, Y-J. and Ning, W. (2016). Adjusted jackknife empirical likelihood. arXiv:1603.04093.

Cheng, C. H., Liu, Y., Liu, Z. and Zhou, W. (2018). Balanced augmented jackknife empirical likelihood for two sample U-statistics. *Science China Mathematics*, 61: 1129-1138.

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

Cook, R. J. and Lawless, J. F. (2007). The Statistical Analysis of Recurrent Events. *New York: Springer-Verlag*.

Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika*, 85(4): 967-972.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B. Statistical Methodology*, 34: 187-220.

Dabrowska D. M. and Doksum K. A. (1988). Estimation and testing in the two-sample generalized odds rate model. *Journal of the American Statistical Association*, 83, 744–749.

D'Angio G. J., Breslow N., Beckwith J. B., Evans A., Baum H., de Lorimier A., Ferbach D., Hrabovsky E., Jones G. and Kelalis P. (1989). Treatment of Wilms' tumor. Results of the third national Wilms' tumor study. *Cancer*, 64: 349-360.

Dauxois J-V., Flesch, A. and Varron, D. (2016). Empirical likelihood confidence bands for mean functions of recurrent events with competing risks and a terminal event. *ESAIM: PS*, 20: 66-94.

DiCiccio, T., Hall, P. and Romano, J. (1991). Empirical likelihood is bartlett-correctable. *The Annals of Statistics*, 19(2): 1053-1061.

Dobbie, M. J. and Welsh, A. H. (2001). Modelling correlated zero-inflated count data. *Australian and New Zealand Journal of Statistics*, 43: 431-444.

Duan, N., Manning, W. G. Jr., Morris, C. N., and Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics*, 1(2): 115-126.

Emerson, S. C. and Owen A. B. (2009). Calibration of the empirical likelihood method for a vector mean. *Electronic Journal of Statistics*, 3: 1161-1192.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. *The Annals of Statistics*, 12(1): 261-268.

Fang, S., Zhang, H., Sun, L. and Wang , D. (2017). Analysis of panel count data with time-dependent covariates and informative observation process. *Acta Mathematicae Applicatae Sinica, English Series*, 33(1): 147-156.

Feng, H. and Peng, L. (2012). Jackknife empirical likelihood tests for distribution functions. *Journal of Statistical Planning and Inference*, 142(6): 1571-1585.

Fine, J., Ying, Z. and Wei, L. J. (1998). On the linear transformation model for censored data. *Biometrika*, 85(4): 980–986.

Fleming, T. R. and Harrington, D. P. (1991). Counting processes and survival analysis. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley and Sons Inc., New York.*

Fletcher, D., MacKenzie, D. and Villouta, E. (2005). Modelling skewed data with many zeros: A simple approach combining ordinary and logistic regression. *Environmental and Ecological Statistics*, 12: 45-54.

Green D. M., Breslow N. E. , Beckwith J. B., Finklestein J. Z., Grundy P. G., Thomas P. R. M., Kim T., Shochat S., Haase G. M., Ritchey M. L., Kelalis P. P. and D'Angio G. J. (1998). Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms' tumor: a report from the national Wilms' tumor study group. *J Clin Oncol*, 16: 237-245.

Gong, Y., Peng, L., Qi, Y. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis*, 101(6): 1520-1531.

Hall, D. B. (2000). Zero-inflated poisson and binomial regression with random effects: a case study. *Biometrics*, 56: 1030-1039.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58(2): 109-127.

Hasan M. S. and Krishnamoorthy, K. (2018). Confidence intervals for the mean and a percentile based on zero-inflated lognormal data. *Journal of Statistical Computation and Simulation*, 88(8): 1499-1514.

He, X., Tong, X., and Sun, J. (2009). Semiparametric analysis of panel count data with correlated observation and follow-up times. *Lifetime Data Analysis*, 15: 177-196.

He, X., Tong, X., Sun, J. and Cook, R. (2008). Regression analysis of multivariate panel count data. *Biostatistics*, 9: 234-248.

Hu, M-C., Pavlicova, M. and Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American Journal of Drug and Alcohol Abuse*, 37(5): 367-375.

Hu, S. and Lin, L. (2012). Empirical likelihood analysis of longitudinal data involving within-subject correlation. *Acta Math. Appl. Sin. Engl. Ser.*, 28(4): 731-744.

Huang, C-Y., Wang, M-C. and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika*, 93(4): 763-775.

Jing, B-Y., Tsao, M. and Zhou, W. (2017). Transforming the empirical likelihood towards better accuracy. *The Canadian Journal of Statistics*, 45(3): 340-352.

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

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80: 863-871.

Kang, L., Vexler A., Tian L., Cooney M., Louis B. and Germaine M. (2010). Empirical and parametric likelihood interval estimation for populations with many zero values: application for assessing environmental chemical concentrations and reproductive health. *Epidemiology*, 21(4): S58-S63.

Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics*, 65: 135–142.

Kong, L., Cai, J. and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika*, 91(2): 305-319.

Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika*, 87(1): 73-87.

Lachenbruch, P. A. (1976). Analysis of data with clumping at zero. *Biometrical Journal*, 18: 351-356.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34: 1-14.

Li, G. and Wang, Q. (2003). Empirical likelihood regression analysis for right censored data. *Statistical Sinica*, 13: 51-68.

Li, M., Peng, L. and Qi, Y. (2011). Reduce computation in profile empirical likelihood method. *The Canadian Journal of Statistics*, 39(2): 370-384.

Li, N., Sun, L. and Sun, J. (2010). Semiparametric transformation models for panel count data with dependent observation process. *Statistics in Biosciences*, 2(2): 191-210.

Liang, W., Dai, H. and He, S. (2019). Mean empirical likelihood. *Computational Statistics and Data Analysis*, 138: 155-169.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society. Series B,* 62(4), 711-730.

Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*, 88: 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, W., Lu, X. and Xie, C. (2014). Empirical likelihood for the additive hazards model with current status data. *Communications in Statistics - Simulation and Computation,* 45(8), 2720-2732.

Liu, Y. and Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *The Annals of Statistics*, 38(3): 1341-1362.

Liu, Y. and Yu, C. W. (2010). Bartlett correctable two-sample adjusted empirical likelihood. *Journal of Multivariate Analysis*, 101: 1701-1711.

Liu, Z., Xia X. and Zhou W. (2015). A test for equality of two distributions via jackknife empirical likelihood and characteristic functions. *Computational Statistics and Data Analysis*, 92: 97-114.

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

Lu, W. and Tsiatis, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika*, 93(1): 207–214.

Ma, H., Qiu, Z. and Zhou, Y. (2016). Semiparametric transformation models with length-biased and right-censored data under the case-cohort design. *Statistics and Its Inference*, 9: 213-222.

Min, Y. and Agresti, A. (2002). Modeling nonnegative data with clumping at zero: a survey. *Journal of the Iranian Statistical Society*, 1(1-2): 7-33.

Morrison-Beedy, D., Carey, M. P., Crean, H. F. and Jones, H. S. (2011). Risk behaviors among adolescent girls in an HIV prevention trial. *Western Journal of Nursing Research*, 33(5): 690-711.

Nan, B., Yu, M. and Kalbfleisch, J. D. (2006). Censored linear regression for case-cohort studies. *Biometrika*, 93: 747–762.

Neuhäuser, M. (2012). Nonparametric Statistical Tests: A Computational Approach. *Chapman & Hall/CRC, Boca Raton, FL.*

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

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

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

Owen, A. B. (2001). Empirical Likelihood, Chapman & Hall/ CRC, New York.

Pailden, J. and Chen. H. (2013). Empirical likelihood ratio test for difference of the means of zero-inflated populations. *Advances and Applications in Statistics*, 35(1): 29-55.

Peng, L. (2012). Approximate jackknife empirical likelihood method for estimating equations. *Canadian Journal of Statistics*, 40(1): 110-123.

Prentice, R. L. (1986). A case-cohort design for epidemiological cohort studies and disease prevention trials. *Biometrika*, 73: 1-11.

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

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

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89: 846-866.

Sang, Y., Dang, X. and Zhao, Y. (2019). Jackknife empirical likelihood methods for Gini correlations and their equality testing. *Journal of Statistical Planning and Inference*, 199: 45-59.

Satter, F. and Zhao, Y. (2020). Nonparametric interval estimation for the mean of a zero-inflated population. *Communication in Statistics- Simulation and Computation*, 49(8): 2059-2067.

Satter, F. and Zhao, Y. (2021). Jackknife empirical likelihood for the mean difference of two zero-inflated skewed populations. *Journal of Statistical Planning and Inference*, 211: 414-422.

Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*, 16: 64-81.

Sheather, S. J. and Matron, J. S. (1990). Kernel quantile estimators. *Journal of the American Statistical Association*, 85(410): 410-416.

Shen, P-S. (2012). Empirical likelihood ratio for linear transformation models with doubly censored data. *Communications in Statistics - Simulation and Computation*, 41(4): 531-543.

Shi, X. (1984). The Approximate independence of jackknife pseudo-values and the bootstrap methods. *Journal of Wuhan Institute Hydra-Electric Engineering*, 2: 83–90.

Shi, X., Ibrahim, J. G., Lieberman, J., Styner, M., Li, Y. and Zhu, H. (2011). Two-stage empirical likelihood for longitudinal neuroimaging data. *The Annals of Applied Statistics*, 5: 1132-1159.

Siegmund K. D., Laird P. W. and Laird-Offringa I. A. (2004). A comparison of cluster analysis methods using DNA methylation data. *Bioinformatics*, 20(12): 1896-1904.

Stewart, P. and Ning, W. (2020). Modified empirical likelihood-based confidence intervals for data containing many zero observations. *Computational Statistics*, 35:2019-2042.

Subramanian, S. (2007). Censored median regression and profile empirical likelihood. *Statistical Methodology*, 4(4): 493-503.

Sun, J. and Fang, H. B. (2003). A nonparametric test for panel count data. *Biometrika*, 90: 199-208.

Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, 5: 279-289.

Sun, J., Tong, X. and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics*, 63: 1053-1059.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring time. *Journal of the Royal Statistical Society, Series B*, 62: 293-302.

Sun, J. and Zhao, X. (2013). Statistical Analysis of Panel Count Data. *New York: Springer*.

Sun, Y., Yu, W. and Zheng, M. (2016). Case-cohort analysis with general additive-multiplicative hazard models. *Acta Mathematicae Applicatae Sinica, English Series*, 32(4): 851-866.

Tang, W., He, H. and Tu, X. M. (2012). Applied Categorical and Count Data Analysis. *Chapman & Hall/CRC, New York*.

Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, 83: 339-347.

Therneau, T. M. and Grambsch, P. M. (1990). Martingale-based residuals for survival models. *Biometrica*, 77(1): 147-160.

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

Tsao. M. and Wu, F. (2013). Empirical likelihood on the full parameter space. *The Annals of Statistics*, 41(4): 2176–2196.

Wang, D. and Zhao, Y. (2016). Jackknife empirical likelihood for comparing two Gini indices. *Canadian Journal of Statistics*, 46: 102-119.

Wang R., Peng L. and Qi Y. (2013). Jackknife empirical likelihood test for equality of two high dimensional means. *Statistica Sinica*, 23: 667-690.

Wang, S., Qian. L. and Carrol, R. J. (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika*, 97(1): 79-93.

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.

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count. *Annals of Statistics*, 28: 779-814.

Welsh, A. H., Cunningham, R. B., Donnelly, C. F. and Lindenmayer, D.B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, 88: 297-308.

Yang, H., Liu, S. and Zhao, Y. (2017). Jackknife empirical likelihood for linear transformation models with right censoring. *Annals of the Institute of Statistical Mathematics*, 69: 1059–1073.

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

Yang, H. and Zhao, Y. (2016). Smoothed jackknife empirical likelihood for the difference of two quantiles. *Computational Statistics and Data Analysis*, 120: 58-69.

Yang, S-S. (1985). A smooth nonparametric estimator of a quantile function. *Journal of the American Statistical Association*, 80(392): 1004-1011.

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

Yu, X. and Zhao, Y. (2019). Empirical likelihood inference for semi-parametric transformation models with length-biased sampling. *Computational Statistics and Data Analysis*, 132: 115-125.

Zhang, H., Zhao, H., Sun, J., Wang, D. and Kim, K. (2013). Regression analysis of multivariate panel count data with an informative observation process. *Journal of Multivariate Analysis*, 119: 71-80.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, 89: 39-48.

Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika*, 93: 777-790.

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, H., Li, Y. and Sun, J. (2013). Analyzing panel count data with a dependent observation process and a terminal event. *The Canadian Journal of Statistics*, 41(1): 174-191.

Zhao, X., Tong, X. and Sun, J. (2013). Robust estimation for panel count data with informative observation times. *Computational Statistics and Data Analysis*, 57: 33-40.

Zhao, Y. (2011). Empirical likelihood inference for the accelerated failure time model. *Statistics and Probability Letters*, 81(5): 603-610.

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

Zhao, Y., Su, Y. and Yang, H. (2020). Jackknife empirical likelihood inference for the Pietra ratio. *Computational Statistics and Data Analysis*, 152: 107049.

Zhou, M. (2016). Empirical Likelihood Method in Survival Analysis. *New York: Chapman & Hall/CRC.*

Zhou, W. and Jing, B-Y. (2003a). Adjusted empirical likelihood method for quantiles. *Annals Institute of Statistical Mathematics*, 55: 689-703.

Zhou, W. and Jing, B-Y. (2003b). Smoothed empirical likelihood confidence intervals for the difference of quantiles. *Statistica Sinica*, 13(1): 83-95.

Zhou, W. and Zhou, X-H. (2005). Empirical likelihood intervals for the mean difference of two skewed populations with additional zero values. *UW Biostatistics Working Paper Series. Working Paper 246*. http://biostats.bepress.com/uwbiostat/paper246.

Zhou, X. H. and Tu, W. (1999). Comparison of several independent population means when their samples contain lognormal and possibly zero observations. *Biometrics*, 55(2): 645-651.

Zhou, X-H. and Tu, W. (2000). Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics*, 56(4):1118-1125.

Appendix A

PROOFS OF CHAPTER 2

Proof of Theorem 2.1. The proof is similar to Owen (1988, 1990). From equation (2.1), we have

$$\begin{aligned} 0 &= \tfrac{1}{n_1}\left|\sum_{i=1}^{n_1}(\hat{V}_i-\mu)-\lambda\sum_{i=1}^{n_1}\tfrac{(\hat{V}_i-\mu)^2}{1+\lambda(\hat{V}_i-\mu)}\right| \\ &\geq \tfrac{|\lambda|}{n_1}\sum_{i=1}^{n_1}\tfrac{(\hat{V}_i-\mu)^2}{1+\lambda(\hat{V}_i-\mu)}-\tfrac{1}{n_1}\left|\sum_{i=1}^{n_1}(\hat{V}_i-\mu)\right| \\ &\geq \tfrac{|\lambda|S}{1+|\lambda|Z_{n_1}}-\left|\tfrac{1}{n_1}\sum_{i=1}^{n_1}(\hat{V}_i-\mu)\right|, \end{aligned}$$

where $S=\frac{1}{n_1}\sum_{i=1}^{n_1}(\hat{V}_i-\mu)^2$ and $Z_{n_1}=\max_{1\leq i\leq n_1}|\hat{V}_i-\mu|$.

$$\begin{aligned} \text{Then, } S &= \tfrac{1}{n_1}\sum_{i=1}^{n_1}(\hat{V}_i-\mu)^2 \\ &= \tfrac{1}{n_1}\sum_{i=1}^{n_1}\left(\tfrac{n_1x_i}{n}-\mu\right)^2 \\ &= \tfrac{1}{n_1}\sum_{i=1}^{n_1}\left(\tfrac{n_1^2x_i^2}{n^2}-2\tfrac{n_1x_i}{n}\mu-\mu^2\right) \\ &= \tfrac{1}{n_1}\sum_{i=1}^{n_1}\tfrac{n_1^2x_i^2}{n^2}-2\tfrac{1}{n_1}\sum_{i=1}^{n_1}\tfrac{n_1x_i}{n}\mu-\tfrac{1}{n_1}\sum_{i=1}^{n_1}\mu^2 \\ &= \tfrac{n_1^2}{n^2}\tfrac{\sum_{i=1}^{n_1}x_i^2}{n_1}-2\tfrac{\sum_{i=1}^{n_1}x_i}{n}\mu+\mu^2 \\ &\to (1-\delta)^2E(X^2|X>0)-2\mu^2+\mu^2 \\ &= (1-\delta)^2E(X^2|X>0)-(1-\delta)^2(E(X|X>0))^2 \\ &= (1-\delta)^2(E(X^2|X>0)-(E(X|X>0))^2) \\ &= (1-\delta)^2Var(X|X>0) \\ &< Var(X) \end{aligned}$$

for $0<\delta<1$, and the second term is $O_p(n^{-1/2})$. Also, $Z_{n_1}=\max_{1\leq i\leq n_1}|\hat{V}_i-\mu|\leq\max_{1\leq i\leq n_1}|\hat{V}_i|+\mu\leq\max_{1\leq i\leq n_1}|x_i|+\mu=\max_{1\leq i\leq n}|X_i|+\mu=o(n^{1/2})$ using Lemma 3 of Owen (1990). Then it follows

$$\tfrac{|\lambda|S}{1+|\lambda|Z_{n_1}}=O_p(n^{-1/2}).$$

Therefore, we have $|\lambda|=O_p(n^{-1/2})$. Let $\gamma_i=\lambda(\hat{V}_i-\mu)$. Then

$$\max_{1\leq i\leq n_1}|\gamma_i|=O_p(n^{-1/2})o(n^{1/2})=o_p(1).$$

Expanding equation (2.1), one has

$$\begin{aligned} 0 &= \tfrac{1}{n_1}\textstyle\sum_{i=1}^{n_1}(\hat{V}_i-\mu)(1-\gamma_i+\gamma_i^2/(1+\gamma_i)) \\ &= \tfrac{1}{n_1}\textstyle\sum_{i=1}^{n_1}\hat{V}_i-\mu-S\lambda+\tfrac{1}{n_1}\sum_{i=1}^{n_1}(\hat{V}_i-\mu)\gamma_i^2/(1+\gamma_i), \end{aligned}$$

where the final term is $o(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2})$. We can write,

$$\lambda = \frac{1}{S}\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\hat{V}_i-\mu\right)+\beta = \frac{1}{S}(T_{n_1}-\mu)+\beta,$$

where $\beta = o_p(n^{-1/2})$. By the Taylor expansion, one has

$$\begin{aligned} -2\log R(\mu_0) &= 2n_1\lambda(T_{n_1}-\mu)-n_1S\lambda^2+2\textstyle\sum_{i=1}^{n_1}\eta_i \\ &= \tfrac{n_1(T_{n_1}-\mu)^2}{S}-n_1S\beta^2+2\textstyle\sum_{i=1}^{n_1}\eta_i, \end{aligned}$$

where for some finite $B>0, P(|\eta_i|\leq B|\gamma_i|^3, 1\leq i\leq n_1)\to 1$. Since $n_1(T_{n_1}-\mu)^2/S$ converges to χ_1^2, $|-n_1S\beta^2| = n_1(\sigma_p^2+o(1))o_p(n^{-1}) = o_p(1)$ and $|2\sum_{i=1}^{n_1}\eta_i| \leq 2B|\lambda|^3\sum_{i=1}^{n_1}|\hat{V}_i-\mu|^3 = O_p(n^{-3/2})o(n^{3/2}) = o_p(1)$. Therefore, the proof is completed. □

Proof of Theorem 2.2. The proof is similar to Chen et al. (2008), Zhao et al. (2015) and Wang and Zhao (2016). Denote $g_i = \hat{V}_i-\mu$ and $\bar{g}_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{V}_i-\mu$. Let λ be the solution to

$$\sum_{i=1}^{n_1+1}\frac{g_i}{1+\lambda g_i} = 0$$

Denote $g^\star = \max_{1\leq i\leq n_1}|\hat{V}_i-\mu|$. One has $g^\star = \max_{1\leq i\leq n_1}|\hat{V}_i-\mu| \leq \max_{1\leq i\leq n_1}|\hat{V}_i|+\mu = \max_{1\leq i\leq n_1}|x_i|+\mu = o_p(n_1^{1/2})$ and $\bar{g}_{n_1} = O_p(n_1^{-1/2})$. Let $\rho = ||\lambda||$ and $\hat{\lambda} = \lambda/\rho$,

$$\begin{aligned} 0 &= \tfrac{\hat{\lambda}}{n_1}\textstyle\sum_{i=1}^{n_1+1}g_i-\rho\sum_{i=1}^{n_1+1}\frac{(\hat{\lambda}g_i)^2}{1+\rho\hat{\lambda}g_i} \\ &\leq \hat{\lambda}\bar{g}_{n_1}(1-\tfrac{a_{n_1}}{n_1})-\tfrac{\rho}{n_1(1+\rho g^\star)}\textstyle\sum_{i=1}^{n_1}(\hat{\lambda}g_i)^2. \end{aligned}$$

With $a_{n_1} = o_p(n_1)$, we have

$$\begin{aligned} \tfrac{\rho}{1+\rho g^\star} &\leq \hat{\lambda}\bar{g}_{n_1}\left(1-\tfrac{1}{n_1}a_{n_1}\right)\left(\tfrac{1}{n_1}\textstyle\sum_{i=1}^{n_1}(\hat{\lambda}g_i)^2\right)^{-1} \\ &= \hat{\lambda}\bar{g}_{n_1}\left(1-\tfrac{1}{n_1}a_{n_1}\right)\left(\tfrac{1}{n_1}\textstyle\sum_{i=1}^{n_1}g_i^2\right)^{-1} \\ &= O_p({n_1}^{-1/2})(1-o_p(1))(\sigma_p^2+o_p(1))^{-1} \\ &= O_p({n_1}^{-1/2}), \end{aligned}$$

which implies $\rho = O_p({n_1}^{-1/2})$ and $\lambda = O_p(n_1^{-1/2})$. Denote $\hat{V}_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}g_i^2$. One has that

$$0 = \tfrac{1}{n_1}\textstyle\sum_{i=1}^{n_1+1}\frac{g_i}{1+\lambda g_i}$$

$$= \bar{g}_{n_1} - \lambda \hat{V}_{n_1} + o_p({n_1}^{-1/2}).$$

As $n_1 \to \infty, \lambda = \frac{\bar{g}_{n_1}}{\hat{V}_{n_1}} + o_p({n_1}^{-1/2})$.

$$-2\log R^\star(\mu_0) = 2\sum_{i=1}^{n_1+1} \log(1 + \lambda g_i)$$

$$= 2\sum_{i=1}^{n_1+1} \left\{ \lambda g_i - \frac{(\lambda g_i)^2}{2} \right\} + o_p(1).$$

Replacing λ by $\frac{\bar{g}_{n_1}}{\hat{V}_{n_1}} + o_p({n_1}^{-1/2})$, we have

$$-2\log R^\star(\mu_0) = \frac{n_1 \bar{g}_{n_1}^2}{\hat{V}_{n_1}} + o_p(1).$$

Thus, we finish the proof. □

Appendix B

PROOFS OF CHAPTER 3

Lemma 3.1. *Let* $Eh_0^2(X,Y)^2 < \infty$ *and* $\sigma_1^2 > 0$ *and* $\sigma_2^2 > 0$. *Then, as* $m_1 \to \infty$, $P(\underset{1\le i\le t}{min}(\hat{V}_i - \theta_0) < 0 < \underset{1\le i\le t}{max}(\hat{V}_i - \theta_0)) \to 1$.

Proof of Lemma 3.1. Let $\phi(x,y) = h(x,y) - h_1(x) - h_2(y) + \theta_0$, where $h(x,y) = (1-\hat{\delta}_1)x - (1-\hat{\delta}_2)y$, $h_1(x) = Eh(x,Y) = (1-\hat{\delta}_1)x - (1-\delta_2)E(Y|Y>0)$ and $h_2(y) = Eh(X,y) = (1-\delta_1)E(X|X>0) - (1-\hat{\delta}_2)y$. Then by the Hoeffding decomposition,

$$\begin{aligned} T_t &= (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\ &= \theta_0 + \frac{1}{m_1}\sum_{i=1}^{m_1}(h_1(x_i) - \theta_0) + \frac{1}{n_1}\sum_{j=1}^{n_1}(h_2(y_j) - \theta_0) + \frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}\phi(x_i, y_j). \end{aligned}$$

Let T_t^0 be a two-sample U-statistic, defined as

$$\begin{aligned} T_t^0 &= \frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}((1-\delta_1)x_i - (1-\delta_2)y_j) \\ &= \frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1} h_0(x_i, y_j). \end{aligned}$$

Also let $\phi_0(x,y) = h_0(x,y) - h_{01}(x) - h_{02}(y) + \theta_0$, where $h_0(x,y) = (1-\delta_1)x - (1-\delta_2)y$, $h_{01}(x) = (1-\delta_1)x - (1-\delta_2)E(Y|Y>0)$ and $h_{02}(y) = (1-\delta_1)E(X|X>0) - (1-\delta_2)y$. Then, again by the Hoeffding decomposition,

$$\begin{aligned} T_t^0 &= \frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}((1-\delta_1)x_i - (1-\delta_2)y_j) \\ &= \theta_0 + \frac{1}{m_1}\sum_{i=1}^{m_1}(h_{01}(x_i) - \theta_0) + \frac{1}{n_1}\sum_{j=1}^{n_1}(h_{02}(y_j) - \theta_0) + \frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}\phi_0(x_i, y_j). \end{aligned}$$

As $m \to \infty, (1-\hat{\delta}_1) \to (1-\delta_1)$ and $n \to \infty, (1-\hat{\delta}_2) \to (1-\delta_2)$. Then $\hat{\delta}_1 - \delta_1 \to 0$ and

$\hat{\delta}_2 - \delta_2 \to 0$. In addition,

$$\frac{\sum_{i=1}^{m_1} x_i}{m_1} \to E(X|X>0) \text{ and } \frac{\sum_{j=1}^{n_1} y_j}{n_1} \to E(Y|Y>0).$$

One has

$$\begin{aligned}
T_t - T_t^0 &= (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&\quad -\frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}((1-\delta_1)x_i - (1-\delta_2)y_j) \\
&= (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} - \frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}(1-\delta_1)x_i \\
&\quad +\frac{1}{m_1 n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}(1-\delta_2)y_j \\
&= (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} - \frac{1}{m_1 n_1}\sum_{i=1}^{m_1} n_1(1-\delta_1)x_i \\
&\quad +\frac{1}{m_1 n_1}m_1\sum_{j=1}^{n_1} m_1(1-\delta_2)y_j \\
&= (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} - \frac{1}{m_1}\sum_{i=1}^{m_1}(1-\delta_1)x_i + \frac{1}{n_1}\sum_{j=1}^{n_1}(1-\delta_2)y_j \\
&= (\hat{\delta}_1-\delta_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (\hat{\delta}_2-\delta_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&\to 0.
\end{aligned}$$

Also let $T_{m_1-1,n_1}^{-k,0}$ be the statistics after deleting $x_k, k = 1, 2, \ldots, m_1$. Then

$$\begin{aligned}
V_{k,0} &= m_1 T_t - (m_1-1)T_{m_1-1,n_1}^{-k,0} \\
&= m_1\left[(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}\right] \\
&\quad -(m_1-1)\left[(1-\hat{\delta}_1)\frac{\sum_{i\neq k,i=1}^{m_1} x_i}{m_1-1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}\right] \\
&= m_1(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - m_1(1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&\quad -(m_1-1)(1-\hat{\delta}_1)\frac{\sum_{i\neq k,i=1}^{m_1} x_i}{m_1-1} + (m_1-1)(1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}
\end{aligned}$$

$$\begin{aligned}
&= (1-\hat{\delta}_1)\sum_{i=1}^{m_1} x_i - (1-\hat{\delta}_1)\sum_{i\neq k, i=1}^{m_1} x_i - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&= (1-\hat{\delta}_1)x_k - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}, \quad k = 1, 2, \ldots, m_1.
\end{aligned}$$

By the central limit theorem,

$$\begin{aligned}
\sqrt{m}|\hat{\delta}_1 - \delta_1| &= \sqrt{m}\left|\frac{1(X_1 = 0) + \cdots + 1(X_m = 0)}{m} - P(X = 0)\right| \\
&= O_p(1),
\end{aligned}$$

$$\begin{aligned}
\sqrt{n}|\hat{\delta}_2 - \delta_2| &= \sqrt{n}\left|\frac{1(Y_1 = 0) + \cdots + 1(Y_n = 0)}{n} - P(Y = 0)\right| \\
&= O_p(1),
\end{aligned}$$

and $\max\limits_{1\leq k\leq m_1}|x_k| = o_p(m^{1/2})$, $\max\limits_{1\leq k\leq n_1}|y_k| = o_p(n^{1/2})$ and $n_1^{-1}\sum_{j=1}^{n_1} y_j \to E(Y|Y>0)$ as $n \to \infty$. Therefore,

$$\begin{aligned}
|V_{k,0} - h_{01}(x_k)| &= |(1-\hat{\delta}_1)x_k - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} - (1-\delta_1)x_k + (1-\delta_2)E(Y|Y>0)| \\
&= |(1-\hat{\delta}_1)x_k - (1-\hat{\delta}_2)[E(Y|Y>0) + o_p(1)] \\
&\quad -(1-\delta_1)x_k + (1-\delta_2)E(Y|Y>0)| \\
&\leq |(\hat{\delta}_1 - \delta_1)x_k - (\hat{\delta}_2 - \delta_2)E(Y|Y>0)| + |1-\hat{\delta}_2|o_p(1) \\
&= |(\hat{\delta}_1 - \delta_1)x_k| + |(\hat{\delta}_2 - \delta_2)E(Y|Y>0)| + o_p(1) \\
&\leq |\hat{\delta}_1 - \delta_1|\max_{1\leq k\leq m_1}|x_k| + |\hat{\delta}_2 - \delta_2|\max_{1\leq k\leq n_1}E(Y|Y>0) + o_p(1) \\
&= O_p(m^{-1/2})o_p(m^{1/2}) + O_p(n^{-1/2})o_p(n^{1/2}) + o_p(1) \\
&= o_p(1) + o_p(1) + o_p(1) \\
&= o_p(1).
\end{aligned}$$

Therefore, $|V_{k,0} - h_{01}(x_k)| \to 0$ in probability.

Now we make a connection between $\hat{V}_k$ and $V_{k,0}$. From equation (3.1), for $k = 1, \ldots, m_1$,

$$
\begin{aligned}
\hat{V}_k &= (1-\hat{\delta}_1)\left[t\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (t-1)\frac{\sum_{i\neq k,i=1}^{m_1} x_i}{m_1-1}\right] - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&= t(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (t-1)(1-\hat{\delta}_1)\frac{\sum_{i\neq k,i=1}^{m_1} x_i}{m_1-1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&= t(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} + (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} \\
&\quad -(t-1)(1-\hat{\delta}_1)\frac{\sum_{i\neq k,i=1}^{m_1} x_i}{m_1-1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&= (t-1)(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (t-1)(1-\hat{\delta}_1)\frac{\sum_{i\neq k,i=1}^{m_1} x_i}{m_1-1} + (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} \\
&\quad -(1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\
&= (t-1)(1-\hat{\delta}_1)\left[\frac{\sum_{i=1}^{m_1} x_i}{m_1} - \frac{\sum_{i\neq k,i=1}^{m_1} x_i}{m_1-1}\right] + T_t \\
&= (t-1)(1-\hat{\delta}_1)\left[\frac{1}{m_1(m_1-1)}\{(m_1-1)\sum_{i=1}^{m_1} x_i - m_1\sum_{i\neq k,i=1}^{m_1} x_i\}\right] + T_t \\
&= (t-1)(1-\hat{\delta}_1)\left[\frac{1}{m_1(m_1-1)}\{m_1\sum_{i=1}^{m_1} x_i - \sum_{i=1}^{m_1} x_i - m_1\sum_{i\neq k,i=1}^{m_1} x_i\}\right] + T_t \\
&= (t-1)(1-\hat{\delta}_1)\left[\frac{1}{m_1(m_1-1)}\{m_1\sum_{i=1}^{m_1} x_i - m_1\sum_{i\neq k,i=1}^{m_1} x_i - \sum_{i=1}^{m_1} x_i\}\right] + T_t \\
&= (t-1)(1-\hat{\delta}_1)\left[\frac{1}{m_1(m_1-1)}\{m_1 x_k - \sum_{i=1}^{m_1} x_i\}\right] + T_t \\
&= (t-1)(1-\hat{\delta}_1)\left[\frac{1}{m_1(m_1-1)}m_1 x_k - \frac{1}{m_1(m_1-1)}\sum_{i=1}^{m_1} x_i\}\right] + T_t \\
&= (t-1)(1-\hat{\delta}_1)\frac{1}{(m_1-1)}x_k - (t-1)(1-\hat{\delta}_1)\frac{1}{m_1(m_1-1)}\sum_{i=1}^{m_1} x_i + T_t \\
&= \frac{t-1}{m_1-1}(1-\hat{\delta}_1)x_k - \frac{t-1}{m_1-1}(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} + T_t \\
&= \frac{t-1}{m_1-1}(1-\hat{\delta}_1)x_k - \frac{t-1}{m_1-1}(1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} + \frac{t-1}{m_1-1}(1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} - \\
&\quad \frac{t-1}{m_1-1}(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} + T_t \\
&= \frac{t-1}{m_1-1}\left[(1-\hat{\delta}_1)x_k - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}\right]
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{t-1}{m_1-1}\left[(1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1}\right] + T_t \\
= \; & \frac{t-1}{m_1-1}V_{k,0} - \frac{t-1}{m_1-1}T_t + T_t \\
= \; & \frac{t-1}{m_1-1}V_{k,0} - \frac{t-m_1}{m_1-1}T_t \\
= \; & \frac{t-1}{m_1-1}V_{k,0} - \frac{n_1}{m_1-1}T_t.
\end{aligned}
$$

Then for $k = 1, \ldots, m_1$,

$$
\begin{aligned}
\hat{V}_k - \theta_0 &= \frac{t-1}{m_1-1}V_{k,0} - \frac{n_1}{m_1-1}T_t - \theta_0 \\
&= \frac{t-1}{m_1-1}V_{k,0} - \frac{n_1}{m_1-1}T_t - \left(\frac{t-1}{m_1-1} - \frac{n_1}{m_1-1}\right)\theta_0 \\
&= \frac{t-1}{m_1-1}(V_{k,0} - \theta_0) - \frac{n_1}{m_1-1}(T_t - \theta_0) \\
&= \frac{t-1}{m_1-1}[h_{01}(x_k) - \theta_0] - \frac{n_1}{m_1-1}(T_t - \theta_0) + o_p(1) \\
&= g_1(x_k) - R_{1k} + o_p(1),
\end{aligned}
$$

where $g_1(x_k) = (t-1)(m_1-1)^{-1}[h_{01}(x_k) - \theta_0]$ and $R_{1k} = n_1(m_1-1)^{-1}(T_t - \theta_0)$.

Let $\zeta_{tk} = \Psi(\hat{V}_k - \theta_0)$, where $\Psi(x)$ is a nondecreasing, twice differentiable function with bounded first and second derivatives such that

$$
\Psi(x) = \begin{cases} 0, & \text{if} \quad x \leq 0 \\ a(x), & \text{if} \quad 0 < x < \epsilon \\ 1, & \text{if} \quad x \geq \epsilon, \end{cases}
$$

with $0 < a(x) < 1$ for $0 < x < \epsilon$. A Taylor expansion gives

$$
\zeta_{tk} = \Psi(g_1(x_k)) + \Psi'(g_1(x_k))R_{1k} + \eta_k(R_{1k})^2,
$$

where $|\eta_k| < C$ for some generic constant $C > 0$.

Since

$$\begin{aligned}
E[R_{1k}] &= E\left(\frac{n_1}{m_1-1}(T_t-\theta_0)\right)\\
&= E\left[E\left(\frac{n_1}{m_1-1}(T_t-\theta_0)\middle| m_1,n_1\right)\right]\\
&= E\left[\frac{n_1}{m_1-1}\ E((T_t-\theta_0)|\, m_1,n_1)\right]\\
&= E\left[\frac{n_1}{m_1-1}\{(1-\hat{\delta}_1)E(X|X>0)-(1-\hat{\delta}_2)E(Y|Y>0)-\theta_0\}\right]\\
&\to E\left[\frac{n_1}{m_1-1}\{(1-\delta_1)E(X|X>0)-(1-\delta_2)E(Y|Y>0)-\theta_0\}\right]\\
&= E\left[\frac{n_1}{m_1-1}(\theta_0-\theta_0)\right]\\
&= 0.
\end{aligned}$$

and

$$\begin{aligned}
E[R_{1k}^2] &= E\left[\left(\frac{n_1}{m_1-1}\right)^2(T_t-\theta_0)^2\right]\\
&= E\left[E\left\{\left(\frac{n_1}{m_1-1}\right)^2(T_t-\theta_0)^2\middle| m_1,n_1\right\}\right]\\
&= E\left[\left(\frac{n_1}{m_1-1}\right)^2E\big\{(T_t-\theta_0)^2|m_1,n_1\big\}\right]\\
&= E\left[\left(\frac{n_1}{m_1-1}\right)^2\{(1-\hat{\delta}_1)E(X|X>0)-(1-\hat{\delta}_2)E(Y|Y>0)-\theta_0\}^2\right]\\
&\to E\left[\left(\frac{n_1}{m_1-1}\right)^2(\theta_0-\theta_0)^2\right]\\
&= 0.
\end{aligned}$$

Therefore,

$$E\zeta_{tk}=E[\Psi(g_1(x_k))]+E[\Psi'(g_1(x_k))R_{1k}]+E[\varphi(R_{1k})^2]\to E[\Psi(g_1(x_k))].$$

Also,

$$\begin{aligned} E(g_1(x_k)) &= E\left[\frac{t-1}{m_1-1}\{h_{01}(x_k)-\theta_0\}\right] \\ &= E\left[\frac{t-1}{m_1-1}[(1-\delta_1)x_k-(1-\delta_2)E(Y|Y>0)-\theta_0]\right] \\ &= E\left[E\left\{\frac{t-1}{m_1-1}[(1-\delta_1)x_k-(1-\delta_2)E(Y|Y>0)-\theta_0]\middle| m_1,n_1\right\}\right] \\ &= E\left[\frac{t-1}{m_1-1}\{(1-\delta_1)E(x_k|m_1)-(1-\delta_2)E(Y|Y>0)-\theta_0]\}\right] \\ &= E\left[\frac{t-1}{m_1-1}\{(1-\delta_1)E(X|X>0)-(1-\delta_2)E(Y|Y>0)-\theta_0\}\right] \\ &= E\left[\frac{t-1}{m_1-1}(\theta_0-\theta_0)\right] \\ &= 0, \end{aligned}$$

and as $\sigma_1^2 > 0$, we have $P(g_1(x_k) > 0) > 0$. Then $E[\Psi(g_1(x_k))] > 0$.

Similarly, for $k = m_1 + 1, \ldots, t$, it can be shown that $E[\Psi(g_2(y_k))] > 0$, where $g_2(y_k) = t_1(n_1-1)^{-1}[h_{02}(y_k) - \theta_0]$.

Then, using the same arguments from Lemma A.6 of Jing et al. (2009), Lemma 3.1 can be proved. □

Lemma 3.2. *Let* $S^2_{m_1,n_1} = \sigma_1^2/m_1 + \sigma_2^2/n_1$. *If* $Eh_0^2(X,Y) < \infty, \sigma_1^2 > 0, \sigma_2^2 > 0$. *Then*

$$\tfrac{T_t-\theta_0}{S_{m_1,n_1}} \xrightarrow{\mathfrak{D}} N(0,1) \text{ as } t \to \infty.$$

Proof of Lemma 3.2. For our method, we have the consistent estimator of θ,

$$T_t = (1-\hat{\delta}_1)\tfrac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\tfrac{\sum_{j=1}^{n_1} y_j}{n_1}.$$

We can write,

$$T_t = (T_t - T_t^0) + T_t^0.$$

Since T_t^0 is a U-statistic for fixed δ_1 and δ_2, we have

$$\frac{T_t^0-\theta_0}{S_{m_1,n_1}} \xrightarrow{\mathfrak{D}} N(0,1) \text{ as } t \to \infty.$$

Also, $T_t - T_t^0 \to 0$ from Lemma 3.1. Thus, we prove Lemma 3.2. $\square$

Lemma 3.3. *Let* $S_t = t^{-1}\sum_{i=1}^{t}(\hat{V}_i - \theta_0)^2$ *and* $Eh_0^2(X,Y)^2 < \infty$. *Then with probability one, we have* $S_t = tS^2_{m_1,n_1} + o(1)$.

Proof of Lemma 3.3. Let $\hat{V}_i^0$ denote the pseudo-values from equations (3.1) and (3.2) when $\hat{\delta}_1$ and $\hat{\delta}_2$ are replaced by fixed δ_1 and δ_2, respectively. Define $S_t^0 = t^{-1}\sum_{i=1}^{t}(\hat{V}_i^0 - \theta_0)^2$. Then we can rewrite,

$$S_t = S_t^0 + [S_t - S_t^0].$$

Using Lemma A.7 of Jing et al. (2009), $S_t^0 = tS^2_{m_1,n_1} + o(1)$. We also need to show that $S_t - S_t^0$ is negligible. To this end, we have

$$\begin{aligned} T_t &= (1-\hat{\delta}_1)\frac{\sum_{i=1}^{m_1} x_i}{m_1} - (1-\hat{\delta}_2)\frac{\sum_{j=1}^{n_1} y_j}{n_1} \\ &= (1-\hat{\delta}_1)U_{m_1} - (1-\hat{\delta}_2)U_{n_1}, \end{aligned}$$

where $U_{m_1} = 1/m_1\sum_{i=1}^{m_1} x_i$ and $U_{n_1} = 1/n_1\sum_{j=1}^{n_1} y_j$ are one sample U-statistics for $x_1, x_2, \ldots, x_{m_1}$ and $y_1, y_2, \ldots, y_{n_1}$, respectively. We can define

$$\hat{V}_k^I = tU_{m_1} - (t-1)U_{m_1-1}^{(-k)} \text{ and } \hat{V}_k^{II} = tU_{n_1} - (t-1)U_{n_1-1}^{(-k)}.$$

Then,

$$\hat{V}_k = (1-\hat{\delta}_1)\hat{V}_k^I - (1-\hat{\delta}_2)U_{n_1}; \quad \text{for } k = 1, 2, \ldots, m_1,$$

and

$$\hat{V}_k = (1-\hat{\delta}_1)U_{m_1} - (1-\hat{\delta}_2)\hat{V}_k^{II}; \quad \text{for } k = m_1+1, m_1+2, \ldots, t.$$

Now,

$$\begin{aligned} S_t - S_t^0 &= \frac{1}{t}\sum_{k=1}^{t}[(\hat{V}_k - \theta_0)^2 - (\hat{V}_k^0 - \theta_0)^2] \\ &= \frac{1}{t}\sum_{k=1}^{t}[((\hat{V}_k)^2 - (\hat{V}_k^0)^2) - 2\theta_0(\hat{V}_k - \hat{V}_k^0)] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{t}\sum_{k=1}^{t}((\hat{V}_k)^2-(\hat{V}_k^0)^2)-2\theta_0\frac{1}{t}\sum_{k=1}^{t}(\hat{V}_k-\hat{V}_k^0)\\
&= \frac{1}{t}\left[\sum_{k=1}^{m_1}((\hat{V}_k)^2-(\hat{V}_k^0)^2)+\sum_{k=m_1+1}^{t}((\hat{V}_k)^2-(\hat{V}_k^0)^2)\right]\\
&\quad -2\theta_0\frac{1}{t}\left[\sum_{k=1}^{m_1}(\hat{V}_k-\hat{V}_k^0)+\sum_{k=m_1+1}^{t}(\hat{V}_k-\hat{V}_k^0)\right]\\
&= \frac{1}{t}\left[\sum_{k=1}^{m_1}\{((1-\hat{\delta}_1)\hat{V}_k^I-(1-\hat{\delta}_2)U_{n_1})^2-((1-\delta_1)\hat{V}_k^I-(1-\delta_2)U_{n_1})^2\}\right.\\
&\quad \left.+\sum_{k=m_1+1}^{t}\{((1-\hat{\delta}_1)U_{m_1}-(1-\hat{\delta}_2)\hat{V}_k^{II})^2-((1-\delta_1)U_{m_1}-(1-\delta_2)\hat{V}_k^{II})^2\}\right]\\
&\quad -2\theta_0\frac{1}{t}\left[\sum_{k=1}^{m_1}\{(1-\hat{\delta}_1)\hat{V}_k^I-(1-\hat{\delta}_2)U_{n_1}-(1-\delta_1)\hat{V}_k^I+(1-\delta_2)U_{n_1}\}\right.\\
&\quad \left.+\sum_{k=m_1+1}^{t}\{(1-\hat{\delta}_1)U_{m_1}-(1-\hat{\delta}_2)\hat{V}_k^{II}-(1-\delta_1)U_{m_1}+(1-\delta_2)\hat{V}_k^{II}\}\right]\\
&= \frac{m_1}{t}\frac{1}{m_1}\left[\sum_{k=1}^{m_1}\{(\hat{V}_k^I)^2((1-\hat{\delta}_1)^2-(1-\delta_1)^2)-2\hat{V}_k^IU_{n_1}((1-\hat{\delta}_1)(1-\hat{\delta}_2)\right.\\
&\quad \left.-(1-\delta_1)(1-\delta_2))+U_{n_1}^2((1-\hat{\delta}_2)^2-(1-\delta_2)^2)\}\right]\\
&\quad +\frac{n_1}{t}\frac{1}{n_1}\left[\sum_{k=m_1+1}^{t}\{((1-\hat{\delta}_1)^2U_{m_1}^2-2\hat{V}_k^{II}U_{m_1}((1-\hat{\delta}_1)(1-\hat{\delta}_2)\right.\\
&\quad \left.-(1-\delta_1)(1-\delta_2))+(\hat{V}_k^{II})^2((1-\hat{\delta}_2)^2-(1-\delta_2)^2)\}\right]\\
&\quad -2\theta_0\left[\frac{m_1}{t}\frac{1}{m_1}\sum_{k=1}^{m_1}\{\hat{V}_k^I(\delta_1-\hat{\delta}_1)-U_{n_1}(\delta_2-\hat{\delta}_2)\}\right.\\
&\quad \left.+\frac{n_1}{t}\frac{1}{n_1}\sum_{k=m_1+1}^{t}\{U_{m_1}(\delta_1-\hat{\delta}_1)-\hat{V}_k^{II}(\delta_2-\hat{\delta}_2)\}\right]\\
&\to 0 \text{ as } t\to\infty.
\end{aligned}
$$

This is true because

$$
\begin{aligned}
\frac{1}{m_1}\sum_{k=1}^{m_1}\hat{V}_k^I &= \frac{1}{m_1}\sum_{k=1}^{m_1}tU_{m_1}-\frac{1}{m_1}\sum_{k=1}^{m_1}(t-1)U_{m_1-1}^{(-k)}\\
&= tU_{m_1}-\frac{1}{m_1}(t-1)m_1U_{m_1}
\end{aligned}
$$

$$= \quad U_{m_1} < \infty.$$

Similarly, $n_1^{-1}\sum_{k=m_1+1}^{t}\hat{V}_k^{II} < \infty$. Also

$$\begin{aligned}
\frac{1}{m_1}\sum_{k=1}^{m_1}(\hat{V}_k^I)^2 &= \frac{1}{m_1}\sum_{k=1}^{m_1}(\hat{V}_k^I - E(U_{m_1}) + E(U_{m_1}))^2 \\
&= \frac{1}{m_1}\sum_{k=1}^{m_1}(\hat{V}_k^I - E(U_{m_1}))^2 + 2E(U_{m_1})\frac{1}{m_1}\sum_{k=1}^{m_1}\hat{V}_k^I - E(U_{m_1})^2 \\
&= \frac{1}{m_1}\sum_{k=1}^{m_1}(\hat{V}_k^I - E(U_{m_1}))^2 + 2U_{m_1}E(U_{m_1}) - E(U_{m_1})^2,
\end{aligned}$$

in which both $1/m_1\sum_{k=1}^{m_1}(\hat{V}_k^I - E(U_{m_1}))^2$ and $2U_{m_1}E(U_{m_1}) - E(U_{m_1})^2$ go to a finite number, i.e., $1/m_1\sum_{k=1}^{m_1}(\hat{V}_k^I)^2 = C + o(1)$ for some constant C. Similarly, $1/n_1\sum_{k=m_1+1}^{t}(\hat{V}_k^{II})^2 = C + o(1)$ for some constant C. One has $0 < 1/m_1 < 1,\ 0 < 1/n_1 < 1,\ (\hat{\delta}_1 - \delta_1) \to 0$ and $(\hat{\delta}_2 - \delta_2) \to 0$. Therefore, $S_t - S_t^0 \to 0$. □

Lemma 3.4. *Let* $H_t = \underset{1\le i\le m_1, 1\le j\le n_1, i\ne j}{max}|(1-\hat{\delta}_1)x_i - (1-\hat{\delta}_2)y_j|$. *If* $EX_1^2 < \infty$ *and* $EY_1^2 < \infty$, *then* $H_t = o_p(t^{1/2})$, *as* $t \to \infty$.

Proof of Lemma 3.4. We can write

$$\begin{aligned}
H_t &= \max_{1\le i\le m_1, 1\le j\le n_1, i\ne j}|(1-\hat{\delta}_1)x_i - (1-\delta_1)x_i + (1-\delta_1)x_i - (1-\hat{\delta}_2)y_j \\
&\quad +(1-\delta_2)y_j - (1-\delta_2)y_j| \\
&= \max_{1\le i\le m_1}|(\hat{\delta}_1 - \delta_1)x_i| + \max_{1\le j\le n_1}|(\hat{\delta}_2 - \delta_2)y_j| \\
&\quad + \max_{1\le i\le m_1, 1\le j\le n_1, i\ne j}|(1-\delta_1)x_i - (1-\delta_2)y_j| \\
&= |\hat{\delta}_1 - \delta_1|\max_{1\le i\le m_1}|x_i| + |\hat{\delta}_2 - \delta_2|\max_{1\le j\le n_1}|y_j| \\
&\quad + \max_{1\le i\le m_1, 1\le j\le n_1, i\ne j}|(1-\delta_1)x_i - (1-\delta_2)y_j|.
\end{aligned}$$

As $|\hat{\delta}_1 - \delta_1| = O_p(m^{-1/2})$ and $|\hat{\delta}_2 - \delta_2| = O_p(n^{-1/2})$,

$$H_t = O_p(m^{-1/2})o_p(m^{1/2}) + O_p(n^{-1/2})o_p(n^{1/2})$$

$$
\begin{aligned}
& + \max_{1\le i\le m_1, 1\le j\le n_1, i\ne j} |(1-\delta_1)x_i - (1-\delta_2)y_j| \\
= & \ o_p(1) + o_p(1) + \max_{1\le i\le m_1, 1\le j\le n_1, i\ne j} |(1-\delta_1)x_i - (1-\delta_2)y_j| \\
= & \max_{1\le i\le m_1, 1\le j\le n_1, i\ne j} |(1-\delta_1)x_i - (1-\delta_2)y_j| + o_p(1).
\end{aligned}
$$

By using Lemma A.8 in Jing et al. (2009), $H_t = o_p(t^{1/2})$, as $t \to \infty$. □

Proof of Theorem 3.1.Using the same arguments in the proof of Theorem 1 in Jing et al. (2009), Theorem 3.1 can be proved. The details are omitted here. □

Proof of Theorem 3.2. The proof follows Chen et al. (2008) and Zhao et al. (2015). We give a sketch of the proof following Satter and Zhao (2020). From equation (3.3), let $\lambda(\theta_0)$ be the solution to

$$\sum_{i=1}^{t+1} \frac{g_i(\theta_0)}{1+\lambda(\theta_0)g_i(\theta_0)} = 0,$$

where $g_i(\theta_0)$ is defined as $g_i(\theta_0) = \hat{V}_i - \theta_0, i = 1, 2\ldots, t$, and $g_{t+1}(\theta_0) = -(a_t/t)\sum_{i=1}^{t} g_i(\theta_0) = -a_t\bar{g}(\theta_0)$.

We have $g^\star(\theta_0) = \max_{1\le i\le t}|\hat{V}_i - \theta_0| \le \max_{1\le i\le t}|\hat{V}_i| + \theta_0 = \max_{1\le i\le t}|x_i| + \theta_0 = o_p(t^{1/2})$, and $\bar{g}_t(\theta_0) = O_p(t^{-1/2})$. Let $\rho(\theta_0) = ||\lambda(\theta_0)||$ and $\hat{\lambda}(\theta_0) = \lambda(\theta_0)/\rho(\theta_0)$. Then

$$
\begin{aligned}
0 &= \frac{\hat{\lambda}(\theta_0)}{t}\sum_{i=1}^{t+1} \frac{g_i(\theta_0)}{1+\lambda(\theta_0)g_i(\theta_0)} \\
&= \frac{\hat{\lambda}(\theta_0)}{t}\sum_{i=1}^{t+1} g_i(\theta_0) - \frac{\rho(\theta_0)}{t}\sum_{i=1}^{t+1}\frac{(\hat{\lambda}(\theta_0)g_i(\theta_0))^2}{1+\rho(\theta_0)\hat{\lambda}(\theta_0)g_i(\theta_0)} \\
&\le \hat{\lambda}(\theta_0)\bar{g}_t(\theta_0)(1-\frac{a_t}{t}) - \frac{\rho(\theta_0)}{t(1+\rho(\theta_0)g^\star(\theta_0))}\sum_{i=1}^{t}(\hat{\lambda}(\theta_0)g_i(\theta_0))^2.
\end{aligned}
$$

With $a_t = o_p(t)$, one has

$$
\begin{aligned}
\frac{\rho(\theta_0)}{1+\rho(\theta_0)g^\star(\theta_0)} &\le \hat{\lambda}(\theta_0)\bar{g}_t(\theta_0)\left(1-\frac{a_t}{t}\right)\left(\frac{1}{t}\sum_{i=1}^{t}(\hat{\lambda}(\theta_0)g_i(\theta_0))^2\right)^{-1} \\
&= O_p(t^{-1/2}).
\end{aligned}
$$

It implies $\lambda(\theta_0) = O_p(t^{-1/2})$. With $\hat{V}_t(\theta_0) = {}^1\!/\!_t \sum_{i=1}^{t} g_i^2(\theta_0)$, we have

$$\begin{aligned} 0 &= \frac{1}{t}\sum_{i=1}^{t+1} \frac{g_i(\theta_0)}{1+\mu(\theta_0)g_i(\theta_0)} \\ &= \bar{g}_t(\theta_0) - \lambda(\theta_0)\hat{V}_t(\theta_0) + o_p(t^{-1/2}). \end{aligned}$$

As $t \to \infty$, we have, $\lambda(\theta_0) = \bar{g}_t(\theta_0) \Big/ \hat{V}_t(\theta_0) + o_p(t^{-1/2})$. Therefore, by substituting $\lambda(\theta_0)$ by the previous equation

$$\begin{aligned} -2\log R^{\star}(\theta_0) &= 2\sum_{i=1}^{t+1} \log(1+\lambda(\theta_0)g_i(\theta_0)) \\ &= 2\sum_{i=1}^{t+1} \left\{ \lambda(\theta_0)g_i(\theta_0) - \frac{(\lambda(\theta_0)g_i(\theta_0))^2}{2} \right\} + o_p(1). \\ &= \frac{t\bar{g}_t^2(\theta_0)}{\hat{V}_t(\theta_0)} + o_p(1). \end{aligned}$$

Hence, $-2\log R^{\star}(\theta_0) \to \chi_1^2$ as $t \to \infty$. $\square$

Appendix C

PROOFS OF CHAPTER 4

Let $D(\theta; s,t) = F\left[\theta + F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right] - \left(\frac{t-\delta}{1-\delta}\right)$. For the theoretical development, we assume the following regularity conditions:

(C1) $w(u)$ is a symmetric density function with support $[-1,1]$ and $w'(u)$ is bounded and continuous for $u \in [-1,1]$;

(C2) $h = h(n) \to 0, nh^2/\log(n)$ as $n \to \infty$, and $nh^4 \to 0$ as $n \to \infty$;

(C3) $F(x)$ and its first derivative $f(x)$ are bounded and continuous. Assume that $f\{F^{-1}(s)\} > 0$ and $f\{F^{-1}(t)\} > 0$;

(C4) $D(\theta; s,t)$ and its first first derivation with respect to s, $D'(\theta; s,t)$ are bounded and continuous.

Lemma 4.1. *Assume that $0 < \delta < 1$ and the regularity conditions $C1 - C4$ hold. Then for θ_0, the true quantile difference between s and t,*

$$T_{n_1}(\theta_0; s,t) \xrightarrow{P} 0,$$

Proof of Lemma 4.1. Let

$$\widetilde{T}_{n_1}(\theta_0; s,t) = \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F(x_i - \theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}.$$

We can decompose $T_{n_1}(\theta_0; s,t)$ as

$$T_{n_1}(\theta_0; s,t) = \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i - \theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}} - \widetilde{T}_{n_1}(\theta_0; s,t) + \widetilde{T}_{n_1}(\theta_0; s,t)$$

We can further simplify $\widetilde{T}_{n_1}(\theta_0; s, t)$ as

$$
\begin{aligned}
\widetilde{T}_{n_1}(\theta_0; s, t) &= \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F(x_i - \theta_0)}{h}\right] - \frac{t-\delta}{1-\delta} \\
&= \int_{-\infty}^{\infty} K\left[\frac{\frac{s-\delta}{1-\delta} - F(x - \theta_0)}{h}\right] \mathrm{d}F_{n_1}(x) - \frac{t-\delta}{1-\delta} \\
&= K\left[\frac{\frac{s-\delta}{1-\delta} - F(x - \theta_0)}{h}\right] F_{n_1}(x)|_{-\infty}^{\infty} \\
&\quad - \int_{-\infty}^{\infty} F_{n_1}(x) \mathrm{d}K\left[\frac{\frac{s-\delta}{1-\delta} - F(x - \theta_0)}{h}\right] - \frac{t-\delta}{1-\delta} \\
&= \frac{1}{h}\int_{-\infty}^{\infty} F_{n_1}(x) w\left[\frac{\frac{s-\delta}{1-\delta} - F(x - \theta_0)}{h}\right] \mathrm{d}F(x-\theta_0) - \frac{t-\delta}{1-\delta} \\
&= -\int_{-1}^{1} F_{n_1}\left[F^{-1}\left(\frac{s-\delta}{1-\hat{\delta}} - uh\right) + \theta_0\right] w(u)\mathrm{d}u - -\frac{t-\delta}{1-\delta} \\
&= -\int_{-1}^{1} \left\{F_{n_1}\left[F^{-1}\left(\frac{s-\delta}{1-\delta} - uh\right) + \theta_0\right] - F\left[F^{-1}\left(\frac{s-\delta}{1-\delta} - uh\right) + \theta_0\right]\right. \\
&\quad \left. + F\left[F^{-1}\left(\frac{s-\delta}{1-\delta} - uh\right) + \theta_0\right] - F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right) + \theta_0\right]\right\} w(u)\mathrm{d}u \\
&= o_p(1).
\end{aligned}
$$

The above equation is obtained by the Glivenko–Cantelli theorem of F and the bounded derivative of $D(\theta_0; s, t)$.

As $n \to \infty$, $(1-\hat{\delta}) \to (1-\delta)$. Then $(\hat{\delta} - \delta) \to 0$. Also

$$
\begin{aligned}
\frac{s-\hat{\delta}}{1-\hat{\delta}} &= \frac{s-\hat{\delta}}{1-\hat{\delta}} - \frac{s-\delta}{1-\delta} + \frac{s-\delta}{1-\delta} \\
&= s\left[\frac{1}{1-\hat{\delta}} - \frac{1}{1-\delta}\right] - \left[\frac{\hat{\delta}}{1-\hat{\delta}} - \frac{\delta}{1-\delta}\right] + \frac{s-\delta}{1-\delta} \\
&= s\left[\frac{1-\delta-1+\hat{\delta}}{(1-\hat{\delta})(1-\delta)}\right] - \left[\frac{\hat{\delta}(1-\delta) - \delta(1-\hat{\delta})}{(1-\hat{\delta})(1-\delta)}\right] + \frac{s-\delta}{1-\delta} \\
&= s\left[\frac{\hat{\delta}-\delta}{(1-\hat{\delta})(1-\delta)}\right] - \left[\frac{\hat{\delta}-\delta}{(1-\hat{\delta})(1-\delta)}\right] + \frac{s-\delta}{1-\delta} \\
&= \frac{s-\delta}{1-\delta} + o_p(1). \qquad \text{(C.1)}
\end{aligned}
$$

Similarly,

$$\frac{t-\hat{\delta}}{1-\hat{\delta}} = \frac{t-\delta}{1-\delta} + o_p(1). \tag{C.2}$$

Using eqns. $(C.1)$ and $(C.2)$, we have

$$\begin{aligned} T_{n_1}(\theta_0; s, t) &= \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i-\theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}} \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_i-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta} + o_p(1). \end{aligned}$$

Now, extending eqns. (10) and (11) in Gong et al. (2010), we have

$$\begin{aligned} T_{n_1}(\theta_0; s, t) - \widetilde{T}_n(\theta_0; s, t) &= \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_i-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta} \\ &\quad -\frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F(x_i-\theta_0)}{h}\right] + \frac{t-\delta}{1-\delta} + o_p(1) \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_i-\theta_0)}{h}\right] \\ &\quad -\frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F(x_i-\theta_0)}{h}\right] + o_p(1) \\ &= o_p(1) + o_p(1) \\ &= o_p(1). \end{aligned}$$

Then

$$T_{n_1}(\theta_0; s, t) \xrightarrow{P} 0.$$

□

Lemma 4.2. *Assume that $0 < \delta < 1$ and the regularity conditions $C1 - C4$ hold. Then for θ_0, the true quantile difference between s and t,*

$$\sqrt{n_1}T_{n_1}(\theta_0; s, t) \xrightarrow{\mathfrak{D}} N(0, \sigma_1^2),$$

where

$$\sigma_1^2 = \left(1-\frac{s-\delta}{1-\delta}\right)\left(\frac{s-\delta}{1-\delta}\right)\{D'(\theta_0;s,t)\}^2+2\left\{\frac{s-\delta}{1-\delta}\wedge\frac{t-\delta}{1-\delta}-\left(\frac{s-\delta}{1-\delta}\right)\left(\frac{t-\delta}{1-\delta}\right)\right.$$
$$\left.D'(\theta_0;s,t)\right\}+\left(1-\frac{t-\delta}{1-\delta}\right)\left(\frac{t-\delta}{1-\delta}\right).$$

Proof of Lemma 4.2. We can write

$$\begin{aligned}\sqrt{n_1}T_{n_1}(\theta_0;s,t) &= \sqrt{n_1}\widetilde{T}_{n_1}(\theta_0;s,t)+\sqrt{n_1}[T_{n_1}(\theta_0;s,t)-\widetilde{T}_{n_1}(\theta_0;s,t)]\\ &:= I+II.\end{aligned}$$

For term I,

$$\begin{aligned}\sqrt{n_1}\widetilde{T}_{n_1}(\theta_0;s,t) &= \int_{-1}^{1}\sqrt{n_1}\left\{F_{n_1}\left[F^{-1}\left(\frac{s-\delta}{1-\delta}-uh\right)+\theta_0\right]\right.\\ &\quad -F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}-uh\right)+\theta_0\right]\\ &\quad \left.+F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}-uh\right)+\theta_0\right]-F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)+\theta_0\right]\right\}w(u)\mathrm{d}u\\ &= \int_{-1}^{1}W_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}-uh\right)+\theta_0\right]w(u)du\\ &\quad +\sqrt{n_1}\int_{-1}^{1}D'(\theta_0;s,t)\frac{uh}{1-\delta}\ w(u)du+O_p(\sqrt{n_1}h^2)\\ &= \int_{-1}^{1}W_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)+\theta_0\right]w(u)du+o_p(1)\\ &= W_F\left[F^{-1}\left(\frac{t-\delta}{1-\delta}\right)\right]+o_p(1),\end{aligned}$$

where $W_F(t)=\sqrt{n_1}[F_{n_1}(t)-F(t)]$. Note that $\sqrt{n_1}\int_{-1}^{1}D'(\theta_0;s,t)uh\ w(u)du=0$ because of the symmetric property of kernel function.

Using similar steps as Gong et al. (2010), we can write term II as

$$\sqrt{n_1}[T_{n_1}(\theta_0;s,t)-\widetilde{T}_{n_1}(\theta_0;s,t)] = \sqrt{n_1}\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}}-F_{n_1}(x_i-\theta_0)}{h}\right]\right.$$

$$
\begin{aligned}
& -\frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta}-F(x_i-\theta_0)}{h}\right]\Bigg\} \\
= \; & \sqrt{n_1}\Bigg\{\frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_i-\theta_0)}{h}\right] \\
& -\frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta}-F(x_i-\theta_0)}{h}\right]\Bigg\} + o_p(1) \\
= \; & -\int_{-\infty}^{\infty} W_F(x-\theta_0) w\left[\frac{\frac{s-\delta}{1-\delta}-F(x-\theta_0)}{h}\right] dF(x) \\
& + O_p(n_1^{-1/2}h^{-1}) + o_p(1) \\
= \; & \int_{-1}^{1} W_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right] w(u) D'(\theta_0; s,t) du \\
& + o_p(1) + o_p(1) \\
= \; & W_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right] D'(\theta_0; s,t) + o_p(1).
\end{aligned}
$$

By the covariance properties of Wiener process, we have

$$
\text{Cov}\left\{W_F\left(\frac{s-\delta}{1-\delta}\right), W_F\left(\frac{t-\delta}{1-\delta}\right)\right\} = F\left(\frac{s-\delta}{1-\delta}\wedge\frac{t-\delta}{1-\delta}\right) - F\left(\frac{s-\delta}{1-\delta}\right)F\left(\frac{t-\delta}{1-\delta}\right),
$$

where $\sqrt{n_1}\{F_{n_1}(x)-F(x)\} \stackrel{\mathfrak{D}}{\to} W_F(x)$. Then, the cross-effect of $F_{n_1}\left(\frac{s-\delta}{1-\delta}\right)$ and $F_{n_1}\left(\frac{t-\delta}{1-\delta}\right)$ is given by $2\left(\frac{s-\delta}{1-\delta}\wedge\frac{t-\delta}{1-\delta} - \frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right)D'(\theta_0; s,t)$. Therefore, by the Donsker Theorem,

$$
\begin{aligned}
\sqrt{n_1} T_{n_1}(\theta_0; s,t) \; &= \; W_F\left[F^{-1}\left(\frac{t-\delta}{1-\delta}\right)\right] + W_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right] D'(\theta_0; s,t) + o_p(1) \\
& \stackrel{\mathfrak{D}}{\to} \; B_F\left[F^{-1}\left(\frac{t-\delta}{1-\delta}\right)\right] + B_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right] D'(\theta_0; s,t) \\
&= \; N(0, \sigma_1^2), \qquad \text{(C.3)}
\end{aligned}
$$

where $B_F(\cdot)$ is a Brownian bridge for F and

$$
\begin{aligned}
& \text{Var}\left\{B_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right] D'(\theta_0; s,t) + B_F\left[F^{-1}\left(\frac{t-\delta}{1-\delta}\right)\right]\right\} \\
= \; & \text{Var}\left\{B_F\left[F^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right] D'(\theta_0; s,t)\right\} + \text{Var}\left\{B_F\left[F^{-1}\left(\frac{t-\delta}{1-\delta}\right)\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
&+2\ \mathrm{Cov}\bigg\{B_F\bigg[F^{-1}\bigg(\frac{s-\delta}{1-\delta}\bigg)\bigg]D'(\theta_0;s,t), B_F\bigg[F^{-1}\bigg(\frac{t-\delta}{1-\delta}\bigg)\bigg]\bigg\} \\
= &\bigg(1-\frac{s-\delta}{1-\delta}\bigg)\bigg(\frac{s-\delta}{1-\delta}\bigg)\{D'(\theta_0;s,t)\}^2 + 2\bigg\{\frac{s-\delta}{1-\delta}\wedge\frac{t-\delta}{1-\delta} - \bigg(\frac{s-\delta}{1-\delta}\bigg)\bigg(\frac{t-\delta}{1-\delta}\bigg) \\
&D'(\theta_0;s,t)\bigg\} + \bigg(1-\frac{t-\delta}{1-\delta}\bigg)\bigg(\frac{t-\delta}{1-\delta}\bigg) \\
= &\ \sigma_1^2.
\end{aligned}
$$

□

Lemma 4.3. *Assume that $0 < \delta < 1$ and the regularity conditions $C1 - C4$ hold. Then for θ_0, the true quantile difference between s and t,*

$$
\sqrt{n_1}\bigg\{\frac{1}{n_1}\sum_{i=1}^{n_1} V_i(\theta_0;s,t)\bigg\} \xrightarrow{\mathfrak{D}} N(0,\sigma_1^2).
$$

Proof of Lemma 4.3. One has

$$
\begin{aligned}
F_{n_1}^{-i}(x_j) - F_{n_1}(x_j) &= \frac{1}{n_1-1}\sum_{k=1,k\neq i}^{n_1} I(x_k\le x_j) - \frac{1}{n_1}\sum_{k=1}^{n_1} I(x_k\le x_j) \\
&= \frac{1}{n_1-1}\sum_{k=1}^{n_1} I(x_k\le x_j) - \frac{1}{n_1}\sum_{k=1}^{n_1} I(x_k\le x_j) \\
&\quad -\frac{1}{n_1-1}\sum_{k=1}^{n_1} I(x_k\le x_j) + \frac{1}{n_1-1}\sum_{k=1,k\neq i}^{n_1} I(x_k\le x_j) \\
&= \bigg(\frac{1}{n_1-1}-\frac{1}{n_1}\bigg)\sum_{k=1}^{n_1} I(x_k\le x_j) \\
&\quad -\frac{1}{n_1-1}\bigg\{\sum_{k=1}^{n_1} I(x_k\le x_j) - \sum_{k=1,k\neq i}^{n_1} I(x_k\le x_j)\bigg\} \\
&= \bigg(\frac{1}{n_1-1}-\frac{1}{n_1}\bigg)\sum_{k=1}^{n_1} I(x_k\le x_j) - \frac{1}{n_1-1} I(x_i\le x_j) \\
&= \frac{1}{n_1(n_1-1)}\sum_{k=1}^{n_1} I(x_k\le x_j) - \frac{1}{n_1-1} I(x_i\le x_j) \\
&= \frac{1}{n_1-1}\bigg(\frac{1}{n_1}\sum_{k=1}^{n_1} I(x_k\le x_j)\bigg) - \frac{1}{n_1-1} I(x_i\le x_j) \\
&= \frac{1}{n_1-1}\{F_{n_1}(x_j) - I(x_i\le x_j)\}.
\end{aligned}
$$

Now, we establish some properties of $F_{n_1}^{-i}$ as:

$$\begin{aligned} F_{n_1}^{-i}(x_j) - F_{n_1}(x_j) &= \frac{1}{n_1-1}\{F_{n_1}(x_j) - I(x_i \le x_j)\} \\ &= O_p\left(\frac{1}{n_1-1}\right), j = 1, 2, \ldots, n_1. \end{aligned}$$

And

$$\begin{aligned} \sum_{i=1}^{n_1}\{F_{n_1}^{-i}(x_j) - F_{n_1}(x_j)\} &= \sum_{i=1}^{n_1}\frac{1}{n_1-1}\{F_{n_1}(x_j) - I(x_i \le x_j)\} \\ &= \frac{1}{n_1-1}\sum_{i=1}^{n_1}F_{n_1}(x_j) - \frac{1}{n_1-1}\sum_{i=1}^{n_1}I(x_i \le x_j) \\ &= \frac{n_1}{n_1-1}F_{n_1}(x_j) - \frac{n_1}{n_1-1}\frac{1}{n_1}\sum_{i=1}^{n_1}I(x_i \le x_j) \\ &= \frac{n_1}{n_1-1}F_{n_1}(x_j) - \frac{n_1}{n_1-1}F_{n_1}(x_j) \\ &= 0. \end{aligned}$$

Using equation (16) in Gong et al. (2010), we have

$$\begin{aligned} &\frac{1}{n_1}\sum_{i=1}^{n_1}V_i(\theta_0; s, t) \\ =\ & n_1 T_{n_1}(\theta_0; s, t) - \frac{n_1-1}{n_1}\sum_{i=1}^{n_1}T_{n_1}^{-i}(\theta_0; s, t) \\ =\ & n_1\left(\frac{1}{n_1}\sum_{i=1}^{n_1}K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i-\theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}}\right) \\ & -\frac{n_1-1}{n_1}\left(\frac{1}{n_1-1}\sum_{i=1, i\neq j}^{n_1}K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}^{-j}(x_i-\theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}}\right) \\ =\ & \sum_{i=1}^{n_1}K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i-\theta_0)}{h}\right] - \frac{1}{n_1}\sum_{i=1}^{n_1}\sum_{i=1, i\neq j}^{n_1}K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}^{-j}(x_i-\theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}} \\ =\ & \sum_{i=1}^{n_1}K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i-\theta_0)}{h}\right] - \frac{1}{n_1}\sum_{i=1}^{n_1}\sum_{i=1, i\neq j}^{n_1}K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}^{-j}(x_i-\theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i - \theta_0)}{h}\right] - \frac{1}{n_1}\sum_{i=1}^{n_1}\sum_{i=1,i\neq j}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i - \theta_0)}{h}\right] \\
&\quad + \frac{1}{n_1}\sum_{i=1}^{n_1}\sum_{i=1,i\neq j}^{n_1}\left(K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i - \theta_0)}{h}\right] - K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}^{-i}(x_i - \theta_0)}{h}\right]\right) - \frac{t-\hat{\delta}}{1-\hat{\delta}} \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i - \theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}} + O_p\left(\frac{n_1^2}{n_1(n_1-1)^2 h}\right) \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i - \theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}} + o_p(1) \\
&= T_{n_1}(\theta_0; s, t) + o_p(1). \qquad\qquad (C.4)
\end{aligned}$$

From equations $(C.3)$ and $(C.4)$,

$$\begin{aligned}
\sqrt{n_1}\left\{\frac{1}{n_1}\sum_{i=1}^{n_1} V_i(\theta_0; s, t)\right\} &= \sqrt{n_1} T_{n_1}(\theta_0; s, t) + o_p(1) \\
&\xrightarrow{\mathfrak{D}} N(0, \sigma_1^2).
\end{aligned}$$

□

We define the pseudo-sample variance as

$$v_n^2(\theta_0; s, t) = \frac{1}{n_1}\sum_{i=1}^{n_1}\{V_i(\theta_0; s, t) - \frac{1}{n_1}\sum_{i=1}^{n_1} V_i(\theta_0; s, t)\}^2.$$

Lemma 4.4. *Assume that $0 < \delta < 1$ and the regularity conditions $C1 - C4$ hold. Then for θ_0, the true quantile difference between s and t,*

$$v_n^2(\theta_0; s, t) \xrightarrow{P} \sigma_1^2.$$

Proof of Lemma 4.4. For $1 \leq i \leq n_1$,

$$V_i(\theta_0; s, t) = \sum_{i=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_i - \theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}}$$

$$= \sum_{j=1}^{n_1}\left\{K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_j-\theta_0)}{h}\right] - K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}^{-i}(x_j-\theta_0)}{h}\right]\right\}$$
$$+K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}}.$$

Then,

$$\begin{aligned}
\frac{1}{n_1}\sum_{i=1}^{n_1} V_i^2(\theta_0;s,t) &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1} K\left[\frac{\frac{s-\hat{\delta}}{1-\hat{\delta}} - F_{n_1}(x_j-\theta_0)}{h}\right] - \frac{t-\hat{\delta}}{1-\hat{\delta}}\right\}^2 \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right\}^2 + o_p(1) \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1}\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right] - K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_j-\theta_0)}{h}\right]\right)\right\}^2 \\
&\quad + \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right\}^2 \\
&\quad + \frac{2}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1}\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right] - K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right]\right)\right. \\
&\quad \left.\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right)\right\} + o_p(1) \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1}\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right] - K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_j-\theta_0)}{h}\right]\right)\right\}^2 \\
&\quad + \left\{\frac{1}{n_1}\sum_{i=1}^{n_1} K^2\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right] - \left(\frac{t-\delta}{1-\delta}\right)^2\right. \\
&\quad \left. - \frac{2\frac{t-\delta}{1-\delta}}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right]\right\} \\
&\quad + \frac{2}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1}\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right] - K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right]\right)\right. \\
&\quad \left.\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right)\right\} + o_p(1) \\
&:= T_1 + T_2 + T_2 + o_p(1). \qquad \text{(C.5)}
\end{aligned}$$

Now,

$$\begin{aligned}
T_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1}\left(K\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_j-\theta_0)}{h}\right]-K\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}^{-i}(x_j-\theta_0)}{h}\right]\right)\right\}^2\\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1}w\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_j-\theta_0)}{h}\right]\frac{F_{n_1}(x_j-\theta_0)-F_{n_1}^{-i}(x_j-\theta_0)}{h}\right\}^2+o_p(1)\\
&= \frac{1}{n_1(n_1-1)^2h^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\sum_{l=1}^{n_1}\{F_{n_1}(x_j-\theta_0)F_{n_1}(x_l-\theta_0)\\
&\quad -F_{n_1}(x_j-\theta_0)I(x_i\le x_l-\theta_0)-F_{n_1}(x_l-\theta_0)I(x_i\le x_j-\theta_0)\\
&\quad +I(x_i\le x_j-\theta_0)I(xi\le x_l-\theta_0)\}w\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_j-\theta_0)}{h}\right]\\
&\quad w\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_l-\theta_0)}{h}\right]+o_p(1)\\
&= \frac{1}{h^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F_{n_1}(x_1\wedge x_2-\theta_0)-F_{n_1}(x_1-\theta_0)F_{n_1}(x_2-\theta_0)\}\\
&\quad w\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_1-\theta_0)}{h}\right]w\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_2-\theta_0)}{h}\right]dF_{n_1}(x_1)dF_{n_1}(x_2)+o_p(1)\\
&= \frac{1}{h^2}\int_{-1}^{1}\int_{-1}^{1}\left[F\left\{F^{-1}\left(\frac{s-\delta}{1-\delta}-u_1h\right)\wedge F^{-1}\left(\frac{s-\delta}{1-\delta}-u_2h\right)\right\}\right.\\
&\quad \left.-F\left\{F^{-1}\left(\frac{s-\delta}{1-\delta}-u_1h\right)\right\}F\left\{F^{-1}\left(\frac{s-\delta}{1-\delta}-u_2h\right)\right\}\right]w(u_1)w(u_2)\\
&\quad dF\left\{F^{-1}\left(\frac{s-\delta}{1-\delta}-u_1h\right)+\theta_0\right\}dF\left\{F^{-1}\left(\frac{s-\delta}{1-\delta}-u_2h\right)+\theta_0\right\}+o_p(1)\\
&= \int_{-1}^{1}\int_{-1}^{1}\frac{s-\delta}{1-\delta}\left(1-\frac{s-\delta}{1-\delta}\right)\{D'(\theta_0;s,t)\}^2w(u_1)w(u_2)du_1du_2\\
&= \frac{s-\delta}{1-\delta}\left(1-\frac{s-\delta}{1-\delta}\right)\{D'(\theta_0;s,t)\}^2+o_p(1). \qquad \text{(C.6)}
\end{aligned}$$

For T_2, following eqn. (18) of Gong et al. (2010), we have

$$\begin{aligned}
T_2 &= \frac{1}{n_1}\sum_{i=1}^{n_1}K^2\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}^{-i}(x_i-\theta_0)}{h}\right]-\left(\frac{t-\delta}{1-\delta}\right)^2\\
&\quad -\frac{2\frac{t-\delta}{1-\delta}}{n_1}\sum_{i=1}^{n_1}K\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}^{-i}(x_i-\theta_0)}{h}\right]\\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}K^2\left[\frac{\frac{s-\delta}{1-\delta}-F_{n_1}(x_i-\theta_0)}{h}\right]-\left(\frac{t-\delta}{1-\delta}\right)^2
\end{aligned}$$

$$
\begin{aligned}
& -\frac{2\frac{t-\delta}{1-\delta}}{n_1}\sum_{i=1}^{n_1} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right] + o_p(1) \\
\xrightarrow{P} \quad & \frac{t-\delta}{1-\delta}\left(1-\frac{t-\delta}{1-\delta}\right). \qquad \text{(C.7)}
\end{aligned}
$$

And the third term,

$$
\begin{aligned}
T_3 \quad = \quad & \frac{2}{n_1}\sum_{i=1}^{n_1}\left\{\sum_{j=1}^{n_1}\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right] - K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right]\right)\right. \\
& \left.\left(K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}^{-i}(x_i-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right)\right\} \\
= \quad & \frac{2}{n_1}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\left\{w\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right]\frac{F_{n_1}^{-i}(x_j-\theta_0) - F_{n_1}(x_j-\theta_0)}{h}\right\} \\
& \left\{K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h} - \frac{t-\delta}{1-\delta}\right\} + o_p(1)\right. \\
= \quad & \frac{2}{n_1 h}\sum_{i=1}^{n_1}\left(\left\{K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right\}\frac{1}{n_1-1}\sum_{j=1}^{n_1}\{F_{n_1}(x_j-\theta_0)\right. \\
& \left. I(x_i < x_j-\theta_0)\}w\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}\right]\right) + o_p(1) \\
= \quad & \frac{2}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[F_{n_1}(x_2-\theta_0) - I(x_1 < x_2-\theta_0)]w\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_2-\theta_0)}{h}\right]dF_{n_1}(x_2) \\
& \left\{K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right\}dF_{n_1}(x_1) + o_p(1).
\end{aligned}
$$

Replacing $\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_j-\theta_0)}{h}$ by u_2,

$$
\begin{aligned}
& \frac{2}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\left(\frac{s-\delta}{1-\delta} - u_2 h\right) - I\left\{x_1 < F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta} - u_2 h\right)\right\}\right]w(u_2) \\
& dF_{n_1}\left(F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta} - u_2 h\right) + \theta_0\right) \\
& \left\{K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right] - \frac{t-\delta}{1-\delta}\right\}dF_{n_1}(x_1) + o_p(1) \\
= \quad & 2D'(\theta_0; s-u_2 h, t)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(u_2)\left[\frac{s-\delta}{1-\delta} - u_2 h - I\left\{x_1 < F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta} - u_2 h\right)\right\}\right] \\
& du_2\left\{\frac{t-\delta}{1-\delta} - K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right]\right\}dF_{n_1}(x_1) + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= 2D'(\theta_0;s,t)\int_{-\infty}^{\infty} w(u_2)du_2 \int_{-\infty}^{\infty}\left[\frac{s-\delta}{1-\delta} - I\left\{x_1 < F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right\}\right] \\
&\quad \left\{\frac{t-\delta}{1-\delta} - K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right]\right\}dF_{n_1}(x_1) + o_p(1) \\
&= 2D'(\theta_0;s,t)\left\{\frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta} - \frac{s-\delta}{1-\delta}\int_{-\infty}^{\infty} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right]dF_{n_1}(x_1)\right. \\
&\quad -\frac{t-\delta}{1-\delta}\int_{-\infty}^{\infty} I\left(x_1 - F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right)dF_{n_1}(x_1) \\
&\quad \left. + \int_{-\infty}^{F_{n_1}^{-1}(\frac{s-\delta}{1-\delta})} K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right]dF_{n_1}(x_1) + o_p(1)\right. \\
&= 2D'(\theta_0;s,t)\left\{F_{n_1}(x_1)K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right]\Bigg|_{\infty}^{F_{n_1}^{-1}(\frac{s-\delta}{1-\delta})}\right. \\
&\quad \left. - \int_{-\infty}^{F_{n_1}^{-1}(\frac{s-\delta}{1-\delta})} F_{n_1}(x_1)\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(x_1-\theta_0)}{h}\right]\right\} + 2D'(\theta_0;s,t)\left\{\frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right. \\
&\quad \left. -\frac{s-\delta}{1-\delta}F_{n_1}\left((F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right) + \theta_0\right) + \frac{t-\delta}{1-\delta}F_{n_1}\left(F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right)\right)\right\} + o_p(1) \\
&= 2D'(\theta_0;s,t)\left\{\frac{s-\delta}{1-\delta}K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}\left(F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right) - \theta_0\right)}{h}\right]\right. \\
&\quad \left. + \int_{\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(F_{n_1}^{-1}(\frac{s-\delta}{1-\delta}) - \theta_0)}{h}}^{\infty} F_{n_1}\left(F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta} - hu\right) + \theta_0\right)dK(u) - \frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right\} + o_p(1) \\
&= 2D'(\theta_0;s,t)\left\{\frac{s-\delta}{1-\delta}K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}\left(F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right) - \theta_0\right)}{h}\right]\right. \\
&\quad \left. + \int_{\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(F_{n_1}^{-1}(\frac{s-\delta}{1-\delta}) - \theta_0)}{h}}^{\infty} \frac{t-\delta}{1-\delta}dK(u) - \frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right\} + o_p(1) \\
&= 2D'(\theta_0;s,t)\left\{\frac{s-\delta}{1-\delta}K\left[\frac{\frac{s-\delta}{1-\delta} - F_{n_1}\left(F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right) - \theta_0\right)}{h}\right]\right. \\
&\quad \left. + \frac{t-\delta}{1-\delta}K(u)\Bigg|_{\frac{\frac{s-\delta}{1-\delta} - F_{n_1}(F_{n_1}^{-1}(\frac{s-\delta}{1-\delta}) - \theta_0)}{h}}^{\infty} - \frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right\} + o_p(1).
\end{aligned}
$$

As $h \to 0$, the function $K\left\{\frac{s-\delta}{1-\delta} - F_{n_1}\left(F_{n_1}^{-1}\left(\frac{s-\delta}{1-\delta}\right) - \theta_0\right)\Big/ h\right\} \to I(\theta_0 > 0)$. Then

$$
\begin{aligned}
T_3 &= 2D'(\theta_0;s,t)\left\{\frac{s-\delta}{1-\delta}I(\theta_0 > 0) + \frac{t-\delta}{1-\delta}(1 - I(\theta_0 > 0) - \frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right\} + o_p(1) \\
&= 2D'(\theta_0;s,t)\left\{\frac{s-\delta}{1-\delta}I\left(\frac{t-\delta}{1-\delta} > \frac{s-\delta}{1-\delta}\right) + \frac{t-\delta}{1-\delta}I\left(\frac{t-\delta}{1-\delta} \leq \frac{s-\delta}{1-\delta}\right)\right. \\
&\quad \left. -\frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right\} + o_p(1)
\end{aligned}
$$

$$\xrightarrow{P} \quad 2\left(\frac{s-\delta}{1-\delta} \wedge \frac{t-\delta}{1-\delta} - \frac{s-\delta}{1-\delta}\frac{t-\delta}{1-\delta}\right) D'(\theta_0; s, t). \tag{C.8}$$

Plugging the results from eqns. $(C.6)$, $(C.7)$, $(C.8)$ in Eqn. $(C.5)$, one has

$$\begin{aligned} v_n^2(\theta_0; s, t) &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{V_i(\theta_0; s, t) - \frac{1}{n_1}\sum_{i=1}^{n_1} V_i(\theta_0; s, t)\right\}^2 \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1} V_i^2(\theta_0; s, t) - \left(\frac{1}{n_1}\sum_{i=1}^{n_1} V_i(\theta_0; s, t)\right)^2 \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1} V_i^2(\theta_0; s, t) + o_p(1) \\ &\xrightarrow{P} \sigma_1^2. \end{aligned}$$

□

Proof of Theorem 4.1. Using the standard arguments in Owen (1988), combining with Lemmas 4.3 and Lemma 4.4, Theorem 4.1 can be proved. The details are omitted here.

□

Proof of Theorem 4.2. Using the similar arguments in Chen et al. (2008), Theorem 4.2 can be proved.

□

Proof of Theorem 4.3. Since transformed jackknife empirical likelihood shares the same asymptotic properties as jackknife empirical likelihood [cf. Jing et al. (2017)]. The proof of Theorem 4.3 is established.

□

Proof of Theorem 4.4. Following Jing et al. (2017), the transformed jackknife empirical log-likelihood ratio, $l^{ta}(\theta_0; s, t)$, should have the following properties so that it maintains the asymptotic properties of $l^a(\theta_0; s, t)$ [cf. Jing et al. (2017)] :

(P1) $0 \leq l^{ta}(\theta_0; s, t) \leq l^a(\theta_0; s, t)$;

(P2) $l^{ta}(\theta_0; s, t)$ is a monotonically increasing function of $l^a(\theta_0; s, t)$;

(P3) $l^{ta}(\theta_0; s, t) = l^a(\theta_0; s, t) + o_p(1)$;

(P4) The level-τ_1 contour of $l^{ta}(\theta_0; s, t)$, $\{\theta_0 : l^{ta}(\theta_0; s, t) = \tau_1; \forall\ \tau_1 \in [0, +\infty)\}$ is the same in shape as some level-τ_2 contour of $l^{ta}(\theta_0; s, t), \{\theta_0 : l^{ta}(\theta_0; s, t) = \tau_2\}$; and $l^{ta}(\tilde{\theta_0}; s, t) < l^{ta}(\theta_0; s, t)$ for $\theta_0 \neq \tilde{\theta}_0$.

As in Jing et al. (2017), we set $\gamma = 0.5$ and check if $l^{ta}(\theta_0; s, t)$ maintains the above mentioned properties:

(P1) Note that $l^a(\theta_0; s, t) \geq 0$, which implies

$$0 < \max\{1 - l^a(\theta_0; s, t)/(n_1 + 1), \gamma = 0.5\} \leq 1.$$

Then $0 \leq l^{ta}(\theta_0; s, t) \leq l^a(\theta_0; s, t)$.

(P2) The function $l^{ta}(\theta_0; s, t) = l^a(\theta_0; s, t)[1 - l^a(\theta_0; s, t)/(n_1 + 1)]$ is a strictly monotonically increasing function of $l^a(\theta_0; s, t)$ over the interval for $[0, (n_1 + 1)/2]$. For $l^a(\theta_0; s, t) > (n_1 + 1)/2, l^{ta}(\theta_0; s, t) = l^a(\theta_0; s, t)/2$ is also a strictly monotonically increasing function of $l^a(\theta_0; s, t)$. $l^{ta}(\theta_0; s, t)$ is also continuous over the entire interval of $[0, \infty)$. Therefore, $l^{ta}(\theta_0; s, t)$ is non-negative, continuous, and strictly monotonically increasing over $l^a(\theta_0; s, t) \in [0, \infty)$.

(P3) (i) Since the limiting distribution of $l^a(\theta_0; s, t)$ is χ^2_1 distribution, we have $l^a(\theta_0; s, t) = O_p(1)$.

(ii) From (i), we have $l^a(\theta_0; s, t) \leq (n_1 + 1)/2$ with probability tending to unity. Thus, for all asymptotic discussions, we may assume that $l^{ta}(\theta_0; s, t) = l^a(\theta_0; s, t)[1 - l^a(\theta_0; s, t)/(n_1 + 1)]$.

Combining (i) and (ii), we have $l^{ta}(\theta_0; s, t) = l^a(\theta_0; s, t) + o_p(1)$.

(P4) For a level-τ_1 contour of $l^{ta}(\theta_0; s, t), \{\theta_0 : l^{ta}(\theta_0; s, t) = \tau_1\}$ (by P2). Let $\tau_2 = (l^{ta})^{-1}(\theta_0; s, t)(\tau_1)$. Then, $\{\theta_0 : l^{ta}(\theta_0; s, t) = \tau_1\} = \{\theta_0 : l^a(\theta_0; s, t) = \tau_2\}$. Also, as $l^a(\theta_0; s, t)$ typically has a unique minimum at some $\tilde{\theta}_0$, the second part of (P4) also follows from the monotonicity of $l^{ta}(\theta_0; s, t)$.

Therefore, all four desired properties are satisfied by $l^{ta}(\theta_0; s, t)$. They make this transformation a good candidate transformation that in turns preserves the asymptotic properties of $l^a(\theta_0; s, t)$. It follows as $n \to \infty$,

$$-2l^{ta}(\theta_0; s, t) \xrightarrow{\mathfrak{D}} \chi^2_1.$$

□

Appendix D

PROOFS OF CHAPTER 5

We assume the following regularity conditions throughout the chapter.

(D1) The covariate vector $X(t)$ and the weight function $W(t)$ are bounded for all $t \in [0, \tau]$.

(D2) The function g is twice continuously differentiable.

(D3) The function $W(t)$ converges uniformly to a deterministic function $w(t)$ for all $t \in [0, \tau]$ a.s.

(D4) There exists a $\tau > 0$ such that $Pr(C_i \geq \tau) > 0, i = 1, \cdots, n$.

(D5) The matrix Σ is positive definite for all $t \in [0, \tau]$.

(D6) $A_* = E\left[\int_0^\tau \{Z(t) - \bar{z}(t; \gamma_0)\}^{\otimes 2} \Delta_i(t) e^{\gamma_0' Z(t)} \lambda_0(t) dt\right]$ is positive definite, where E is the expectation.

Lemma 5.1. *Assume that the conditions (D1)-(D6) hold. If* $\beta_0 = (\beta_{10}', \beta_{20}')'$ *are the true values of the parameters, then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{ni}(\beta_0; \hat{\gamma}) \xrightarrow{\mathfrak{D}} N(0, \Sigma).$$

Proof. Note that

$$\sum_{i=1}^{n} d\hat{M}_i(t; \beta_0, \hat{\gamma}) = \sum_{i=1}^{n} \left[Y_i(t)\Delta_i(t) dO_i(t) - g\{\hat{\mu}_0(t) e^{\beta_0' X_i(t)}\}\Delta_i(t) e^{\hat{\gamma}' Z_i(t)} d\hat{\Lambda}_0(t; \hat{\gamma})\right] = 0,$$

and

$$\sum_{i=1}^{n} d\hat{M}_i^*(t; \hat{\gamma}) \quad = \quad \sum_{i=1}^{n} \left[\Delta_i(t) dO_i(t) - \Delta_i(t) e^{\hat{\gamma}' Z_i(t)} d\hat{\Lambda}_0(t; \hat{\gamma})\right]$$

$$\begin{aligned}
&= \sum_{i=1}^{n}\left[\Delta_i(t)dO_i(t) - \Delta_i(t)e^{\hat{\gamma}'Z_i(t)}\frac{\sum_{i=1}^{n}\Delta_i(t)dO_i(t)}{\sum_{i=1}^{n}\Delta_i(t)e^{\hat{\gamma}'Z_i(t)}}\right] \\
&= \sum_{i=1}^{n}\Delta_i(t)dO_i(t) - \sum_{i=1}^{n}\Delta_i(t)e^{\hat{\gamma}'Z_i(t)}\frac{\sum_{i=1}^{n}\Delta_i(t)dO_i(t)}{\sum_{i=1}^{n}\Delta_i(t)e^{\hat{\gamma}'Z_i(t)}} \\
&= \sum_{i=1}^{n}\Delta_i(t)dO_i(t) - \sum_{i=1}^{n}\Delta_i(t)dO_i(t) \\
&= 0.
\end{aligned}$$

Since $\hat{\gamma}$ is the maximum partial likelihood estimate of γ obtained by solving $V(\gamma) = 0$ at $\gamma = \hat{\gamma}$, we have [cf. p. 148-149, Therneau and Grambsch (1990)]

$$\begin{aligned}
0 &= V(\hat{\gamma}) \\
&= \sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}\Delta_i(t)dO_i(t) \\
&= \sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}d\hat{M}_i^*(t;\hat{\gamma}).
\end{aligned}$$

Some simple algebra lead to

$$\begin{aligned}
\sum_{i=1}^{n}U_{ni}(\beta_0;\hat{\gamma}) &= \sum_{i=1}^{n}\Bigg[\int_0^{\tau}W(t)\{X_i(t) - \hat{E}_X(t;\beta_0,\hat{\gamma})\}d\hat{M}_i(t;\beta_0,\hat{\gamma}) \\
&\quad -\int_0^{\tau}\frac{W(t)\hat{R}(t;\beta_0,\hat{\gamma})}{S^{(0)}(t,\hat{\gamma})}d\hat{M}_i^*(t;\hat{\gamma}) \\
&\quad -\hat{P}(\beta_0,\hat{\gamma})\hat{D}^{-1}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}d\hat{M}_i^*(t;\hat{\gamma})\Bigg] \\
&= \sum_{i=1}^{n}\int_0^{\tau}W(t)\{X_i(t) - \hat{E}_X(t;\beta_0,\hat{\gamma})\}d\hat{M}_i(t;\beta_0,\hat{\gamma}) \\
&\quad -\int_0^{\tau}\frac{W(t)\hat{R}(t;\beta_0,\hat{\gamma})}{S^{(0)}(t,\hat{\gamma})}\sum_{i=1}^{n}d\hat{M}_i^*(t;\hat{\gamma}) \\
&\quad -\hat{P}(\beta_0,\hat{\gamma})\hat{D}^{-1}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}d\hat{M}_i^*(t;\hat{\gamma}) \\
&= \sum_{i=1}^{n}\int_0^{\tau}W(t)X_i(t)d\hat{M}_i(t;\beta_0,\hat{\gamma})
\end{aligned}$$

$$
\begin{aligned}
& -\sum_{i=1}^{n}\int_0^\tau W(t)\hat{E}_X(t;\beta_0,\hat{\gamma})d\hat{M}_i(t;\beta_0,\hat{\gamma}) \\
= & \sum_{i=1}^{n}\int_0^\tau W(t)X_i(t)d\hat{M}_i(t;\beta_0,\hat{\gamma}) \\
& -\int_0^\tau W(t)\hat{E}_X(t;\beta_0,\hat{\gamma})\sum_{i=1}^{n}d\hat{M}_i(t;\beta_0,\hat{\gamma}) \\
= & \sum_{i=1}^{n}\int_0^\tau W(t)X_i(t)d\hat{M}_i(t;\beta_0,\hat{\gamma}) \\
= & \sum_{i=1}^{n}\int_0^\tau W(t)X_i(t)\Big[Y_i(t)\Delta_i(t)dO_i(t) - g\{\hat{\mu}_0(t;\beta_0,\hat{\gamma})e^{\beta_0' X_i(t)}\}\Delta_i(t) \\
& e^{\hat{\gamma}' Z_i(t)}d\hat{\Lambda}_0(t;\hat{\gamma})\Big] \\
= & U(\beta_0,\hat{\gamma}).
\end{aligned}
$$

As shown in the appendix of Li et al. (2010) that $n^{-1/2}U(\beta_0;\hat{\gamma})$ converges to a zero mean Gaussian distribution with covariance matrix Σ, we have

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_{ni}(\beta_0;\hat{\gamma}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}U(\beta_0;\hat{\gamma}) \xrightarrow{\mathfrak{D}} N(0,\Sigma).
$$

Lemma 5.2. *Assume that the conditions (D1)-(D6) hold. If* $\beta_0 = (\beta_{10}', \beta_{20}')'$ *are the true values of the parameters, then*

$$
\frac{1}{n}\sum_{i=1}^{n}U_{ni}(\beta_0;\hat{\gamma})U_{ni}'(\beta_0;\hat{\gamma}) \xrightarrow{p} \Sigma.
$$

Proof. It can be shown that

$$
\begin{aligned}
U_{ni}(\beta_0;\hat{\gamma}) = & \; U_i(\beta_0;\hat{\gamma}) + \int_0^\tau \{W(t) - w(t)\}X_i(t)Y_i(t)\Delta_i(t)dO_i(t) \\
& +(-1)\int_0^\tau \Big(W(t)\hat{E}_X(t;\beta_0,\hat{\gamma}) - w(t)e_x(t)\Big)Y_i(t)\Delta_i(t)dO_i(t) \\
& +(-1)\Big(\int_0^\tau W(t)X_i(t)g\{\hat{\mu}_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\hat{\gamma}' Z_i(t)}d\hat{\Lambda}_0(t;\hat{\gamma}) - \\
& \int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t)\Big)
\end{aligned}
$$

$$
\begin{aligned}
&+\Big(\int_0^\tau W(t)\hat{E}_X(t;\beta_0,\hat\gamma)g\{\hat\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\hat\gamma' Z_i(t)}d\hat\Lambda_0(t;\hat\gamma) - \\
&\quad\int_0^\tau w(t)e_x(t)g\{\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t)\Big) \\
&+(-1)\int_0^\tau \Big(\frac{W(t)\hat{R}(t;\beta_0,\hat\gamma)}{S^{(0)}(t;\hat\gamma)} - \frac{w(t)r(t)}{s^{(0)}(t;\gamma_0)}\Big)\Delta_i(t)dO_i(t) \\
&+\int_0^\tau \Big(\hat{P}(\beta_0,\hat\gamma)\hat{D}^{-1}\{Z_i(t)-\bar{Z}(t;\hat\gamma)\} - P(\beta_0,\gamma_0)D^{-1}\{Z_i(t)-\bar{z}(t;\gamma_0)\}\Big)\Big] \\
&\quad\Delta_i(t)dO_i(t) \\
&+\Big(\int_0^\tau \Big[\frac{W(t)\hat{R}(t;\beta_0,\hat\gamma)}{S^{(0)}(t,\hat\gamma)} - \hat{P}(\beta_0,\hat\gamma)\hat{D}^{-1}\{Z_i(t)-\bar{Z}(t;\hat\gamma)\}\Big] \\
&\quad\Delta_i(t)e^{\hat\gamma' Z_i(t)}d\hat\Lambda_0(t;\hat\gamma) \\
&\quad-\int_0^\tau \Big[\frac{w(t)r(t)}{s^{(0)}(t;\gamma_0)} - P(\beta_0,\gamma)D^{-1}\{Z_i(t)-\bar{z}(t;\gamma_0)\}\Big]\Delta_i(t)e^{\hat\gamma' Z_i(t)}d\Lambda_0(t)\Big) \\
:= \;& U_i(\beta_0;\hat\gamma)+\epsilon_{i1}+\epsilon_{i2}+\epsilon_{i3}+\epsilon_{i4}+\epsilon_{i5}+\epsilon_{i6}+\epsilon_{i7}.
\end{aligned}
$$

Under condition (D1) and (D3), $||\epsilon_{i1}|| = o_p(1), ||\epsilon_{i2}|| = o_p(1), ||\epsilon_{i5}|| = o_p(1), ||\epsilon_{i6}|| = o_p(1)$ hold.

Note that [cf. p. 1103-1104,Andersen and Gill (1982)]

$$
\hat\Lambda_0(t;\gamma_0) - \Lambda_0(t) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{dM_i^*(u)}{s^{(0)}} + o_p(n^{-1/2}).
$$

Then under conditions (D1)-(D3) and by the consistency of $\hat\mu_0(t)$ to $\mu_0(t)$, we have

$$
\begin{aligned}
\epsilon_{i3} \;=\;& \int_0^\tau W(t)X_i(t)g\{\hat\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\hat\gamma' Z_i(t)}d\hat\Lambda_0(t;\hat\gamma) \\
&-\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t) \\
=\;& \int_0^\tau W(t)X_i(t)g\{\hat\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\hat\gamma' Z_i(t)}d\hat\Lambda_0(t;\hat\gamma) \\
&-\int_0^\tau W(t)X_i(t)g\{\hat\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\hat\gamma' Z_i(t)}d\Lambda_0(t) \\
&+\int_0^\tau W(t)X_i(t)g\{\hat\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\hat\gamma' Z_i(t)}d\Lambda_0(t) \\
&-\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0'X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^\tau W(t)X_i(t)g\{\hat{\mu}_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\hat{\gamma}' Z_i(t)}d[\hat{\Lambda}_0(t;\hat{\gamma}) - \Lambda_0(t)] \\
&\quad +(1+o_p(1))\int_0^\tau W(t)X_i(t)g\{\hat{\mu}_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t) \\
&\quad -\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t) \\
&= (1+o_p(1))\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d[\hat{\Lambda}_0(t;\hat{\gamma}) - \Lambda_0(t)] \\
&\quad +\int_0^\tau W(t)X_i(t)g\{\hat{\mu}_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t) \\
&\quad -\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t) + o_p(1) \\
&= (1+o_p(1))\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d[\hat{\Lambda}_0(t;\hat{\gamma}) - \Lambda_0(t)] \\
&\quad +\int_0^\tau \left(W(t)g\{\hat{\mu}_0(t)e^{\beta_0' X_i(t)}\} - w(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\right) X_i(t)\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t) + o_p(1) \\
&= (1+o_p(1))\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d[\hat{\Lambda}_0(t;\gamma_0) - \Lambda_0(t)] \\
&\quad +(1+o_p(1))\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d[\hat{\Lambda}_0(t;\hat{\gamma}) - \hat{\Lambda}_0(t;\gamma_0)] \\
&\quad +\int_0^\tau \left(W(t)g\{\hat{\mu}_0(t)e^{\beta_0' X_i(t)}\} - w(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\right) X_i(t)\Delta_i(t)e^{\gamma_0' Z_i(t)}d\Lambda_0(t) + o_p(1) \\
&= (1+o_p(1))\frac{1}{n}\int_0^\tau \frac{w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}\sum_{i=1}^n dM_i^*(t)}{s^{(0)}} \\
&\quad +(1+o_p(1))\int_0^\tau w(t)X_i(t)g\{\mu_0(t)e^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}d[\hat{\Lambda}_0(t;\hat{\gamma}) - \hat{\Lambda}_0(t;\gamma_0)] \\
&\quad +o_p(1) \\
&:= E_1 + E_2 + o_p(1).
\end{aligned}
$$

In the term E_1, for each i, $w(t)X_i(t)g\{\mu_0(t)^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}\Big/ s^{(0)}$ is predictable and finite, and $M_i^*(t)$ is a martingale. Then

$$
\int_0^\tau \frac{w(t)X_i(t)g\{\mu_0(t)^{\beta_0' X_i(t)}\}\Delta_i(t)e^{\gamma_0' Z_i(t)}\sum_{i=1}^n dM_i^*(u)}{s^{(0)}}
$$

is a martingale integral and converges to zero in probability. Also, similar to p. 300 of Fleming and Harrington (1991), the Taylor series expansion of $\hat{\Lambda}_0(t;\hat{\gamma}) - \hat{\Lambda}_0(t;\gamma_0)$ about

$\gamma = \gamma_0$ in the term E_2 can be written as

$$H'(t;\gamma^*)(\hat{\gamma} - \gamma_0),$$

where γ^* is on the line segment between $\hat{\gamma}$ and γ_0 and H is the column vector defined as

$$H(t;\gamma) = -\sum_{i=1}^{n} \int_0^t \frac{S^{(1)}(u;\gamma)\Delta_i(u)dO_i(u)}{n\{S^{(0)}(u:\gamma)\}^2}.$$

Following Lin et al. (2000), it can be shown that $H(t;\gamma_0)$ converges almost surely to the deterministic function

$$h(t;\gamma_0) = -\int_0^t \frac{s^{(0)}(u;\gamma_0)}{s^{(1)}(u;\gamma_0)} \lambda_0(t)du$$

uniformly in t and

$$\hat{\gamma} - \gamma_0 = n^{-1} A_*^{-1} \sum_{i=1}^{n} \int_0^\tau \{Z_i(t) - \bar{z}(t;\gamma_0)\} dM_i^*(t) + o_p(n^{-1/2}).$$

Therefore, $\hat{\Lambda}_0(t;\hat{\gamma}) - \hat{\Lambda}_0(t;\gamma_0)$ in the term E_2 is tight and therefore equal to

$$n^{-1} h'(t;\gamma_0) A_*^{-1} \sum_{i=1}^{n} \int_0^\tau \{Z_i(t) - \bar{z}(t;\gamma_0)\} dM_i^*(t) + o_p(n^{-1/2}).$$

Hence,

$$\begin{aligned} E_2 &= (1 + o_p(1)) \int_0^\tau w(t) X_i(t) g\{\mu_0(t) e^{\beta_0' X_i(t)}\} \Delta_i(t) e^{\gamma_0' Z_i(t)} \\ &\quad [h'(t;\gamma_0) A_*^{-1} \sum_{i=1}^{n} \frac{1}{n} \{Z_i(t) - \bar{z}(t;\gamma_0)\} dM_i^*(t)] + o_p(n^{-1/2}), \end{aligned}$$

where $\int_0^\tau w(t) X_i(t) g\{\mu_0(t) e^{\beta_0' X_i(t)}\} \Delta_i(t) e^{\gamma_0' Z_i(t)} [h'(t;\gamma_0) A_*^{-1} \sum_{i=1}^{n} \frac{1}{n} \{Z_i(t) - \bar{z}(t;\gamma_0)\} dM_i^*(t)]$ is a martingale integral and converges to zero in probability. Thus $\epsilon_{i3} = o_p(1)$. Using similar arguments, it can be shown that $\epsilon_{i4} = o_p(1)$ and $\epsilon_{i7} = o_p(1)$. Therefore,

$$U_{ni}(\beta_0;\hat{\gamma}) = U_i(\beta_0;\hat{\gamma}) + o_p(1), \quad i = 1, 2, \cdots, n. \tag{D.1}$$

Let $\hat{Q}_n = \frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0)$ and $Q_n = \frac{1}{n}\sum_{i=1}^n W_i(\beta_0)W'_i(\beta_0)$. For any $c \in R^p$, the following decomposition holds:

$$\begin{aligned} c'(\hat{Q}_n - Q_n)c &= c'\Big(\frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0) - \frac{1}{n}\sum_{i=1}^n W_i(\beta_0)W'_i(\beta_0)\Big) \\ &= \frac{1}{n}\sum_{i=1}^n [c'U_{ni}(\beta_0;\hat{\gamma}) - W_i(\beta_0)]^2 + \frac{2}{n}\sum_{i=1}^n [c'W_i(\beta_0)][c'(U_{ni}(\beta_0;\hat{\gamma}) - U_i(\beta_0))] \\ &:= I_1 + 2I_2. \end{aligned}$$

Both I_1 and I_2 are $o_p(1)$ by D.1. As a result,

$$\frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0) = \frac{1}{n}\sum_{i=1}^n W_i(\beta_0)W'_i(\beta_0) + o_p(1).$$

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

$$\frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0) \xrightarrow{p} E[W_i(\beta_0)W'_i(\beta_0)],$$

i.e.,

$$\frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0) \xrightarrow{p} \Sigma.$$

Lemma 5.3. *Assume that the conditions (D1)-(D6) hold. If* $\beta_0 = (\beta'_{10}, \beta'_{20})'$ *are true value of the parameters, then*

$$Z_n = \underset{1\le i\le n}{max}||U_{ni}(\beta_0;\hat{\gamma})|| = o_p(n^{1/2}).$$

Proof. In eqn. (D.1), we have $U_{ni}(\beta_0;\hat{\gamma}) = U_i(\beta_0) + o_p(1),\ i = 1, 2, \cdots, n$. As $U_i(\beta_0)$ are i.i.d. r.v. and $E[W_i(\beta_0)W'_i(\beta_0) = \Sigma < \infty$, $U_i(\beta_0)$ has finite second moment. Then by Lemma 11.2 of Owen (2001), Lemma 5.3 can be proven.

Proof of Theorem 5.1. Following Owen (1990), Let $q(\lambda) = \frac{1}{n}\sum_{i=1}^n \frac{U_{ni}(\beta_0;\hat{\gamma})}{1+\lambda' U_{ni}(\beta_0;\hat{\gamma})}$ and $\lambda = \rho\theta$, where $\rho \ge 0$ and $||\theta|| = 1$. Then

$$0 = ||q(\lambda)||$$

$$
\begin{aligned}
&= ||q(\rho\theta)|| \\
&\geq |\theta' q(\rho\theta)| \\
&= \frac{1}{n}\left|\theta' \sum_{i=1}^{n} \frac{U_{ni}(\beta_0;\hat{\gamma})}{1+\lambda' U_{ni}(\beta_0;\hat{\gamma})}\right| \\
&= \frac{1}{n}\left|\theta' \left\{ \sum_{i=1}^{n} U_{ni}(\beta_0;\hat{\gamma}) - \rho \sum_{i=1}^{n} \frac{U_{ni}(\beta_0;\hat{\gamma})\theta' U_{ni}(\beta_0;\hat{\gamma})}{1+\rho\theta' U_{ni}(\beta_0;\hat{\gamma})} \right\}\right| \\
&\geq \frac{\rho}{n}\theta' \sum_{i=1}^{n} \frac{U_{ni}(\beta_0;\hat{\gamma})\theta' U_{ni}(\beta_0;\hat{\gamma})}{1+\rho\theta' U_{ni}(\beta_0;\hat{\gamma})}\theta - \frac{1}{n}\left|\sum_{i=1}^{n} U_{ni}(\beta_0;\hat{\gamma})\right| \\
&\geq \frac{\rho\theta'\hat{Q}_n\theta}{1+\rho Z_n} - \frac{1}{n}\left|\sum_{i=1}^{n} U_{ni}(\beta_0;\hat{\gamma})\right|.
\end{aligned}
$$

We have showed that $\hat{Q}_n = \frac{1}{n}\sum_{i=1}^{n} U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} W_i(\beta_0)W'_i(\beta_0) + o_p(1)$ in Lemma 4.2. Then as $n \to \infty$, $\lim_{n\to\infty} \hat{Q}_n = \Sigma$. Also $Z_n = o_p(n^{1/2})$ by Lemma 5.3. Then, based on Lemma 5.2,

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

Then, it follows from eqn. (5.6) and Owen (1990) that

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

Combining $||\lambda|| = O_p(n^{1/2})$ and $Z_n = o_p(n^{1/2})$, it can be shown that

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

By Taylor expansion of eqn. (5.5), one obtains

$$
\begin{aligned}
l(\beta_0) &= 2\sum_{i=1}^{n} \lambda' U_{ni}(\beta_0;\hat{\gamma}) - \sum_{i=1}^{n} \lambda' U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0)\lambda + o_p(1) \\
&= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0;\hat{\gamma})\right)'\left(\frac{1}{n}\sum_{i=1}^{n} U_{ni}(\beta_0;\hat{\gamma})U'_{ni}(\beta_0)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0;\hat{\gamma})\right) + o_p(1)
\end{aligned}
$$

By Lemma 5.2 and the Slutsky theorem, Theorem 5.1 is proved.

Proof of Theorem 5.2. The proof is similar to the proof of Proposition 3 of Yu et al. (2011). We can write $Z = ((Z^{(1)})', (Z^{(2)})')'$ so that $X = ((X^{(1)})', (X^{(2)}))'$, which is corresponding to $\beta_0 = ((\beta_0^{(1)})', (\beta_0^{(2)})')$. Define

$$\begin{aligned} A^*(\beta_0) &= E\Big[\int_0^\tau w(t)\Delta_i(t)\dot{g}\{\mu_0(t,\beta_0)e^{\beta_0{}' X_i(t)}\}\{X_i^{(2)}(t) - e_x(t)\}^{\otimes 2} \\ &\qquad e^{\beta_0{}' X_i(t) + \gamma_0{}' Z_i(t)}\mu_0(t,\beta_0)d\Lambda_0(t)\Big]. \end{aligned}$$

$A(\beta_0)$ is assumed to be positive definite. Therefore the rank of $A^*(\beta_0)$ is $p - q$. Denote $\hat{\beta}_2(\beta_0^{(1)'}) = \arg\inf_{\beta^{(2)}} l((\beta_0^{(1)})', (\beta^{(2)})')'$. We also denote

$$\Psi(\beta_0) = [A^*(\beta_0)]'[\Sigma(\beta_0)]^{-1}A^*(\beta_0).$$

Using the similar arguments from Qin and Lawless (1994) and Yu et al. (2011), we have that

$$\sqrt{n}(\hat{\beta}_2 - \beta_0^{(2)'}) = -[\Psi(\beta_0)]^{-1}[A^*(\beta_0)]'[\Sigma(\beta_0)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma}) + o_p(1),$$

and the Lagrange multiplier satisfies that

$$\sqrt{n}\lambda_2 = \left\{I - [\Sigma(\beta_0)]^{-1}A^*(\beta_0)[\Psi(\beta_0)]^{-1}[A^*(\beta_0)]'\right\}[\Sigma(\beta_0)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma}) + o_p(1).$$

By Taylor expansion,

$$\begin{aligned} l_{profile}(\beta_0^{(1)}) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})\right)'\Big([\Sigma(\beta_0)]^{-1} - [\Sigma(\beta_0)]^{-1}A^*(\beta_0)[\Psi(\beta_0)]^{-1} \\ &\qquad [A^*(\beta_0)]'[\Sigma(\beta_0)]^{-1}\Big)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})\right) + o_p(1) \\ &= \left([\Sigma(\beta_0)]^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})\right)' S\left([\Sigma(\beta_0)]^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{ni}(\beta_0;\hat{\gamma})\right) + o_p(1), \end{aligned}$$

where

$$S = I - [\Sigma(\beta_0)]^{-1/2} A^*(\beta_0)[\Psi(\beta_0)]^{-1}[A^*(\beta_0)]'[\Sigma(\beta_0)]^{-1/2}.$$

One can easily verify that S is a symmetric and idempotent matrix and $\mathrm{tr}(S) = q$. Then by Lemma 5.2,

$$[\Sigma(\beta_0)]^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{ni}(\beta_0; \hat{\gamma}) \xrightarrow{\mathfrak{D}} N(0, I_{p\times p}).$$

Theorem 5.2 is completed.

Appendix E

PROOFS OF CHAPTER 6

We assume the following regularity conditions as Lu and Tsiatis (2006) did.

(R.1) The covariate vector Z is bounded, i.e., $P(|Z| < C) = 1$ for some constant C.

(R.1) p converges to a positive constant $\tilde{p} \in (0, 1)$ as $n, n \to \infty$.

(R.3) $\lambda(\cdot)$ is positive and $\dot{\lambda}(\cdot)$ is bounded and continuous on $(-\infty, C)$, where C is any finite constant.

(R.4) $H_0(\cdot)$ has continuous and positive derivatives on $[0, \tau]$.

(R.5) The matrix A and Σ are finite and nonsingular.

Additionally, we assume one more regularity condition.

(R.6) $E\big[\pi Z Y(t) d\Lambda\{H_0(t) + \beta_0' Z\}\big]^2 < \infty$.

Lemma 6.1. *Under the regularity conditions in the Appendix E, if β_0 is the true values of β,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{ni}(\beta_0) \overset{\mathfrak{D}}{\to} N(0, \Sigma(\beta_0)).$$

Proof. Note that

$$\begin{aligned} \sum_{i=1}^{n} \pi_i d\hat{M}_i(\beta_0, t) &= \sum_{i}^{n} \pi_i [dN_i(t) - Y_i(t) d\Lambda\{\hat{H}(\beta_0, t) + \beta_0' Z_i\}] \\ &= 0. \end{aligned}$$

Denote

$$U_i(\beta_0) = \int_0^\tau \pi_i (Z_i - \mu_Z(\beta_0, t)) dM_i(\beta_0, t), \quad i = 1, 2, \ldots, n.$$

We can show that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_{ni}(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \pi_i(Z_i-\bar{Z}(\beta,t))d\hat{M}_i(\beta_0,t)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \pi_i Z_i d\hat{M}_i(\beta_0,t) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \pi_i\bar{Z}(\beta,t)d\hat{M}_i(\beta_0,t)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \pi_i Z_i d\hat{M}_i(\beta_0,t) - \frac{1}{\sqrt{n}}\int_0^\tau \bar{Z}(\beta,t)\sum_{i=1}^{n}\pi_i d\hat{M}_i(\beta_0,t)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \pi_i Z_i d\hat{M}_i(\beta_0,t)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \pi_i Z_i[dN_i(t)-Y_i(t)d\Lambda\{\hat{H}(\beta_0,t)+\beta_0'Z_i\}]\\
&= \frac{1}{\sqrt{n}}\Bigg[\sum_{i=1}^{n}\int_0^\tau \pi_i Z_i dM_i(\beta_0,t)\\
&\quad -\sum_{i=1}^{n}\pi_i Z_i[\Lambda\{\hat{H}(\beta_0,t)+\beta_0'Z_i\}-\Lambda\{H(\beta_0,t)+\beta_0'Z_i\}]\Bigg]\\
&= \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\int_0^\tau \pi_i(Z_i-\mu_Z(\beta_0,t))dM_i(\beta_0,t)\right]+o_p(1)\\
&:= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\beta_0)+o_p(1).
\end{aligned}$$

Here, the second last equation comes from page 213 in the Appendix of Lu and Tsiatis (2006). Then following the steps in the Appendix of Lu and Tsiatis (2006), as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_{ni}(\beta_0)\xrightarrow{\mathfrak{D}} N(0,\Sigma(\beta_0)).$$

Lemma 6.2. *Under the regularity conditions in the Appendix E, if β_0 is the true values of β,*

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

Proof. It can be shown that for $i=1,2,\ldots,n$,

$$U_{ni}(\beta_0) = U_i(\beta_0)+\int_0^\tau \pi_i(\bar{Z}(\beta,t)-\mu_Z(\beta,t))dN_i(t)$$

$$
\begin{aligned}
&+Z_i\Big(\int_0^\tau \pi_i Y_i(t) d\Lambda\{\hat{H}(\beta_0,t)+\beta_0' Z_i\} - \int_0^\tau \pi_i Y_i(t) d\Lambda\{H_0(t)+\beta_0' Z_i\}\Big) \\
&+(-1)\Big(\int_0^\tau \pi_i \bar{Z}(\beta,t) Y_i(t) d\Lambda\{\hat{H}(\beta_0,t)+\beta_0' Z_i\} \\
&-\int_0^\tau \pi_i \mu_Z(\beta,t) Y_i(t) d\Lambda\{H_0(t)+\beta_0' Z_i\}\Big) \\
:= \;& U_i(\beta_0)+\epsilon_{i1}+\epsilon_{i2}+\epsilon_{i3}.
\end{aligned}
$$

It is easy to obtain the uniform consistency of $\bar{Z}(\beta,t)$, i.e.,

$$
\sup_{0\le t\le \tau} |\bar{Z}(\beta,t)-\mu_Z(\beta,t)| \xrightarrow{p} 0. \tag{E.1}
$$

Using the similar arguments from Ma et al. (2016), we can show that

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

We can write

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

Here, $M_j(t), j=1,2,\ldots,n$, is a martingale, and $B(\beta_0,u,t)/B_2(\beta_0,u)$ is bounded. Also, by condition (R.6), $\left|\int_0^\tau \pi_i Y(t) d\Lambda\{H_0(t)+\beta_0' Z_i\}\right|$ has a finite second moment. This follows

$$\left|\left|\int_0^\tau \pi_i Y_i(t) d\Lambda\{\hat{H}(\beta_0,t)+\beta_0' Z_i\} - \int_0^\tau \pi_i Y_i(t) d\Lambda\{H_0(t)+\beta_0' Z_i\}\right|\right| = o_p(1). \tag{E.2}$$

Similarly, with (R.1) and (E.1), we can show that

$$\left|\left|\int_0^\tau \pi_i \bar{Z}(\beta,t) Y_i(t) d\Lambda\{\hat{H}(\beta_0,t)+\beta_0' Z_i\} - \int_0^\tau \pi_i \mu_Z(\beta,t) Y_i(t) d\Lambda\{H_0(t)+\beta_0' Z_i\}\right|\right| = o_p(1). \tag{E.3}$$

Thus, by (E.1), (E.2), (E.3), we have $||\epsilon_{i1}|| = o_p(1), ||\epsilon_{i2}|| = o_p(1), ||\epsilon_{i3}|| = o_p(1)$. Then

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

For any $c \in R^p$, the following decomposition holds:

$$\begin{aligned} & c'\Big(\frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0)(U_{ni}(\beta_0)' - \frac{1}{n}\sum_{i=1}^n U_i(\beta_0)(U_i(\beta_0)'\Big)c \\ = \; & \frac{1}{n}\sum_{i=1}^n [c'U_{ni}(\beta_0) - U_i(\beta_0)]^2 + \frac{2}{n}\sum_{i=1}^n [c'U_i(\beta_0)][c'(U_{ni}(\beta_0) - U_i(\beta_0))] \\ := \; & I_1 + 2I_2. \end{aligned}$$

By eqn. (E.4), both I_1 and I_2 of the above equation are $o_p(1)$.Then

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

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

$$\frac{1}{n}\sum_{i=1}^n U_{ni}(\beta_0)(U_{ni}(\beta_0))' \xrightarrow{p} E[U_i(\beta_0)(U_i(\beta_0))'].$$

Therefore,

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

□

Lemma 6.3. *Under the regularity conditions in the Appendix E, if β_0 is the true value of β, then*

$$M_n = \underset{1\leq i\leq n}{max}||U_{ni}(\beta_0)|| = o_p(n^{1/2}).$$

Proof. By Lemma 6.2, we have $U_{ni}(\beta_0) = U_i(\beta_0) + o_p(1)$, for $i = 1, 2, \ldots, n$. Now, $E[U_i(\beta_0)U_i'(\beta_0)] = \Sigma(\beta_0) < \infty$ and the fact that $U_i(\beta_0)$ are i.i.d. random variables, imply that $U_i(\beta_0)$ has a finite second moment. Then by Lemma 11.2 of Owen (2001), we have,

$$M_n = \max_{1\leq i\leq n}||U_{ni}(\beta_0)|| = o_p(n^{1/2}).$$

□

Proof of Theorem 6.1. Let $q(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{U_{ni}(\beta_0)}{1+(\theta(\beta_0))'U_{ni}(\beta_0)}$ and $\theta = \rho\eta$, where $\rho \geq 0$ and $||\eta|| = 1$. Then following the arguments from Owen (1990),

$$\begin{aligned}
0 &= ||q(\theta)|| \\
&= ||q(\rho\eta)|| \\
&\geq |\eta' q(\rho\eta)| \\
&= \frac{1}{n}\left|\eta'\sum_{i=1}^{n}\frac{U_{ni}(\beta_0)}{1+(\theta(\beta_0))'U_{ni}(\beta_0)}\right| \\
&= \frac{1}{n}\left|\eta'\left\{\sum_{i=1}^{n}U_{ni}(\beta_0) - \rho\sum_{i=1}^{n}\frac{U_{ni}(\beta_0)\eta' U_{ni}(\beta_0)}{1+\rho\eta' U_{ni}(\beta_0)}\right\}\right| \\
&\geq \frac{\rho}{n}\eta'\sum_{i=1}^{n}\frac{U_{ni}(\beta_0)(U_{ni}(\beta_0))'}{1+\rho\eta' U_{ni}(\beta_0)}\eta - \frac{1}{n}\left|\sum_{i=1}^{n}U_{ni}(\beta_0)\right| \\
&\geq \frac{\rho\eta' W_n(\beta_0)\eta}{1+\rho M_n} - \frac{1}{n}\left|\sum_{i=1}^{n}U_{ni}(\beta_0)\right|, \qquad \text{(E.5)}
\end{aligned}$$

where $W_n(\beta_0) = n^{-1}\sum_{i=1}^{n}U_{ni}(\beta_0)(U_{ni}(\beta_0))'$.

From Lemma 6.2, we have $W_n(\beta_0) = n^{-1}\sum_{i=1}^{n}U_i(\beta_0)(U_i(\beta_0))' + o_p(1)$. Then, $\lim_{n\to\infty}W_n(\beta_0) = E[U_i(\beta_0)(U_i(\beta_0))'] = \Sigma(\beta_0)$ as $n \to \infty$. Also $M_n = o_p(n^{1/2})$ by Lemma

6.3. By Lemma 6.1, we have

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

Then, it follows from eqn. (E.5) and Owen (1990) that

$$\rho = ||\theta|| = O_p(n^{-1/2}).$$

Combining $||\theta|| = O_p(n^{1/2})$ and $M_n = o_p(n^{1/2})$, it can be shown that

$$\theta = \left(\frac{1}{n}\sum_{i=1}^{n} U_{ni}(\beta_0)(U_{ni}(\beta_0))'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} U_{ni}(\beta_0)\right) + o_p(n^{-1/2}). \quad \text{(E.6)}$$

By Taylor expansion to $l(\beta_0)$, one obtains

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

By Lemma 6.1 and the Slutsky theorem, $l(\beta_0)$ converges to χ^2_p in distribution. Hence, Theorem 6.1 is proved. □

Proof of Theorem 6.2. Theorem 6.2 is proved using the similar arguments in Yu et al. (2011) and Yu and Zhao (2019). Let $\beta_0 = ((\beta_0^{(1)})', (\beta_0^{(2)})')$, corresponding to $Z = ((Z^{(1)})', (Z^{(2)})')'$. Define

$$\tilde{A}(\beta_0) = \int_0^{\tau} E[\{Z - \mu_Z(\beta_0, t)\}(Z^{(2)})'\dot{\lambda}\{H_0(t) + \beta_0' Z\}Y(t)]dH_0(t).$$

As $A(\beta_0)$ is assumed to be positive definite, the rank of $\tilde{A}(\beta_0)$ is $p - q$. Let $\hat{\beta}_{II}(\beta_0^{(1)'}) = \arg\inf_{\beta^{(2)}} l((\beta_0^{(1)})', (\beta^{(2)})')'$. Using the arguments from Qin and Lawless (1994) and Yu et al.

(2011),

$$\sqrt{n}(\hat{\beta}_{II} - \beta_0^{(2)'}) = -[\Psi(\beta_0)]^{-1}[\tilde{A}(\beta_0)]'[\Sigma(\beta_0)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0) + o_p(1),$$

$$\sqrt{n}\theta_2 = \left\{I - [\Sigma(\beta_0)]^{-1}\tilde{A}(\beta_0)[\Psi(\beta_0)]^{-1}[\tilde{A}(\beta_0)]'\right\}[\Sigma(\beta_0)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0) + o_p(1),$$

where θ_2 is the corresponding Lagrange multiplier, and

$$\Psi(\beta_0) = [\tilde{A}(\beta_0)]'[\Sigma(\beta_0)]^{-1}\tilde{A}(\beta_0).$$

By Taylor expansion, we have

$$\begin{aligned} l_{\text{profile}}(\beta_0^{(1)}) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0)\right)' \left([\Sigma(\beta_0)]^{-1} - [\Sigma(\beta_0)]^{-1}\tilde{A}(\beta_0)[\Psi(\beta_0)]^{-1}[\tilde{A}(\beta_0)]'[\Sigma(\beta_0)]^{-1}\right) \\ &\quad \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0)\right) + o_p(1) \\ &= \left([\Sigma(\beta_0)]^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0)\right)' S \left([\Sigma(\beta_0)]^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0)\right) + o_p(1), \end{aligned}$$

where

$$S = I - [\Sigma(\beta_0)]^{-1/2}\tilde{A}(\beta_0)[\Psi(\beta_0)]^{-1}[\tilde{A}(\beta_0)]'[\Sigma(\beta_0)]^{-1/2}$$

is a symmetric and idempotent matrix with trace q. Then by Lemma 6.1,

$$[\Sigma(\beta_0)]^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_{ni}(\beta_0) \xrightarrow{\mathfrak{D}} N(0, I_{p\times p}).$$

Hence, we prove Theorem 6.2. □

Proof of Theorem 6.3. This proof is motivated by Chen et al. (2008) and Yu and Zhao (2019). Define

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

and $\theta^{ad} = \rho^{ad}\eta^{ad}$, where $\rho^{ad} > 0$ and $||\eta^{ad}|| = 1$. First, we want to show that $\theta^{ad} = O_p(n^{-1/2})$.

From the results of Theorem 6.1 and eqn. (6.5), we can write

$$\begin{aligned}
0 &= ||g(\theta^{ad})|| \\
&\geq |(\eta^{ad})' g(\rho^{ad}\eta^{ad})| \\
&\geq \frac{1}{n}\left|(\eta^{ad})' \sum_{i=1}^{n+1} \frac{U_{ni}^{ad}(\beta_0)}{1+(\theta^{ad})' U_{ni}^{ad}(\beta_0)}\right| \\
&\geq \frac{\rho^{ad}}{n(1+\rho^{ad} M_n^{ad})} \sum_{i=1}^{n} [(\eta^{ad})' U_{ni}^{ad}(\beta_0)]^2 - \frac{1}{n}\left|\sum_{i=1}^{n} U_{ni}^{ad}(\beta_0)\right| \left(1 - \frac{a_n}{n}\right) \\
&= \frac{\rho^{ad}}{n(1+\rho^{ad} M_n^{ad})} \sum_{i=1}^{n} [(\eta^{ad})' U_{ni}^{ad}(\beta_0)]^2 - \frac{1}{n}\left|\sum_{i=1}^{n} U_{ni}^{ad}(\beta_0)\right| + O_p(n^{-2/3} a_n),
\end{aligned}$$

where $M_n^{ad} = \max_{1\leq i\leq n} ||U_{ni}^{ad}(\beta_0)||$. Using the similar arguments from Chen et al. (2008) and following the proof of Theorem 6.2, we have $\theta^{ad} = O_p(n^{-1/2})$ as $a_n = o_p(n)$. Also, similar to eqn. (E.6), we have

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

By Taylor expansion to $l^{ad}(\beta_0)$, one obtains

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

Substituting the expansion of θ^{ad}, we have

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

Therefore,

$$l^{ad}(\beta_0) \xrightarrow{\mathfrak{D}} \chi_p^2.$$

□

Proof of Theorem 6.4. Combining Theorems 6.2 and 6.3, we can prove Theorem 6.4 easily. □