SUBGROUP ANALYSIS FOR HETEROGENEOUS COX MODEL AND STATISTICAL INFERENCE FOR PANEL COUNT DATA WITH TERMINAL EVENT

HU XIANGBIN

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Applied Mathematics

Subgroup Analysis for Heterogeneous Cox Model and Statistical Inference for Panel Count Data with Terminal Event

Hu Xiangbin

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Feb. 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_______________________________(Signed)

Hu Xiangbin (Name of student)

Abstract

Survival analysis is an essential branch of statistics, which focuses on analyzing the duration time until one or more events happen. It is applied widely to medicine, economics, engineering, and sociology. Despite the rapid development during the past several decades, there are still many interesting researches to study in this area. This thesis considers two topics: subgroup analysis for the heterogeneous Cox model and statistical inference for panel count data with an informative terminal event.

In survival analysis, Cox model is commonly used to study the covariate effects. Nevertheless, the homogeneous effect assumption in the classical Cox model is usually not satisfied in many applications due to the differences among underlying groups of individuals. Then the homogeneous model will lead to inaccurate estimation results. To remove the bias, we conduct the subgroup analysis and build the Cox model with individual-specific coefficients. We introduce the pairwise fusion penalty function and minimize the penalized criterion function by the majorized alternating direction method of multipliers (ADMM) algorithm. Then our estimation procedure automatically clusters individuals having similar treatment effects into the same subgroup and estimates the treatment effects simultaneously. For the asymptotic theory, we first verify that the oracle estimator, the estimator with prior information about the subgroup structure, is asymptotically consistent and has the asymptotic normal distribution. Then we prove that under some mild conditions, the oracle estimator is a local minimizer of our criterion function with high probability. This implies the asymptotic consistency and normality of our estimator. We show the finite sample estimation results by the simulation studies. Furthermore, we use our method to analyze the breast cancer data collected by the Netherlands Cancer Institute (NKI).

In the long-term follow-up study of recurrent events, panel count data occurs when the observations of individuals are some discrete time points such that only the occurrence numbers of recurrent events between the adjacent time points are available. In general, the follow-up study often ends with some events having intricate interactions with the recurrent events, which motivates us to study the statistical inference approaches for panel count data with an informative terminal event. This thesis builds the nonparametric and semiparametric models for this problem with the least squares-based loss functions. Treating the distribution of terminal event time as a nuisance functional parameter, we consider the two-stage estimation procedures. We approximate the nonparametric function by the monotone I-spline function because the spline estimation converges faster than the estimation approximated by the step function. Using the empirical process theories, we verify the asymptotic properties for the proposed estimators. We also conduct the two-sample hypothesis test for the mean function in the nonparametric model. Our simulation studies demonstrate that the proposed estimations perform well. Finally, we use our methods to analyze the dataset of the Chinese Longitudinal Healthy Longevity Survey (CLHLS).

Key Words: Survival analysis; Subgroup analysis; Cox model; Penalization; Majorized ADMM; Panel count data; Terminal event; Two-stage estimation; Monotone I-spline; Empirical process.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my chief supervisor Prof. Xingqiu Zhao for her detailed guidance, full support, strict regulation, and warm encouragement. Her dedication and enthusiasm for research and life always motivate me to overcome the challenges and difficulties during my PhD study. When I start to slacken my effort, her rigorous requirement makes me powerful and full of energy.

Next, I need to convey my sincerest appreciation to my co-supervisor Prof. Defeng Sun for his explicit instruction on the techniques of optimization algorithm and the approaches to debugging the program. I also admire his profound knowledge and rigorous scholarship that I model for my future study and work.

I would like to express my sincerest gratitude to Dr. Li Liu at Wuhan University for her patient guidance and advice. I will never forget the days that we discuss the technical solutions to the problems in theoretical analysis and the program for the simulation studies. She also teaches me a lot in writing papers and thesis in English. Without her help, I cannot complete this thesis.

Then I want to thank my former co-supervisor Prof. Jian Huang at the University of Iowa for his ideas and comments on many statistical problems I encountered in this thesis. His unique perspectives on statistics always inspire me.

Finally, I am grateful to my families and friends for their companies and supports.

Contents

List of Figures

List of Tables

List of Notations

Abbreviations

ADMM	Alternating Direction Method of Multipliers
BIC	Bayesian Information Criterion
CLHLS	Chinese Longitudinal Healthy Longevity Survey
CP	Coverage Probability
ESE	Estimated Standard Error
KM	Kaplan-Meier
MCP	Minimax Concave Penalty
NKI	Netherlands Cancer Institute
SCAD	Smoothly Clipped Absolute Deviation
SD	Standard Deviation
SELO	Seamless-L_0
SSE	Sample Standard Error
TPR	True Positive Rate

Scalars, Vectors and Matrixes

$a \wedge b$	Minimum value of a and b
$\boldsymbol{X}^T$	Transpose of $\boldsymbol{X}$
$\|\|\boldsymbol{a}\|\|$	L_2 norm of vector $\boldsymbol{a}$, i.e. $\|\|\boldsymbol{a}\|\| = \sqrt{\boldsymbol{a}^T\boldsymbol{a}}$
$\langle \boldsymbol{a}, \boldsymbol{b} \rangle$	Inner product of $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e. $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^T\boldsymbol{b}$
$\boldsymbol{a}^{\otimes 2}$	Outer product of vector $\boldsymbol{a}$, i.e. $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$
$\|\|\boldsymbol{a}\|\|_\infty$	Maximum norm of $\boldsymbol{a}$, i.e. $\|\|\boldsymbol{a}\|\|_\infty = \max_i \|a_i\|$
$\|\|\boldsymbol{a}\|\|_{\boldsymbol{G}}^2$	Norm induced by the inner product, i.e. $\|\|\boldsymbol{a}\|\|_{\boldsymbol{G}}^2 = \langle \boldsymbol{a}, G\boldsymbol{a} \rangle$
$\boldsymbol{e}_i$	Vector with ith entry 1 and other ones 0, i.e. $\boldsymbol{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)^T$

$\boldsymbol{I}_p$	Identity matrix of size p
$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$

Functions

1_A	Indicator function of A, i.e. $1_A = 1$ when A holds, and $1_A = 0$ when A does not hold
f'	First order derivative of f
$f^{(r)}$	rth order derivative of f
$\|\|f\|\|_2$	L_2 norm of f, i.e. $\|\|f\|\|_2 = \{\int f(s)^2 ds\}^{1/2}$
$\|\|f\|\|_{L_2(\mu)}$	L_2 norm of f w.r.t. measure μ, i.e. $\|\|f\|\|_{L_2(u)} = \{\int f(s)^2 d\mu(s)\}^{1/2}$
$\|\|f\|\|_\infty$	L_∞ norm of f, i.e. $\|\|f\|\|_\infty = \sup_s \|f(s)\|$
$\|\|f\|\|_{P,B}$	Bernstein norm of f, i.e. $\|\|f\|\|_{P,B} = \{2P(e^{\|f\|} - 1 - \|f\|)\}^{1/2}$
$\rho_P(f)$	Seminorm defined by $\rho_P(f) = \{\mathcal{P}(f - \mathcal{P}f)^2\}^{1/2}$

Moments and Operators

$E(X)$	Expectation of X
$E(Y\|X)$	Conditional expectation of Y given X
$Var(X)$	Variance of X
$\mathcal{P}f(X)$	Integral of f w.r.t. P, i.e. $\mathcal{P}f(X) = \int f(x)dP$
$\mathbb{P}_n f(X)$	Empirical measure of $f(X)$, i.e. $\mathbb{P}_n f(X) = \frac{1}{n}\sum_{i=1}^n f(X_i)$

Others

$\mathbb{R}$	Set of real numbers
$\mathbb{R}^+$	Set of positive real numbers
$\mathbb{N}$	Set of natural numbers
$\mathbb{Z}$	Set of integers
$\mathcal{B}$	Collection of Borel sets
$\xrightarrow{a.s.}$	Almost surely converge
$\xrightarrow{p}$	Converge in probability
$\xrightarrow{d}$	Converge in distribution
$N_{[]}(\varepsilon, \mathcal{F}, \|\| \cdot \|\|)$	ε-bracketing number of $\mathcal{F}$

Chapter 1

Introduction

This thesis studies two topics in survival analysis. That is the subgroup analysis for the heterogeneous Cox model and the statistical inference for panel count data with an informative terminal event. In this chapter, we first introduce the research area and some common concepts of survival analysis, such as right censoring, Cox model, and recurrent events. Then we demonstrate some existing studies closely related to the two topics in this thesis. Finally, we show the organization of this thesis.

1.1 Background

The purpose of survival analysis is to investigate the duration time from a starting point to the occurrences of one or more events. When the event of interest only happens once for each individual, the duration time is referred to as survival time, which is usually denoted by a non-negative continuous random variable U. Then the behavior of survival time is determined by its distribution function $F(u) = P(U \leq u)$ (or equivalently the probability density function $f(u) = F'(u)$). In survival analysis, we usually adopt three other quantities to describe the behavior of survival time. That is the survival function $S(u) = 1 - F(u)$, the cumulative hazard function $H(u) = -ln(1 - F(u))$ (or equivalently the hazard function $h(u) = H'(u)$), and the mean residual life function $mrl(u) = \int_u^\infty S(t)dt/S(u)$. Because there are some one-to-one maps among the above four quantities, we can focus on any one of them for statistical inference. The relationships between $F(u)$, $S(u)$, $H(u)$ and $mrl(u)$ were summarized in Theoretical Notes 2 in Section 2.4 of Klein and Moeschberger (2006).

In applications, the survival time is sometimes partly observed. For example, in the right-censored data, because of some censoring events such as loss-of-connection, drop-out of participants, and the end of study, the remainder of survival time after the censoring events is not observed. Hence, taking the censoring time to be C, the right-censored data is recorded as (Y, Δ), where $Y = U \wedge C$ is the right-censored survival time and $\Delta = 1_{\{U \leq C\}}$ is the indicator of uncensored. Due to the loss of information from the partial observation of survival time in right-censored data, we need special approaches to analyze the behavior of U. Based on the right-censored data, Kaplan and Meier (1958) proposed the Kaplan-Meier (KM) estimator for the survival function of U. The asymptotic consistency and normality of the KM estimator were established by Breslow and Crowley (1974), and the estimations for the variance of the KM estimator were studied by Aalen and Johansen (1978) and Klein (1991). Using the idea of the KM estimator, Nelson (1972) provided the Nelson-Aalen estimator for the cumulative hazard function of U with right-censored data. Aalen (1978) restudied the Nelson-Aalen estimator using the counting process theory. The estimations for the density function of U with right-censored data were investigated by Földes, Rejtő, and Winter (1981), Mielniczuk (1986), and McNichols and Padgett (1986). Gehan (1965), Breslow (1970), Tarone and Ware (1977), and Harrington and Fleming (1982) considered two- or multiple-sample tests for the distributions of the survival time.

Many semiparametric models, such as the Cox model (Cox, 1972; Cox, 1975; Breslow, 1972), the additive hazard model (Aranda-Ordaz, 1983; Buckley, 1984; Lin and Ying, 1994), and the accelerated failure time model (Buckley and James, 1979;

Wei, 1992) were developed to analyze the covariate effects in survival analysis. The Cox model, which is the most popular model among them, is also known as the proportional hazard model because it supposes that a unit change on the covariate generates the exponential paces of changes on the hazard function. Cox (1972, 1975) proposed the partial likelihood estimation for the coefficients of the covariates. Breslow (1972) provided the Breslow estimator for the baseline hazard function. Using counting process theory, Andersen and Gill (1982) investigated the Cox model with time-dependent covariates and verified the asymptotic properties of their estimation. Christensen et al. (1986) further estimated the baseline hazard function for the Cox model with the time-dependent covariates. There were also many statistical studies based on other important topics on the Cox model, for example, the Cox model with covariates measured with error (Prentice, 1982; Hu, Tsiatis, and Davidian, 1998; Hu and Lin, 2002), variable selection approaches for the Cox model (Fan and Li, 2002; Zhang and Lu, 2007; Zhao et al., 2020), asymptotic properties of estimators for high dimensional Cox model (Bradic, Fan, and Jiang, 2011; Huang et al., 2013; Fang, Ning, and Liu, 2017), and functional Cox model (Chen et al., 2011; Qu, Wang, and Wang, 2016; Kong et al., 2018).

Most studies mentioned before considered the situation that the event of interest happened only once. Another situation is that one individual may experience the same type of events many times, and such events are referred to as recurrent events. Based on the idea of the KM estimator, Gill (1980) proposed a nonparametric model for the recurrent event, and Gill (1981) verified the asymptotic consistency. Sellke (1988) extended the Nelson–Aalen estimator and investigated the estimation for the intensity function when the recurrent events were observed in an infinite interval. Aalen and Husebye (1991) provided a renewal process model and studied the waiting time between two adjacent events. Ignoring the last event time, Wang and Chang (1999) considered a weighed moment estimation for the marginal survival function of the time between two adjacent events. For the semiparametric model of recurrent events, based on the Poisson process assumption, Andersen and Gill (1982) studied the proportional intensity model. Lin, Wei, and Ying (1998) investigated the accelerated failure time model for the counting process. Fleming and Harrington (1991) summarized the nonparametric and semiparametric estimations for survival analysis using the counting process and martingale techniques.

1.2 Literature Review

In this section, we introduce some existing researches about the two topics this thesis focusing on.

1.2.1 Subgroup Analysis

The classical models presented in Section 1.1 supposed that the treatment effects were homogeneous for all the individuals. However, in clinical trials, the treatment effects might be different among patients with different characters, which causes the homogeneous effect model to be misleading. (Sorensen, 1996; Kravitz, Duan, and Braslow; 2004) Precision medicine focuses on investigating the treatment heterogeneity, and the corresponding methods are referred to as subgroup analysis.

Early subgroup analysis methods supposed that the treatment heterogeneity only depends on the observed covariates, and the patients were clustered by the descriptive statistics. (Kravitz, Duan, and Braslow, 2004; Rothwell, 2005; Lagakos, 2006) Due to the lack of statistical frameworks, these approaches were inaccurate. The finite mixture model (Everitt and Hand, 1981) was one of the most popular statistical model in classification. In particular, Banfield and Raftery (1993), Hastie and Tibshirani (1996), and McNicholas (2010) studied the gaussian mixture model, and Wong and Li (2001), Muthén and Shedden (1999), and Muthén and Asparouhov (2009) investigated the logistic mixture model. Shen and He (2015) first applied the logistic mixture model to test the existence of subgroups and estimate the treatment effects in the linear regression model. Wu Zheng and Yu (2016) extended the approach of Shen and He (2015) to the subgroup analysis in the Cox model. Although the coefficient estimation in the finite mixture model for subgroup analysis performs well, we need to specify the number of subgroups before the analysis, which is a challenge in reality. Furthermore, the selected model in the finite mixture model supposes that the subgroup structure is only dependent on the observed covariates, which may be also not satisfied.

Recently, Ma and Huang (2017) and Ma et al. (2019) considered the model with individual-specific coefficients and proposed a data-driven subgroup identification procedure for the linear model. Introducing the pairwise fusion penalty, their methods clustered the individuals and estimated the coefficients simultaneously. Yan, Yin, and Zhao (2020) and Zhang, Wang, and Zhu (2019) extended this subgroup analysis method to the accelerated failure time model and the quantile regression model, respectively.

1.2.2 Panel Count Data Analysis

In long-term follow-up studies, when the observation time is discontinuous and only the occurrence numbers of the events of interest between two adjacent observation time points are available, such data is referred to as panel count data. Due to the missing information caused by the discontinuous observation time, the methods for survival analysis mentioned before are not applicable. Therefore, we need to develop the statistical inference approaches for panel count data specifically.

Some early studies assumed that the event only happened once for each individual, so they applied the methods for interval-censored data to analyze the panel count data. (Diamond, McDonald, and Shah, 1986; Sun and Kalbfleisch, 1993) For the studies of recurrent events, Kalbfleisch and Lawless (1985) investigated the panel count data under the Markov assumption. Sun and Kalbfleisch (1995) built a model on the mean function of counting processes with panel count data and considered the isotonic regression to obtain the estimation. Using monotone step function approximation, Wellner and Zhang (2000) provided the maximum pseudo-likelihood and maximum likelihood estimations for the mean function of counting processes with panel count data.

Considering the covariate effects, Sun and Wei (2000), Hu, Sun, and Wei (2003), and Zhang (2002) constructed the semiparametric model for panel count data. Wellner and Zhang (2007) studied the maximum pseudo-likelihood and maximum likelihood estimations under the proportional mean model with panel count data. Using the empirical process theory, they verified the convergence rate and the asymptotic normality of the estimations. Lu, Zhang, and Huang (2007, 2009) further investigated the model in Wellner and Zhang (2007) and improved the convergence rate of the estimations by the monotone spline approximation. There were also some studies focusing on the variable selection problem in the semiparametric model with panel count data. For example, Tong et al. (2009) and Zhang, Sun, and Wang (2013) proposed the penalized estimations for the coefficients using the nonconcave penalty function and the seamless-L_0 (SELO) penalty function, respectively.

For the hypothesis test with panel count data, Sun and Fang (2003) conducted the k-sample log-rank test for the mean function of counting processes with treatment indicator. Based on the maximum pseudo-likelihood and likelihood estimations in Wellner and Zhang (2000), Zhang (2006) and Balakrishnan and Zhao (2009) established the asymptotic normality for some functions of the estimators and proposed k-sample test statistics. Zhao and Sun (2011) studied the k-sample nonparametric hypothesis test when the distribution of observation processes were not equal in different groups. Following Lu, Zhang, and Huang (2007, 2009), Zhao and Zhang (2017) considered the B-spline approximation and conducted the two-sample test for the mean function of counting processes with panel count data.

The researches for panel count data were summarized in Sun and Zhao (2013).

1.2.3 Terminal Event

In survival analysis, the observation of individuals may be stopped by some terminal events which potentially affect the survival time. In some early studies, the terminal event was treated as dependent censoring, and the models were built under the repeated measure desine. (De Gruttola and Tu, 1994; Little, 1995; Sun and Song 2001) Based on this idea, two types of models for the analysis of recurrent events with the terminal event were developed, i.e. the marginal model and the frailty model.

The marginal model focuses on the marginal distribution of the rate function of the recurrent events given the condition of the terminal event time. Based on the marginal model of Wei, Lin, and Weissfeld (1989), Li and Lagakos (1997) studied the marginal model for recurrent events with the terminal event. Cook and Lawless (1997) proposed an estimation for the conditional mean function of the recurrent event using the marginal model. Ghosh and Lin (2002) investigated the semiparametric marginal effect model by the inverse probability censoring weight and the inverse probability survival weight methods. Ghosh and Lin (2003) combined the marginal model with the accelerated failure time model when the terminal event existed. Zhao, Zhou, and Sun (2011) considered the marginal model for the semiparametric regression with time-varying coefficients and an informative terminal event. Using the idea of the marginal model, Zhao, Li, and Sun (2013a, 2013b) provided semiparametric statistical analysis for panel count data with the terminal event.

The frailty model introduces a latent variable to describe the correlation between the recurrent events and the terminal event. It supposes that the recurrent event and the terminal event are conditionally independent given the latent variable. Under the Poisson process assumption, Lancaster and Intrator (1998) studied the joint distribution of the recurrent events and the terminal event. Wang, Qin, and Chiang (2001) proposed the frailty model under the assumption that the recurrent events were from the nonstationary Poisson process. Huang and Wang (2004) and Liu, Wolfe, and Huang (2004) investigated the extension of the proportional hazard model for recurrent events with an informative terminal event. Ye, Kalbfleisch, and Schaubel (2007) developed a time-varying coefficient model when an informative terminal event exists. Zeng and Cai (2010) applied the frailty model and the additive rate model to estimate the rate function of recurrent events. Sun, Tong, and He (2007) and Zhou et al. (2017) considered the frailty model for panel count data with the terminal event.

Both of the above two methods, however, are failed to explain the explicit interactions between the recurrent events and the terminal event. The marginal model is appropriate only when the recurrent events are not observed but indeed occur after the terminal event, and the frailty model describes the relationship between the recurrent events and the terminal event indirectly through the latent variable. (Kong et al., 2018) Recently, to detect the explicit interactions, Chan and Wang (2010) considered the situation that the terminal event stops the occurrences of recurrent events and built a time-backward model describing the occurrence rate of recurrent events before the terminal event. Treating the terminal event time as a fixed effect covariate, Kong et al. (2018) extended the time-backward model in Chan and Wang (2010) and developed the mixed effect model for longitudinal data with an informative terminal event.

1.3 Motivation and Outline

For the first topic, as mentioned in Section 1.2, the existing methods for the subgroup analysis with the Cox model are based on the logistic-Cox mixture model

(Wu Zheng and Yu, 2016), which still needs to be improved. This motivates us to consider the pairwise fusion function in Ma and Huang (2017) and Ma et al. (2019) to do the subgroup analysis with the Cox model. The proposed subgroup analysis procedure clusters the individuals with similar treatment effects into the same group, which not only does not need to specify the total number of subgroups but also clusters the individuals according to their latent characters. The majorized ADMM algorithm (Li, Sun, and Toh, 2016) considered in this thesis is implemented more easily and more accurately than the standard ADMM algorithm suggested by Ma and Huang (2017) and Ma et al. (2019).

For the analysis of panel count data with an informative terminal event, we consider the time backward model to study the explicit effect from the terminal event with the least squares-based loss function. Since it is difficult to minimize the loss function with respect to all the unknown parameters and functions, we treat the distribution function of the terminal event as a nuisance functional parameter and proposed a two-stage estimation. Furthermore, using the monotone I-spline approximation, the overall convergence rate of our estimation is slower than $n^{1/2}$, and the classical approach for the asymptotic normality is not applicable. Hence, we establish two general theorems for the asymptotic normality of M-estimation with nuisance parameter under the nonparametric model and the semiparametric model, respectively. In the two-sample hypothesis test for the nonparametric model, it is difficult to construct the statistics when the latent distribution functions of the terminal event are different in two groups. Based on the techniques introduced by Zhao and Sun (2011), we overcome this challenge which is by no means a straightforward extension from the asymptotic normality theorem.

The remainder of this thesis is organized as follows.

Chapter 2 proposes the subgroup analysis for the Cox model, which identifies the subgroup structure and estimates the coefficients simultaneously. We introduce the individual-specific parameter to represent the heterogeneous treatment effects. The individuals with similar treatment effects are clustered into the same subgroup by the concave pairwise fusion penalty function. To minimize the penalized criterion function, we introduce the majorized ADMM algorithm. Using the oracle estimation as a bridge, we verified the asymptotic consistency and the asymptotic normality of the proposed estimator. The finite sample performances of our estimator are demonstrated by the simulation studies. Finally, we applied our approach to analyzing the treatment effects in the breast cancer data.

Chapter 3 studies the nonparametric model for the panel count data with an informative terminal event. Based on the least squares-based loss function and treating the distribution function of the terminal event as a nuisance parameter, we consider the two-stage estimation procedure. Under some mild conditions, we prove the asymptotic consistency and the convergence rate of our estimation. We establish a general theorem for the asymptotic normality of M-estimation with nuisance parameter and verify the asymptotic normality for the proposed estimation based on this general theorem. Then the two-sample test statistics is constructed for comparison. We also use the simulation studies to show the finite sample performances of our estimation and the test statistics. At the end of Chapter 3, we use our method to analyze the rate of occurrences of severe diseases by the data of the Chinese Longitudinal Healthy Longevity Survey (CLHLS).

As a straightforward extension of Chapter 3, Chapter 4 investigates the semiparametric model for the panel count data with an informative terminal event. The estimation is still based on the least squares-based loss function with a two-stage estimation procedure. The asymptotic consistency and the convergence rate are verified under some mild conditions, and we also establish a general theorem for the asymptotic normality of M-estimation with nuisance parameter for the semiparametric model. We conduct simulation studies to show the finite sample estimation results. In the real data analysis, we further study the occurrence rate of severe diseases by the CLHLS data.

In Chapter 5, we summarise the research results in this thesis and provide some researches for future study.

Chapter 2

Subgroup Analysis in the Cox Model

2.1 Introduction

In this chapter, we consider the subgroup analysis in the heterogenous Cox model under the assumption of sparsity subgroup structure. Based on the objective function constructed through combining the negative logarithmic partial likelihood function and a concave fusion penalty function, we can identify the subgroup structure and estimate treatment effects simultaneously without any prior knowledge about the group structure. The likelihood-based regularization approaches make the statistical inference of identifying the subgroup structure and estimating treatment effects become an automated procedure and so it is easy to implement.

To overcome the computational difficulties caused from the complicated nature of the likelihood-based objective function, we borrow the ideas of the majorized alternating direction method of multipliers (ADMM) algorithm. (Li, Sun, and Toh; 2016) Compared to the classical ADMM algorithm suggested by Ma and Huang (2017), this algorithm is able to efficiently handle large scale problems to get more accurate solutions by transforming an objective function into a majorized convex function with a pairwise fusion penalty. We take the ridge solution of the negative log-likelihood function as the initial solution of the algorithm, and find that the initial solution performs well in identifying the subgroup structure in our simulation studies.

Using the oracle estimator as a bridge, we obtain the oracle property of the proposed estimator. Concretely, we obtain the consistency and asymptotic normality of the oracle estimator at first. Then we show that the oracle estimator and the proposed estimator are asymptotically equivalent. Thus, the latter is consistent and possesses the asymptotic normality. This property also illustrates that the proposed method can identify the subgroup structure of the model as if we knew it in advance.

The rest of this chapter is organized as follows. In Section 2.2, we introduce the heterogenous Cox model with right censored data and propose a penalized estimation approach. Section 2.3 presents the majorized ADMM algorithm for computing the proposed estimators. In Section 2.4, we establish the consistency and the asymptotic normality of the proposed estimator. We then conduct simulation studies to demonstrate the performance of the proposed method in Section 2.5, and use the method to analyze a real data example in Section 2.6. The proofs of the theoretical results are relegated to the Appendix.

2.2 Heterogenous Cox Model and Estimation Procedure

Consider a survival study containing n independent subjects. For subject i, let U_i and C_i denote the failure time and the censoring time, respectively. Then the observed data consist of $\{(T_i, \Delta_i) : i = 1, \cdots, n\}$, where $T_i = U_i \wedge C_i$ and $\Delta_i = 1_{\{U_i \leq C_i\}}$. Let X_i and Z_i denote covariates with dimensions p and q, respectively. Let $\lambda(t|X_i, Z_i)$ be the conditional hazard rate function of U given X_i and Z_i. Then the homogeneous Cox model is

$$\lambda(t|X_i, Z_i) = \lambda_0(t)\exp(Z_i^T\eta + X_i^T\beta), i = 1, \cdots, n, \tag{2.1}$$

where $\lambda_0(t)$ is the baseline hazard function, η and β are unknown regression parameters denoting the average effects. However, the homogeneous assumption about covariate effects is not satisfied when the effects of X_i are different among subjects. To describe the treatment heterogeneity, we propose the heterogeneous Cox model as follows:

$$\lambda(t|X_i, Z_i) = \lambda_0(t)\exp(Z_i^T\eta + X_i^T\beta_i), i = 1, \cdots, n, \tag{2.2}$$

where β_i is subject-specific effect of X_i on the hazard function. We suppose that n subjects are divided into K potential subgroups according to set $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_K)$, and $\beta_i \equiv \alpha_k$ for all $i \in \mathcal{G}_k$, $k = 1, \cdots, K$. For this model, we focus on identifying the subgroup set $\mathcal{G}$ and estimating parameters $\{\alpha_1, \cdots, \alpha_K\}$ and η.

For the coefficient of X, define $\boldsymbol{\alpha} = (\alpha_1^T, \ldots, \alpha_K^T)^T$ and $\boldsymbol{\beta} = (\beta_1^T, \ldots, \beta_n^T)^T$. The negative partial log-likelihood function is

$$\ell_n(\eta, \boldsymbol{\beta}) = -\sum_{i=1}^{n} \Delta_i(Z_i^T\eta + X_i^T\beta_i) + \sum_{i=1}^{n} \Delta_i \log\Big(\sum_{j\in R(T_i)} \exp(Z_j^T\eta + X_j^T\beta_j)\Big), \tag{2.3}$$

where $R(T_i) = \{j : T_j \geq T_i\}$ is the risk set. For the purpose of identifying the subgroup structure, we use a concave pairwise penalty $p_\gamma(||\beta_i - \beta_j||, \lambda)$ to shrink small value of $||\beta_i - \beta_j||$ to 0, where $||\cdot||$ is the L_2-norm of a vector. Then the criterion function is

$$Q_n(\eta, \boldsymbol{\beta}) = \ell_n(\eta, \boldsymbol{\beta}) + \sum_{i<j} p_\gamma(||\beta_i - \beta_j||, \lambda), \tag{2.4}$$

where $\lambda \geq 0$ is a tuning parameter. Thus, we can obtain the estimator $(\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda))$ by minimizing the objective function (2.4) with a given turning parameter λ. Finally, the estimator for $\boldsymbol{\alpha}$ is the distinct value of $\widehat{\boldsymbol{\beta}}(\lambda)$, denoted by $\widehat{\boldsymbol{\alpha}}(\lambda) = (\hat{\alpha}_1^T(\lambda), \ldots, \hat{\alpha}_{\widehat{K}}^T(\lambda))^T$. The identified subgroup structure is $\widehat{\mathcal{G}}_k(\lambda) = \{i : \hat{\beta}_i(\lambda) = \hat{\alpha}_k(\lambda), 1 \leq i \leq n\}$, where $1 \leq k \leq \widehat{K}(\lambda)$.

The penalty function can be naively chosen as the L_1 penalty function $p_\gamma(t, \lambda) = \lambda|t|$, but L_1 penalty tends to choose too many subgroups. Following Ma and Huang (2017), a better choice of the penalty function is the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001) with

$$p_\gamma(t, \lambda) = \lambda \int_0^{|t|} \min\{1, (\gamma - x/\lambda)_+/(\gamma - 1)\} dx,$$

or the minimax concave penalty (MCP) (Zhang, 2010) with

$$p_\gamma(t, \lambda) = \lambda \int_0^{|t|} (1 - x/(\gamma\lambda))_+ dx.$$

2.3 Majorized ADMM Algorithm

In this section, we present the algorithm to find the solution path $(\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda))$. Introducing a new set of parameters $u_{ij} = \beta_i - \beta_j$, we can reformulate the criterion function $Q_n(\eta, \boldsymbol{\beta})$ as

$$Q_n(\eta, \boldsymbol{\beta}, \boldsymbol{u}) = \ell_n(\eta, \boldsymbol{\beta}) + \sum_{i<j} p_\gamma(||u_{ij}||, \lambda)$$

subject to $\beta_i - \beta_j - u_{ij} = 0$, where $\boldsymbol{u} = (u_{ij}^T, i < j)^T$. Following Ma et al. (2019), we can solve this minimization problem using the standard ADMM algorithm by approximating $\ell_n(\eta, \boldsymbol{\beta})$ as the quadratic function

$$\ell_n(\eta, \boldsymbol{\beta}) \approx \ell_n(\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)}) + \nabla \ell_n(\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)})^T \big((\eta, \boldsymbol{\beta}) - (\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)})\big)$$
$$+\frac{1}{2}\big((\eta, \boldsymbol{\beta})^T - (\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)})^T\big) \nabla^2 \ell_n(\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)}) \big((\eta, \boldsymbol{\beta}) - (\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)})\big),$$

where $(\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)})$ is the value of parameter in the mth iteration step. However, the quadratic approximation is only accurate when $(\eta, \boldsymbol{\beta})$ is close to $(\eta^{(m-1)}, \boldsymbol{\beta}^{(m-1)})$, and the calculation of the second order derivative $\nabla^2 \ell_n(\eta, \boldsymbol{\beta})$ is time consuming. Hence, it motivates us to utilize the idea of the majorized ADMM algorithm. (Li, Sun and Toh, 2016)

Introduce another set of parameters $Y_i = Z_i^T \eta + X_i^T \beta_i$, and let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$. The negative log partial-likelihood function $l_n(\eta, \boldsymbol{\beta})$ can be rewritten as

$$g(\boldsymbol{Y}) = -\sum_{i=1}^{n} \Delta_i Y_i + \sum_{i=1}^{n} \Delta_i \log\Big(\sum_{j \in R(T_i)} \exp(Y_j)\Big).$$

Then we need to minimize

$$Q_n(\eta, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{Y}) = g(\boldsymbol{Y}) + \sum_{i<j} p_\gamma(||u_{ij}||, \lambda) \tag{2.5}$$

subject to $\beta_i - \beta_j - u_{ij} = 0$ and $Y_i = Z_i^T \eta + X_i^T \beta_i$. Since $\nabla^2 g(\boldsymbol{Y}) \preceq \widetilde{G}$ for $\widetilde{G} = \frac{1}{2}\mathrm{diag}\{\tilde{\mathrm{g}}_1, \ldots, \tilde{\mathrm{g}}_\mathrm{n}\}$ and $\tilde{g}_j = \sum_{i=1}^{n} \Delta_i I_{j \in R(T_i)}$, we have

$$g(\boldsymbol{Y}) \leq \tilde{g}(\boldsymbol{Y}; \boldsymbol{Y}') := g(\boldsymbol{Y}') + \langle \boldsymbol{Y} - \boldsymbol{Y}', \nabla g(\boldsymbol{Y}') \rangle + \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{Y}'\|_{\widetilde{G}}^2$$

for any $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ with $\|\boldsymbol{x}\|_{\widetilde{G}}^2 = \langle \boldsymbol{x}, \widetilde{G}\boldsymbol{x} \rangle$. The objective function (2.5) is then transformed to the majorized augmented Lagrangian function as follows

$$Q_n'(\eta, \boldsymbol{\beta}, \boldsymbol{Y}, \boldsymbol{u}; \boldsymbol{w}, \boldsymbol{v}, \boldsymbol{Y}') = \tilde{g}(\boldsymbol{Y}; \boldsymbol{Y}') + \sum_{i<j} p_\gamma(\|u_{ij}\|, \lambda) + \sum_{i=1}^{n} \langle w_i, Y_i - Z_i^T \eta - X_i^T \beta_i \rangle$$

$$+ \sum_{i<j} \langle v_{ij}, \beta_i - \beta_j - u_{ij} \rangle + \frac{\vartheta}{2} \sum_{i=1}^{n} (Y_i - Z_i^T \eta - X_i^T \beta_i)^2 + \frac{\vartheta}{2} \sum_{i<j} ||\beta_i - \beta_j - u_{ij}||^2,$$

where the dual variables $\boldsymbol{w} = (w_i, i = 1, \ldots, n)^T$ and $\boldsymbol{v} = (v_{ij}^T, i < j)^T$ are the Lagrange multipliers, and ϑ is the penalty parameter. We then compute the estimators $\widehat{\boldsymbol{\beta}}$ and $\hat{\eta}$ through the following majorized ADMM algorithm.

At the mth iteration, for a given value of parameter

$$(\eta^{(m-1)}, \beta^{(m-1)}, \boldsymbol{Y}^{(m-1)}, \boldsymbol{u}^{(m-1)}; \boldsymbol{w}^{(m-1)}, \boldsymbol{v}^{(m-1)}, \boldsymbol{Y}'^{(m-1)}),$$

cluster size $K^{(m-1)}$, and subgroup set $\mathcal{G}^{(m-1)}$, the iteration goes as follows:

Step 1. Update $(\eta^{(m)}, \boldsymbol{\beta}^{(m)})$ by minimizing

$$Q'_n(\eta, \boldsymbol{\beta}, \boldsymbol{Y}^{(m-1)}, \boldsymbol{u}^{(m-1)}; \boldsymbol{w}^{(m-1)}, \boldsymbol{v}^{(m-1)}, \boldsymbol{Y}'^{(m-1)});$$

Step 2. Update $(\boldsymbol{Y}^{(m)}, \boldsymbol{u}^{(m)})$ by minimizing

$$Q'_n(\eta^{(m)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{Y}, \boldsymbol{u}; \boldsymbol{w}^{(m-1)}, \boldsymbol{v}^{(m-1)}, \boldsymbol{Y}'^{(m-1)})$$

and update

$$Y_i'^{(m)} = Z_i^T \eta^{(m)} + X_i^T \beta_i^{(m)} \tag{2.6}$$

for $i = 1, \ldots, n$;

Step 3. Update $\boldsymbol{w}^{(m)}$ and $\boldsymbol{v}^{(m)}$ by

$$\begin{aligned} w_i^{(m)} =& w_i^{(m-1)} + \varrho\vartheta(Y_i^{(m)} - Z_i^T \eta^{(m)} - X_i^T \beta_i^{(m)}), \\ v_{ij}^{(m)} =& v_{ij}^{(m-1)} + \varrho\vartheta(\beta_i^{(m)} - \beta_j^{(m)} - u_{ij}^{(m)}), \end{aligned} \tag{2.7}$$

where the constant $\varrho \in (0, (1+\sqrt{5})/2)$;

Step 4. Update $K^{(m)}$ and $\mathcal{G}^{(m)}$ by clustering $\boldsymbol{\beta}^{(m)}$.

At Step 1, for fixed $(\boldsymbol{Y}, \boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}, \boldsymbol{Y}')$, it suffices to minimize the following objective function in order to update $\boldsymbol{\beta}$ and η:

$$\begin{aligned} &\sum_{i=1}^n \langle w_i, Y_i - Z_i^T \eta - X_i^T \beta_i \rangle + \sum_{i<j} \langle v_{ij}, \beta_i - \beta_j - u_{ij} \rangle \\ &+ \frac{\vartheta}{2} \sum_{i=1}^n (Y_i - Z_i^T \eta - X_i^T \beta_i)^2 + \frac{\vartheta}{2} \sum_{i<j} ||\beta_i - \beta_j - u_{ij}||^2. \end{aligned} \tag{2.8}$$

Define $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$, $\boldsymbol{X} = \text{diag}(\mathrm{X}_1^{\mathrm{T}}, \ldots, \mathrm{X}_{\mathrm{n}}^{\mathrm{T}})$ and $\boldsymbol{A} = \boldsymbol{D} \otimes I_p$, where $\boldsymbol{D} = \{(e_i - e_j), i < j\}^T$ with e_i being an $n \times 1$ vector whose ith entry is 1 and the remaining ones are 0, I_p is a $p \times p$ identity matrix, and $\otimes$ is a Kronecker product. For given K and $\mathcal{G}$, let $\boldsymbol{W}_{\mathcal{G}} = \{\omega_{ik}\}$ be an $n \times K$ matrix, where the entry ω_{ik} takes 1 if $i \in \mathcal{G}_k$ and 0 otherwise. In addition, we define $\widetilde{\boldsymbol{W}}_{\mathcal{G}} = \boldsymbol{W}_{\mathcal{G}} \otimes I_p$, $\widetilde{\boldsymbol{X}} = \boldsymbol{X}\widetilde{\boldsymbol{W}}_{\mathcal{G}}$ and $\widetilde{\boldsymbol{A}} = \boldsymbol{A}\widetilde{\boldsymbol{W}}_{\mathcal{G}}$. Thus, after removing the terms irrelevant to $\boldsymbol{\beta}$ and η, the minimal point of (2.8) is obtained equivalently by minimizing

$$\frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{Z}\eta - \widetilde{\boldsymbol{X}}\boldsymbol{\alpha} + \frac{\boldsymbol{w}}{\vartheta}\|^2 + \frac{1}{2}\|\widetilde{\boldsymbol{A}}\boldsymbol{\alpha} - \boldsymbol{u} + \frac{\boldsymbol{v}}{\vartheta}\|^2.$$

At the mth iteration, setting $\boldsymbol{Q}_Z = I_n - \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$, the parameters $\boldsymbol{\beta}$ and η are updated through the following equations

$$\begin{aligned}
\boldsymbol{\alpha}^{(m)} &= \boldsymbol{H}_{\mathcal{G}}^{-1}\boldsymbol{S}_{\mathcal{G}}^{(m-1)}, \\
\boldsymbol{\beta}^{(m)} &= \widetilde{\boldsymbol{W}}_{\mathcal{G}}\boldsymbol{\alpha}^{(m)}, \qquad (2.9)\\
\eta^{(m)} &= (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T(\boldsymbol{Y}^{(m-1)} - \boldsymbol{X}\boldsymbol{\beta}^{(m)} + \vartheta^{-1}\boldsymbol{w}^{(m-1)}),
\end{aligned}$$

where $\boldsymbol{H}_{\mathcal{G}} = \widetilde{\boldsymbol{X}}^T\boldsymbol{Q}_Z\widetilde{\boldsymbol{X}} + \widetilde{\boldsymbol{A}}^T\widetilde{\boldsymbol{A}}$, and $\boldsymbol{S}_{\mathcal{G}}^{(m-1)} = \widetilde{\boldsymbol{X}}^T\boldsymbol{Q}_Z(\boldsymbol{Y}^{(m-1)} + \vartheta^{-1}\boldsymbol{w}^{(m-1)}) + \widetilde{\boldsymbol{A}}^T(\boldsymbol{u}^{(m-1)} - \vartheta^{-1}\boldsymbol{v}^{(m-1)})$. It deserves to note that the updated solution of parameter $\boldsymbol{\beta}^{(m)}$ includes the integrated information of $\boldsymbol{\alpha}^{(m)}$, $\mathcal{G}^{(m-1)}$ and $K^{(m-1)}$.

At Step 2, for fixed $(\eta, \beta, \boldsymbol{w}, \boldsymbol{v}, \boldsymbol{Y}')$, we need to get the minimal points

$$\begin{aligned}
\arg\min_{\boldsymbol{Y}} \quad & \langle \boldsymbol{Y}, \nabla g(\boldsymbol{Y}')\rangle + \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{Y}'\|_{\widetilde{G}}^2 \qquad (2.10)\\
& + \sum_{i=1}^{n}\langle w_i, Y_i - Z_i^T\eta - X_i^T\beta_i\rangle + \frac{\vartheta}{2}\sum_{i=1}^{n}(Y_i - Z_i^T\eta - X_i^T\beta_i)^2, \\
\arg\min_{u_{ij}} \quad & \frac{1}{2}\|\beta_i - \beta_j + \frac{v_{ij}}{\vartheta} - u_{ij}\|^2 + \frac{1}{\vartheta}p_\gamma(\|u_{ij}\|, \lambda). \qquad (2.11)
\end{aligned}$$

At the mth iteration, for (2.10), it can be solved that for $i = 1, \ldots, n$,

$$Y_i^{(m)} = (\tilde{g}_i + \vartheta)^{-1} \Big[-\nabla_i g(\boldsymbol{Y}'^{(m-1)}) + \tilde{g}_i Y_i'^{(m-1)} - w_i^{(m-1)} + \vartheta(Z_i^T \eta^{(m)} + X_i^T \beta_i^{(m)}) \Big]. \quad (2.12)$$

For (2.11), we can get the closed form of $u_{ij}^{(m)}$ for some commonly used penalties, such as group MCP and group SCAD. For the group SCAD penalty with parameter γ, i.e.,

$$p'_\gamma(\|u_{ij}\|, \lambda) = \lambda I(\|u_{ij}\| \leq \lambda) + \frac{(\gamma\lambda - \|u_{ij}\|)_+}{\gamma - 1} I(\|u_{ij}\| > \lambda),$$

we have

$$u_{ij}^{(m)} = \begin{cases} S(c_{ij}^{(m-1)}; \lambda/\vartheta), & \|c_{ij}^{(m-1)}\| \leq \lambda + \lambda/\vartheta, \\ \frac{\left(\vartheta(\gamma-1) - \lambda\gamma/\|c_{ij}^{(m-1)}\|\right) c_{ij}^{(m-1)}}{\vartheta\gamma - \vartheta - 1}, & \lambda + \lambda/\vartheta < \|c_{ij}^{(m-1)}\| \leq \lambda\gamma, \\ c_{ij}^{(m-1)}, & \|c_{ij}^{(m-1)}\| > \lambda\gamma, \end{cases} \quad (2.13)$$

where $c_{ij}^{(m-1)} = \beta_i^{(m)} - \beta_j^{(m)} + \frac{v_{ij}^{(m-1)}}{\vartheta}$ and $S(c; \lambda) = (1 - \lambda/\|c\|)_+ c$. For group MCP penalty with parameter γ, i.e.,

$$p'_\gamma(\|u_{ij}\|, \lambda) = \frac{(\gamma\lambda - \|u_{ij}\|)_+}{\gamma},$$

we have

$$u_{ij}^{(m)} = \begin{cases} S\left(\frac{\vartheta c_{ij}^{(m-1)}}{\vartheta - 1/\gamma}; \frac{\lambda}{\vartheta - 1/\gamma}\right), & \|c_{ij}^{(m-1)}\| \leq \lambda\gamma, \\ c_{ij}^{(m-1)}, & \|c_{ij}^{(m-1)}\| > \lambda\gamma. \end{cases} \quad (2.14)$$

At Step 4, we first solve the following optimization problem

$$\tilde{u}_{ij}^{(m)} = \arg\min_{\tilde{u}_{ij}} \frac{1}{2} \|\beta_i^{(m)} - \beta_j^{(m)} - \tilde{u}_{ij}\|^2 + p_\gamma(\|\tilde{u}_{ij}\|, \lambda), \quad (2.15)$$

and then update $K^{(m)}$ and $\mathcal{G}^{(m)}$ by clustering individuals i and j into the same group if $\tilde{u}_{ij} = 0$. This step is critical to clustering analysis of the regression coefficient $\boldsymbol{\beta}$ so that Step 1 can be carried out smoothly in the recursive process. The performance of the algorithm depends on the choice of the penalty function and the tuning parameter λ.

The initial points in the algorithm are taken as follows. Since covariate Z has no subgroup effect, we simply take the estimator $\hat{\eta}$ as $\eta^{(0)}$ by treating the hazard function as a homogeneous effect model. As a reasonable initial point of parameter $\boldsymbol{\beta}$, it should reflect not only the form of the assumed hazard function but also the subgroup relation among different individuals. So we consider the ridge solution of the negative log-likelihood function as $\boldsymbol{\beta}^{(0)}$. Concretely, we define

$$\boldsymbol{\beta}^{(0)} = \arg\min_{\boldsymbol{\beta}} l_n(\eta^{(0)}, \boldsymbol{\beta}) + \frac{\lambda^*}{2} \sum_{i<j} \|\beta_i - \beta_j\|^2,$$

where tuning parameter λ^* is taken as 0.001 in our simulation studies, and utilize a majorized algorithm to find the solution of $\boldsymbol{\beta}^{(0)}$ through (2.5). We take $K^{(0)} = \lfloor \sqrt{n} \rfloor$ to ensure that there are enough groups at the beginning of the iteration. A cluster analysis method can then be applied to $\boldsymbol{\beta}^{(0)}$ for determining $\mathcal{G}^{(0)} = (\mathcal{G}_1^{(0)}, \dots, \mathcal{G}_{K^{(0)}}^{(0)})$. Take $\boldsymbol{Y}^{(0)} = \boldsymbol{Y}'^{(0)} = \boldsymbol{Z}\eta^{(0)} + \boldsymbol{X}\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)} = \boldsymbol{A}\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{w}^{(0)} = \boldsymbol{v}^{(0)} = \underline{0}$.

Denote the primal residual as

$$r^{(m)} = \sum_{i=1}^{n} (y_i^{(m)} - \mathbf{z}_i^T \eta^{(m)} - \mathbf{x}_i^T \beta_i^{(m)})^2 + \sum_{i<j} ||\beta_i^{(m)} - \beta_j^{(m)} - \mathbf{u}_{ij}^{(m)}||^2. \tag{2.16}$$

We stop the iteration when $r^{(m)}$ is small enough.

We summarize the above descriptions in Algorithm 1.

2.4 Asymptotic Results

Let $N_i(t) = 1_{(T_i \le t, \Delta_i = 1)}$, $Y_i(t) = 1_{(T_i \ge t)}$, and τ be the end time of study. Suppose that $\int_0^\tau \lambda_0(t)dt < \infty$. The negative partial log-likelihood function can be rewritten

Algorithm 1 Majorized ADMM algorithm

Initialize $(\boldsymbol{\eta}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{Y}^{(0)}, \boldsymbol{u}^{(0)}; \boldsymbol{w}^{(0)}, \boldsymbol{v}^{(0)}, \boldsymbol{Y}'^{(0)})$, $K^{(0)}$, and $\mathcal{G}^{(0)}$
for $m = 1, 2, \cdots$ **do**
  Update $(\boldsymbol{\beta}^{(m)}, \boldsymbol{\eta}^{(m)})$ using (2.9)
  Update $(\boldsymbol{Y}^{(m)}, \boldsymbol{u}^{(m)})$ using (2.12) (2.13), and (2.14)
  Update $\boldsymbol{Y}'^{(m)}$ using (2.6)
  Update $(\boldsymbol{w}^{(m)}, \boldsymbol{v}^{(m)})$ using (2.7)
  Compute $\tilde{\boldsymbol{u}}_{ij}^{(m)}$ using (2.15), and update $(K^{(m)}, \mathcal{G}^{(m)})$ according to $\tilde{\boldsymbol{u}}_{ij}^{(m)}$
  Compute $r^{(m)}$ using (2.16)
  if $r^{(m)}$ is small enough **then**
    Stop and denote the last iteration by $(\hat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\eta}})$
  end if
end for

as

$$\ell_n(\eta, \boldsymbol{\beta}) = -\sum_{i=1}^{n} \int_0^{\tau} \left[\left(Z_i^T \eta + X_i^T \beta_i \right) - \log \left\{ \sum_{j=1}^{n} Y_j(t) \exp(Z_j^T \eta + X_j^T \beta_j) \right\} \right] dN_i(t).$$

The objective function is $Q_n(\eta, \boldsymbol{\beta}) = \ell_n(\eta, \boldsymbol{\beta}) + P_n(\boldsymbol{\beta})$, where $P_n(\boldsymbol{\beta}) = \sum_{i<j} p_\gamma(||\beta_i - \beta_j||, \lambda)$. Denote the true subgroup set as $\mathcal{G}_0 = (\mathcal{G}_{0,1}, \cdots, \mathcal{G}_{0,K_0})$. Define $\widetilde{\boldsymbol{W}}_{\mathcal{G}_0} = \mathbf{W}_{\boldsymbol{\mathcal{G}_0}} \otimes I_p$, $\widetilde{\boldsymbol{X}}_{\mathcal{G}_0} = \mathbf{X}\widetilde{\boldsymbol{W}}_{\mathcal{G}_0}$, $\mathbf{B} = (\mathbf{Z}, \widetilde{\boldsymbol{X}}_{\mathcal{G}_0})$, and let $\mathbf{B}_i$ be the i-th column of $\mathbf{B}^T$. Let $\boldsymbol{\theta} = (\eta^T, \boldsymbol{\alpha}^T)^T$, and $S^{(0)}(\boldsymbol{\theta}, \mathbf{B}, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \exp(\mathbf{B}_i^T \boldsymbol{\theta})$. Thus, with the prior information of $\mathcal{G}_0$, we write the negative partial log-likelihood function as

$$\tilde{\ell}_n(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{B}_i^T \boldsymbol{\theta} - \log[n S^{(0)}(\boldsymbol{\theta}, \mathbf{B}, t)] \right] dN_i(t).$$

Then the oracle estimator $\widehat{\boldsymbol{\theta}}^{or} = (\hat{\eta}^{or}, \widehat{\boldsymbol{\alpha}}^{or})$ is the minimizer of $\tilde{\ell}_n(\boldsymbol{\theta})$.

Now we present the asymptotic results of the proposed estimators.

Theorem 2.1. *Suppose that Conditions* (C1)-(C3) *given in the Appendix hold. Let* $\boldsymbol{\theta}_0$ *be the true value of parameter* $\boldsymbol{\theta}$. *Then*

(i) $\widehat{\boldsymbol{\theta}}^{or} \xrightarrow{p} \boldsymbol{\theta}_0$;

(ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}^{or} - \boldsymbol{\theta}_0)$ *converges in distribution to the multivariate normal distribution with zero mean and covariance matrix* $\Sigma^{-1}(\boldsymbol{\theta}_0)$*, where* $\Sigma(\boldsymbol{\theta}_0)$ *is given in the Appendix.*

Theorem 2.1 shows that when the grouping structure is known, the oracle estimator is consistent and asymptotically normal. Next, when the true subgroup set $\mathcal{G}_0$ is known, we define the oracle parameter space of $\boldsymbol{\beta}$ as

$$\mathcal{M}_{\mathcal{G}_0} = \{\boldsymbol{\beta} \in R^{np} : \beta_i = \beta_j = \alpha_k, \text{ for any } i, j \in \mathcal{G}_{0,k}, 1 \leq k \leq K_0\}.$$

Define $(\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})$ as the minimizer of $\ell_n(\eta, \boldsymbol{\beta})$ with subject to $\boldsymbol{\beta} \in \mathcal{M}_{\mathcal{G}_0}$. Set $\boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0$ to be the true parameter. We first consider the case of $K_0 \geq 2$ and have the following result.

Theorem 2.2. *Suppose that Conditions* (C1)-(C4) *given in the Appendix hold. Let* $b = \min_{i\in\mathcal{G}_{0,k}, j\in\mathcal{G}_{0,k'}, k\neq k'} ||\beta_{0i} - \beta_{0j}|| = \min_{k\neq k'} ||\alpha_{0k} - \alpha_{0k'}||$*. Assume that* $b > a\lambda$ *for constant* a *in Condition* (C4)*. Then there exists a local minimizer* $(\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda))$ *of the objective function* $Q_n(\eta, \boldsymbol{\beta}; \lambda)$ *satisfying* $P\big((\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda)) = (\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})\big) \to 1$.

Next, we consider the case of a homogeneous model in which $K_0 = 1$ and $\beta_{01} = \cdots = \beta_{0n} \equiv \boldsymbol{\alpha}_0$.

Theorem 2.3. *Suppose that Conditions* (C1)-(C4) *given in the Appendix hold. When there is only one group, we define the oracle parameter space of* $\boldsymbol{\beta}$ *as* $\mathcal{M} = \{\boldsymbol{\beta} \in R^{np} : \beta_i \equiv \boldsymbol{\alpha}, i = 1, \cdots, n\}$*, and the oracle estimator* $(\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})$ *as the minimizer of* $\ell_n(\eta, \boldsymbol{\beta})$ *with* $\boldsymbol{\beta} \in \mathcal{M}$*. Then there exists a local minimizer* $(\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda))$ *of the objective function* $Q_n(\eta, \boldsymbol{\beta}; \lambda)$ *satisfying* $P\big((\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda)) = (\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})\big) \to 1$.

Let $\widehat{\boldsymbol{\alpha}}(\lambda)$ be the distinct value of $\widehat{\boldsymbol{\beta}}(\lambda)$ and $\widehat{\boldsymbol{\alpha}}^{or}$ be the distinct value of $\widehat{\boldsymbol{\beta}}^{or}$. By Theorems 2.1–2.3, we conclude that $n^{1/2}(\widehat{\boldsymbol{\theta}}(\lambda) - \boldsymbol{\theta}_0)$ converges in distribution to the multivariate normal distribution with mean 0 and covariance matrix $\Sigma^{-1}(\boldsymbol{\theta}_0)$.

2.5 Simulation Studies

We conducted simulation studies to evaluate the performance of the proposed method. The data were generated from model (2.2) with censoring rate 0.20, where $\lambda_0(t) = 1$, $\eta = (-1, 1)^T$, and $Z_i = (Z_{i1}, Z_{i2})^T$ was generated from multivariate normal with mean 0, variance 1 and correlation 0.4. We considered four examples: (i) one treatment variable with two latent subgroups of equal size; (ii) multi-treatment variable with two subgroups of unequal size; (iii) one treatment variable with three latent subgroups of equal size; (iv) one treatment variable with a homogeneous effect. Two penalties, group SCAD and group MCP, were used in the examples to compare their performance with oracle estimators. The parameter γ was taken as 3.7 and 2.5 for SCAD and MCP, respectively. We set sample size $n = 100$ or 200 in Examples 1, 2 and 4 and $n = 150$ or 300 in Example 3, and let $\vartheta = 1$ in the majorized ADMM algorithm.

To implement the algorithm, we adopt the warm start to update the solution path of $\boldsymbol{\beta}$ and η along different values of λ, and use the modified BIC criterion in Lee, Noh, and Park (2014) to select the optimal tuning parameter λ by minimizing

$$BIC(\lambda) = l_n(\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda)) + C_n \frac{\log n}{n} (\widehat{K}(\lambda)p + q),$$

where $C_n = \log(n\widehat{K}(\lambda) + q)$. The simulation results are based on 100 replications.

Example 1. We first generated X_i from Bernoulli(0.5)+1. Let $\mathcal{G}_1 = \{1, \ldots, n/2\}$ and $\mathcal{G}_2 = \{n/2 + 1, \ldots, n\}$, and the effects of variable X on the survival time were divided into 2 groups with equal size. We considered the following two cases to investigate the effect of the size of the difference between the subgroup-specific treatment effects:

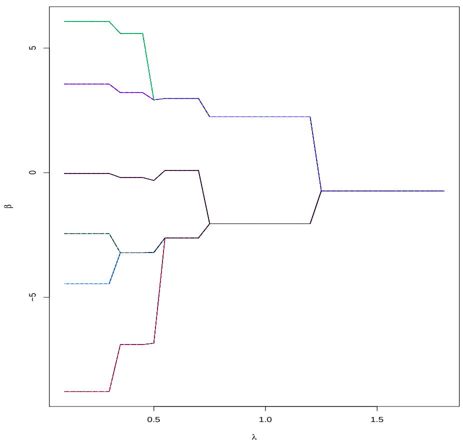
Case 1: $\beta_i = -1.5$ for $i \in \mathcal{G}_1$ and $\beta_i = 1.5$ for $i \in \mathcal{G}_2$, that is, $\boldsymbol{\alpha} = (-1.5, 1.5)^T$.

Case 2: $\beta_i = -3$ for $i \in \mathcal{G}_1$ and $\beta_i = 3$ for $i \in \mathcal{G}_2$, that is $\boldsymbol{\alpha} = (-3, 3)^T$.

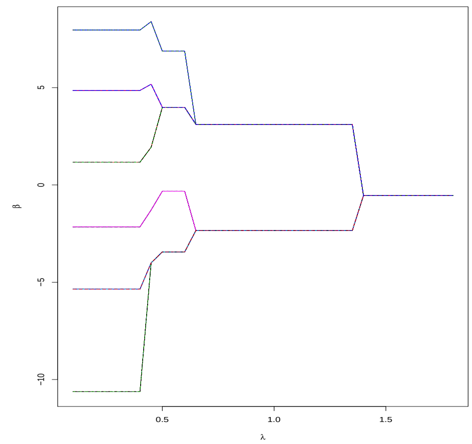
We also compared our approach with the subgroup analysis approach under the logistic-Cox mixture model (Wu, Zheng and Yu, 2016) in Example 1.

(a) Fusiongram based on one dataset

Case 1: $\alpha = (-1.5, 1.5)$

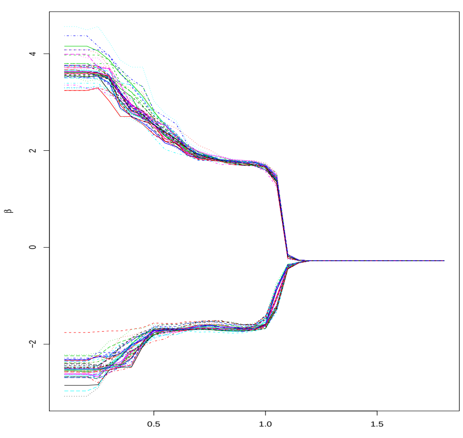


Case 2: $\alpha = (-3, 3)$



(b) Fusiongram based on 100 replications

Case 1: $\alpha = (-1.5, 1.5)$



Case 2: $\alpha = (-3, 3)$

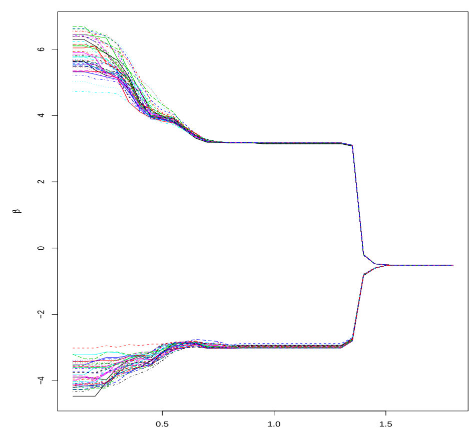


Figure 2.1: Fusiongram for estimation of parameter $\boldsymbol{\beta}$ for GMCP in Example 1 when $n = 100$.

The simulation results for Exapmle 1 are summarized in Tables 2.1 and 2.2 and Figure 2.1. Figure 2.1 includes two kinds of fusiongrams for GMCP when $n = 100$,

Table 2.1: Simulation results for estimation of group size K in Example 1.

n	METHOD	MEAN	MEDIAN	SD	TPR
		Case 1: $\boldsymbol{\alpha} = (-1.5, 1.5)$ and $\boldsymbol{\eta} = (-1, 1)$			
100	GMCP	2.10	2	0.333	0.911
	GSCAD	2.09	2	0.321	0.909
200	GMCP	2.13	2	0.367	0.922
	GSCAD	2.08	2	0.273	0.923
		Case 2: $\boldsymbol{\alpha} = (-3, 3)$ and $\boldsymbol{\eta} = (-1, 1)$			
100	GMCP	2	2	0	0.978
	GSCAD	2	2	0	0.979
200	GMCP	2	2	0	0.980
	GSCAD	2	2	0	0.984

The true value of K is $K = 2$. SD represents standard deviation; TPR represents rate of individuals selected into the subgroups correctly.

where one is from one simulated dataset and the other is based on the median estimate of 100 replications for each fixed tuning parameter. The plots from one dataset show how the group size and estimates change as the tuning parameter value increases. It is clear that regression coefficients will be estimated as one group for large enough value of the tuning parameter. As a comparison, the estimates in the fusiongram based on 100 replications are more concentrated. This implies that our ridge initial solution can statistically subgroup the regression coefficients to some degree. The fusiongram for GSCAD and the fusiongram for $n = 200$ are similar and so omitted here. Table 2.1 reports the estimates of group size K in Example 1. The means and medians of $\widehat{K}$ under both GMCP and GSCAD selectors are close to the true value. When the difference of treatment effects between two subgroups increases, the true positive rate (TPR) becomes larger and are closer to 1, indicating identification of the subgroup structure more accurate. Table 2.2 further shows the estimates of regression coefficients. We can see that the MEANs and MEDIANs are close to the true values of the parameters, and the standard deviations reduce as the sample size increases. Noting that the logistic-Cox mixture model assumes that the

Table 2.2: Simulation results for estimation of regression coefficients in Example 1.

n	PARAMETER	METHOD	MEAN	MEDIAN	SD
			Case 1: $\boldsymbol{\alpha} = (-1.5, 1.5)$ and $\boldsymbol{\eta} = (-1, 1)$		
100	α	GMCP	$(-1.760, 1.773)$	$(-1.782, 1.789)$	$(0.413, 0.421)$
		GSCAD	$(-1.735, 1.791)$	$(-1.773, 1.778)$	$(0.409, 0.406)$
		MIXTURE	$(-1.545, 1.543)$	$(-1.520, 1.570)$	$(0.500, 0.487)$
		Oracle	$(-1.518, 1.586)$	$(-1.505, 1.594)$	$(0.339, 0.287)$
	$\boldsymbol{\eta}$	GMCP	$(-0.850, 0.844)$	$(-0.869, 0.854)$	$(0.232, 0.236)$
		GSCAD	$(-0.843, 0.841)$	$(-0.859, 0.852)$	$(0.229, 0.234)$
		MIXTURE	$(-1.021, 1.013)$	$(-1.020, 1.026)$	$(0.235, 0.242)$
		Oracle	$(-1.025, 1.027)$	$(-1.015, 1.012)$	$(0.175, 0.168)$
200	α	GMCP	$(-1.704, 1.667)$	$(-1.715, 1.698)$	$(0.299, 0.287)$
		GSCAD	$(-1.750, 1.625)$	$(-1.773, 1.671)$	$(0.307, 0.328)$
		MIXTURE	$(-1.521, 1.550)$	$(-1.516, 1.554)$	$(0.271, 0.256)$
		Oracle	$(-1.532, 1.538)$	$(-1.531, 1.522)$	$(0.215, 0.215)$
	$\boldsymbol{\eta}$	GMCP	$(-0.924, 0.925)$	$(-0.917, 0.910)$	$(0.158, 0.149)$
		GSCAD	$(-0.918, 0.910)$	$(-0.916, 0.903)$	$(0.162, 0.162)$
		MIXTURE	$(-1.032, 1.033)$	$(-1.031, 1.033)$	$(0.140, 0.137)$
		Oracle	$(-1.020, 1.019)$	$(-1.014, 1.013)$	$(0.112, 0.113)$
			Case 2: $\boldsymbol{\alpha} = (-3, 3)$ and $\boldsymbol{\eta} = (-1, 1)$		
100	α	GMCP	$(-2.969, 3.171)$	$(-3.013, 3.175)$	$(0.642, 0.471)$
		GSCAD	$(-2.976, 3.175)$	$(-3.019, 3.175)$	$(0.645, 0.475)$
		MIXTURE	$(-2.846, 2.879)$	$(-2.944, 3.109)$	$(0.896, 1.124)$
		Oracle	$(-3.077, 3.167)$	$(-3.013, 3.137)$	$(0.545, 0.450)$
	$\boldsymbol{\eta}$	GMCP	$(-0.957, 0.965)$	$(-0.932, 0.969)$	$(0.217, 0.215)$
		GSCAD	$(-0.960, 0.968)$	$(-0.936, 0.969)$	$(0.215, 0.218)$
		MIXTURE	$(-0.965, 0.968)$	$(-0.975, 1.004)$	$(0.281, 0.296)$
		Oracle	$(-1.025, 1.028)$	$(-1.015, 1.015)$	$(0.178, 0.172)$
200	α	GMCP	$(-2.815, 2.931)$	$(-2.831, 2.955)$	$(0.493, 0.436)$
		GSCAD	$(-2.856, 2.944)$	$(-2.897, 2.987)$	$(0.487, 0.464)$
		MIXTURE	$(-3.002, 3.013)$	$(-3.008, 3.058)$	$(0.538, 0.572)$
		Oracle	$(-3.077, 3.069)$	$(-3.081, 3.040)$	$(0.339, 0.328)$
	$\boldsymbol{\eta}$	GMCP	$(-0.977, 0.966)$	$(-1.000, 1.006)$	$(0.186, 0.199)$
		GSCAD	$(-0.994, 0.976)$	$(-1.005, 1.006)$	$(0.171, 0.186)$
		MIXTURE	$(-1.012, 1.015)$	$(-1.024, 1.020)$	$(0.158, 0.164)$
		Oracle	$(-1.021, 1.018)$	$(-1.018, 1.009)$	$(0.114, 0.114)$

SD represents standard deviation

parameter $K = 2$ is given and the grouping membership satisfies a logistic model, its parameter space is much smaller than our model. Table 2.2 shows the biases and standard errors of our estimators are comparable to those obtained by fitting the logistic-Cox mixture model.

Table 2.3: Simulation results for estimation of group size K in Example 2.

n	METHOD	MEAN	MEDIAN	SD	TPR
100	GMCP	2.10	2	0.362	0.982
	GSCAD	2.08	2	0.339	0.984
200	GMCP	2.07	2	0.256	0.991
	GSCAD	2.07	2	0.256	0.991

The true value of K is $K = 2$. SD represents standard deviation; TPR represents rate of individuals selected into the subgroups correctly.

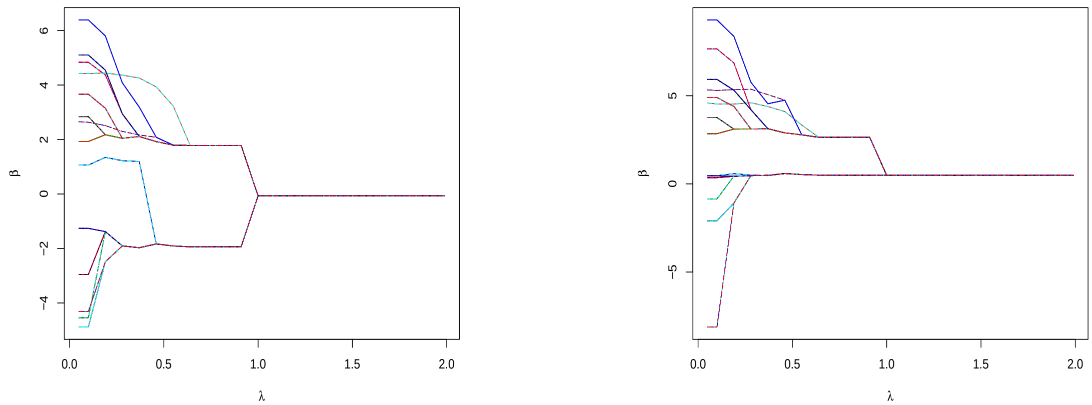
Table 2.4: Simulation results for estimation of regression coefficients in Example 2.

n	PARAMETER	METHOD	MEAN	MEDIAN	SD
			Truth: $\boldsymbol{\alpha}_1 = (-2, 0.5)$, $\boldsymbol{\alpha}_2 = (2, 3)$, $\boldsymbol{\eta} = (-1, 1)$		
100	$\boldsymbol{\alpha}_1$	GMCP	$(-2.062, 0.496)$	$(-1.996, 0.507)$	$(0.558, 0.475)$
		GSCAD	$(-2.059, 0.497)$	$(-1.996, 0.507)$	$(0.555, 0.474)$
		Oracle	$(-2.108, 0.486)$	$(-2.064, 0.529)$	$(0.606, 0.453)$
	$\boldsymbol{\alpha}_2$	GMCP	$(2.110, 2.995)$	$(2.107, 3.004)$	$(0.465, 0.755)$
		GSCAD	$(2.108, 3.012)$	$(2.120, 3.004)$	$(0.463, 0.727)$
		Oracle	$(2.111, 3.244)$	$(2.058, 3.247)$	$(0.351, 0.470)$
	$\boldsymbol{\eta}$	GMCP	$(-0.942, 0.957)$	$(-0.972, 0.978)$	$(0.279, 0.283)$
		GSCAD	$(-0.946, 0.959)$	$(-0.977, 0.978)$	$(0.271, 0.274)$
		Oracle	$(-1.058, 1.060)$	$(-1.053, 1.071)$	$(0.150, 0.173)$
200	$\boldsymbol{\alpha}_1$	GMCP	$(-1.989, 0.485)$	$(-1.976, 0.479)$	$(0.389, 0.329)$
		GSCAD	$(-1.989, 0.485)$	$(-1.976, 0.479)$	$(0.389, 0.329)$
		Oracle	$(-2.108, 0.516)$	$(-2.095, 0.501)$	$(0.413, 0.273)$
	$\boldsymbol{\alpha}_2$	GMCP	$(1.973, 2.929)$	$(1.974, 2.981)$	$(0.250, 0.456)$
		GSCAD	$(1.973, 2.929)$	$(1.974, 2.981)$	$(0.250, 0.456)$
		Oracle	$(2.070, 3.090)$	$(2.065, 3.066)$	$(0.249, 0.275)$
	$\boldsymbol{\eta}$	GMCP	$(-0.984, 1.003)$	$(-0.989, 1.024)$	$(0.163, 0.165)$
		GSCAD	$(-0.984, 1.003)$	$(-0.989, 1.024)$	$(0.163, 0.165)$
		Oracle	$(-1.021, 1.009)$	$(-1.001, 0.996)$	$(0.123, 0.114)$

SD represents standard deviation

Example 2. Suppose $X_i = (X_{i1}, X_{i2})^T$, where X_{i1} and X_{i2} were generated from Bernoulli(0.5) + 1 and Uniform(1, 3), respectively. Set $\beta_i = (-2, 0.5)^T$ for $i \in \mathcal{G}_1$,

(a) Fusiongram based on one dataset



(b) Fusiongram based on 100 replications

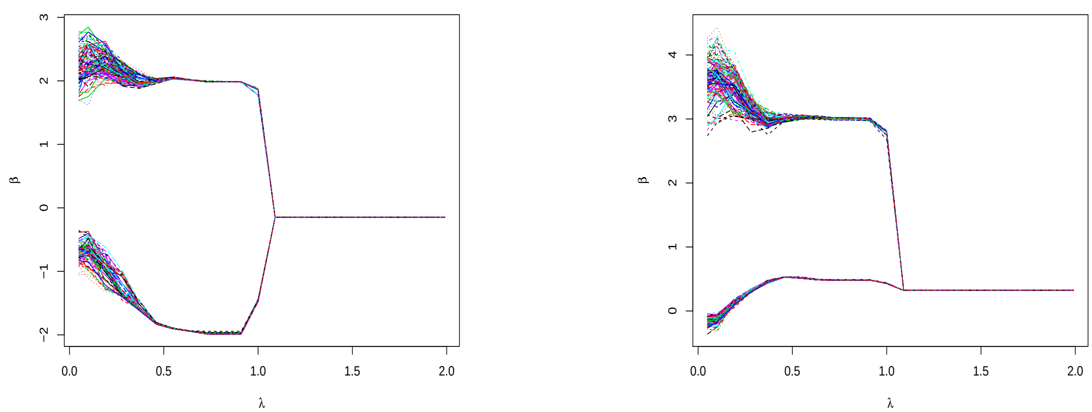


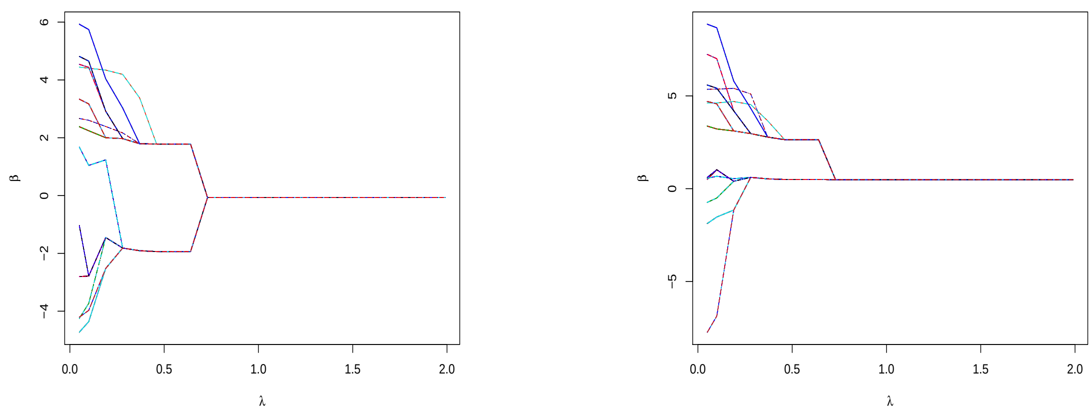
Figure 2.2: Fusiongram for estimation of parameter $\boldsymbol{\beta}$ by GMCP selector in Example 2 when $n = 200$.

and $\beta_i = (2,3)^T$ for $i \in \mathcal{G}_2$, where $\mathcal{G}_1 = \{1,\ldots,2n/5\}$, and $\mathcal{G}_2 = \{2n/5+1,\ldots,n\}$. Thus, $\boldsymbol{\alpha} = (\alpha_1^T, \alpha_2^T)^T$ with $\alpha_1 = (-2, 0.5)^T$ and $\alpha_2 = (2,3)^T$.

Example 3. Suppose that X_i was generated from Bernoulli(0.5) + 1. Set $\mathcal{G}_1 = \{1,\ldots,n/3\}$, $\mathcal{G}_2 = \{n/3+1,\ldots,2n/3\}$, and $\mathcal{G}_3 = \{2n/3+1,\ldots,n\}$. We set $\beta_i = -3$ for $i \in \mathcal{G}_1$, $\beta_i = 0$ for $i \in \mathcal{G}_2$, and $\beta_i = 3$ for $i \in \mathcal{G}_3$. That is $\boldsymbol{\alpha} = (-3,0,3)^T$.

Example 4. Consider the homogeneous model where X_i was generated from

(a) Fusiongram based on one dataset



(b) Fusiongram based on 100 replications

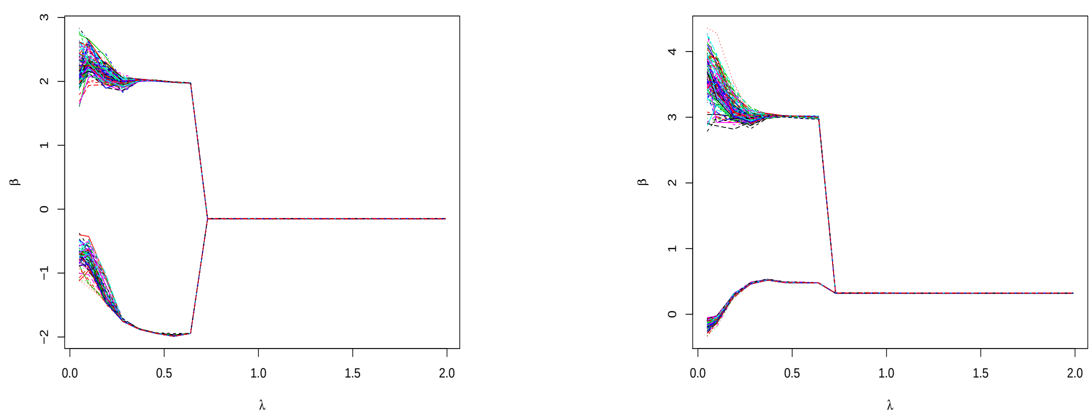


Figure 2.3: Fusiongram for estimation of parameter $\boldsymbol{\beta}$ by GSCAD selector in Example 2 when $n = 200$.

Bernoulli(0.5) + 1, and $\beta_i \equiv 1$ for all i.

The simulation results for Examples 2–4 are summarized in Tables 2.3–2.7 and Figures 2.2–2.5 in the online supplementary material. The figures show the fusiongram for estimation in Examples 2–4, respectively. Tables 2.3, 2.5 and 2.7 display the estimates of group size K and the TPR in Examples 2–4, respectively. The means and medians of $\widehat{K}$ under both GMCP and GSCAD selectors are close to the true

(a) Fusiongram based on one dataset (b) Fusiongram based on 100 replications

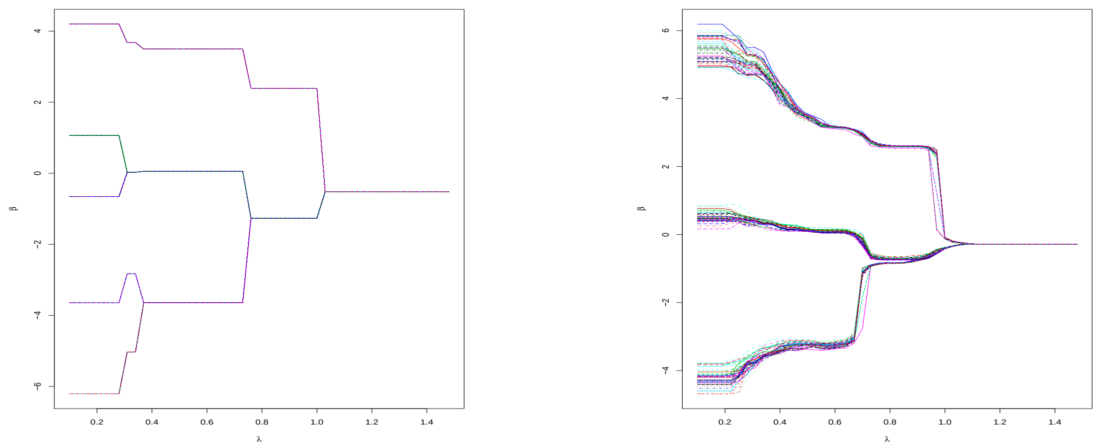


Figure 2.4: Fusiongram for estimation of parameter $\boldsymbol{\beta}$ for GMCP in Example 3 when $n = 150$.

(a) Fusiongram based on one dataset (b) Fusiongram based on 100 replications

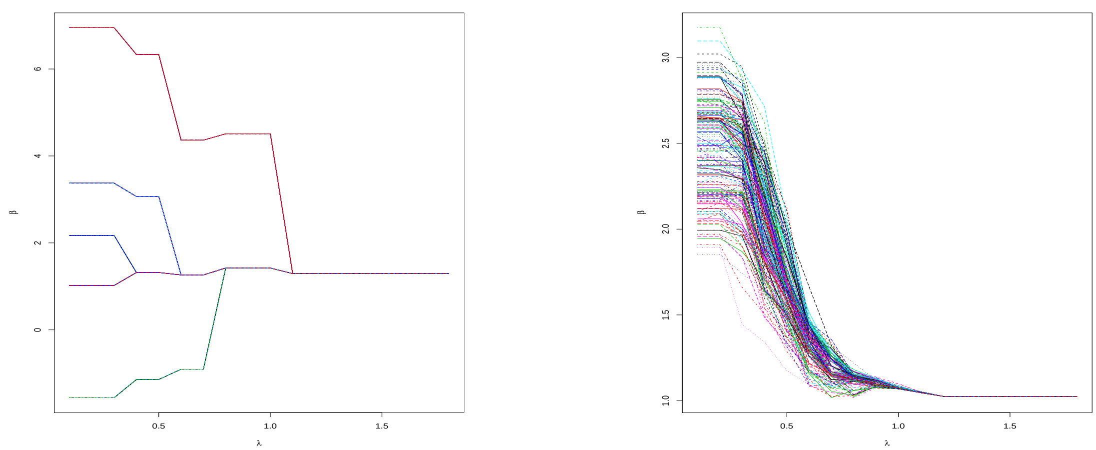


Figure 2.5: Fusiongram for estimation of parameter $\boldsymbol{\beta}$ for GMCP in Example 4 when $n = 200$.

Table 2.5: Simulation results for estimation of group size K in Example 3.

n	METHOD	MEAN	MEDIAN	SD	TPR
150	GMCP	2.99	3	0.225	0.866
	GSCAD	3.03	3	0.264	0.866
300	GMCP	3.01	3	0.100	0.874
	GSCAD	3	3	0	0.876

The true value of K is $K = 3$. SD represents standard deviation; TPR represents rate of individuals selected into the subgroups correctly.

Table 2.6: Simulation results for estimation of regression coefficients in Example 3.

n	PARAMETER	METHOD	MEAN	MEDIAN	SD
			Truth: $\boldsymbol{\alpha} = (-3, 0, 3)$, $\boldsymbol{\eta} = (-1, 1)$		
150	$\boldsymbol{\alpha}_1$	GMCP	-3.422	-3.453	0.556
		GSCAD	-3.429	-3.387	0.553
		Oracle	-3.087	-3.043	0.429
	$\boldsymbol{\alpha}_2$	GMCP	0.062	0.074	0.393
		GSCAD	0.033	0.041	0.420
		Oracle	-0.007	0.001	0.233
	$\boldsymbol{\alpha}_3$	GMCP	3.309	3.219	0.552
		GSCAD	3.280	3.204	0.578
		Oracle	3.076	3.040	0.364
	$\boldsymbol{\eta}$	GMCP	$(-0.747, 0.727)$	$(-0.777, 0.720)$	$(0.228, 0.224)$
		GSCAD	$(-0.716, 0.713)$	$(-0.714, 0.716)$	$(0.234, 0.233)$
		Oracle	$(-1.033, 1.024)$	$(-1.029, 1.016)$	$(0.129, 0.133)$
300	$\boldsymbol{\alpha}_1$	GMCP	-3.283	-3.289	0.388
		GSCAD	-3.288	-3.283	0.411
		Oracle	-3.065	-3.031	0.298
	$\boldsymbol{\alpha}_2$	GMCP	-0.068	-0.067	0.314
		GSCAD	-0.068	-0.072	0.298
		Oracle	-0.007	-0.007	0.152
	$\boldsymbol{\alpha}_3$	GMCP	3.088	3.016	0.525
		GSCAD	3.133	3.158	0.523
		Oracle	3.033	3.038	0.238
	$\boldsymbol{\eta}$	GMCP	$(-0.775, 0.788)$	$(-0.797, 0.816)$	$(0.182, 0.186)$
		GSCAD	$(-0.785, 0.790)$	$(-0.795, 0.805)$	$(0.183, 0.185)$
		Oracle	$(-1.012, 1.017)$	$(-1.009, 1.013)$	$(0.099, 0.094)$

SD represents standard deviation.

Table 2.7: Simulation results for estimation of K and regression coefficients in Example 4.

n	PARAMETER	METHOD	MEAN	MEDIAN	SD
100	K	GMCP	1.08	1	0.273
		GSCAD	1.09	1	0.288
		Oracle	–	–	–
	$\boldsymbol{\alpha}$	GMCP	1.022	0.999	0.221
		GSCAD	1.017	0.995	0.221
		Oracle	1.042	1.026	0.245
	$\boldsymbol{\eta}$	GMCP	$(-1.004, 0.999)$	$(-0.995, 0.991)$	$(0.178, 0.171)$
		GSCAD	$(-1.000, 0.997)$	$(-0.990, 0.988)$	$(0.180, 0.170)$
		Oracle	$(-1.026, 1.026)$	$(-1.017, 1.008)$	$(0.170, 0.170)$
200	K	GMCP	1.04	1	0.197
		GSCAD	1.01	1	0.100
		Oracle	–	–	–
	$\boldsymbol{\alpha}$	GMCP	1.022	1.021	0.184
		GSCAD	1.024	1.023	0.183
		Oracle	1.020	1.019	0.172
	$\boldsymbol{\eta}$	GMCP	$(-1.029, 1.030)$	$(-1.039, 1.027)$	$(0.114, 0.118)$
		GSCAD	$(-1.029, 1.030)$	$(-1.032, 1.024)$	$(0.113, 0.117)$
		Oracle	$(-1.019, 1.018)$	$(-1.015, 1.008)$	$(0.109, 0.111)$

SD represents standard deviation.

value, and the TPR are close to 1, which reflect that our methods can identify the group structure correctly with high probability. As the sample size increases, the standard deviation of $\widehat{K}$ decreases and the TPR increases, which demonstrate the good performances of our approaches. Furthermore, Tables 2.4, 2.6 and 2.7 report the estimates of the regression coefficients. The MEAN and MEDIAN of estimators are very close to the true value, and standard deviation (SD) for parameters reduce as the sample size increases.

2.6 Real Data Analysis

We applied the proposed method to analyzing the breast cancer data (van de Vijver et al., 2002; van't Veer el al., 2002), which can be found in the "nki" data set in the R package "dynpred". This trail was carried out in the Dutch Cancer Institute, where 295 patients with breast cancer were put into two treatment groups by the type of surgery (excision and mastectomy), some of them accompanying with two kinds of adjuvant therapies, chemotherapy or hormonal therapy. The main goal is to investigate effects of different surgical treatments on patients' hazard. Hence we focused on the observed data from 255 patients who were not treated with the hormonal therapy for the analysis. Let U_i and C_i be survival and censoring times for the ith patient, $i = 1, \ldots, n$ where $n = 255$. Let X denote the treatment group indicator defined as 1 for patients treated with excision and 0 for patients treated with mastectomy. According to the iterative sure independence screening result (Fan and Lv, 2008), we took 5 additional baseline covariates $Z_1, \cdots, Z_5$ into consideration, including age (age), the logarithmic intensity ratio for estrogen-receptor status ($mlratio$), histological grade ($histolgrade = 1$ if well differentiated; 0 otherwise), vascular invasion ($vasc.inv = 1$ for more than 3 vessels; 0 otherwise), and the cross-validated version of the prognostic index ($PICV$). All the continuous covariates were standardized for convenience.

To check for the possible heterogeneity of treatment effects, we first fitted the homogeneous Cox model based on the excision treatment group. Figure 2.6 displays the plot of the kernel density estimate of the martingale residual. We observed that the distribution has multiple modes, indicating the existence of heterogeneous treatment effects.

To demonstrate the heterogeneity of treatment effects, we fitted the proposed heterogeneous Cox model in (2.2) using our subgroup analysis procedure with group MCP and group SCAD penalties, where the optimal tuning parameter was determined by the modified BIC criterion. Figure 2.7 displays the fusiongram for the estimate of $\boldsymbol{\beta}$. The grouping and parameter estimation results with GMCP are summarized in Table 2.8, while the results with GSCAD are similar and so are omitted.

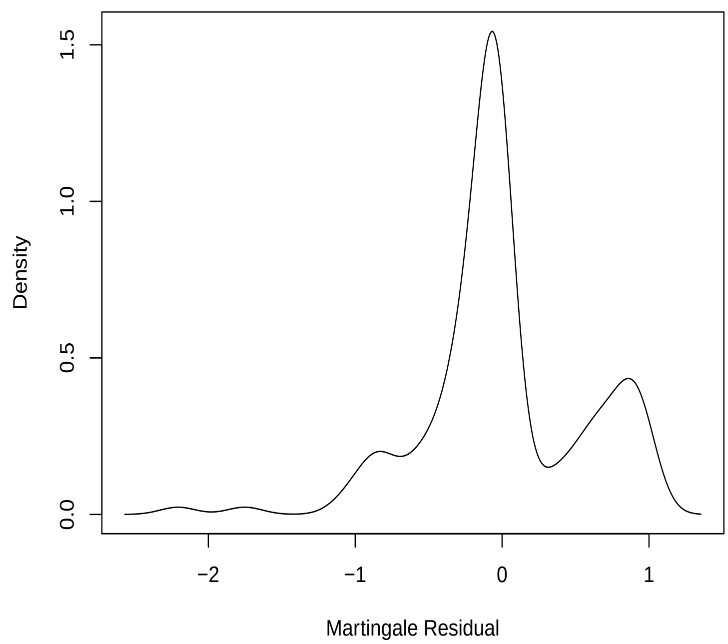


Figure 2.6: The kernel density plot of the residuals after controlling for the effects of the 5 covariates for the patients treated with excision in the Breast Cancer data.

For comparison, we also provide the estimation results by fitting both the homogeneous Cox model and the logistic-Cox mixture model in the table. It can be seen from the table that the fitted homogenous Cox model could not detect any significant treatment effect, while both the logistic-Cox mixture approach and the proposed subgroup analysis approach identified the significant subgroup-specific treatment effects.

Furthermore, we present the grouping result in Table 2.9 according to the type of surgery. It can be seen from the table that our subgroup analysis approach identifies 90% of the patients with the excision and 4% of the patients with the mastectomy as one subgroup and 96% of the patients with the mastectomy and 10% of the patients with excision as another subgroup. For the patients in subgroup 1, the excision

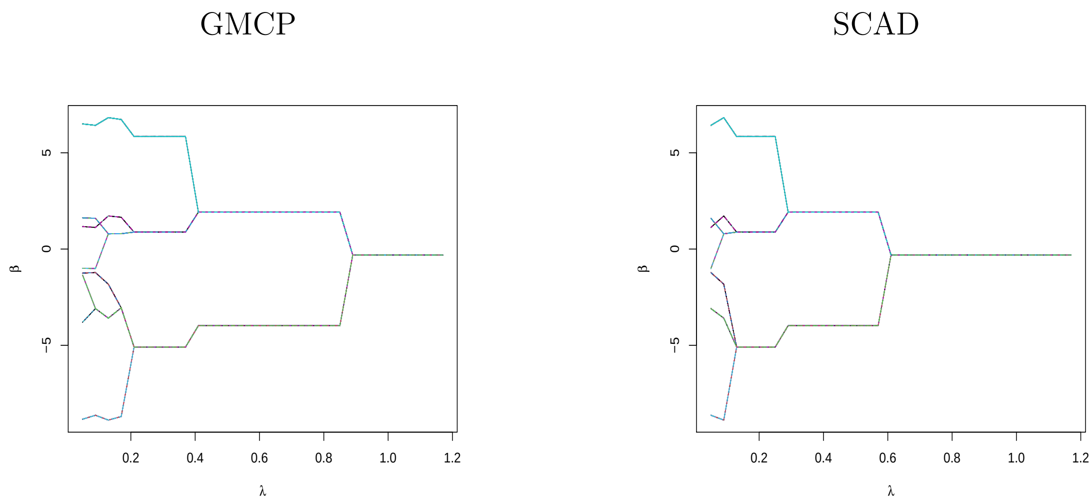


Figure 2.7: Fusiongram for estimation of parameter $\boldsymbol{\beta}$ in Breast Cancer Data analysis.

Table 2.8: Analysis results for Breast Cancer Data.

	PL		MIXTURE		GMCP	
Covariate	Estimate(ESE)	$p-$value	Estimate(ESE)	$p-$value	Estimate(ESE)	$p-$value
$typesurgery1$	$-0.311(0.244)$	.203	$-1.571(0.409)$	$<.001^*$	$-3.981(0.575)$	$<.001^*$
$typesurgery2$	$-(-)$	$-$	$1.374(0.425)$	$.001^*$	$1.917(0.343)$	$<.001^*$
age	$-0.323(0.110)$	$.003^*$	$-0.058(0.146)$	.689	$-0.320(0.109)$	$.003^*$
$mlratio$	$-0.285(0.152)$	.060	$-0.347(0.179)$	.053	$-0.420(0.155)$	$.006^*$
$histolgrade$	$-1.110(0.542)$	$.041^*$	$-1.004(0.587)$	.087	$-1.289(0.551)$	$.019^*$
$vasc.inv$	$0.642(0.250)$	$.010^*$	$0.046(0.324)$	.889	$1.081(0.274)$	$<.001^*$
$PICV$	$0.421(0.165)$	$.011^*$	$0.534(0.166)$	$.001^*$	$0.505(0.171)$	$.003^*$

PL represents partial likelihood estimators; MIXTURE represents the logistic-Cox mixture estimators; $typesurgery1$, $typesurgery2$ represent the different subgroup variables of $typesurgery$; * represents significant at 0.05 level.

can reduce the hazard and prolong the lifetime significantly; while for the patients in subgroup 2, the mastectomy is better than the excision. The subgroup analysis approach (Wu, Zheng and Yu, 2016) provides the estimates of the probabilities that patients belong to each subgroup under the logistic model.

The key difference between our approach and the subgroup analysis approach (Wu, Zheng and Yu, 2016) is that the number of the potential subgroups K and the grouping structure are left completely unspecified in our proposed model, while Wu, Zheng, and Yu (2016) assumed that $K = 2$ and the subgroup membership satisfies a logistic model. Our subgroup analysis method is more flexible and applicable.

Table 2.9: The number of patients with different type of surgery and subgroups.

	Subgroup 1	Subgroup 2	Total
Excision	128	15	143
Mastectomy	5	107	112
Total	133	122	255

2.7 Appendix: Proofs of Theorems

To establish the asymptotic properties of the proposed estimator, we need the following regularity conditions.

(C1) The end time of study τ satisfies that $\int_0^\tau \lambda_0(t)dt < \infty$.

(C2) The covariates X_i and Z_i satisfy that $||X_i|| \leq c_1$ and $||Z_i|| \leq c_2$ with probability 1.

(C3) The dimension of covariates p, q and the true cluster size K_0 are constants. The sizes of $\mathcal{G}_{0,k}$ satisfy that $|\mathcal{G}_{0,k}|/n \to p_k$ for $k = 1, \cdots, K_0$ when n goes to infinity.

(C4) Set the penalty function $\rho_\gamma(t) = \lambda^{-1} p_\gamma(t, \lambda)$. Suppose that $\rho_\gamma(t)$ is symmetric, non-decreasing and concave on $[0, \infty)$. $\rho_\gamma(t)$ is constant when $t \geq a\lambda$, where a is a positive constant. Furthermore, $\rho_\gamma(0) = 0$ and the derivative $\rho'_\gamma(t)$ satisfies that $\rho'_\gamma(0^+) = 1$.

We introduce more notation before proving the theorems.

Let $S^{(l)}(\boldsymbol{\theta}, \mathbf{B}, t) = n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{B}_i^{\otimes l} \exp(\mathbf{B}_i^T \boldsymbol{\theta})$, where $\mathbf{a}^{\otimes l} = 1, \mathbf{a}, \mathbf{a}\mathbf{a}^T$ for $l =$

0, 1, 2. Define the score function

$$\widetilde{U}_n(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \int_0^{\tau} \left[\mathbf{B}_i - \frac{S^{(1)}(\boldsymbol{\theta}, \mathbf{B}, t)}{S^{(0)}(\boldsymbol{\theta}, \mathbf{B}, t)} \right] dN_i(t),$$

and the Hessian matrix

$$\widetilde{H}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\tau} \left[\frac{S^{(2)}(\boldsymbol{\theta}, \mathbf{B}, t)}{S^{(0)}(\boldsymbol{\theta}, \mathbf{B}, t)} - \left\{ \frac{S^{(1)}(\boldsymbol{\theta}, \mathbf{B}, t)}{S^{(0)}(\boldsymbol{\theta}, \mathbf{B}, t)} \right\}^{\otimes 2} \right] dN_i(t).$$

Let $S^{(k,l)}(\boldsymbol{\theta}, \mathbf{B}, t) = \frac{1}{|\mathcal{G}_{0,k}|} \sum_{i \in \mathcal{G}_{0,k}} Y_i(t) \mathbf{B}_i^{\otimes l} \exp(\mathbf{B}_i^T \boldsymbol{\theta})$, where $l = 0, 1, 2$ and $k = 1, \cdots, K_0$. Then we have

$$S^{(l)}(\boldsymbol{\theta}, \mathbf{B}, t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) \mathbf{B}_i^{\otimes l} \exp(\mathbf{B}_i^T \boldsymbol{\theta}) = \sum_{k=1}^{K_0} \frac{|\mathcal{G}_{0,k}|}{n} S^{(k,l)}(\boldsymbol{\theta}, \mathbf{B}, t).$$

Note that $\mathbf{B}_i$, $i \in \mathcal{G}_{0,k}$ are independent and identically distributed random vectors. Denote the expectation of $S^{(k,l)}(\boldsymbol{\theta}, \mathbf{B}, t)$ by $s^{(k,l)}(\boldsymbol{\theta}, t)$, and $s^{(l)}(\boldsymbol{\theta}, t) = \sum_{k=1}^{K} p_k s^{(k,l)}(\boldsymbol{\theta}, t)$, where $|\mathcal{G}_{0,k}|/n \to p_k$ when $n \to \infty$. Then we have

$$\sup_{t \in [0,\tau]} |S^{(k,l)}(\boldsymbol{\theta}, \mathbf{B}, t) - s^{(k,l)}(\boldsymbol{\theta}, t)|_\infty \xrightarrow{p} 0,$$

and $\sup_{t \in [0,\tau]} |S^{(l)}(\boldsymbol{\theta}, \mathbf{B}, t) - s^{(l)}(\boldsymbol{\theta}, t)|_\infty \xrightarrow{p} 0$, where $|\cdot|_\infty$ denotes the maximum norm. Define

$$\Sigma(\boldsymbol{\theta}_0) = \int_0^{\tau} \left\{ \frac{s^{(2)}(\boldsymbol{\theta}_0, t)}{s^{(0)}(\boldsymbol{\theta}_0, t)} - \left(\frac{s^{(1)}(\boldsymbol{\theta}_0, t)}{s^{(0)}(\boldsymbol{\theta}_0, t)} \right)^{\otimes 2} \right\} s^{(0)}(\boldsymbol{\theta}_0, t) \lambda_0(t) dt.$$

2.7.1 Proof of Theorem 2.1

(i) The proof of the first part is based on the techniques for the consistency of the M-estimator. Note that

$$\frac{1}{n}\left(\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\boldsymbol{\theta}_0)\right) = -\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{B}_i^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \log \frac{S^{(0)}(\boldsymbol{\theta}, \boldsymbol{B}, t)}{S^{(0)}(\boldsymbol{\theta}_0, \boldsymbol{B}, t)} \right] dN_i(t).$$

Define

$$A_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{B}_i^T(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \log\frac{S^{(0)}(\boldsymbol{\theta},\boldsymbol{B},t)}{S^{(0)}(\boldsymbol{\theta}_0,\boldsymbol{B},t)}\right] Y_i(t)\exp(\boldsymbol{B}_i^T\boldsymbol{\theta}_0)\lambda_0(t)dt$$

$$= -\int_0^{\tau}\left[S^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{B},t)^T(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \log\left\{\frac{S^{(0)}(\boldsymbol{\theta},\boldsymbol{B},t)}{S^{(0)}(\boldsymbol{\theta}_0,\boldsymbol{B},t)}\right\} S^{(0)}(\boldsymbol{\theta}_0,\boldsymbol{B},t)\right]\lambda_0(t)dt$$

as the compensator of $\frac{1}{n}\big(\tilde{\ell}_n(\boldsymbol{\theta})-\tilde{\ell}_n(\boldsymbol{\theta}_0)\big)$, and $M_i(t) = N_i(t)-\int_0^t Y_i(u)\exp(\boldsymbol{B}_i^T\boldsymbol{\theta}_0)\lambda_0(u)du$. Since $M_i(t)$ is a locally square integrable martingale, then

$$\frac{1}{n}\big(\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\boldsymbol{\theta}_0)\big) - A_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\Big[\boldsymbol{B}_i^T(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \log\frac{S^{(0)}(\boldsymbol{\theta},\boldsymbol{B},t)}{S^{(0)}(\boldsymbol{\theta}_0,\boldsymbol{B},t)}\Big]dM_i(t)$$

is also a locally square integrable martingale. Hence $\frac{1}{n}\big(\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\boldsymbol{\theta}_0)\big) - A_n(\boldsymbol{\theta})$ has a predictable variation process

$$\Big\langle\frac{1}{n}\big(\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\boldsymbol{\theta}_0)\big) - A_n(\boldsymbol{\theta}), \frac{1}{n}\big(\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\boldsymbol{\theta}_0)\big) - A_n(\boldsymbol{\theta})\Big\rangle$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\int_0^{\tau}\left[\left\{\boldsymbol{B}_i^T(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \log\frac{\sum_{i=1}^{n}Y_i(t)\exp(\boldsymbol{B}_i^T\boldsymbol{\theta})}{\sum_{i=1}^{n}Y_i(t)\exp(\boldsymbol{B}_i^T\boldsymbol{\theta})}\right\}^2 Y_i(t)\exp(\boldsymbol{B}_i^T\boldsymbol{\theta}_0)\lambda_0(t)\right]dt$$

$$=\frac{1}{n}\int_0^{\tau}\left[(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T S^{(2)}(\boldsymbol{\theta},\boldsymbol{B},t)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - 2(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T S^{(1)}(\boldsymbol{\theta},\boldsymbol{B},t)\log\frac{S^{(0)}(\boldsymbol{\theta},\boldsymbol{B},t)}{S^{(0)}(\boldsymbol{\theta}_0,\boldsymbol{B},t)}\right.$$

$$\left.+\left\{\log\frac{S^{(0)}(\boldsymbol{\theta},\boldsymbol{B},t)}{S^{(0)}(\boldsymbol{\theta}_0,\boldsymbol{B},t)}\right\}^2\right]\lambda_0(t)dt.$$

By Conditions (C2) and (C3), for any k and l, $s^{(k,l)}(\boldsymbol{\theta},t)$ and $s^{(l)}(\boldsymbol{\theta},t)$ are bounded. Then, by Condition (C1), the predictable variation process has a finite limit. This gives that $\lim_{n\to\infty}\frac{1}{n}\big(\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\boldsymbol{\theta}_0)\big) = A(\boldsymbol{\theta})$, where

$$A(\boldsymbol{\theta}) = \lim_{n\to\infty}A_n(\boldsymbol{\theta}) = -\int_0^{\tau}\left[s^{(1)}(\boldsymbol{\theta}_0,t)^T(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \log\left\{\frac{s^{(0)}(\boldsymbol{\theta},t)}{s^{(0)}(\boldsymbol{\theta}_0,t)}\right\}s^{(0)}(\boldsymbol{\theta}_0,t)\right]\lambda_0(t)dt.$$

Noting that $\widehat{\boldsymbol{\theta}}^{or}$ is the global minimizer of $\tilde{\ell}_n(\boldsymbol{\theta})$, it is also the global minimizer of $\frac{1}{n}\big(\tilde{\ell}_n(\boldsymbol{\theta})-\tilde{\ell}_n(\boldsymbol{\theta}_0)\big)$. Since $A(\boldsymbol{\theta})$ is a convex function about $\boldsymbol{\theta}$ and has a global minimizer $\boldsymbol{\theta}_0$, it follows that $\widehat{\boldsymbol{\theta}}^{or} \xrightarrow{p} \boldsymbol{\theta}_0$.

(ii) To prove this part, it suffices to show that $\frac{1}{\sqrt{n}}\widetilde{U}_n(\boldsymbol{\theta}_0)$ converges to a zero mean multivariate normal distribution with covariance matrix $\Sigma(\boldsymbol{\theta}_0)$, and $|\frac{1}{n}\widetilde{H}_n(\widehat{\boldsymbol{\theta}}^{or}) - \Sigma(\boldsymbol{\theta}_0)|_\infty \xrightarrow{p} 0$. For this, we only need to verify the conditions of Theorem 8.2.1 of Fleming and Harrington (1991). Recall that

$$\sup_{0\le t\le\tau} |S^{(l)}(\boldsymbol{\theta}_0,\mathbf{B},t) - s^{(l)}(\boldsymbol{\theta}_0,t)|_\infty \xrightarrow{p} 0.$$

Noting that $\frac{\partial}{\partial\boldsymbol{\theta}}S^{(k,0)}(\boldsymbol{\theta},\mathbf{B},t) = S^{(k,1)}(\boldsymbol{\theta},\mathbf{B},t)$ and $\frac{\partial}{\partial\boldsymbol{\theta}}S^{(k,1)}(\boldsymbol{\theta},\mathbf{B},t) = S^{(k,2)}(\boldsymbol{\theta},\mathbf{B},t)$, we have $\frac{\partial}{\partial\boldsymbol{\theta}}s^{(k,0)}(\boldsymbol{\theta},t) = s^{(k,1)}(\boldsymbol{\theta},t)$ and $\frac{\partial}{\partial\boldsymbol{\theta}}s^{(k,1)}(\boldsymbol{\theta},t) = s^{(k,2)}(\boldsymbol{\theta},t)$, $k=1,\cdots,K$. Since $s^{(l)}(\boldsymbol{\theta},t)$ is a linear combination of $s^{(k,l)}(\boldsymbol{\theta},t)$, it follows that $\frac{\partial}{\partial\boldsymbol{\theta}}s^{(0)}(\boldsymbol{\theta},t) = s^{(1)}(\boldsymbol{\theta},t)$ and $\frac{\partial}{\partial\boldsymbol{\theta}}s^{(1)}(\boldsymbol{\theta},t) = s^{(2)}(\boldsymbol{\theta},t)$. By Condition (C2), $s^{(l)}(\boldsymbol{\theta},t)$ is bounded. In addition, as the composition of continuous functions is continuous, we then get that $s^{(l)}(\boldsymbol{\theta}_0,t)$, $0<t<\tau$ are equicontinuous for $l=0,1,2$.

Condition (C2) gives that $||\mathbf{B}_i|| \le \sqrt{c_1^2+c_2^2}$ with probability 1. Noting that Y_i is a decreasing counting process from 1 to 0, and $\mathbf{B}_i^T\boldsymbol{\theta}_0 > -||\mathbf{B}_i||\cdot||\boldsymbol{\theta}_0||$, we have

$$n^{-1/2}\sup_{1\le i\le n, 0\le t\le\tau} ||\mathbf{B}_i||Y_i(t)1_{\{\mathbf{B}_i^T\boldsymbol{\theta}_0 > -||\mathbf{B}_i||\cdot||\boldsymbol{\theta}_0||\}} \xrightarrow{p} 0.$$

Finally, the convexity of negative partial log-likelihood ensures that $\frac{1}{n}\widetilde{H}_n(\boldsymbol{\theta}_0)$ is positive definite and so its limit is

$$\Sigma(\boldsymbol{\theta}_0) = \int_0^\tau \left\{ \frac{s^{(2)}(\boldsymbol{\theta}_0,t)}{s^{(0)}(\boldsymbol{\theta}_0,t)} - \left(\frac{s^{(1)}(\boldsymbol{\theta}_0,t)}{s^{(0)}(\boldsymbol{\theta}_0,t)}\right)^{\otimes 2} \right\} s^{(0)}(\boldsymbol{\theta}_0,t)\lambda_0(t)dt.$$

By Theorem 8.2.1 in Fleming and Harrington(1991), we conclude the asymptotic normality of $\frac{1}{\sqrt{n}}\widetilde{U}_n(\boldsymbol{\theta}_0)$ and $|\frac{1}{n}\widetilde{H}_n(\widehat{\boldsymbol{\theta}}^{or}) - \Sigma(\boldsymbol{\theta}_0)|_\infty \xrightarrow{p} 0$.

By the Taylor's expansion, we get that $\widetilde{U}_n(\widehat{\boldsymbol{\theta}}^{or}) = \widetilde{U}_n(\boldsymbol{\theta}_0) - \widetilde{H}_n(\widetilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}^{or} - \boldsymbol{\theta}_0)$, where $\widetilde{\boldsymbol{\theta}}$ is a vector between $\widehat{\boldsymbol{\theta}}^{or}$ and $\boldsymbol{\theta}_0$. Noting that $\widetilde{U}_n(\widehat{\boldsymbol{\theta}}^{or}) = 0$, we have

$$\frac{1}{n}\widetilde{H}_n(\widetilde{\boldsymbol{\theta}})\sqrt{n}(\widehat{\boldsymbol{\theta}}^{or} - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\widetilde{U}_n(\boldsymbol{\theta}_0).$$

Using the fact that both $\frac{1}{n}\widetilde{H}_n(\widehat{\boldsymbol{\theta}}^{or})$ and $\frac{1}{n}\widetilde{H}_n(\boldsymbol{\theta}_0)$ converge to $\Sigma(\boldsymbol{\theta}_0)$ in probability, $\widetilde{H}_n(\widetilde{\boldsymbol{\theta}})$ also converges to $\Sigma(\boldsymbol{\theta}_0)$ in probability. Besides, as $\frac{1}{\sqrt{n}}\widetilde{U}_n(\boldsymbol{\theta}_0)$ converges to a zero mean normal distribution with covariance matrix $\Sigma(\boldsymbol{\theta}_0)$, we conclude that $\sqrt{n}(\widehat{\boldsymbol{\theta}}^{or} - \boldsymbol{\theta}_0)$ converges to a normal distribution with zero mean and covariance matrix $\Sigma^{-1}(\boldsymbol{\theta}_0)$.

2.7.2 Proof of Theorem 2.2

Define the mapping $T^* : R^{np} \to R^{K_0 p}$ as

$$T^*(\boldsymbol{\beta}) = \{|\mathcal{G}_{0,k}|^{-1} \sum_{i \in \mathcal{G}_{0,k}} \beta_i^T, k = 1, \cdots, K_0\}^T,$$

and let the one-to-one mapping $T : \mathcal{M}_{\mathcal{G}_0} \to R^{K_0 p}$ satisfying $T(\boldsymbol{\beta}) = T^*(\boldsymbol{\beta})$. For any vector $\boldsymbol{\beta} \in R^{np}$, set $\boldsymbol{\alpha} = T^*(\boldsymbol{\beta})$ and $\boldsymbol{\beta}^* = T^{-1}(T^*(\boldsymbol{\beta})) = T^{-1}(\boldsymbol{\alpha})$. Noting that for any vector $\eta \in R^q$ and $\boldsymbol{\beta}^* \in \mathcal{M}_{\mathcal{G}_0}$, we have $\ell_n(\eta, \boldsymbol{\beta}^*) = \tilde{\ell}_n((\eta^T, \boldsymbol{\alpha}^T)^T)$. Hence, $\widehat{\boldsymbol{\theta}}^{or}$ defined in Theorem 2.1 equals to $\left((\hat{\eta}^{or})^T, T(\widehat{\boldsymbol{\beta}}^{or})^T\right)^T$. Consider the neighbourhood of $(\eta_0, \boldsymbol{\beta}_0)$, i.e.,

$$\Theta = \{\eta \in R^q, \boldsymbol{\beta} \in R^{np} : ||\eta - \eta_0|| \leq \phi_n, \max_i ||\beta_i - \beta_{0i}|| \leq \phi_n\},$$

where $\phi_n \to 0$ as n goes to infinity. To conclude the theorem, it suffices to verify the following two steps.

(i) For any $(\eta^T, \boldsymbol{\beta}^T)^T \in \Theta$, if $(\eta^T, (\boldsymbol{\beta}^*)^T)^T \neq ((\hat{\eta}^{or})^T, (\widehat{\boldsymbol{\beta}}^{or})^T)^T$, then $Q_n(\eta, \boldsymbol{\beta}^*) > Q_n(\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})$.

(ii) For any $(\eta^T, \boldsymbol{\beta}^T)^T \in \Theta$ and large enough n, $Q_n(\eta, \boldsymbol{\beta}) \geq Q_n(\eta, \boldsymbol{\beta}^*)$.

In fact, by Theorem 2.1, we have $P\big((\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or}) \in \Theta\big) \to 1$. If (i) and (ii) hold, for any $(\eta^T, \boldsymbol{\beta}^T)^T \in \Theta$ satisfying $(\eta^T, (\boldsymbol{\beta}^*)^T)^T \neq ((\hat{\eta}^{or})^T, (\widehat{\boldsymbol{\beta}}^{or})^T)^T$ and large enough n, we have $Q_n(\eta, \boldsymbol{\beta}) > Q_n(\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})$. That means that there is a local minimizer of $Q_n(\eta, \boldsymbol{\beta}; \lambda)$ satisfying that $(\hat{\eta}(\lambda), \widehat{\boldsymbol{\beta}}(\lambda)) = (\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})$ with probability tend to 1.

For (i), since $\ell_n(\eta, \boldsymbol{\beta}^*) = \tilde{\ell}_n\big((\eta^T, \boldsymbol{\alpha}^T)^T\big) > \tilde{\ell}_n\big(((\hat{\eta}^{or})^T, (\widehat{\boldsymbol{\alpha}}^{or})^T)^T\big) = \ell_n(\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})$, we only need to consider the penalty function $P_n(\beta) = \lambda \sum_{i<j} \rho_\gamma(||\beta_i - \beta_j||)$. Note that $\beta_i^* = \beta_j^*$ when subjects i and j are from the same group. Thus,

$$P_n(\beta^*) = \lambda \sum_{i<j, i \in \mathcal{G}_{0,k}, j \in \mathcal{G}_{0,k'}} \rho_\gamma(||\beta_i^* - \beta_j^*||) = \lambda \sum_{k \neq k'} \frac{|\mathcal{G}_{0,k}||\mathcal{G}_{0,k'}|}{2} \rho_\gamma(||\alpha_k - \alpha_{k'}||).$$

For any $(\eta^T, \boldsymbol{\beta}^T)^T \in \Theta$, we have $\max_i ||\beta_i - \beta_{0i}|| \leq \phi_n$. Then for any $k \neq k'$,

$$\begin{aligned}
&||\alpha_k - \alpha_{k'}|| \\
\geq &||\alpha_{0k} - \alpha_{0k'}|| - ||\alpha_k - \alpha_{0k}|| - ||\alpha_{0k'} - \alpha_{k'}|| \geq ||\alpha_{0k} - \alpha_{0k'}|| - 2 \max_k ||\alpha_k - \alpha_{0k}|| \\
\geq &b - 2 \max_k \left|\left| |\mathcal{G}_{0,k}|^{-1} \sum_{i \in \mathcal{G}_{0,k}} (\beta_i - \beta_{0i}) \right|\right| \geq b - 2|\mathcal{G}_{0,k}|^{-1} \max_k \sum_{i \in \mathcal{G}_{0,k}} ||(\beta_i - \beta_{0i})|| \\
\geq &b - 2 \max_i ||\beta_i - \beta_{0i}|| \geq b - 2\phi_n > a\lambda.
\end{aligned} \tag{2.17}$$

The last inequality follows since $b > a\lambda$ and $b \gg \phi_n$. By Condition (C4), $\rho_\gamma(||\alpha_k - \alpha_{k'}||)$ is a constant, and $P_n(\boldsymbol{\beta}^*)$ is only dependent on sample size n for any $(\eta^T, \boldsymbol{\beta}^T)^T \in \Theta$, which can be denoted as C_n. By the fact that $(\hat{\eta}^{or}, \widehat{\boldsymbol{\alpha}}^{or})$ is the unique global minimizer of $\tilde{\ell}_n(\eta, \boldsymbol{\alpha})$, we get

$$Q_n(\eta, \boldsymbol{\beta}^*) = \ell_n(\eta, \boldsymbol{\beta}^*) + C_n > \ell_n(\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or}) + C_n = Q_n(\hat{\eta}^{or}, \widehat{\boldsymbol{\beta}}^{or})$$

when $(\eta^T, (\boldsymbol{\beta}^*)^T)^T \neq ((\hat{\eta}^{or})^T, (\widehat{\boldsymbol{\beta}}^{or})^T)^T$. Thus, (i) is concluded.

For (ii), by the Taylor's expansion, we have

$$Q_n(\eta,\boldsymbol{\beta}) - Q_n(\eta,\boldsymbol{\beta}^*) = \frac{\partial \ell_n(\eta,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) + \frac{\partial P_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) =: \Gamma_1 + \Gamma_2,$$

where $\widetilde{\boldsymbol{\beta}}$ is a vector between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$.

We first consider the second term Γ_2. Note that $P_n(\boldsymbol{\beta}) = \lambda \sum_{i<j, i\in\mathcal{G}_{0,k}, j\in\mathcal{G}_{0,k'}} \rho_\gamma(||\beta_i - \beta_j||)$. Then

$$\begin{aligned}
\Gamma_2 = & \frac{\partial P_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \\
=&\lambda \sum_{n\geq j>i\geq 1} \rho'_\gamma(||\tilde{\beta}_i - \tilde{\beta}_j||)\frac{(\tilde{\beta}_i-\tilde{\beta}_j)^T}{||\tilde{\beta}_i-\tilde{\beta}_j||}(\beta_i-\beta_i^*) \\
&+\lambda \sum_{1\leq j<i\leq n} \rho'_\gamma(||\tilde{\beta}_j - \tilde{\beta}_i||)\frac{-(\tilde{\beta}_j-\tilde{\beta}_i)^T}{||\tilde{\beta}_j-\tilde{\beta}_i||}(\beta_i-\beta_i^*) \\
=&\lambda \sum_{1\leq i<j\leq n} \rho'_\gamma(||\tilde{\beta}_i - \tilde{\beta}_j||)\frac{(\tilde{\beta}_i-\tilde{\beta}_j)^T}{||\tilde{\beta}_i-\tilde{\beta}_j||}\big\{(\beta_i-\beta_i^*)-(\beta_j-\beta_j^*)\big\}.
\end{aligned}$$

On one hand, when subjects i and j are from different groups, that is $i \in \mathcal{G}_{0,k}$ and $j \in \mathcal{G}_{0,k'}$, $k \neq k'$, we have

$$||\tilde{\beta}_i - \tilde{\beta}_j|| \geq ||\beta_{0i} - \beta_{0j}|| - 2\max_i ||\tilde{\beta}_i - \beta_{0i}|| = ||\alpha_{0k} - \alpha_{0k'}|| - 2\max_i ||\tilde{\beta}_i - \beta_{0i}||.$$

Since $(\eta,\boldsymbol{\beta}) \in \Theta$, we can see that $\max_i ||\beta_i - \beta_{0i}|| \leq \phi_n$. By (2.17), we have $\max_k ||\alpha_k - \alpha_{0k}|| \leq \phi_n$ for $\boldsymbol{\alpha} = T^*(\boldsymbol{\beta})$. Then $\boldsymbol{\beta}^*$ satisfies that $\max_i ||\beta_i^* - \beta_{0i}|| \leq \phi_n$. By the definition of $\widetilde{\boldsymbol{\beta}}$, we have $\max_i ||\tilde{\beta}_i - \beta_{0i}|| \leq \phi_n$, and $||\tilde{\beta}_i - \tilde{\beta}_j|| \geq b - 2\phi_n > a\lambda$. By Condition (C4), $\rho_\gamma(t)$ is a constant when $t > a\lambda$ and $\rho'_\gamma(t) \equiv 0$ when $t > a\lambda$. Thus, when subjects i and j are from different groups, $\rho'_\gamma(||\tilde{\beta}_i - \tilde{\beta}_j||) \equiv 0$. On the other hand, $\beta_i^* = \beta_j^*$ when i and j are from the same group. Hence $\frac{(\tilde{\beta}_i-\tilde{\beta}_j)^T}{||\tilde{\beta}_i-\tilde{\beta}_j||} = \frac{(\beta_i-\beta_j)^T}{||\beta_i-\beta_j||}$ and

$$\rho'_\gamma(||\tilde{\beta}_i - \tilde{\beta}_j||)\frac{(\tilde{\beta}_i-\tilde{\beta}_j)^T}{||\tilde{\beta}_i-\tilde{\beta}_j||}\big\{(\beta_i-\beta_i^*)-(\beta_j-\beta_j^*)\big\} = \rho'_\gamma(||\tilde{\beta}_i - \tilde{\beta}_j||)||\tilde{\beta}_i - \tilde{\beta}_j||.$$

Note that

$$\max_k \max_{i,j\in\mathcal{G}_{0,k}} ||\tilde{\beta}_i - \tilde{\beta}_j|| = \max_k \max_{i,j\in\mathcal{G}_{0,k}} ||\tilde{\beta}_i - \beta_i^* + \beta_i^* - \beta_j^* + \beta_j^* - \tilde{\beta}_j||$$

$$\leq 2\max_i ||\tilde{\beta}_i - \beta_i^*|| \leq 2\max_i(||\tilde{\beta}_i - \beta_{0i}|| + ||\beta_i^* - \beta_{0i}||) \leq 4\phi_n.$$

By Condition (C4), we have

$$\Gamma_2 = \sum_{k=1}^{K_0} \sum_{\{i,j\in\mathcal{G}_{0,k}, i<j\}} \lambda\rho_\gamma'(||\tilde{\beta}_i - \tilde{\beta}_j||)||\beta_i - \beta_j|| \geq \sum_{k=1}^{K_0} \sum_{\{i,j\in\mathcal{G}_{0,k}, i<j\}} \lambda\rho_\gamma'(4\phi_n)||\beta_i - \beta_j||.$$

Now we turn to the first term Γ_1. Let

$$\mathbf{U}_i = \frac{\partial \ell_n(\eta, \boldsymbol{\beta})}{\partial \beta_i}\bigg|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} = -\int_0^\tau X_i dN_i(t) + \int_0^\tau \frac{Y_i(t)X_i\exp(Z_i^T\eta + X_i^T\tilde{\beta}_i)}{\frac{1}{n}\sum_{j=1}^n Y_j(t)\exp(Z_j^T\eta + X_j^T\tilde{\beta}_j)} d\bar{N}(t), \tag{2.18}$$

where $\bar{N}(t) = \frac{1}{n}\sum_{i=1}^n N_i(t)$. Then after some calculation, we have

$$\Gamma_1 = \sum_{i=1}^n \mathbf{U}_i^T(\beta_i - \beta_i^*) = \sum_{k=1}^{K_0} \sum_{i\in\mathcal{G}_{0,k}} \mathbf{U}_i^T(\beta_i - \beta_i^*)$$

$$= \sum_{k=1}^{K_0} \sum_{i,j\in\mathcal{G}_{0,k}} \frac{\mathbf{U}_i^T(\beta_i - \beta_j)}{|\mathcal{G}_{0,k}|} = \sum_{k=1}^{K_0} \sum_{i,j\in\mathcal{G}_{0,k}} \frac{\mathbf{U}_i^T(\beta_i - \beta_j)}{2|\mathcal{G}_{0,k}|} + \sum_{k=1}^{K_0} \sum_{i,j\in\mathcal{G}_{0,k}} \frac{\mathbf{U}_j^T(\beta_j - \beta_i)}{2|\mathcal{G}_{0,k}|}$$

$$= \sum_{k=1}^{K_0} \sum_{i,j\in\mathcal{G}_{0,k}} \frac{(\mathbf{U}_i - \mathbf{U}_j)^T(\beta_i - \beta_j)}{2|\mathcal{G}_{0,k}|} = \sum_{k=1}^{K_0} \sum_{\{i,j\in\mathcal{G}_{0,k}, i<j\}} \frac{(\mathbf{U}_i - \mathbf{U}_j)^T(\beta_i - \beta_j)}{|\mathcal{G}_{0,k}|}$$

$$\geq -\sum_{k=1}^{K_0} \sum_{\{i,j\in\mathcal{G}_{0,k}, i<j\}} \frac{2\max_i ||\mathbf{U}_i|| \cdot ||\beta_i - \beta_j||}{|\mathcal{G}_{\min}|},$$

where $|\mathcal{G}_{\min}| = \min_{k=1,\cdots,K_0} |\mathcal{G}_{0,k}|$. Following the same clues as before, for any $(\eta, \boldsymbol{\beta}) \in \Theta$, we have $(\eta, \widetilde{\boldsymbol{\beta}}) \in \Theta$. Then, by Condition (C2) and (2.18), we can find a constant C_U such that $\max_i ||\mathbf{U}_i|| \leq C_U$ with probability 1.

Note that $\lim_{n\to\infty}\rho'_\gamma(4\phi_n)=1$ and $|\mathcal{G}_{\min}|$ goes to infinity as $n\to\infty$. For large enough n, we can get that

$$Q_n(\eta,\boldsymbol{\beta})-Q_n(\eta,\boldsymbol{\beta}^*)=\Gamma_1+\Gamma_2\geq\sum_{k=1}^{K_0}\sum_{\{i,j\in\mathcal{G}_{0,k},i<j\}}||\beta_i-\beta_j||\big[\lambda\rho'_\gamma(4\phi_n)-2C_U/|\mathcal{G}_{\min}|\big]\geq 0.$$

Thus, (ii) is concluded.

2.7.3 Proof of Theorem 2.3

Proof. Similar to the proof of Theorem 2.2, we define the mapping T and T^* when $K_0=1$ and $\mathcal{M}_{\mathcal{G}_0}=\mathcal{M}$. For any vector $\boldsymbol{\beta}\in R^{np}$, set $\boldsymbol{\alpha}=T^*(\boldsymbol{\beta})\in R^p$ and $\boldsymbol{\beta}^*=T^{-1}(\boldsymbol{\alpha})\in\mathcal{M}$. The neighbourhood of true parameter Θ and ϕ_n are the same as those in Theorem 2.2. Then we only need to show the following two steps.

(i) For any $(\eta^T,\boldsymbol{\beta}^T)^T\in\Theta$, if $(\eta^T,(\boldsymbol{\beta}^*)^T)^T\neq((\hat{\eta}^{or})^T,(\widehat{\boldsymbol{\beta}}^{or})^T)^T$, then $Q_n(\eta,\boldsymbol{\beta}^*)>Q_n(\hat{\eta}^{or},\widehat{\boldsymbol{\beta}}^{or})$.

(ii) For any $(\eta^T,\boldsymbol{\beta}^T)^T\in\Theta$ and large enough n, $Q_n(\eta,\boldsymbol{\beta})\geq Q_n(\eta,\boldsymbol{\beta}^*)$.

For (i), when there is only one group, we have $\beta_i^*\equiv\boldsymbol{\alpha}$ and so $P_n(\boldsymbol{\beta}^*)=P_n(\widehat{\boldsymbol{\beta}}^{or})\equiv 0$. Since $\ell_n(\eta,\boldsymbol{\beta}^*)=\ell_n(\hat{\eta}^{or},\widehat{\boldsymbol{\beta}}^{or})$, it follows that $Q_n(\eta,\boldsymbol{\beta}^*)>Q_n(\hat{\eta}^{or},\widehat{\boldsymbol{\beta}}^{or})$.

For (ii),

$$Q_n(\eta,\boldsymbol{\beta})-Q_n(\eta,\boldsymbol{\beta}^*)=\frac{\partial\ell_n(\eta,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)+\frac{\partial P_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)=:\Gamma_1+\Gamma_2,$$

where $\widetilde{\boldsymbol{\beta}}$ is a vector between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$. We first consider the second term

$$\Gamma_2=\frac{\partial P_n(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)=\lambda\sum_{1\leq i<j\leq n}\rho'_\gamma(||\tilde{\beta}_i-\tilde{\beta}_j||)\frac{(\tilde{\beta}_i-\tilde{\beta}_j)^T}{||\tilde{\beta}_i-\tilde{\beta}_j||}\big\{(\beta_i-\beta_i^*)-(\beta_j-\beta_j^*)\big\}.$$

Since i and j are from the same group, we have $\beta_i^*=\beta_j^*$ and $\frac{(\tilde{\beta}_i-\tilde{\beta}_j)^T}{||\tilde{\beta}_i-\tilde{\beta}_j||}=\frac{(\beta_i-\beta_j)^T}{||\beta_i-\beta_j||}$.

Furthermore, $\max_{i,j} ||\tilde{\beta}_i - \tilde{\beta}_j|| \leq 4\phi_n$. Then by Condition (C4), we get that

$$\Gamma_2 = \lambda \sum_{1\leq i<j\leq n} \rho'_\gamma(||\tilde{\beta}_i - \tilde{\beta}_j||)||\tilde{\beta}_i - \tilde{\beta}_j|| \geq \lambda \sum_{1\leq i<j\leq n} \rho'_\gamma(4\phi_n)||\tilde{\beta}_i - \tilde{\beta}_j||.$$

For the first term Γ_1, we have

$$\mathbf{U}_i = \frac{\partial \ell_n(\eta, \boldsymbol{\beta})}{\partial \beta_i}\bigg|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} = -\int_0^\tau X_i dN_i(t) + \int_0^\tau \frac{Y_i(t) X_i \exp(Z_i^T \eta + X_i^T \tilde{\beta}_i)}{\frac{1}{n}\sum_{j=1}^n Y_j(t)\exp(Z_j^T \eta + X_j^T \tilde{\beta}_j)} d\bar{N}(t),$$

where $\bar{N}(t) = \frac{1}{n}\sum_{i=1}^n N_i(t)$. Since there is a constant C_U such that $\max_i ||\mathbf{U}_i|| \leq C_U$ with probability 1, it yields that

$$\Gamma_1 \geq -\sum_{1\leq i<j\leq n} \frac{2\max_i ||\mathbf{U}_i|| \cdot ||\beta_i - \beta_j||}{|n|} \geq -\sum_{1\leq i<j\leq n} \frac{2C_U ||\beta_i - \beta_j||}{|n|}.$$

Noting that $\lim_{n\to\infty} \rho'_\gamma(4\phi_n) = 1$, we obtain that for large enough n,

$$Q_n(\eta, \boldsymbol{\beta}) - Q_n(\eta, \boldsymbol{\beta}^*) = \Gamma_1 + \Gamma_2 \geq \sum_{1\leq i<j\leq n} ||\beta_i - \beta_j|| \big[\lambda \rho'_\gamma(4\phi_n) - 2C_U/|n|\big] \geq 0.$$

Hence, (ii) is concluded. □

Chapter 3

Nonparametric Statistical Inference for Panel Count Data with Terminal Event

3.1 Introduction

The main contributions of this chapter are fourfold. First, we propose a reversed nonparametric mean model for panel count data with a right-censored terminal event, where the nonparametric mean function is increasingly dependent on the terminal event time. Thus, the proposed model provides an intuitive interpretation of effects of terminal events on recurrent event processes. Second, we develop a two-stage sieve-based nonparametric estimation procedure by treating the distribution function of the terminal event time as a nuisance functional parameter. Third, we establish the asymptotic properties of the proposed estimator. In particular, we develop a general theorem for the asymptotic normality of nonparametric M-estimators with nuisance parameter when estimators have a convergence rate slower than the standard rate $n^{-1/2}$. Fourth, we develop a class of nonparametric tests for nonparametric comparison of mean functions of reversed recurrent event processes with panel count data in the presence of informative terminal event.

The remainder of this chapter is organized as follows. In Section 3.2, we present a reversed mean model anchoring at a terminal event and propose a two-stage nonparametric sieve-based estimation procedure. In Section 3.3, we establish the asymptotic properties of the proposed estimator. We first show the consistency and the convergence rate of the proposed estimators. Then we provide a general theorem for the asymptotic normality of nonparametric M-estimators with nuisance parameter. Section 3.4 presents a class of new test statistics for two sample test and establish their asymptotic normality. In Section 3.5, we conduct simulation studies to demonstrate the finite-sample performance of the proposed methods. In Section 3.6, we use the proposed methods to a set of panel count data from Chinese Longitudinal Healthy Longevity study. The proofs of the main results are given in the Appendix.

3.2 Model Setting and Estimation Procedure

Suppose that a counting process $\{N(t) : 0 \leq t \leq \tau\}$ denotes the number of recurrent events occurring up to time t, where τ is fixed time point. Let $\underline{T} = (T_1, T_2, \cdots, T_K)$ be the observation times of $N(t)$, where K represents the total number of observation times. Then the observed counting process is

$$\underline{N} = (N_1, N_2, \cdots, N_K) = (N(T_1), N(T_2), \cdots, N(T_K)).$$

Let U and C be the terminal event time and the censoring times, respectively. The observed right censored terminal event time is $Y = U \wedge C$ and the indicator whether the terminal time is uncensored is $\Delta = 1_{\{U \leq C\}}$. The observed data for subject i consists of $X_i = (Y_i, \Delta_i, K_i, \underline{T_i}, \underline{N_i})$, $i = 1, \cdots, n$, where $\underline{T_i} = (T_{i1}, T_{i2}, \cdots, T_{iK_i})$ and $\underline{N_i} = (N(T_{i1}), N(T_{i2}), \cdots, N(T_{iK_i}))$ with sample size n.

To investigate the effect of the terminal event on the recurrent event process, we consider a counting process $\tilde{N}(t; U)$ denoting the event counts from time t to the terminal event U, and propose a reversed nonparametric mean model anchoring at the terminal event:

$$E(\tilde{N}(t;U)|U=u) = \Lambda(u-t), \tag{3.1}$$

where $\Lambda(\cdot)$ is an unknown nondecreasing function with $\Lambda(0)=0$ to ensure the identifiability of this model. This model implies that

$$E(\tilde{N}(t_1;U)|U=u) - E(\tilde{N}(t_2;U)|U=u) = \Lambda(u-t_1) - \Lambda(u-t_2),$$

where $0 \le t_1 \le t_2 \le u$. Noting that $N(t_2)-N(t_1) = \tilde{N}(t_1;U) - \tilde{N}(t_2;U)$ and $N(0)=0$, we obtain $E(N(t)|U=u) = \Lambda(u) - \Lambda(u-t)$.

Let F denote the underlying distribution function of U. To make a valid inference on our model, we need the following basic conditions: (i) U and C are independent; (ii) The censoring event time C is noninformative to Λ. (iii) Given (Y,Δ), the distribution of $(K,\underline{T})$ is non-informative to Λ. Define $\triangle N_j = N(T_j) - N(T_{j-1})$ and $\triangle\Lambda_j(u) = \Lambda(u-T_{j-1}) - \Lambda(u-T_j)$ for $j=1,\cdots,K$ with $T_0=0$. Motivated by

$$\begin{aligned} &E\left[\sum_{j=1}^{K}\{\triangle N_j - \triangle\Lambda_j(U)\}^2|Y,\Delta,K,\underline{T},\underline{N}\right] \\ =&\sum_{j=1}^{K}\Delta\{\triangle N_j - \triangle\Lambda_j(Y)\}^2 + \sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}^2 dF(u)}{1-F(Y)}, \end{aligned}$$

with $X=(Y,\Delta,K,\underline{T},\underline{N})$, we propose a least squares-based loss function

$$\begin{aligned} \ell_n(\Lambda,F;X) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Big[&\Delta_i\{\triangle N_{i,j} - \triangle\Lambda_{i,j}(Y_i)\}^2 \\ &+ (1-\Delta_i)\frac{\int_{Y_i}^\infty\{\triangle N_{i,j} - \triangle\Lambda_{i,j}(u)\}^2 dF(u)}{1-F(Y_i)}\Big], \end{aligned} \tag{3.2}$$

where $\triangle N_{i,j} = N_i(T_{ij}) - N_i(T_{i(j-1)})$ and $\triangle\Lambda_{i,j}(u) = \Lambda(u-T_{i(j-1)}) - \Lambda(u-T_{ij})$. A natural idea is to take the minimizer of $l_n(\Lambda,F;X)$ defined in (3.2) as the estimator of the parameter. However, since the loss function involves an unknown distribution function F, it is difficult to estimate Λ and F simultaneously. To tackle the problem, we propose a two-stage approach. Concretely, in stage 1, we estimate F by using the Kaplan-Meier (KM) estimator $\hat{F}_n(u)$ (Kaplan and Meier, 1958). In stage 2, $\hat{\Lambda}_n$ is obtained by minimizing the loss function $l_n(\Lambda, \hat{F}_n; X)$ with respect to Λ. Since the observed data of (Y, Δ) is used in both stages 1 and 2, to distinguish them, we use the notation $(\tilde{Y}, \tilde{\Delta})$ to represent the data when we obtain the KM estimator in stage 1 without any ambiguity.

We adapt spline sieve estimator to estimate Λ since it converges fast and is easy to implement according to Lu, Zhang, and Huang (2007, 2009). Let $\{t_i : i = 1, \cdots, m_n + 2d\}$ be a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals, where

$$0 = t_1 = \cdots = t_d < t_{d+1} < \cdots < t_{m_n+d} < t_{m_n+d+1} = \cdots = t_{m_n+2d} = \tau.$$

Let $q_n = m_n + d$ and $\{I_l(s), l = 1, \cdots, q_n\}$ be the I-spline basis functions of order d (Ramsay, 1988). We then define the functional space of the estimator for Λ to be

$$\Phi_n = \left\{ \sum_{l=1}^{q_n} \alpha_l I_l(s) : \alpha_l \geq 0, l = 1, \cdots, q_n \right\}.$$

Usually, we take $d = 3$ corresponding to the cubic I-spline. Since $\hat{F}_n$ is a monotone step function, as shown in the Appendix 3.7.1, minimizing the loss function $l_n(\Lambda, \hat{F}_n; X)$ is a quadratic programming with the constraint that $\alpha_l \geq 0$ for $l = 1, \cdots, q_n$. Let $\boldsymbol{I}(s) = (I_1(s), \cdots, I_{q_n}(s))^T$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{q_n})^T$, and the solution of quadratic programming be $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_{q_n})^T$. Then the spline estimator of $\Lambda(s)$ is $\hat{\Lambda}_n(s) = \boldsymbol{I}(s)^T \hat{\boldsymbol{\alpha}}$.

3.3 Asymptotic Properties of the Estimator

To present the asymptotic results, we introduce some notations. Let $g^{(r)}$ be the rth derivative. For $r \geq 1$, define

$$
\begin{aligned}
\mathcal{H}_r =& \left\{ g : |g^{(r-1)}(s) - g^{(r-1)}(t)| \leq c_0 |s-t| \ for\ all\ 0 \leq s,t \leq \tau \right\}, \\
\Phi =& \{\Lambda \in \mathcal{H}_r : \Lambda \text{ is a nondecreasing continuous function on } [0,\tau] \text{ with } \Lambda(0)=0\}, \\
\mathcal{F} =& \{F : F \text{ is a distribution function on } [0,\infty)\}.
\end{aligned}
$$

Denote the true value of (Λ, F) to be $(\Lambda_0, F_0) \in \Phi \times \mathcal{F}$. For $B_1, B_2 \in \mathcal{B}_{[0,\tau]} =: \{B \cap [0,\tau] : B \in \mathcal{B}\}$, where $\mathcal{B}$ denotes the collection of Borel sets, set

$$
\begin{aligned}
\mu_1(B_1 \times B_2) =& \int \sum_{k=1}^{\infty} P(K=k|U=u) \\
& \times \sum_{j=1}^{k} P\Big((u-T_j) \in B_1, (u-T_{j-1}) \in B_2 | K=k, U=u\Big) dF_0(u), \\
\mu_2(B_1 \times B_2) =& \int \sum_{k=1}^{\infty} P(K=k|U=u) \\
& \times P\Big((u-T_K) \in B_1, u \in B_2 | K=k, U=u\Big) dF_0(u).
\end{aligned} \tag{3.3}
$$

For any functions $\Lambda_1, \Lambda_2 \in \Phi$, we define the metric as

$$
d_1(\Lambda_1,\Lambda_2)^2 = ||\triangle\Lambda_1(s_1,s_2) - \triangle\Lambda_2(s_1,s_2)||^2_{L_2(\mu_1)} = E\left[\sum_{j=1}^{K}\Big(\triangle\Lambda_{1,j}(U) - \triangle\Lambda_{2,j}(U)\Big)^2\right]
$$

$$
=E\left[\sum_{j=1}^{K}\left\{\Delta\Big(\triangle\Lambda_{1,j}(Y)-\triangle\Lambda_{2,j}(Y)\Big)^2 + (1-\Delta)\frac{\int_Y^\infty \Big(\triangle\Lambda_{1,j}(u)-\triangle\Lambda_{2,j}(u)\Big)^2 dF_0(u)}{1-F_0(Y)}\right\}\right],
$$

where $\triangle\Lambda(s_1,s_2) = \Lambda(s_2) - \Lambda(s_1)$. For any $F_1, F_2 \in \mathcal{F}$, we define the metric as

$$
d_2(F_1,F_2) = ||F_1 - F_2||_\infty,
$$

where $||\cdot||_\infty$ represents the L_∞ norm. Write $\mathcal{F}_\delta = \{F \in \mathcal{F} : d_2(F, F_0) \leq \delta\}$ for small $\delta > 0$.

To establish the asymptotic properties of the proposed estimator, we need the following conditions.

(C1) $0 < \Lambda_0(\tau) < \infty$.

(C2) $0 < F_0(\tau) < 1$. F_0 is absolutely continuous with respect to Lebesgue measure. Moreover, the density function $f_0(s)$ has a uniform positive lower bound for all $s \in [M_1, \tau]$, where M_1 is a constant representing the minimum value of the support of F_0.

(C3) $E\big[\sum_{j=1}^K \{\triangle N_j - \triangle\Lambda_{0,j}(U)\}^2\big] < \infty$.

(C4) The probability of censoring $\varrho = P(Y < U)$ satisfies that $0 < \varrho < 1$.

(C5) The number of subinterval in $[0, \tau]$ satisfies $m_n = O(n^\nu)$ for $0 < \nu < 1/2$. Moreover, we suppose that

$$\max_{d+1 \leq i \leq m_n+d+1} |t_i - t_{i-1}| = O(n^{-\nu}),$$

and there is a constant $M_2 > 0$ such that

$$\frac{\max_{d+1 \leq i \leq m_n+d+1} |t_i - t_{i-1}|}{\min_{d+1 \leq i \leq m_n+d+1} |t_i - t_{i-1}|} \leq M_2$$

uniformly for n.

(C6) There is a constant $M_3 > 0$ such that $P(K \leq M_3) = 1$.

(C7) $P(T_j - T_{j-1} \geq M_4$ for all $j = 1, \cdots, K) = 1$ with some constant $M_4 > 0$.

Remark 1. Condition (C1) is standard in the literature of nonparametric estimation. Condition (C2) holds for the most cumulated distribution functions of continuous random variables. Condition (C3) demands that $\sum_{j=1}^{K} \triangle N_j$ has finite second order central moment. Condition (C4) is regular in survival analysis to ensure the censoring rate between 0 and 1. Condition (C5) is required to guarantee a monotone spline approximation for a monotone function, see Lu, Zhang, and Huang (2007, 2009). Condition (C6) is similar to Condition (C2) in Wellner and Zhang (2007), which indicates that the number of observations is bounded. According to Wellner and Zhang (2007), Condition (C7) is common in practice, which requires that the adjacent observation times are separable.

Theorem 3.1 (Consistency for Two-Stage Estimator)**.** *Suppose that Conditions* (C1)–(C7) *hold. Then, for every* $0 \le b_1 \le b_2 \le \tau$ *satisfying* $\mu_2([0,b_1]\times[b_2,\tau]) > 0$, *we have*

$$||\triangle\hat{\Lambda}_n(s_1,s_2)1_{\{(s_1,s_2)\in[b_1,b_2]\times[b_1,b_2]\}} - \triangle\Lambda_0(s_1,s_2)1_{\{(s_1,s_2)\in[b_1,b_2]\times[b_1,b_2]\}}||^2_{L_2(\mu_1)} = o_p(1).$$

In particular, if $\mu_2(\{0\}\times\{\tau\}) > 0$, *then* $d_1(\hat{\Lambda}_n, \Lambda_0) = o_p(1)$.

To establish the rate of convergence and the asymptotic normality, we need the following additional conditions.

(C8) μ_1 is absolutely continuous with respect to Lebesgue measure with a derivative $\dot{\mu}_1$, and $\dot{\mu}_1$ has a uniform positive lower bound.

(C9) There is a positive constant M_5 such that $1/M_5 < \Lambda_0'(s) < M_5$ for all $s \in [\tau', \tau]$ with $0 < \tau' \le \tau$ such that $\Lambda_0(\tau') > 0$.

(C10) $P(U \ge \tau) = \omega_1 > 0$ and $P(C \ge \tau) = \omega_2 > 0$.

(C11) $E(e^{cN(t)})$ is uniformly bounded for $t \in [0,\tau]$ and some constant c.

Remark 2. Condition (C8) implies that the metric μ_1 defined in (3.3) has a strictly positive intensity. Condition (C9) requires the true conditional mean function being absolutely continuous with bounded intensity function, which is reasonable as explained in Wellner and Zhang (2007). (C10) is used as a technical condition in the proof of the uniform weak convergence rate of the KM estimator according to Kong et al. (2018). Condition (C11) holds when $N(t)$ is from a Poisson-type process or is uniformly bounded conditional on terminal event time, which is often true in clinical trials.

Theorem 3.2 (Rate of Convergence)**.** *Suppose that Conditions* (C1)–(C11) *hold and* $\mu_2(\{0\} \times \{\tau\}) > 0$. *Taking* $\nu = 1/(1+2r)$, *we have* $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(n^{-r/(1+2r)})$.

By Theorem 3.2, the convergence rate of Λ_n is slower than $n^{1/2}$. Thus, the classic approach used for asymptotic normality of estimator with nuisance parameter is not applicable anymore. Here we build a general theorem for the asymptotic normality of a function of nonparametric M-estimator with nuisance parameter.

We write the loss function $l_n(\Lambda, F; X)$ defined in (3.2) as $\mathbb{P}_n m(\Lambda, F; X)$. Suppose that Λ_η is a parameter path satisfying $\Lambda_\eta \in \Phi$, and $\Lambda_\eta|_{\eta=0} = \Lambda_0$. Set $\mathcal{H} = \{h : h = \frac{\partial \Lambda_\eta}{\partial \eta}|_{\eta=0}\}$ and $\psi(\Lambda, F; X)[h] = \frac{\partial}{\partial \eta} m(\Lambda_\eta, F; X)|_{\eta=0}$. For $h \in \mathcal{H}$, we define Q_n and Q by $Q_n(\Lambda, F)[h] = \mathbb{P}_n \psi(\Lambda, F; X)[h]$ and $Q(\Lambda, F)[h] = \mathcal{P}\psi(\Lambda, F; X)[h]$, respectively.

To establish the asymptotic normality, we need the following conditions:

(B1) $Q(\Lambda_0, F_0)[h] = 0$ and $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h] = o_p(n^{-1/2})$.

(B2) $\sqrt{n}(Q_n - Q)(\hat{\Lambda}_n, \hat{F}_n)[h] - \sqrt{n}(Q_n - Q)(\Lambda_0, F_0)[h] = o_p(1)$.

(B3) $Q(\Lambda, F)[h]$ is Fréchet-differentiable with respect to Λ at $(\Lambda_0, \hat{F}_n)$ with a continuous derivative $\dot{Q}^{(1)}_{\Lambda_0, \hat{F}_n}[h]$; $Q(\Lambda, F)[h]$ is Fréchet-differentiable with respect to F at $\theta_0 = (\Lambda_0, F_0)$ with a continuous derivative $\dot{Q}^{(2)}_{\Lambda_0, F_0}[h]$.

(B3') $Q(\Lambda, F)[h]$ is Fréchet-differentiable with respect to Λ at (Λ_0, F_0) with a continuous derivative $\dot{Q}^{(1)}_{\Lambda_0, F_0}[h]$.

(B4) $Q(\hat{\Lambda}_n, \hat{F}_n)[h] - Q(\Lambda_0, F_0)[h] - \dot{Q}^{(1)}_{\Lambda_0, \hat{F}_n}(\hat{\Lambda}_n - \Lambda_0)[h] - \dot{Q}^{(2)}_{\Lambda_0, F_0}(\hat{F}_n - F_0)[h] = o_p(n^{-1/2})$.

(B4') $Q(\hat{\Lambda}_n, \hat{F}_n)[h] - Q(\Lambda_0, F_0)[h] - \dot{Q}^{(1)}_{\Lambda_0, F_0}(\hat{\Lambda}_n - \Lambda_0)[h] - \dot{Q}^{(2)}_{\Lambda_0, F_0}(\hat{F}_n - F_0)[h] = o_p(n^{-1/2})$.

(B5) $\sqrt{n}\dot{Q}^{(2)}_{\Lambda_0, F_0}(\hat{F}_n - F_0)[h] + \sqrt{n} Q_n(\Lambda_0, F_0)[h]$ converges into a tight Gaussian progress.

Remark 3. (B1), (B3) and (B5) are the analytical conditions required in Theorem

3.3.1 of van de Vaart and Wellner (1996). Conditions (B2) and (B4) imply that the remainders of the corresponding Taylor expansions are negligible. (B3') and (B4') are similar to the first part of Condition (B3) and Condition (B4), respectively.

Theorem 3.3 (Asymptotic Normality). *Suppose that Conditions* (B1)–(B5) *hold. Then*

$$-\sqrt{n}\dot{Q}^{(1)}_{\Lambda_0,\hat{F}_n}(\hat{\Lambda}_n-\Lambda_0)[h]=\sqrt{n}\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]+\sqrt{n}Q_n(\Lambda_0,F_0)[h]+o_p(1)$$

converges into a tight Gaussian process. Replacing the first part of (B3) *and* (B4) *by* (B3') *and* (B4'), *we have*

$$-\sqrt{n}\dot{Q}^{(1)}_{\Lambda_0,F_0}(\hat{\Lambda}_n-\Lambda_0)[h]=\sqrt{n}\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]+\sqrt{n}Q_n(\Lambda_0,F_0)[h]+o_p(1)$$

converges into a tight Gaussian process.

Remark 4. The proof clues of Theorem 3.3 mainly follow from multivariate functional delta method. Theorem 3.3 establishes the asymptotic normality of the statistics about two-stage estimator $\hat{\Lambda}_n$, whose convergence rate is not required to be $n^{1/2}$. It can be used to derive the asymptotic normality of nonparametric estimators with a nuisance parameter. As a result, it can be adapted to more general situations compared with Theorem 1 in Zhao and Zhang (2017).

For panel count data, ignoring constant factors, for all $h\in\mathcal{H}_r$ we have

$$\psi(\Lambda,F;X)[h]=\sum_{j=1}^{K}\Big[\Delta\big\{\triangle N_j-\triangle\Lambda_j(Y)\big\}\triangle h_j(Y)$$
$$+(1-\Delta)\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_j(u)\}\triangle h_j(u)dF(u)}{1-F(Y)}\Big],$$

where $\triangle h_j(u)=h(u-T_{j-1})-h(u-T_j)$, $j=1,\cdots,K$.

Theorem 3.4 (Application to Panel Count Data). *Suppose that Conditions* (C1)–(C11) *hold and* $\mu_2(\{0\} \times \{\tau\}) > 0$.

(i) Then for any bounded function $h \in \mathcal{H}_r$, *we have*

$$\sqrt{n}\mathcal{P}\varsigma(\hat{\Lambda}_n, \hat{F}_n; X)[h] = \sqrt{n}\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n - F_0)[h] + \sqrt{n}Q_n(\Lambda_0, F_0)[h] + o_p(1),$$

where

$$\varsigma(\Lambda, F; X)[h] = \sum_{j=1}^{K} \Big[\Delta\{\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y)\}\triangle h_j(Y)$$
$$+ (1-\Delta)\frac{\int_Y^\infty \{\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u)\}\triangle h_j(u)dF(u)}{1-F(Y)}\Big].$$

(ii) Moreover, $\sqrt{n}\mathcal{P}\varsigma(\hat{\Lambda}_n, \hat{F}_n; X)[h] \xrightarrow{d} N(0, \sigma_0^2)$, *where*

$$\sigma_0^2 = E\left[\{\mathcal{P}\varphi(\Lambda_0, F_0; X; \tilde{Y}, \tilde{\Delta})[h] + \psi(\Lambda_0, F_0; X)[h]\}^2\right],$$

and $\varphi(\Lambda, F; X; \tilde{Y}, \tilde{\Delta})[h]$ *is defined as in the Appendix.*

3.4 Two-Sample Test

Suppose that n subjects are from two groups with sample sizes n_1 and n_2, where $n_1 + n_2 = n$. Denote the observed data of the lth group as $\{X_i^{(l)} : i = 1, \cdots, n_l\} = \{(Y_i^{(l)}, \Delta_i^{(l)}, K_i^{(l)}, \underline{T_i}^{(l)}, \underline{N_i}^{(l)}) : i = 1, \cdots, n_l\}$, for $l = 1, 2$. Given the terminal event time $U^{(l)} = u$, the conditional mean function of $\tilde{N}^{(l)}(t)$ is $\Lambda_l(u - t)$.

3.4.1 Terminal Events with Equal Distribution

In this subsection, we assume that terminal event times share the same distribution function F_0 for all the subjects so that its estimator $\hat{F}_n$ can be obtained based on the pooled data. We investigate two-sample test with the null hypothesis $H_0 : \Lambda_1 = \Lambda_2 = \Lambda_0$. Denote $\hat{\Lambda}_l$ and $\hat{\Lambda}_n$ as the estimates of Λ_l and Λ_0 based on the data set of group l and the pooled data, respectively.

Theorem 3.5. *Besides of the conditions in Theorem 3.4, we suppose that $h_n(\cdot)$ is a bounded weight process, and there is a bounded function $h \in \mathcal{H}_r$ such that*

$$d_1^2(h_n, h) = E\left[\sum_{j=1}^{K} \left\{\triangle h_{n,j}(U) - \triangle h_j(U)\right\}^2\right] = o_p(n^{-1/(1+2r)}).$$

Assume that $n_1/n \to p$ as $n \to \infty$, where $0 < p < 1$. Then under $H_0 : \Lambda_1 = \Lambda_2$,

$$U_n = \sqrt{n}\mathbb{P}_n\left(\varsigma(\hat{\Lambda}_1, \hat{F}_n; X)[h_n] - \varsigma(\hat{\Lambda}_2, \hat{F}_n; X)[h_n]\right)$$

converges in distribution to $N(0, (1/p+1/(1-p))\check{\sigma}_0^2)$, where $\check{\sigma}_0^2 = E[\psi^2(\Lambda_0, F_0; X)[h]]$. Moreover, $\check{\sigma}_0^2$ can be consistently estimated by $\hat{\sigma}_n^2 = \mathbb{P}_n[\psi^2(\hat{\Lambda}_n, \hat{F}_n; X)[h_n]]$.

Remark 5. Theorem 3.5 states the asymptotic normality of a new statistics U_n and gives a consistent estimator of its asymptotic variance. Setting $T_n(h_n) = U_n/\hat{\sigma}_n$ to be the statistics with different weight processes h_n, we then can use them to conduct the two-sample hypothesis test. A natural choice is to take $h_n = \hat{\Lambda}_n$. Other possible choices of the weight processes can be found in the simulation studies. It deserves to note that the weight process is only required to be bounded, which is more flexible compared with the monotone condition in Zhang (2006) and Balakrishnan and Zhao (2009).

3.4.2 Terminal Events with Unequal Distributions

In this subsection, we assume the distribution and density functions of $U^{(l)}, l = 1, 2$ to be F_l and f_l, which may be different for two groups. Let $\hat{F}_l$ and $\hat{\Lambda}_l$ be the estimator of F_l and Λ_l, respectively. Given a partition $0 = t_0^{(l)} < t_1^{(l)} < \cdots < t_{\nu_{n_l}}^{(l)} = \tau$, we define the histogram-type estimators of f_l as $\hat{f}_l(u) = (\hat{F}_l(t_{i_l}^{(l)}) - \hat{F}_l(t_{i_l-1}^{(l)}))/(t_i^{(l)} - t_{i-1}^{(l)})$ for $t_{i_l-1}^{(l)} \leq u < t_{i_l}^{(l)}$ following Földes, Rejtő, and Winter (1981). Set $f_T(u) =$

$p_1 f_1(u) + p_2 f_2(u)$,

$$w_l(u-t_1) - w_l(u-t_2) =(h(u-t_1) - h(u-t_2)) f_T(u)/f_l(u),$$

$$w_n^{(l)}(u-t_1) - w_n^{(l)}(u-t_2) =(h_n(u-t_1) - h_n(u-t_2)) \left(\frac{n_l}{n} + \frac{n_r}{n} \frac{\hat{f}_r(u)}{\hat{f}_l(u)} \right),$$

where $p_l = \lim_{n\to\infty} n_l/n$ for $l, r = 1, 2,\ l \neq r$.

Theorem 3.6. *Suppose that the conditions in Theorem 3.5 hold for each group and f_l's are Lipschitz continuous. Then for $\Lambda_0 \in \mathcal{H}_r, r \geq 2$, under the null hypothesis $H_0 : \Lambda_1 = \Lambda_2 = \Lambda_0$ we have*

(i) $\tilde{U}_n = \frac{1}{\sqrt{n}} \sum_{l=1}^{2} \sum_{i=1}^{n_l} (\varsigma(\hat{\Lambda}_1, \hat{F}_l; X_i^{(l)})[h_n] - \varsigma(\hat{\Lambda}_2, \hat{F}_l; X_i^{(l)})[h_n])$ converges in distribution to $N(0, (\sigma_1^2/p_1 + \sigma_2^2/p_2))$, where $\sigma_l^2 = E[\{\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_l] + \psi(\Lambda_0, F_l; X^{(l)})[w_l]\}^2]$ with $\varphi_l(\Lambda, F; X; \tilde{Y}, \tilde{\Delta})[w]$ defined as in the Appendix.

(ii) In addition, suppose the knots of partition satisfy

$$\max_{i=1,\cdots,\nu_{n_l}{}^{(l)}} \{|t_i^{(l)} - t_{i-1}^{(l)}|\} \to 0 \text{ and } \left(\frac{n}{\log n}\right)^{1/4} \min_{i=1,\cdots,\nu_{n_l}{}^{(l)}} \left\{\left|t_i^{(l)} - t_{i-1}^{(l)}\right|\right\} \to \infty$$

as $n \to \infty$. Then σ_l^2 can be consistently estimated by $\hat{\sigma}_l^2$, and the asymptotic variance of $\tilde{U}_n$ can be consistently estimated by $\tilde{\sigma}_n^2 = n(\hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2)$, where

$$\hat{\sigma}_l^2 = \mathbb{P}_{n_l} \left[\left\{ \mathbb{P}_{n_l} \varphi_n^{(l)}(\hat{\Lambda}_l, \hat{F}_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)}] + \psi(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[w_n^{(l)}] \right\}^2 \right]$$

with $\varphi_n^{(l)}(\Lambda, F; X; \tilde{Y}, \tilde{\Delta})[w]$ defined as in the Appendix.

According to Theorem 3.6, we could use the statistics $\tilde{T}_n(h_n) = \tilde{U}_n/\tilde{\sigma}_n$ to test the two sample hypothesis for the mean function of reversed counting processes under the case of different distributions of terminal events.

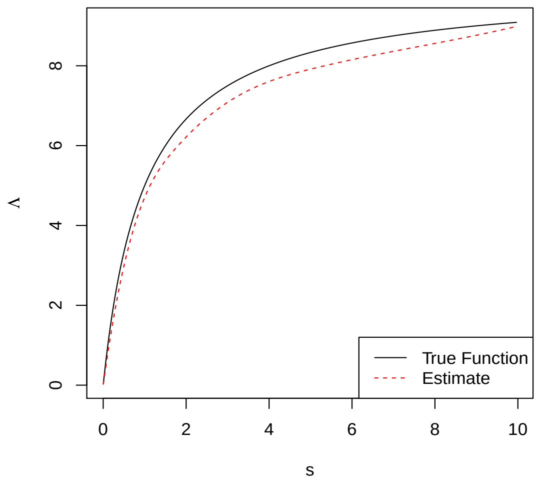
3.5 Simulation Studies

3.5.1 Two-Stage Estimation for Mean Function

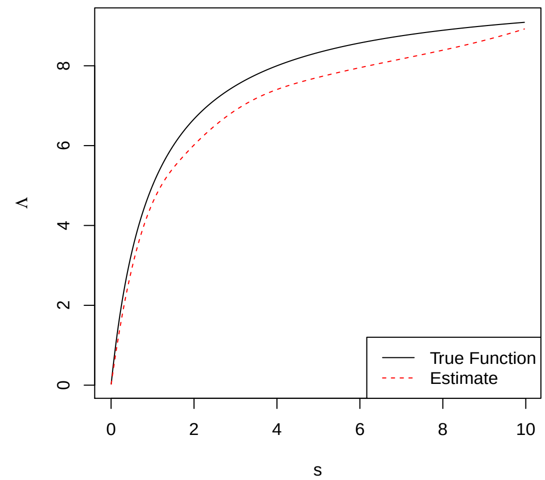
We conducted simulation studies to demonstrate the finite-sample performance of the two-stage estimator $\hat{\Lambda}_n$. The sample size were $n = 50$ and 100, respectively. For subject i, the observation $X_i = (Y_i, \Delta_i, K_i, \underline{T_i}, \underline{N_i})$ was generated as below. The latent terminal event time U_i was from $6 + \exp(1)$, and the censoring time C_i was from $6 + \kappa \exp(3)$, where $\tau = 10$ represented the end time of study. The constant κ was taken such that the censoring rate reached 20% and 40%, respectively. Took $Y_i = U_i \wedge C_i$ and $\Delta_i = 1_{\{U_i \le C_i\}}$. The number of observation K_i was taken from integers between 1 to 6 with equal probability. Given the censored terminal event time Y_i, observation time $\underline{T_i} = (T_{i1}, T_{i2}, \cdots, T_{iK_i})$ was an ordered sample from uniform distribution $\mathrm{Unif}(0, Y_i)$. Let N_i be a Poisson process with $E(N_i(t)|U = u) = \Lambda_0(u) - \Lambda_0(u - t)$ and $\Lambda_0(s) = 10s/(s + 1)$, which means that $N_i(T_{i1})$ was from Poisson distribution with mean $10U_i/(U_i + 1) - 10(U_i - T_{i1})/(U_i - T_{i1} + 1)$, and $N_i(T_{ij}) - N_i(T_{i(j-1)})$ was from Poisson distribution with mean $10(U_i - T_{i(j-1)})/(U_i - T_{i(j-1)} + 1) - 10(U_i - T_{ij})/(U_i - T_{ij} + 1)$. For the knots of spline, took $d = m_n = 3$ and let $t_{d+1}, t_{d+2}, t_{d+3}$ be quartiles of $\{Y_i - T_{ij} : i = 1, \cdots, n; j = 1, \cdots, K_i\}$. All simulation studies were based on 1000 replications.

Figure 3.1 shows the results of two-stage estimators for the conditional mean functions. The solid line represents the true mean function, and dashed line represents the average of estimated mean function based on 1000 replications. From (a)-(d) in Figure 1, we can see that the fitted mean functions are very close to the true ones, which means the proposed nonparametric estimator is nearly unbiased.

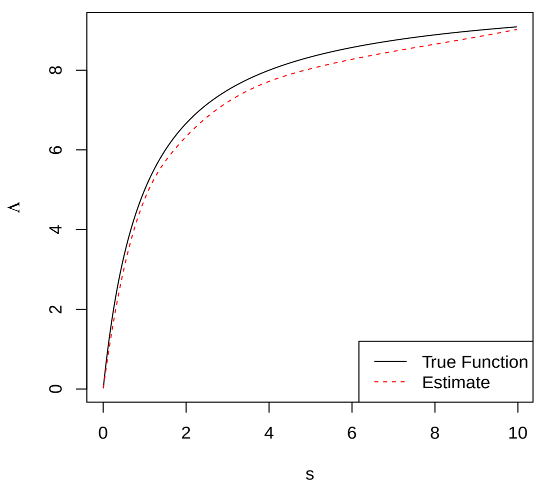
(a) $n = 50$ and censoring rate = 20%



(b) $n = 50$ and censoring rate = 40%



(c) $n = 100$ and censoring rate = 20%



(d) $n = 100$ and censoring rate = 40%

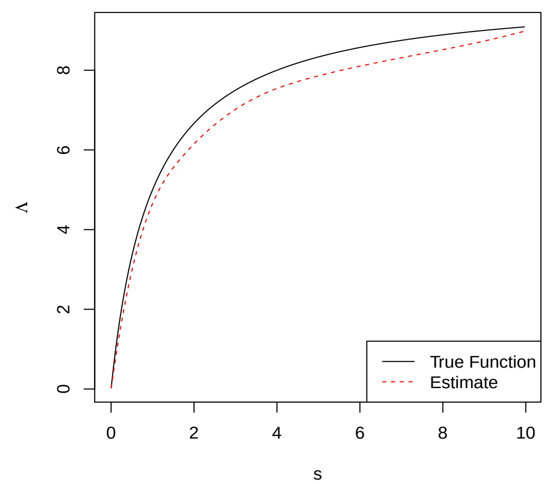


Figure 3.1: Estimates of the mean functions.

3.5.2 Two-Sample Test with the Same Terminal Events

We conducted two sample test for two groups sharing the same distribution of terminal event. We generated two groups of independent and identically distributed sample $\{X_i^{(l)} : i = 1, \cdots, n_l\}$ the same way as in Subsection 3.5.1 for $l = 1, 2$, with sample size $n_1 = n_2 = 50, 100, 150$ or 200. Let $N^{(l)}$ be a Poisson process with $E(N^{(l)}(t)|U^{(l)} = u) = \Lambda_l(u) - \Lambda_l(u-t)$ for the lth group. We considered the following two cases:

$$\text{Case 1} : \Lambda_1(s) = s,\ \Lambda_2(s) = \beta s;$$

$$\text{Case 2} : \Lambda_1(s) = s,\ \Lambda_2(s) = \frac{10s}{s+\beta}.$$

We took $\beta = 1, 1.1, 1.2, 1.3$ in Case 1 and $\beta = 1, 3, 5$ in Case 2.

Figure 3.2 displays the graphs of the true mean functions for the two cases with different values of β. It can be seen that the conditional mean functions of two groups in Case 1 do not overlap but they cross over in Case 2. The weight processes $h_n^{(j)}(t), j = 1, 2, 3$ in Theorem 3.5 are chosen to be

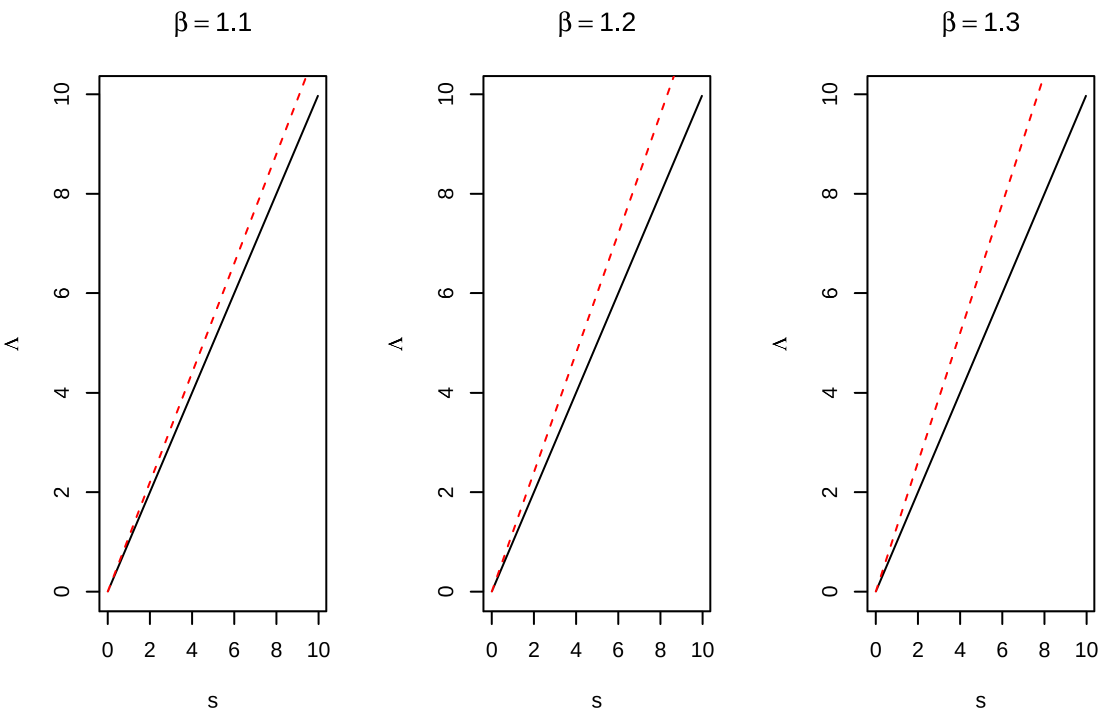
$$h_n^{(1)}(t) = t,\ h_n^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(t \geq T_{iK_i})},\ h_n^{(3)}(t) = \hat{\Lambda}_n(t).$$

The simulation studies are based on 1000 replications.

Figure 3.3 presents the quantile plots of the test statistics against the standard normal distribution in Case 1 with $n_1 = n_2 = 200$, $\beta = 1$ and censoring rate 20% for three weight processes. They reveal that the asymptotic normality given in Theorem 3.5 is satisfied in finite sample size. Similar plots are obtained for other situations and they are omitted here.

Tables 3.1 and 3.2 report the sizes and powers of the proposed statistics $T_n(h_n^{(j)})$ at significance level 0.05 in Cases 1 and 2 for different values of β, where $T_n(h_n^{(j)})$

(a) Case 1



(b) Case 2

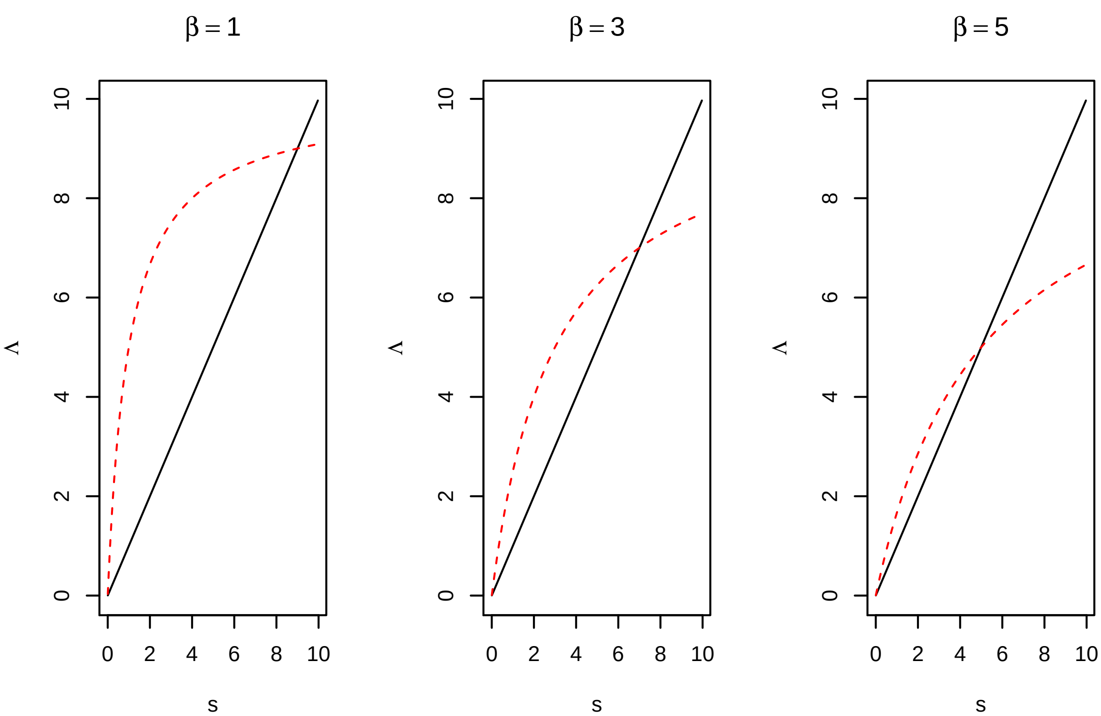


Figure 3.2: The plots of mean functions for Cases 1 and 2.

(a) Q−Q Plot for $T_n(h_n^{(1)})$ (b) Q−Q Plot for $T_n(h_n^{(2)})$ (c) Q−Q Plot for $T_n(h_n^{(3)})$

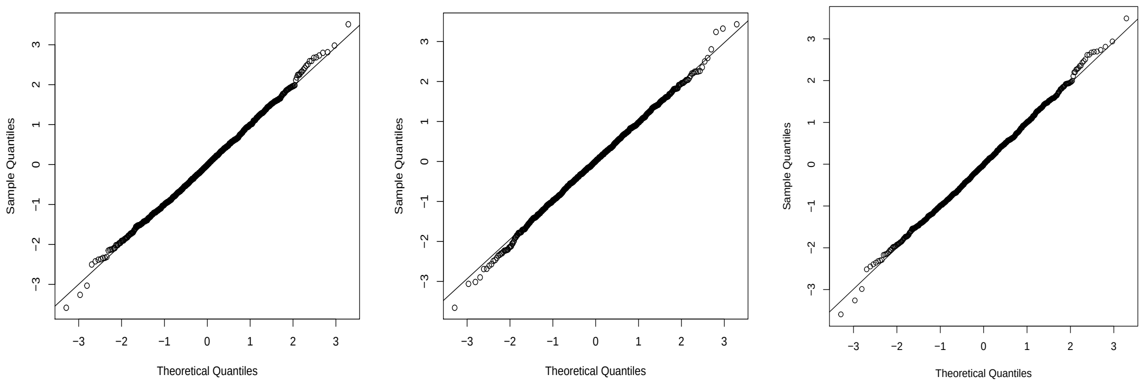


Figure 3.3: Q-Q plots for n_1=n_2=200, β=1, and censoring rate = 20% when the distribution functions of the terminal events are equal in the two groups.

represents the test statistics with the jth weight process for $j = 1, 2, 3$. The simulation results in Tables 3.1 and 3.2 demonstrate that the proposed test possesses good properties: (i) Table 3.1 shows the sizes are all around the significance level of 0.05 and they are closer to 0.05 as the sample size increases; (ii) for fixed values of β in both cases, the power increases when the sample size increases; (iii) power values are closer to 1 when the parameter β takes values farer away from the null hypotheses; (iv) power performance relies on the choice of the weight processes.

3.5.3 Two-Sample Test with Different Terminal Events

We then conducted two-sample test for two groups with different distributions of terminal events, where the sample size $n_1 = n_2 = 50, 100, 150$ or 200. The data sets were generated similar to Subsection 3.5.2 except that the latent terminal event time $U_i^{(1)}$ and $U_i^{(2)}$ were from $6+\exp(1)$ and $6+\exp(2)$, and the censoring time $C_i^{(1)}$ and $C_i^{(2)}$ were from $6+\kappa_1 \exp(3)$ and $6+\kappa_2 \exp(3)$, where κ_1 and κ_2 were taken such that the censoring rates reached 20% or 40%. We also considered Cases 1 and 2 in Subsection 3.5.2 for the poisson process. The weight processes $h_n^{(j)}(t), j = 1, \cdots, 3$

Table 3.1: Simulation results of two-sample tests with different weights for Case 1 when the distribution functions of the terminal events are equal in the two groups.

	Censoring rate 20%			Censoring rate 40%		
β	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$
			n_1=n_2=50			
1	0.045	0.053	0.043	0.056	0.053	0.054
1.1	0.165	0.153	0.166	0.172	0.156	0.171
1.2	0.464	0.417	0.465	0.453	0.419	0.454
1.3	0.776	0.720	0.775	0.763	0.725	0.761
			n_1=n_2=100			
1	0.051	0.047	0.052	0.052	0.048	0.051
1.1	0.242	0.229	0.243	0.258	0.240	0.255
1.2	0.728	0.703	0.728	0.715	0.690	0.717
1.3	0.967	0.956	0.967	0.962	0.956	0.960
			n_1=n_2=150			
1	0.048	0.047	0.048	0.049	0.048	0.048
1.1	0.359	0.344	0.361	0.355	0.338	0.354
1.2	0.873	0.851	0.873	0.877	0.859	0.878
1.3	0.996	0.989	0.996	0.995	0.991	0.995
			n_1=n_2=200			
1	0.039	0.050	0.038	0.043	0.051	0.043
1.1	0.472	0.453	0.475	0.467	0.433	0.467
1.2	0.960	0.948	0.959	0.954	0.939	0.954
1.3	0.999	0.997	0.999	0.999	0.998	0.999

were chosen to be

$$h_n^{(1)}(t) = t,\ h_n^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(t \geq T_{iK_i})},\ h_n^{(3)}(t) = \frac{\hat{\Lambda}_1(t) + \hat{\Lambda}_2(t)}{2}.$$

To calculate the histogram-type estimators of f_l, we divided $[6, \tau]$ into 5 intervals with equal length. The simulation studies were based on 1000 replications.

Figure 3.4 shows the quantile plots of the test statistics against the standard normal distribution with $n_1 = n_2 = 200$, $\beta = 1$ and censoring rate 20% for three weight processes in Case 1, which demonstrates the normality of the proposed statistics in finite sample size. The plots for other situations are similar and omitted here. Tables 3.3 and 3.4 summarize the sizes and powers of proposed statistics $\tilde{T}_n(h_n^{(j)})$, $j = 1, 2, 3$

Table 3.2: Simulation results of two-sample tests with different weights for Case 2 when the distribution functions of the terminal events are equal in the two groups.

	Censoring rate 20%			Censoring rate 40%		
β	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$	$T_n(h_n^{(1)})$	$T_n(h_n^{(2)})$	$T_n(h_n^{(3)})$
			n_1=n_2=50			
1	0.954	0.999	0.512	0.972	1.000	0.601
3	0.801	0.972	0.563	0.852	0.971	0.604
5	0.934	0.986	0.865	0.942	0.984	0.880
			n_1=n_2=100			
1	0.999	1.000	0.752	1.000	1.000	0.868
3	0.982	0.971	0.859	0.987	1.000	0.892
5	0.998	1.000	0.991	0.999	1.000	0.991
			n_1=n_2=150			
1	1.000	1.000	0.900	1.000	1.000	0.960
3	0.998	1.000	0.949	0.998	1.000	0.970
5	1.000	1.000	0.998	1.000	1.000	0.999
			n_1=n_2=200			
1	1.000	1.000	0.941	1.000	1.000	0.977
3	1.000	1.000	0.986	1.000	1.000	0.989
5	1.000	1.000	1.000	1.000	1.000	1.000

in Cases 1 and 2 at significant level 0.05. From Tables 3.3 and 3.4, we could draw the same conclusion as those in Subsection 3.5.2.

3.6 Real Data Analysis

In this section, we applied our method to analyze the dataset from Chinese Longitudinal Healthy Longevity Survey (CLHLS) during the period from 1998 to 2014, which was published in Zeng et al. (2017). This survey was conducted by the Center for Healthy Aging and Development Studies (CHADS) of National School of Development at Peking University and Chinese Center for Disease Control and Prevention (CDC) to shed light on the determinants of healthy human longevity and the oldest-old mortality. The data set was obtained from the seven waves (1998, 2000, 2002, 2005, 2008, 2011 and 2014), collecting the information on 9093 respondents elder

(a) Q−Q Plot for $\tilde{T}_n(h_n^{(1)})$ (b) Q−Q Plot for $\tilde{T}_n(h_n^{(2)})$ (c) Q−Q Plot for $\tilde{T}_n(h_n^{(3)})$

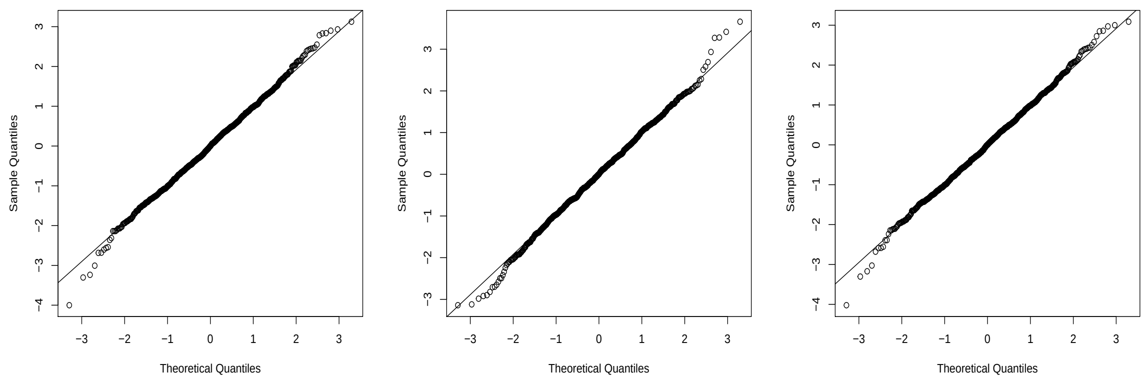


Figure 3.4: Q-Q plots for n_1=n_2=200, β=1, and censoring rate = 20% when the distribution functions of the terminal events are different in the two groups.

than 77-years-old interviewed in the 1998 baseline survey. On each wave, the respondents were asked to provide their information including health, socioeconomic characteristics, family, lifestyle and demographic profile at a random date.

In this study, more than 30% interviewees at least once suffered from serious illness during the survey, and some of them were even bedridden all the time. To analyze the occurrence rate of serious illness among elderly people living in different areas (urban or rural) in the survey, we considered 3050 samples after removing 4262 individuals who were deceased or lost to follow-up in 2000, 1312 individuals who lived in both areas during this period and 469 individuals with missing or typo records. Among them, 1489 lived in the urban and 1561 stayed in the rural. We then took T_{ij} as the number of months that the jth follow-up survey lasted from the baseline survey for the ith individual, $N(t)$ as the times that the respondent suffered from serious illness up to time t and K_i as the total number of follow-up surveys for the ith individual. We took death as terminal event and loss-of-connection as censoring event, and $\tau = 197$, the largest follow-up month. The censoring rate was 27.28%.

Suppose that the recurrent serious illness process follows model (1). To estimate

Table 3.3: Simulation results of two-sample tests with different weights for Case 1 when the distribution functions of the terminal events are different in the two groups.

β	Censoring rate 20%			Censoring rate 40%		
	$\tilde{T}_n(h_n^{(1)})$	$\tilde{T}_n(h_n^{(2)})$	$\tilde{T}_n(h_n^{(3)})$	$\tilde{T}_n(h_n^{(1)})$	$\tilde{T}_n(h_n^{(2)})$	$\tilde{T}_n(h_n^{(3)})$
			n_1=n_2=50			
1	0.070	0.063	0.067	0.077	0.075	0.087
1.1	0.175	0.162	0.171	0.193	0.164	0.192
1.2	0.469	0.416	0.471	0.489	0.439	0.484
1.3	0.753	0.721	0.755	0.789	0.758	0.779
			n_1=n_2=100			
1	0.056	0.061	0.061	0.068	0.056	0.072
1.1	0.231	0.221	0.231	0.260	0.241	0.260
1.2	0.688	0.664	0.693	0.721	0.687	0.717
1.3	0.960	0.956	0.960	0.956	0.951	0.955
			n_1=n_2=150			
1	0.043	0.049	0.041	0.056	0.052	0.054
1.1	0.338	0.311	0.341	0.352	0.330	0.348
1.2	0.863	0.835	0.865	0.872	0.851	0.869
1.3	0.993	0.992	0.994	0.994	0.985	0.992
			n_1=n_2=200			
1	0.046	0.048	0.045	0.058	0.057	0.060
1.1	0.447	0.427	0.450	0.469	0.451	0.469
1.2	0.951	0.933	0.953	0.952	0.941	0.954
1.3	1.000	0.998	1.000	0.999	0.999	0.998

the mean function of the reversed recurrent event process on serious illness and make a comparison between elderly people living in urban and rural, we took the order of I-spline $d = 3$ and $m_n = 3$ to divide $[0, \tau]$ into 4 subintervals, and let the knots of splines $t_{d+1} = \tau/4$, $t_{d+2} = \tau/2$, $t_{d+3} = 3\tau/4$. The estimated mean functions were displayed in Figure 3.5. The solid line in Figure 3.5 plots the estimated mean function based on the whole dataset. It shows that the times of serious illness tend to increase at the end of lifetime, which is reasonable since the physical function of the elder people would decline as their age increases. The dash and dotted-dash lines in Figure 3.5 display the mean function estimates for two groups living in the urban area and the rural area. It can be seen that people living in the former area tend to

Table 3.4: Simulation results of two-sample tests with different weights for Case 2 when the distribution functions of the terminal events are different in the two groups.

	Censoring rate 20%			Censoring rate 40%		
β	$\tilde{T}_n(h_n^{(1)})$	$\tilde{T}_n(h_n^{(2)})$	$\tilde{T}_n(h_n^{(3)})$	$\tilde{T}_n(h_n^{(1)})$	$\tilde{T}_n(h_n^{(2)})$	$\tilde{T}_n(h_n^{(3)})$
			n_1=n_2=50			
1	0.980	1.000	0.622	0.988	1.000	0.751
3	0.860	0.981	0.653	0.921	0.988	0.755
5	0.950	0.987	0.892	0.969	0.992	0.933
			n_1=n_2=100			
1	1.000	1.000	0.834	1.000	1.000	0.926
3	0.986	1.000	0.884	0.999	1.000	0.947
5	0.997	1.000	0.987	1.000	1.000	0.996
			n_1=n_2=150			
1	1.000	1.000	0.940	1.000	1.000	0.985
3	0.998	1.000	0.976	0.999	1.000	1.000
5	1.000	1.000	0.999	1.000	1.000	0.999
			n_1=n_2=200			
1	1.000	1.000	0.972	1.000	1.000	0.990
3	1.000	1.000	0.994	1.000	1.000	0.995
5	1.000	1.000	1.000	1.000	1.000	1.000

experience more serious illnesses than those living in the latter.

We then used the same three weight functions as those in the simulation studies to test the null hypothesis H_0: $\Lambda_U(s) = \Lambda_R(s)$, where Λ_U and Λ_R are the corresponding mean functions for people living in the urban area and the rural area, respectively. To conduct the test, we first draw that the survival functions of people living in urban and rural were significantly different by adopting the rank test (Harrington and Fleming, 1982) with p-value 3×10^{-6}. We divided $[0, \tau]$ into 8 intervals with equal length to obtain the histogram-type estimators, and used the statistics in Theorem 3.6 to compare the mean functions. The test results were summarized in Table 3.5, which suggests that the null hypothesis is rejected significantly at the level 0.05 for all three weights.

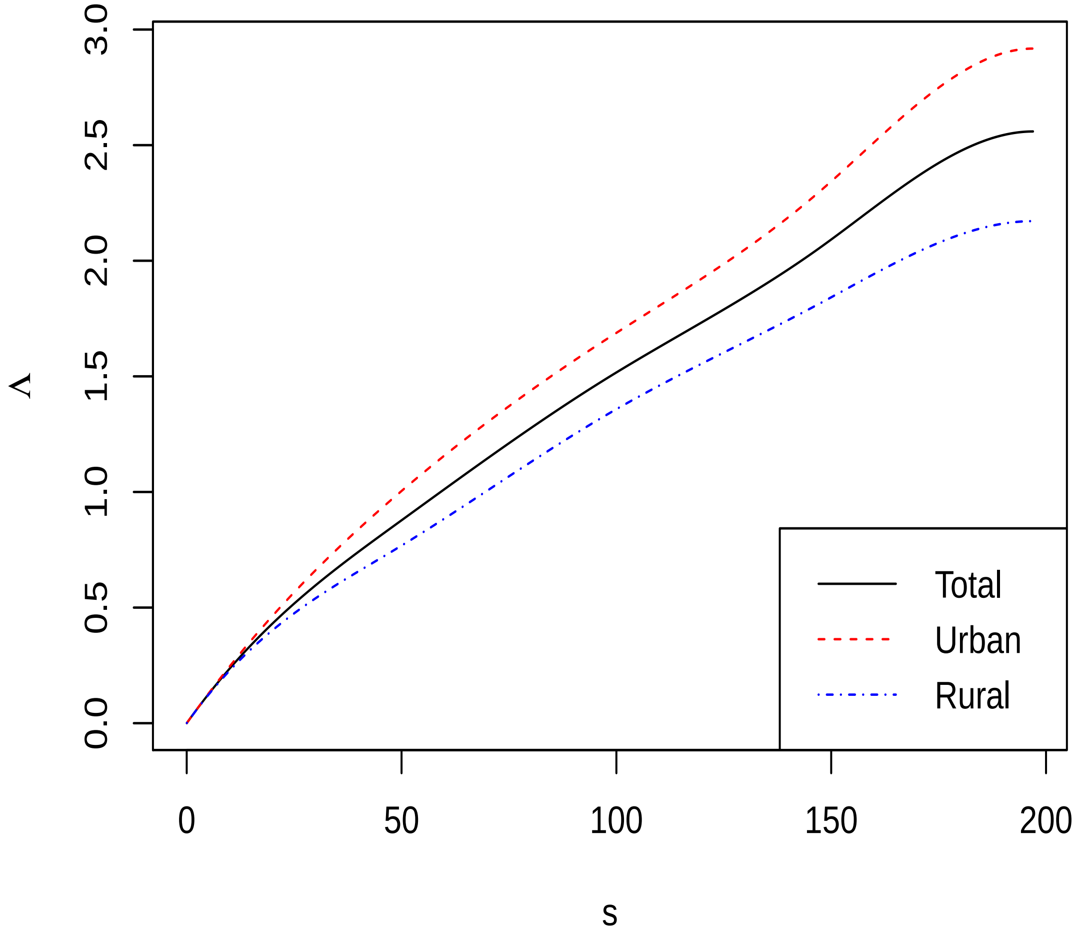


Figure 3.5: Estimates of the mean functions for the CLHLS data.

Table 3.5: Two-sample test results with three weights for the CLHLS data.

	$h_n^{(1)}$	$h_n^{(2)}$	$h_n^{(3)}$
$\tilde{U}_n$	287.788	3.982	4.357
$\tilde{\sigma}_n$	84.445	1.151	1.252
$\tilde{T}_n$	3.408*	3.460*	3.481*

$\tilde{T}_n(h)$ represents the observated value of the test statistic with different weight functions;
* represents significance level of 0.05

3.7 Appendix

3.7.1 Calculation of Loss Function

For the first part of $\ell_n(\Lambda, \hat{F}_n; X)$, replacing $\Lambda(s)$ by $\boldsymbol{I}^T(s)\boldsymbol{\alpha}$, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\left\{\triangle N_{i,j} - \triangle \boldsymbol{I}_j(Y_i)^T\boldsymbol{\alpha}\right\}^2 \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\left\{\boldsymbol{\alpha}^T\triangle \boldsymbol{I}_j(Y_i)\triangle \boldsymbol{I}_j(Y_i)^T\boldsymbol{\alpha} - 2\triangle N_{i,j}\triangle \boldsymbol{I}_j(Y_i)^T\boldsymbol{\alpha} + \triangle N_{i,j}^2\right\} \\
=&\boldsymbol{\alpha}^T\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\triangle \boldsymbol{I}_j(Y_i)\triangle \boldsymbol{I}_j(Y_i)^T\right\}\boldsymbol{\alpha} - \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\triangle N_{i,j}\triangle \boldsymbol{I}_j(Y_i)\right\}^T\boldsymbol{\alpha} \\
+&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\triangle N_{i,j}^2 \\
=:&\,\boldsymbol{\alpha}^T\boldsymbol{A}_1\boldsymbol{\alpha} - 2\boldsymbol{B}_1^T\boldsymbol{\alpha} + C_1.
\end{aligned}
$$

For the second part, the KM estimator is nondecreasing step function

$$
\hat{F}_n(u) = \sum_{l=1}^{L} f_l 1_{[t_l, t_{l+1})}(u),
$$

where $t_1 = 0$, $t_{L+1} = \tau$, and $\{[t_l, t_{l+1}) : l = 1, \cdots, L\}$ is a partition of $[0, \tau)$. Furthermore, for each subject, we assume Y_i to be in the interval $[t_{l_i}, t_{l_i+1})$, and take $f_{L+1} = f_L$. Then we have

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}(1-\Delta_i)\frac{\int_{Y_i}^{\infty}\{\triangle N_{i,j} - \triangle \boldsymbol{I}_j(u)^T\boldsymbol{\alpha}\}^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y_i)}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}(1-\Delta_i)\frac{\sum_{l=l_i}^{L}\left[(f_{l+1}-f_l)\{\triangle N_{i,j} - \triangle \boldsymbol{I}_j(t_{l+1}^{(1)})^T\boldsymbol{\alpha}\}^2\right]}{1-\hat{F}_n(Y_i)}$$

$$=\sum_{i=1}^{n}\frac{1-\Delta_i}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\sum_{l=l_i}^{L}\Big[(f_{l+1}-f_l)\Big\{\boldsymbol{\alpha}^T\triangle \boldsymbol{I}_j(t_{l+1}^{(1)})\triangle \boldsymbol{I}_j(t_{l+1}^{(1)})^T\boldsymbol{\alpha}$$

$$-2\triangle N_{i,j}\triangle \boldsymbol{I}_j(t_{l+1}^{(1)})^T\boldsymbol{\alpha} + \triangle N_{i,j}^2\Big\}\Big]$$

$$=\boldsymbol{\alpha}^T\left[\sum_{i=1}^{n}\frac{1-\Delta_i}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\sum_{l=l_i}^{L}\Big\{(f_{l+1}-f_l)\triangle \boldsymbol{I}_j(t_{l+1}^{(1)})\triangle \boldsymbol{I}_j(t_{l+1}^{(1)})^T\Big\}\right]\boldsymbol{\alpha}$$

$$-2\left[\sum_{i=1}^{n}\frac{1-\Delta_i}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\left\{\triangle N_{i,j}\sum_{l=l_i}^{L}(f_{l+1}-f_l)\triangle \boldsymbol{I}_j(t_{l+1}^{(1)})^T\right\}\right]^T\boldsymbol{\alpha}$$

$$+\sum_{i=1}^{n}\frac{(1-\Delta_i)(f_{L+1}-f_{l_i})}{n(1-\hat{F}_n(Y_i))}\sum_{j=1}^{K_i}\triangle N_{i,j}^2 =: \boldsymbol{\alpha}^T\boldsymbol{A}_2\boldsymbol{\alpha} - 2\boldsymbol{B}_2^T\boldsymbol{\alpha} + C_2.$$

Thus, to obtain the estimator $\hat{\Lambda}_n$, we should minimize $\boldsymbol{\alpha}^T\boldsymbol{A}\boldsymbol{\alpha} - 2\boldsymbol{B}^T\boldsymbol{\alpha} + C$ under the constraints that $\alpha_l \geq 0$ for $l = 1, \cdots, q_n$, where $\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2$, $\boldsymbol{B} = \boldsymbol{B}_1 + \boldsymbol{B}_2$ and $C = C_1 + C_2$.

3.7.2 Lemmas

Lemma 3.1. (i) *Suppose that Condition* (C2) *holds. For sufficiently small δ, any $F \in \mathcal{F}_\delta$ and any differentiable function g, we have*

$$\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\left|\frac{\int_Y^{\infty} g(u-T_j)dF(u)}{1-F(Y)} - \frac{\int_Y^{\infty} g(u-T_j)dF_0(u)}{1-F_0(Y)}\right|\right]$$
$$\lesssim \left(E\left[\sum_{j=1}^{K}|g'(U-T_j)|\right] + E\left[\sum_{j=1}^{K}|g(U-T_j)|\right]\right)||F-F_0||_\infty. \tag{3.4}$$

(ii) *In addition, suppose that Conditions* (C1) *and* (C3) *hold. It follows that for all* $\Lambda \in \Phi$,

$$|\mathcal{P}m(\Lambda, F; X) - \mathcal{P}m(\Lambda, F_0; X)| \lesssim d_2(F, F_0).$$

Thus, we have $\mathcal{P}m(\Lambda, \hat{F}_n; X) - \mathcal{P}m(\Lambda, F_0; X) = o_p(1)$ *for all* $\Lambda \in \Phi$.

Proof. (i) By direct calculations, we have

$$\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\left|\frac{\int_Y^\infty g(u-T_j)dF_0(u)}{1-F_0(Y)} - \frac{\int_Y^\infty g(u-T_j)dF(u)}{1-F(Y)}\right|\right]$$

$$=\mathcal{P}\left[\frac{1-\Delta}{(1-F_0(Y))(1-F(Y))}\sum_{j=1}^{K}\left|(1-F_0(Y))\int_Y^\infty g(u-T_j)d(F_0(u)-F(u))\right.\right.$$

$$\left.\left. + (F_0(Y)-F(Y))\int_Y^\infty g(u-T_j)dF_0(u)\right|\right]$$

$$\leq\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|\int_Y^\infty g(u-T_j)d(F_0(u)-F(u))|}{1-F(Y)}\right]$$

$$+\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty |g(u-T_j)|dF_0(u)}{(1-F_0(Y))(1-F(Y))}\right]||F-F_0||_\infty.$$

Using integration by parts, we have

$$\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|\int_Y^\infty g(u-T_j)d(F_0(u)-F(u))|}{1-F(Y)}\right]$$

$$=\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\left|-\big(F_0(Y)-F(Y)\big)g(Y-T_j) - \int_Y^\infty g'(u-T_j)(F_0(u)-F(u))du\right|}{1-F(Y)}\right]$$

$$\leq\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|g(Y-T_j)|}{1-F(Y)}\right]||F-F_0||_\infty + \mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty |g'(u-T_j)|du}{1-F(Y)}\right]||F-F_0||_\infty.$$

Condition (C2) implies that $1-F_0(Y)$ is larger than a positive constant. Hence, for sufficiently small δ and any $F \in \mathcal{F}_\delta$, $1-F(Y)$ is also larger than a positive constant.

It follows that

$$\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{|g(Y-T_j)|}{1-F(Y)}\right]\lesssim\mathcal{P}\left[\Delta\sum_{j=1}^{K}|g(Y-T_j)|\right]\leq\mathcal{P}\left[\sum_{j=1}^{K}|g(U-T_j)|\right]$$

and

$$\begin{aligned}&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g(u-T_j)|dF_0(u)}{(1-F_0(Y))(1-F(Y))}\right]\\ \lesssim&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g(u-T_j)|dF_0(u)}{1-F_0(Y)}\right]\leq\mathcal{P}\left[\sum_{j=1}^{K}|g(U-T_j)|\right].\end{aligned}$$

Moreover, by the second part of Condition (C2), we obtain

$$\begin{aligned}&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g'(u-T_j)|du}{1-F(Y)}\right]\lesssim\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g'(u-T_j)|/f_0(u)dF_0(u)}{1-F_0(Y)}\right]\\ \lesssim&\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\frac{\int_Y^\infty|g'(u-T_j)|dF_0(u)}{1-F_0(Y)}\right]\leq\mathcal{P}\left[\sum_{j=1}^{K}\left|g'(U-T_j)\right|\right].\end{aligned}$$

Thus, (3.4) holds.

(ii) By the first part of the Lemma, Conditions (C1) and (C3), we have

$$\begin{aligned}&|\mathcal{P}[m(\Lambda,F;X)-m(\Lambda,F_0;X)]|\\ \leq&\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty\left(\triangle N_j-\triangle\Lambda_j(u)\right)^2dF(u)}{1-F(Y)}-\frac{\int_Y^\infty\left(\triangle N_j-\triangle\Lambda_j(u)\right)^2dF_0(u)}{1-F_0(Y)}\right|\right]\\ \lesssim&\left(E\left[\sum_{j=1}^{K}\left\{\left|\Lambda_j'(U)\left(\triangle N_j-\triangle\Lambda_j(U)\right)\right|+\left(\triangle N_j-\triangle\Lambda_j(U)\right)^2\right\}\right]\right)||F-F_0||_\infty\\ \lesssim&d_2(F,F_0).\end{aligned}$$

□

Lemma 3.2. *Suppose that Conditions* (C2), (C4) *and* (C6) *hold. Then for sufficiently small* δ, $\{m(\Lambda, F; X) : \Lambda \in \Phi,\ F \in \mathcal{F}_\delta,\ \Lambda$ *is uniformly bounded*$\}$ *is Donsker, where*

$$m(\Lambda, F; X) = \sum_{j=1}^{K} \Delta\{\triangle N_j - \triangle\Lambda_j(Y)\}^2 + \sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty \{\triangle N_j - \triangle\Lambda_j(u)\}^2 dF(u)}{1-F(Y)}.$$

Proof. Note that functions in $\mathcal{F}_\delta$ and Φ are monotone and uniformly bounded. It follows from Section 3 of van der Vaart (1996) that $\{\Lambda \in \Phi : \Lambda \text{ is uniformly bounded}\}$ and $\mathcal{F}_\delta$ are Donsker. Since $\{\triangle N_j - \triangle\Lambda_j(Y)\}^2$ is Lipschitz for Λ, by Theorem 2.10.6 of van der Vaart and Wellner (1996), $\big\{[\triangle N_j - \triangle\Lambda_j(Y)]^2 : \Lambda \in \Phi \text{ is uniformly bounded}\big\}$ is Donsker. Note that

$$\left\{\int_Y^\infty \big[\triangle N_j - \triangle\Lambda_j(u)\big]^2 dF(u) : \Lambda \in \Phi \text{ is uniformly bounded}\right\}$$

is a subset of the convex combinations of functions in

$$\big\{[\triangle N_j - \triangle\Lambda_j(Y)]^2 : \Lambda \in \Phi \text{ is uniformly bounded}\big\}.$$

By Theorem 2.10.1 and Theorem 2.10.3 of van der Vaart and Wellner (1996),

$$\left\{\int_Y^\infty [\triangle N_j - \triangle\Lambda_j(u)]^2 dF(u) : \Lambda \in \Phi \text{ is uniformly bounded}\right\}$$

is Donsker. Since $1-F(Y)$ and K are bounded from Conditions (C2), (C4) and (C6), $\{m(\Lambda, F; X) : \Lambda \in \Phi,\ F \in \mathcal{F}_\delta,\ \Lambda \text{ is uniformly bounded}\}$ is Donsker by Theorem 2.10.6 of van der Vaart and Wellner (1996). □

Lemma 3.3 (Rate of Convergence of M-estimator with Nuisance Parameter)**.** *Suppose that for every* $\Lambda \in \Phi_n$, *sufficiently large* n *and sufficiently small* η,

$$\mathcal{P}(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \lesssim -d_1(\Lambda, \Lambda_0)^2 + d_1(\Lambda, \Lambda_0)d_2(\hat{F}_n, F_0) + d_2(\hat{F}_n, F_0)^2$$

and

$$E \sup_{\{\Lambda \in \Phi_n : d_1(\Lambda, \Lambda_0) < \eta\}} |(\mathbb{P}_n - \mathcal{P})(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X))| \lesssim \frac{\phi_n(\eta)}{\sqrt{n}}$$

hold, where $\phi_n(\eta)$ *satisfies that* $\eta \mapsto \phi_n(\eta)/\eta^\alpha$ *is decreasing for some* $\alpha < 2$. *Let* $r_n > 0$ *satisfy* $\phi_n(r_n) \lesssim \sqrt{n} r_n^2$. *If the sequence* $\hat{\Lambda}_n$ *satisfies* $\mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \geq \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - O_p(r_n^2)$ *and converges in probability to* Λ_0*, then* $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(r_n + d_2(\hat{F}_n, F_0))$.

Proof. This Lemma is similar to Theorem 5.55 of van der Vaart (1998). In order to verify $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(r_n + d_2(\hat{F}_n, F_0))$, we need to prove that for sufficiently large n, $\lim_{M \to \infty} P(\hat{\Lambda}_n \in \Phi_n : d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M(d_2(\hat{F}_n, F_0) + r_n)) = 0$. Then we divide Φ_n into shells $S_{n,j,M} = \{\Lambda \in \Phi_n : 2^j r_n \leq d_1(\Lambda, \Lambda_0) < 2^{j+1} r_n, 2^M d_2(\hat{F}_n, F_0) \leq d_1(\Lambda, \Lambda_0)\}$. For any $\Lambda \in S_{n,j,M}$, we have $2d_1(\Lambda, \Lambda_0) \geq 2^M d_2(\hat{F}_n, F_0) + 2^j r_n$. Hence,

$$\left\{\hat{\Lambda}_n \in \Phi_n : 2d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M(d_2(\hat{F}_n, F_0) + r_n)\right\} \subseteq \bigcup_{j \geq M} \{\hat{\Lambda}_n \in S_{n,j,M}\}.$$

It follows that

$$P\left(\hat{\Lambda}_n \in \Phi_n : d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^{M-1}(d_2(\hat{F}_n, F_0) + r_n)\right) \leq P\left(\hat{\Lambda}_n \in \bigcup_{j \geq M} S_{n,j,M}\right). \quad (3.5)$$

Furthermore, since $\hat{\Lambda}_n$ satisfies that $\mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \geq \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - O_p(r_n^2)$, for $\hat{\Lambda}_n \in S_{n,j,M}$ we can find a variable $R_n = O_p(r_n^2)$ such that

$$\sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\hat{\Lambda}_n, \hat{F}_n; X)) \geq -R_n.$$

Then for any constant κ, we have

$$P\left(\hat{\Lambda}_n \in \bigcup_{j \geq M} S_{n,j,M}\right) \leq P\left(\sup_{\Lambda \in \bigcup_{j \geq M} S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq -\kappa r_n^2\right) + P(R_n \geq \kappa r_n^2). \quad (3.6)$$

By (3.5) and (3.6), we obtain

$$P\left(\hat{\Lambda}_n \in \Phi_n : d_1(\hat{\Lambda}_n, \Lambda_0) \geq 2^M (d_2(\hat{F}_n, F_0) + r_n)\right) \leq P(d_1(\hat{\Lambda}_n, \Lambda_0) \geq \eta) + P(R_n \geq \kappa r_n^2)$$

$$+ \sum_{j \geq M+1, 2^{j+1} \leq \eta / r_n} P\left(\sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq -\kappa r_n^2 \right).$$

Since $\hat{\Lambda}_n$ is consistent and $R_n = O_p(r_n^2)$, $P(d_1(\hat{\Lambda}_n, \Lambda_0) \geq \eta)$ and $P(R_n \geq \kappa r_n^2)$ can be arbitrarily small for sufficiently large n by the choice of η and κ. Thus, we need to prove the limitation of the summation on the right hand side is 0 as M goes to infinity.

Note that for any positive integer M, $1/4 \leq 1 - 2^{-M} - 2^{-2M} < 1$. Then for all $\Lambda \in S_{n,j,M}$,

$$\begin{aligned}\mathcal{P}(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) &\lesssim -d_1(\Lambda, \Lambda_0)^2 + d_1(\Lambda, \Lambda_0) d_2(\hat{F}_n, F_0) + d_2(\hat{F}_n, F_0)^2 \\ &\leq -(1 - 2^{-M} - 2^{-2M}) d_1(\Lambda, \Lambda_0)^2 \leq -2^{2j} r_n^2.\end{aligned}$$

Hence, $\mathcal{P}(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \leq -c_1 2^{2j} r_n^2$ for some constant c_1. Taking M with $M \geq \frac{1}{2} \log_2(2\kappa / c_1)$, then by the Markov's inequality, for $j \geq M + 1$ and sufficiently large n with $r_n \leq 2^{-(j+1)} \eta$, we have

$$\begin{aligned}&P\left(\sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq -\kappa r_n^2 \right) \\ \leq &P\left(\sup_{\Lambda \in S_{n,j,M}} (\mathbb{P}_n - \mathcal{P})(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq \frac{c_1}{2} 2^{2j} r_n^2 \right) \\ \leq &\frac{2}{c_1 2^{2j} r_n^2} E \sup_{\Lambda \in S_{n,j,M}} \left| (\mathbb{P}_n - \mathcal{P})(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \right| \lesssim \frac{\phi_n(2^{(j+1)} r_n)}{2^{2j} r_n^2 \sqrt{n}}.\end{aligned}$$

Since $\phi_n(\eta)/\eta^\alpha$ is decreasing for some $\alpha < 2$, we have $\phi_n(c\eta) \leq c^\alpha \phi_n(\eta)$ for any $c > 1$. Then $\phi_n(r_n) \lesssim \sqrt{n} r_n^2$ ensures that $\frac{\phi_n(2^{(j+1)} r_n)}{2^{2j} r_n^2 \sqrt{n}} \lesssim \frac{2^{\alpha(j+1)} \sqrt{n} r_n^2}{2^{2j} r_n^2 \sqrt{n}} = \frac{1}{2^{(2-\alpha)j - \alpha}}$. Thus,

$$\sum_{j \geq M+1, 2^{j+1} \leq \eta / r_n} P\left(\sup_{\Lambda \in S_{n,j,M}} \mathbb{P}_n(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)) \geq 0 \right) \lesssim \sum_{j \geq M+1} \frac{1}{2^{(2-\alpha)j - \alpha}}.$$

Noting that $\sum_{j\geq M+1}\frac{1}{2^{(2-\alpha)j-\alpha}}$ tends to 0 as M approaches to infinity, this lemma is concluded. $\square$

Lemma 3.4. *Suppose that Conditions* (C1), (C2), (C4)–(C6), (C8), (C9) *and* (C11) *hold. Define the class* $\mathcal{M}_\eta(F) = \{m(\Lambda, F; X) - m(\Lambda_0, F; X) : \Lambda \in \Phi_n, d_1(\Lambda, \Lambda_0) \leq \eta\}$ *for each* $F \in \mathcal{F}_\delta$. *For any* $\varepsilon < \eta$ *and sufficiently small* δ, *we have*

$$\log N_{[]}(\varepsilon, \mathcal{M}_\eta(F), ||\cdot||_{P,B}) \lesssim q_n \log(\eta/\varepsilon),$$

where the Bernstein norm is defined as $||f||_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$.

Proof. By the calculation in Shen and Wong (1994, page 597), under Condition (C5), for any $\varepsilon < \eta$, there is a set of brackets

$$\left\{[\Lambda_i^L, \Lambda_i^U] : ||\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon, i = 1, \cdots, (\eta/\varepsilon)^{c_0 q_n}\right\}$$

such that for all $\Lambda \in \{\Lambda \in \Phi_n : d_1(\Lambda, \Lambda_0) \leq \eta\}$, we can find an interval $[\triangle\Lambda_i^L, \triangle\Lambda_i^U]$ satisfying $\triangle\Lambda(s_1, s_2) \in [\triangle\Lambda_i^L(s_1, s_2), \triangle\Lambda_i^U(s_1, s_2)]$. This implies that $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon$ and $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_i^U(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon$. Noting that $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq \eta$, we have $||\triangle\Lambda_i^L(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq (\varepsilon^2 + \eta^2)^{1/2}$ and $||\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq (\varepsilon^2 + \eta^2)^{1/2}$. By Lemma 7.1 of Wellner and Zhang (2007), under Conditions (C8) and (C9), for any $\triangle\Lambda(s_1, s_2)$ satisfying $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_{L_2(\mu_1)} \leq \varepsilon^*$, we have $||\triangle\Lambda(s_1, s_2) - \triangle\Lambda_0(s_1, s_2)||_\infty \leq (\varepsilon^*/c_1)^{2/3}$ for some constant c_1. Hence,

$$\begin{aligned} 0 \vee \left(\triangle\Lambda_0(s_1, s_2) - ((\varepsilon^2 + \eta^2)^{1/2}/c_1)^{2/3}\right) &\leq \triangle\Lambda_i^L(s_1, s_2) \\ \leq \triangle\Lambda_i^U(s_1, s_2) &\leq \triangle\Lambda_0(s_1, s_2) + ((\varepsilon^2 + \eta^2)^{1/2}/c_1)^{2/3}, \end{aligned}$$

which implies that $\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2)$ are uniformly bounded by $2((\varepsilon^2 + \eta^2)^{1/2}/c_1)^{2/3}$.

We turn to consider the ε-bracket of $m(\Lambda, F; X) - m(\Lambda_0, F; X)$. Note that

$$m(\Lambda, F; X) - m(\Lambda_0, F; X) = \sum_{j=1}^{K} \Delta \Big[(\triangle N_j - \triangle\Lambda_j(Y))^2 - (\triangle N_j - \triangle\Lambda_{0,j}(Y))^2 \Big]$$

$$+ \sum_{j=1}^{K} (1-\Delta) \left[\frac{\int_Y^\infty (\triangle N_j - \triangle\Lambda_j(u))^2 - (\triangle N_j - \triangle\Lambda_{0,j}(u))^2 \, dF(u)}{1 - F(Y)} \right]$$

$$= \sum_{j=1}^{K} \Delta \left[\{\triangle\Lambda_j(Y)\}^2 - 2\triangle N_j \triangle\Lambda_j(Y) - \{\triangle\Lambda_{0,j}(Y)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(Y) \right]$$

$$+ \sum_{j=1}^{K} (1-\Delta) \frac{\int_Y^\infty \left[\{\triangle\Lambda_j(u)\}^2 - 2\triangle N_j \triangle\Lambda_j(u) - \{\triangle\Lambda_{0,j}(u)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(u) \right] dF(u)}{1 - F(Y)}.$$

By Conditions (C2) and (C4), $1 - F(Y)$ is positive and has uniform upper and lower bounds. Hence, $m_i^L(\Lambda_i^L, \Lambda_i^U, F; X) \le m(\Lambda, F; X) - m(\Lambda_0, F; X) \le m_i^U(\Lambda_i^L, \Lambda_i^U, F; X)$, where

$$m_i^L(\Lambda_i^L, \Lambda_i^U, F; X)$$

$$= \sum_{j=1}^{K} \Delta \Big[\{\triangle\Lambda_{i,j}^L(Y)\}^2 - 2\triangle N_j \triangle\Lambda_{i,j}^U(Y) - \{\triangle\Lambda_{0,j}(Y)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(Y) \Big]$$

$$+ \sum_{j=1}^{K} (1-\Delta) \frac{\int_Y^\infty \Big[\{\triangle\Lambda_{i,j}^L(u)\}^2 - 2\triangle N_j \triangle\Lambda_{i,j}^U(u) - \{\triangle\Lambda_{0,j}(u)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(u) \Big] dF(u)}{1 - F(Y)}$$

and

$$m_i^U(\Lambda_i^L, \Lambda_i^U, F; X)$$

$$= \sum_{j=1}^{K} \Delta \Big[\{\triangle\Lambda_{i,j}^U(Y)\}^2 - 2\triangle N_j \triangle\Lambda_{i,j}^L(Y) - \{\triangle\Lambda_{0,j}(Y)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(Y) \Big]$$

$$+ \sum_{j=1}^{K} (1-\Delta) \frac{\int_Y^\infty \Big[\{\triangle\Lambda_{i,j}^U(u)\}^2 - 2\triangle N_j \triangle\Lambda_{i,j}^L(u) - \{\triangle\Lambda_{0,j}(u)\}^2 + 2\triangle N_j \triangle\Lambda_{0,j}(u) \Big] dF(u)}{1 - F(Y)}$$

for $i = 1, \cdots, (\eta/\varepsilon)^{c_0 q_n}$.

We also need to verify $||m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)||_{P,B}^2 \lesssim \varepsilon^2$. According to

$$||f||_{P,B}^2 = 2P(e^{|f|}-1-|f|) = 2P\left(\sum_{n=2}^{\infty}\frac{|f|^n}{n!}\right) \le 2P\left(|f|^2 e^{|f|}\right),$$

$||f||_{P,B}^2 \lesssim \varepsilon^2$ is followed from $P(|f|^2e^{|f|}) \lesssim \varepsilon^2$. Note that

$$\begin{aligned}
&\{\triangle\Lambda_{i,j}^U(u)\}^2 - \{\triangle\Lambda_{i,j}^L(u)\}^2 + 2\triangle N_j(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)) \\
=&\left(\triangle\Lambda_{i,j}^U(u) + \triangle\Lambda_{i,j}^L(u) + 2\triangle N_j\right)\left(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)\right) \\
\lesssim&\triangle N_j\left(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)\right) + \left(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)\right).
\end{aligned}$$

Since Λ_i^L, Λ_i^U and $1-F(Y)$ are uniformly bounded, we have

$$e^{|m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)|} \lesssim e^{cN(T_K)}$$

with some constant c. By Cauchy-Schwarz inequality and Condition (C11), we obtain

$$\begin{aligned}
&\mathcal{P}\left(|m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)|^2 e^{|m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)|}\right) \\
\lesssim&\mathcal{P}\left(e^{cN(T_K)}\left|m_i^U(\Lambda_i^L,\Lambda_i^U,F;X)-m_i^L(\Lambda_i^L,\Lambda_i^U,F;X)\right|^2\right) \\
\lesssim&\mathcal{P}\left[\left\{\Delta\sum_{j=1}^K(\triangle N_j+1)\left(\triangle\Lambda_{i,j}^U(Y)-\triangle\Lambda_{i,j}^L(Y)\right)\right.\right. \\
&\left.\left.+\ (1-\Delta)\sum_{j=1}^K\frac{\int_Y^\infty(\triangle N_j+1)\left(\triangle\Lambda_{i,j}^U(u)-\triangle\Lambda_{i,j}^L(u)\right)dF(u)}{1-F(Y)}\right\}^2\right] \\
\lesssim&\mathcal{P}\left[\sum_{j=1}^K\left\{\Delta\left(\triangle\Lambda_{i,j}^U(Y)-\Lambda_{i,j}^L(Y)\right)^2+(1-\Delta)\frac{\int_Y^\infty\left(\triangle\Lambda_{i,j}^U(u)-\triangle\Lambda_{i,j}^L(u)\right)^2dF(u)}{1-F(Y)}\right\}\right] \\
\lesssim&\delta+||\triangle\Lambda_i^U(s_1,s_2)-\triangle\Lambda_i^L(s_1,s_2)||_{L_2(\mu_1)}^2 \lesssim \varepsilon^2.
\end{aligned} \tag{3.7}$$

This implies that $||m_i^L-m_i^U||_{P,B}^2 \lesssim \varepsilon^2$. That means $N_{[]}(\varepsilon,\mathcal{M}_\eta(F),||\cdot||_{P,B}) \le (\eta/\varepsilon)^{c_0q_n}$ and $\log N_{[]}(\varepsilon,\mathcal{M}_\eta(F),||\cdot||_{P,B}) \lesssim q_n\log(\eta/\varepsilon)$. □

3.7.3 Proof of Theorem 3.1

Proof. First, for any $\Lambda \in \Phi$, we have

$$
\begin{aligned}
&\mathcal{P}m(\Lambda, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X)\\
=&\mathcal{P}\left[\sum_{j=1}^{K}\Delta\{(\triangle N_j - \triangle\Lambda_j(Y))^2 - (\triangle N_j - \triangle\Lambda_{0,j}(Y))^2\}\right.\\
&\left.+\sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty\{(\triangle N_j - \triangle\Lambda_j(u))^2 - (\triangle N_j - \triangle\Lambda_{0,j}(u))^2\}dF_0(u)}{1-F_0(Y)}\right]\\
=&\mathcal{P}\left[\sum_{j=1}^{K}\Delta\left(2\triangle N_j - \triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y)\right)\left(\triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y)\right)\right.\\
&\left.+\sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty\left(2\triangle N_j - \triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u)\right)\left(\triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u)\right)dF_0(u)}{1-F_0(Y)}\right]\\
=&\mathcal{P}\left[\sum_{j=1}^{K}\Delta\big\{\triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y)\big\}^2 + \sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty\big\{\triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u)\big\}^2 dF_0(u)}{1-F_0(Y)}\right]\\
=&d_1(\Lambda,\Lambda_0)^2.
\end{aligned}
$$

Thus, to draw the conclusion, we only need to consider $\mathcal{P}m(\Lambda, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X)$.

Since $\hat{\Lambda}_n$ is the minimizer of $\mathbb{P}_n m(\Lambda, \hat{F}_n; X)$ with respect to $\Lambda \in \Phi_n$, for any direction function $h \in \Phi_n$, we obtain

$$
\begin{aligned}
0 &= \lim_{\varepsilon\to 0}\frac{\mathbb{P}_n L(\hat{\Lambda}_n + \varepsilon h, \hat{F}_n; X) - \mathbb{P}_n L(\hat{\Lambda}_n, \hat{F}_n; X)}{\varepsilon}\\
&= -2\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle N_j - \triangle\hat{\Lambda}_{n,j}(Y)\big)\triangle h_j(Y)\right.\right.\\
&\quad\left.\left.+\ (1-\Delta)\frac{\int_Y^\infty\big(\triangle N_j - \triangle\hat{\Lambda}_{n,j}(u)\big)\triangle h_j(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right].
\end{aligned}
\tag{3.8}
$$

Taking $h(s) = s$, by Conditions (C1) and (C6), it follows that

$$
\begin{aligned}
&\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{(T_j - T_{j-1})\left(\Delta\triangle\hat{\Lambda}_{n,j}(Y) + (1-\Delta)\frac{\int_Y^\infty \triangle\hat{\Lambda}_{n,j}(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right)\right\}\right] \\
=&\mathbb{P}_n\left[\sum_{j=1}^{K}\{\triangle N_j(T_j - T_{j-1})\}\right] \xrightarrow{a.s.} E\left[\sum_{j=1}^{K}\triangle\Lambda_{0,j}(U)(T_j - T_{j-1})\right] \leq M_3\Lambda_0(\tau)\tau,
\end{aligned}
$$

where the almost surely convergence follows from the strong law of large number. Moreover, for the left hand side, by Condition (C7), we have

$$
\begin{aligned}
&\limsup_{n\to\infty}\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{(T_j - T_{j-1})\left(\Delta\triangle\hat{\Lambda}_{n,j}(Y) + (1-\Delta)\frac{\int_Y^\infty \triangle\hat{\Lambda}_{n,j}(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right)\right\}\right] \\
\gtrsim&\limsup_{n\to\infty}\mathbb{P}_n\left[\Delta\{\hat{\Lambda}_n(Y) - \hat{\Lambda}_n(Y-T_K)\} + (1-\Delta)\frac{\int_Y^\infty\{\hat{\Lambda}_n(u) - \hat{\Lambda}_n(u-T_K)\}d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right] \\
\geq&\limsup_{n\to\infty}\triangle\hat{\Lambda}_n(b_1,b_2)\mathbb{P}_n\left[\Delta 1_{\{Y-T_K\in[0,b_1],Y\in[b_2,\tau]\}} + (1-\Delta)\frac{\int_Y^\infty 1_{\{u-T_K\in[0,b_1],u\in[b_2,\tau]\}}d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right] \\
=&\limsup_{n\to\infty}\triangle\hat{\Lambda}_n(b_1,b_2)E\left[1_{\{U-T_K\in[0,b_1],U\in[b_2,\tau]\}}\right] = \limsup_{n\to\infty}\triangle\hat{\Lambda}_n(b_1,b_2)\mu_2([0,b_1]\times[b_2,\tau]).
\end{aligned}
$$

Hence, for every $0 \leq b_1 \leq b_2 \leq \tau$ satisfying $\mu_2([0,b_1]\times[b_2,\tau]) > 0$, we have

$$
\triangle\hat{\Lambda}_n(s_1,s_2)1_{\{(s_1,s_2)\in[b_1,b_2]\times[b_1,b_2]\}}
$$

is uniformly bounded. In particular, if $\mu_2(\{0\}\times\{\tau\}) > 0$, then $\hat{\Lambda}_n(s)$ is uniformly bounded.

By Lemma A1 of Lu, Zhang, and Huang (2007), under Condition (C5), there is $\Lambda_n^* \in \Phi_n$ such that $||\Lambda_n^* - \Lambda_0||_\infty = O(n^{-\nu r})$. This implies that

$$
\begin{aligned}
&\mathcal{P}m(\Lambda_n^*, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X) \\
=&\mathcal{P}\left[\sum_{j=1}^{K}\Delta\big(\triangle\Lambda_{0,j}(Y) - \triangle\Lambda_{n,j}^*(Y)\big)^2 + \sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty\big(\triangle\Lambda_{0,j}(u) - \triangle\Lambda_{n,j}^*(u)\big)^2dF_0(u)}{1-F_0(Y)}\right] \\
\lesssim&||\Lambda_n^* - \Lambda_0||_\infty^2 = O(n^{-2\nu r}) = o(1).
\end{aligned}
$$

Note that

$$0 \le \mathcal{P}m(\hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X) = \mathcal{P}m(\hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\hat{\Lambda}_n, \hat{F}_n; X)$$
$$+\mathcal{P}m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) + \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X)$$
$$+\mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\Lambda_n^*, \hat{F}_n; X) + \mathcal{P}m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\Lambda_n^*, F_0; X)$$
$$+\mathcal{P}m(\Lambda_n^*, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X).$$

By Lemma 3.1, we have

$$\mathcal{P}m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathcal{P}m(\hat{\Lambda}_n, F_0; X) = o_p(1)$$

and

$$\mathcal{P}m(\Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\Lambda_n^*, F_0; X) = o_p(1).$$

By Lemma 3.2, the class of functions $\{m(\Lambda, F; X) : \Lambda \in \Phi_n, F \in \mathcal{F}_\delta\}$ is Donsker. Hence it is Glivenko-Cantelli, and we have

$$(\mathbb{P}_n - \mathcal{P})m(\Lambda_n^*, \hat{F}_n; X) = o_p(1) \text{ and } (\mathbb{P}_n - \mathcal{P})m(\hat{\Lambda}_n, \hat{F}_n; X) = o_p(1)$$

since $\hat{\Lambda}_n$ is uniformly bounded. According to the definition of $\hat{\Lambda}_n$, $\mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_n^*, \hat{F}_n; X) \le 0$. Hence, $d_1(\hat{\Lambda}_n, \Lambda_0) = \mathcal{P}m(\hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\Lambda_0, F_0; X) = o_p(1)$. □

3.7.4 Proof of Theorem 3.2

Proof. To apply Lemma 3.3, we need to verify that for every $\Lambda \in \Phi_n$ and sufficiently large n, the inequalities

$$\mathcal{P}\Big(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)\Big) \lesssim -d_1(\Lambda, \Lambda_0)^2 + d_1(\Lambda, \Lambda_0)d_2(\hat{F}_n, F_0) + d_2(\hat{F}_n, F_0)^2$$

and

$$E \sup_{\{\Lambda \in \Phi_n : d_1(\Lambda, \Lambda_0) < \eta\}} \Big|\big(\mathbb{P}_n - \mathcal{P}\big)\big(m(\Lambda_0, \hat{F}_n; X) - m(\Lambda, \hat{F}_n; X)\big)\Big| \lesssim \frac{\phi_n(\eta)}{\sqrt{n}}$$

hold. By some calculations,

$$\begin{aligned}
&\mathcal{P}\big(m(\Lambda_0, F; X) - m(\Lambda, F; X)\big)\\
=&\mathcal{P}\left[\sum_{j=1}^{K} \Delta\left\{\big(\triangle N_j - \triangle\Lambda_{0,j}(Y)\big)^2 - \big(\triangle N_j - \triangle\Lambda_j(Y)\big)^2\right\}\right.\\
&+\left.\sum_{j=1}^{K}(1-\Delta)\left\{\frac{\int_Y^\infty \big(\triangle N_j - \triangle\Lambda_{0,j}(u)\big)^2 - \big(\triangle N_j - \triangle\Lambda_j(u)\big)^2 dF(u)}{1-F(Y)}\right\}\right]\\
=&\mathcal{P}\left[\sum_{j=1}^{K} \Delta\big\{2\triangle N_j - \triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y)\big\}\big\{\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y)\big\}\right.\\
&+\left.\sum_{j=1}^{K}(1-\Delta)\frac{\int_Y^\infty \big\{2\triangle N_j - \triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u)\big\}\big\{\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u)\big\}dF(u)}{1-F(Y)}\right]\\
=&-\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y))^2 + (1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF(u)}{1-F(Y)}\right\}\right]\\
=&-d_1(\Lambda,\Lambda_0)^2 + \mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left\{\frac{\int_Y^\infty(\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF_0(u)}{1-F_0(Y)}\right.\right.\\
-&\left.\left.\frac{\int_Y^\infty(\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF(u)}{1-F(Y)}\right\}\right].
\end{aligned}$$

Moreover, by the first part of Lemma 3.1, under Condition (C1), we have

$$\begin{aligned}
&\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left\{\frac{\int_Y^\infty(\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF_0(u)}{1-F_0(Y)} - \frac{\int_Y^\infty(\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 dF(u)}{1-F(Y)}\right\}\right]\\
\lesssim&\mathcal{P}\left[\sum_{j=1}^{K} 2\left|\{\triangle\Lambda'_j(U) - \triangle\Lambda'_{0,j}(U)\}\{\triangle\Lambda_j(U) - \triangle\Lambda_{0,j}(U)\}\right|\right]||F-F_0||_\infty\\
+&\mathcal{P}\left[\sum_{j=1}^{K}\{\triangle\Lambda_j(U) - \triangle\Lambda_{0,j}(U)\}^2\right]||F-F_0||_\infty\\
\lesssim& d_1(\Lambda,\Lambda_0)d_2(F,F_0) + d_1(\Lambda,\Lambda_0)^2 d_2(F,F_0).
\end{aligned}$$

This implies that

$$\mathcal{P}m(\Lambda_0, \hat{F}_n; X) - \mathcal{P}m(\Lambda, \hat{F}_n; X) \lesssim -d_1(\Lambda, \Lambda_0)^2 + d_1(\Lambda, \Lambda_0)d_2(\hat{F}_n, F_0) + d_1(\Lambda, \Lambda_0)^2 d_2(\hat{F}_n, F_0).$$

Second, we need to find a $\phi_n(\eta)$ such that

$$E \sup_{\{\Lambda \in \Phi_n : d_1(\Lambda, \Lambda_0) < \eta\}} |(\mathbb{P}_n - \mathcal{P})(m(\Lambda, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X))| \lesssim \frac{\phi_n(\eta)}{\sqrt{n}}.$$

By Lemma 3.4, we have $\log N_{[]}(\varepsilon, \mathcal{M}_\eta(\hat{F}_n), || \cdot ||_{P,B}) \lesssim q_n \log(\eta/\varepsilon)$, where

$$\mathcal{M}_\eta(\hat{F}_n) = \{m(\Lambda, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X) : \Lambda \in \Phi_n, d_1(\Lambda, \Lambda_0) \leq \eta\}.$$

For all Λ satisfying that $\Lambda \in \Phi_n, d_1(\Lambda, \Lambda_0) \leq \eta$, note that

$$\begin{aligned}
&|m(\Lambda, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)| \\
=&\sum_{j=1}^{K} \Delta \left|\big(\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y)\big)\big(\triangle\Lambda_j(Y) + \triangle\Lambda_{0,j}(Y) - 2\triangle N_j\big)\right| \\
+&\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty \big(\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u)\big)\big(\triangle\Lambda_j(u) + \triangle\Lambda_{0,j}(u) - 2\triangle N_j\big)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right| \\
\lesssim&\sum_{j=1}^{K}\left[(\triangle N_j + 1)\left\{\Delta|\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y)| + (1-\Delta)\frac{\int_Y^\infty |\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u)|d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right].
\end{aligned}$$

Similar to the proof of (3.7), since $\hat{\Lambda}_n$ is uniformly bounded, it follows that

$$e^{|m(\hat{\Lambda}_n, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)|} \lesssim e^{cN(T_K)}$$

and

$$\begin{aligned}
&\mathcal{P}\left[e^{|m(\hat{\Lambda}_n, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)|}|m(\hat{\Lambda}_n, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)|^2\right] \\
\lesssim&\mathcal{P}\left[e^{cN(T_K)}\sum_{j=1}^{K}\left\{\Delta(\triangle N_j + 1)^2(\triangle\Lambda_j(Y) - \triangle\Lambda_{0,j}(Y))^2\right.\right. \\
&\left.\left.+\ (1-\Delta)(\triangle N_j + 1)^2\frac{\int_Y^\infty(\triangle\Lambda_j(u) - \triangle\Lambda_{0,j}(u))^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]
\end{aligned}$$

$$\lesssim \mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_j(Y)-\triangle\Lambda_{0,j}(Y))^2+(1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_j(u)-\triangle\Lambda_{0,j}(u))^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]$$

$$\lesssim d_1^2(\hat{\Lambda}_n,\Lambda_0)+d_1(\hat{\Lambda}_n,\Lambda_0)d_2(\hat{F}_n,F_0).$$

That means that for sufficiently large n with $d_2(\hat{F}_n, F_0) \lesssim \eta$, we have

$$||m(\hat{\Lambda}_n,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X)||_{P,B}^2 \lesssim \eta^2.$$

Then Lemma 3.4.3 of van der Vaart and Wellner (1996) yields that

$$E||n^{1/2}(\mathbb{P}_n-\mathcal{P})||_{\mathcal{M}_\eta(\hat{F}_n)} \lesssim J_{[]}(\eta,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B})\left\{1+\frac{J_{[]}(\eta,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B})}{\eta^2 n^{1/2}}\right\},$$

where $J_{[]}(\eta,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B}) := \int_0^\eta\{1+\log N_{[]}(\varepsilon,\mathcal{M}_\eta(\hat{F}_n),||\cdot||_{P,B})\}^{1/2}d\varepsilon \lesssim q_n^{1/2}\eta$. It follows that

$$E\sup_{\{\Lambda\in\Phi_n:d_1(\Lambda,\Lambda_0)<\eta\}}\sqrt{n}|(\mathbb{P}_n-\mathcal{P})(m(\Lambda,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X))| \lesssim q_n^{1/2}\eta+q_n n^{-1/2}.$$

Set $\phi_n(\eta)=q_n^{1/2}\eta+q_n n^{-1/2}$. It is clear that $\phi_n(\eta)/\eta$ is decreasing about η. Moreover, $r_n^2\phi(1/r_n) = q_n^{1/2}r_n + n^{-1/2}q_n r_n^2$, where $r_n = O(n^a)$. Note that $q_n = O(n^\nu)$ with $0<\nu<1/2$. It follows that

$$r_n^2\phi(\frac{1}{r_n}) = O(n^{a+\frac{\nu}{2}}+n^{2a+\nu-\frac{1}{2}}).$$

Thus, $a \le (1-\nu)/2$ ensures $r_n^2\phi(1/r_n) \lesssim n^{1/2}$. This implies that $r_n = O(n^{(1-\nu)/2})$.

According to the proof of Theorem 3.1 and the definition of $\hat{\Lambda}_n$, we have

$$\begin{aligned}
&\mathbb{P}_n m(\hat{\Lambda}_n,\hat{F}_n;X)-\mathbb{P}_n m(\Lambda_0,\hat{F}_n;X)\\
=&\mathbb{P}_n m(\hat{\Lambda}_n,\hat{F}_n;X)-\mathbb{P}_n m(\Lambda_n^*,\hat{F}_n;X)+\mathbb{P}_n m(\Lambda_n^*,\hat{F}_n;X)-\mathcal{P}m(\Lambda_n^*,\hat{F}_n;X)\\
+&\mathcal{P}m(\Lambda_n^*,\hat{F}_n;X)-\mathcal{P}m(\Lambda_0,\hat{F}_n;X)+\mathcal{P}m(\Lambda_0,\hat{F}_n;X)-\mathbb{P}_n m(\Lambda_0,\hat{F}_n;X)\\
\le&n^{-\nu r+\varepsilon}(\mathbb{P}_n-\mathcal{P})\left(\frac{m(\Lambda_n^*,\hat{F}_n;X)-m(\Lambda_0,\hat{F}_n;X)}{n^{-\nu r+\varepsilon}}\right)+O_p(n^{-2\nu r})
\end{aligned}$$

for any $0 < \varepsilon < 1/2 - \nu r$. Set the class

$$\widetilde{\mathcal{M}}_n = \left\{ \frac{m(\Lambda, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)}{n^{-\nu r + \varepsilon}} : \Lambda \in \Phi_n, ||\Lambda - \Lambda_0||_\infty = O(n^{-\nu r}) \right\}.$$

Similar to the proof of Theorem 2 in Lu, Zhang, and Huang (2009), under Conditions (C2) and (C5), we have $\widetilde{\mathcal{M}}_n$ is Donsker, and $\mathcal{P}\tilde{m}^2 \to 0$ as $n \to \infty$ for any $\tilde{m} \in \widetilde{\mathcal{M}}_n$. Hence,

$$n^{-\nu r + \varepsilon}(\mathbb{P}_n - \mathcal{P}) \left(\frac{m(\Lambda_n^*, \hat{F}_n; X) - m(\Lambda_0, \hat{F}_n; X)}{n^{-\nu r + \varepsilon}} \right) = o_p(n^{-\nu r + \varepsilon - 1/2}) = o_p(n^{-2\nu r}).$$

This implies that $\mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \leq O_p(n^{-2\nu r})$. Noting that $\hat{\Lambda}_n$ is needed to satisfy $\mathbb{P}_n m(\Lambda_0, \hat{F}_n; X) \geq \mathbb{P}_n m(\hat{\Lambda}_n, \hat{F}_n; X) - O_p(r_n^{-2})$, so we should take ν such that $O_p(n^{-2\nu r}) \leq O_p(r_n^{-2})$. Since $r_n = O(n^{(1-\nu)/2})$, it follows that $\nu \geq 1/(1+2r)$. Taking $\nu = 1/(1+2r)$ and by Lemma 3.3, we have $d_1(\hat{\Lambda}_n, \Lambda_0) = O_p(n^{-r/(1+2r)} + n^{-1/2}) = O_p(n^{-r/(1+2r)})$. □

3.7.5 Proof of Theorem 3.3

Proof. Under (B1), (B3) and (B4), we have

$$Q(\hat{\Lambda}_n, \hat{F}_n)[h] - \dot{Q}^{(1)}_{\Lambda_0, \hat{F}_n}(\hat{\Lambda}_n - \Lambda_0)[h] - \dot{Q}^{(2)}_{\Lambda_0, F_0}(\hat{F}_n - F_0)[h] = o_p(n^{-1/2}).$$

According to (B1) and (B2), we have $-Q(\hat{\Lambda}_n, \hat{F}_n)[h] = Q_n(\Lambda_0, F_0)[h] + o_p(n^{-1/2})$. Combining the above two equations, it follows that

$$-\dot{Q}^{(1)}_{\Lambda_0, \hat{F}_n}(\hat{\Lambda}_n - \Lambda_0)[h] = \dot{Q}^{(2)}_{\Lambda_0, F_0}(\hat{F}_n - F_0)[h] + Q_n(\Lambda_0, F_0)[h] + o_p(n^{-1/2}).$$

Similarly, replacing the first part of (B3) and (B4) by (B3') and (B4'), we obtain

$$-\dot{Q}^{(1)}_{\Lambda_0, F_0}(\hat{\Lambda}_n - \Lambda_0)[h] = \dot{Q}^{(2)}_{\Lambda_0, F_0}(\hat{F}_n - F_0)[h] + Q_n(\Lambda_0, F_0)[h] + o_p(n^{-1/2}).$$

□

3.7.6 Proof of Theorem 3.4

Proof. (i) To prove this part, we need to verify (B1)–(B4).

For (B1), since $E(\tilde{N}(t)|U = u) = \Lambda_0(u - t)$, we have $Q(\Lambda_0, F_0)[h] = 0$. Note that $\hat{\Lambda}_n$ is obtained by minimizing the loss function $\mathbb{P}_n m(\Lambda, \hat{F}_n; X)$. By (3.8), we have $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h_n] = 0$ for all $h_n \in \Phi_n$. According to Lemma A1 of Lu, Zhang, and Huang (2007) and the properties of spline functions, for any $h \in \mathcal{H}_r$, there is an $h_n \in \Phi_n$ such that $||h_n - h||_\infty = O(n^{-r\nu}) = O(n^{-r/(1+2r)})$ and $||h'_n - h'||_\infty = o(1)$, where h' is the derivative of h. Next, to prove $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h] = o_p(n^{-1/2})$, we need to show that $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] = \mathbb{P}_n \psi(\hat{\Lambda}_n, \hat{F}_n; X)[h - h_n] = o_p(n^{-1/2})$. Note that

$$Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] = \Big[Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] - Q_n(\hat{\Lambda}_n, F_0)[h - h_n]\Big]$$

$$+ \Big[Q_n(\hat{\Lambda}_n, F_0)[h - h_n] - Q_n(\Lambda_0, F_0)[h - h_n]\Big] + Q_n(\Lambda_0, F_0)[h - h_n] = I_{1n} + I_{2n} + I_{3n}.$$

For the first term, by Lemma 3.1, we have

$$\mathcal{P}|I_{1n}| = \mathcal{P}|Q_n(\hat{\Lambda}_n, \hat{F}_n)[h - h_n] - Q_n(\hat{\Lambda}_n, F_0)[h - h_n]|$$

$$\leq \mathcal{P}\left[\sum_{j=1}^{K}(1 - \Delta)\left|\frac{\int_Y^\infty \big(\triangle N_j - \triangle\hat{\Lambda}_{n,j}(u)\big) \cdot (\triangle h_j(u) - \triangle h_{n,j}(u)) dF_0(u)}{1 - F_0(Y)}\right.\right.$$

$$\left.\left. - \frac{\int_Y^\infty \big(\triangle N_j - \triangle\hat{\Lambda}_{n,j}(u)\big) \cdot (\triangle h_j(u) - \triangle h_{n,j}(u)) d\hat{F}_n(u)}{1 - \hat{F}_n(Y)}\right|\right]$$

$$\lesssim d_2(\hat{F}_n, F_0)(||h - h_n||_\infty + ||h' - h'_n||_\infty) = o_p(n^{-1/2}).$$

For I_{2n}, by the Cauchy-Schwarz inequality, we have

$$\mathcal{P}|I_{2n}| = \mathcal{P}|Q_n(\hat{\Lambda}_n, F_0)[h - h_n] - Q_n(\Lambda_0, F_0)[h - h_n]|$$

$$\leq \mathcal{P}\left[\sum_{k=1}^{K}\Big|\Delta\{\triangle\Lambda_{0,j}(Y) - \triangle\hat{\Lambda}_{n,j}(Y)\} \cdot \{\triangle h_j(Y) - \triangle h_{n,j}(Y)\}\right.$$

$$\left. + (1 - \Delta)\frac{\int_Y^\infty \{\triangle\Lambda_{0,j}(u) - \triangle\hat{\Lambda}_{n,j}(u)\} \cdot \{\triangle h_j(u) - \triangle h_{n,j}(u)\} dF_0(u)}{1 - F_0(Y)}\Big|\right]$$

$$\lesssim \mathcal{P}\left[\sum_{k=1}^{K}\left\{\Delta|\triangle\Lambda_{0,j}(Y)-\triangle\hat{\Lambda}_{n,j}(Y)|+(1-\Delta)\frac{\int_Y^\infty \left|\triangle\Lambda_{0,j}(u)-\triangle\hat{\Lambda}_{n,j}(u)\right| dF_0(u)}{1-F_0(Y)}\right\}\right]$$

$$\times||h-h_n||_\infty$$

$$\leq d_1(\hat{\Lambda}_n,\Lambda_0)||h-h_n||_\infty = o_p(n^{-1/2}).$$

For the third term, note that $\mathcal{P}\psi(\Lambda_0, F_0; X_i)[h-h_n] = 0$. By the independence between X_i and X_j when $i \neq j$, we obtain

$$\mathcal{P}I_{3n}^2 = \mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\psi(\Lambda_0, F_0; X_i)[h-h_n]\right)^2 = \frac{1}{n}\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\psi^2(\Lambda_0, F_0; X_i)[h-h_n]\right)$$

$$\lesssim \frac{1}{n}\mathcal{P}\left[\sum_{j=1}^{K}\left[\Delta|\triangle N_j-\triangle\Lambda_{0,j}(Y)|+(1-\Delta)\frac{\int_Y^\infty |\triangle N_j-\triangle\Lambda_{0,j}(u)|\, dF_0(u)}{1-F_0(Y)}\right]\right]^2 ||h-h_n||_\infty^2$$

$$\lesssim \frac{1}{n}||h-h_n||_\infty^2.$$

Thus, $Q_n(\hat{\Lambda}_n, \hat{F}_n)[h-h_n] = o_p(n^{-1/2})$.

For (B2), note that

$$\sqrt{n}(Q_n - Q)(\hat{\Lambda}_n, \hat{F}_n)[h] - \sqrt{n}(Q_n - Q)(\Lambda_0, F_0)[h]$$

$$= \sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\Lambda}_n, \hat{F}_n; X)[h] - \psi(\Lambda_0, F_0; X)[h]).$$

For a bounded function $h \in \mathcal{H}_r$, define

$$\bar{\Psi}_\eta(h) = \{\psi(\Lambda, F; X)[h] - \psi(\Lambda_0, F_0; X)[h] : \Lambda \in \Phi_n, F \in \mathcal{F}$$

$$d_1(\Lambda, \Lambda_0) < \eta, d_2(F, F_0) < \eta, \Lambda \textit{ is uniformly bounded}\}.$$

Similar to the proof of Lemma 3.2, $\bar{\Psi}_\eta(h)$ is Donsker. According to Condition (C6) and Lemma 3.1, we obtain

$$
\begin{aligned}
&\mathcal{P}(\psi(\Lambda, F; X)[h] - \psi(\Lambda_0, F_0; X)[h])^2 \\
=&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y))\triangle h_j(Y) + (1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u))\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right.\right. \\
&+ (1-\Delta)\left.\left.\left(\frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF(u)}{1-F(Y)} - \frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right)\right\}\right]^2 \\
\lesssim&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_{0,j}(Y) - \triangle\Lambda_j(Y))^2\triangle h_j^2(Y) + (1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_{0,j}(u) - \triangle\Lambda_j(u))^2\triangle h_j^2(u)dF_0(u)}{1-F_0(Y)}\right\}\right] \\
+&\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF(u)}{1-F(Y)} - \frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_j(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right|^2\right] \\
\lesssim& d_1(\Lambda, \Lambda_0)^2 + d_2(F, F_0)^2.
\end{aligned}
$$

It follows that for any function $\bar{\psi} \in \bar{\Psi}_\eta(h)$, $\sup_{\bar{\psi}\in\bar{\Psi}_\eta(h)} \rho_{\mathcal{P}}(\bar{\psi}) \leq \sup_{\bar{\psi}\in\bar{\Psi}_\eta(h)} \mathcal{P}(\bar{\psi}^2)^{1/2} \lesssim \eta \to 0$ as $\eta \to 0$ for the seminorm $\rho_{\mathcal{P}}(\bar{\psi}) = \{\mathcal{P}(\bar{\psi} - \mathcal{P}\bar{\psi})^2\}^{1/2}$. Then by Corollary 2.3.12 of van der Vaart and Wellner (1996), we have

$$
\sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\Lambda}_n, \hat{F}_n; X)[h] - \psi(\Lambda_0, F_0; X)[h]) = o_p(1), \tag{3.9}
$$

and (B2) holds.

For (B3), by the smoothness of $Q_n(\Lambda, F)[h]$ with respect to Λ, we have the Fréchet derivative

$$
\begin{aligned}
&\dot{Q}^{(1)}_{\Lambda_0,\hat{F}_n}(\hat{\Lambda}_n - \Lambda_0)[h] = \frac{d}{d\varepsilon}\left\{Q(\Lambda_0 + \varepsilon(\hat{\Lambda}_n - \Lambda_0), \hat{F}_n)[h]\right\}\Big|_{\varepsilon=0} \\
=&\frac{d}{d\varepsilon}\left\{\mathcal{P}\left[\sum_{j=1}^{K}\Big[\Delta\{\triangle N_j - \triangle\Lambda_{0,j}(Y) - \varepsilon(\triangle\hat{\Lambda}_{n,j}(Y) - \triangle\Lambda_{0,j}(Y))\}\triangle h_j(Y)\right.\right. \\
&+ \left.\left.(1-\Delta)\frac{\int_Y^\infty\{\triangle N_j - \triangle\Lambda_{0,j}(u) - \varepsilon(\triangle\hat{\Lambda}_{n,j}(u) - \triangle\Lambda_{0,j}(u))\}\triangle h_j(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\Big]\right]\right\}\Big|_{\varepsilon=0} \\
=&-\mathcal{P}\varsigma(\hat{\Lambda}_n, \hat{F}_n; X)[h].
\end{aligned}
$$

Similarly, noting that for any integrable function g,

$$d\frac{\int_Y^\infty g(u-T_j)d(F_0+\varepsilon(\hat{F}_n-F_0))(u)}{1-F_0(Y)-\varepsilon(\hat{F}_n-F_0)(Y)}/d\varepsilon\Bigg|_{\varepsilon=0}$$

$$=\frac{(1-F_0(Y))\int_Y^\infty g(u-T_j)d(\hat{F}_n-F_0)(u)+(\hat{F}_n-F_0)(Y)\int_Y^\infty g(u-T_j)dF_0(u)}{(1-F_0(Y))^2}$$

$$=\frac{\int_Y^\infty\left\{g(u-T_j)-\int_Y^\infty\frac{g(s-T_j)}{1-F_0(Y)}dF_0(s)\right\}d(\hat{F}_n-F_0)(u)}{1-F_0(Y)}$$

holds, then the Fréchet derivative with respect to F is

$$\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]=\frac{d}{d\varepsilon}\left\{Q(\Lambda_0,F_0+\varepsilon(\hat{F}_n-F_0))[h]\right\}\Bigg|_{\varepsilon=0}$$

$$=\frac{d}{d\varepsilon}\left\{\mathcal{P}\left[\sum_{j=1}^K(1-\Delta)\frac{\int_Y^\infty\left(\triangle N_j-\triangle\Lambda_{0,j}(u)\right)\cdot\triangle h_j(u)d(F_0+\varepsilon(\hat{F}_n-F_0))(u)}{1-F_0(Y)-\varepsilon(\hat{F}_n-F_0)(Y)}\right]\right\}\Bigg|_{\varepsilon=0}$$

$$=\mathcal{P}\left[\int_Y^\infty\bar{\varphi}_{\Lambda_0,F_0}(u;X)[h]d(\hat{F}_n-F_0)(u)\right],$$

where

$$\bar{\varphi}_{\Lambda,F}(u;X)[h]$$

$$=\frac{1-\Delta}{1-F(Y)}\sum_{j=1}^K\left\{\left(\triangle N_j-\triangle\Lambda_j(u)\right)\cdot\triangle h_j(u)-\frac{\int_Y^\infty\left(\triangle N_j-\triangle\Lambda_j(s)\right)\cdot\triangle h_j(s)dF(s)}{1-F(Y)}\right\}.$$

Next, we verify (B4). Note that

$$\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n-F_0)[h]=\mathcal{P}\left[(1-\Delta)\frac{1-\hat{F}_n(Y)}{1-F_0(Y)}\sum_{j=1}^K\left\{\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right.\right.$$

$$\left.\left.-\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right\}\right].$$

By the conclusion of Lemma 3.1, we have

$$
\begin{aligned}
&|Q(\Lambda_0,\hat F_n)[h]-Q(\Lambda_0,F_0)[h]-\dot Q^{(2)}_{\Lambda_0,F_0}(\hat F_n-F_0)[h]|\\
=&\left|\mathcal{P}\left[(1-\Delta)\frac{\hat F_n(Y)-F_0(Y)}{1-F_0(Y)}\sum_{j=1}^K\left\{\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)d\hat F_n(u)}{1-\hat F_n(Y)}\right.\right.\right.\\
-&\left.\left.\left.\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right\}\right]\right|\\
\lesssim&||\hat F_n-F_0||_\infty\mathcal{P}\left[(1-\Delta)\sum_{j=1}^K\left|\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)d\hat F_n(u)}{1-\hat F_n(Y)}\right.\right.\\
-&\left.\left.\frac{\int_Y^\infty\{\triangle N_j-\triangle\Lambda_{0,j}(u)\}\triangle h_j(u)dF_0(u)}{1-F_0(Y)}\right|\right]\lesssim||\hat F_n-F_0||^2_\infty=o_p(n^{-1/2}).
\end{aligned}
$$

Moreover, $Q(\hat\Lambda_n,\hat F_n)[h]-Q(\Lambda_0,\hat F_n)[h]-\dot Q^{(1)}_{\Lambda_0,\hat F_n}(\hat\Lambda_n-\Lambda_0)[h]=0$. It follows that

$$
Q(\hat\Lambda_n,\hat F_n)[h]-Q(\Lambda_0,F_0)[h]-\dot Q^{(1)}_{\Lambda_0,\hat F_n}(\hat\Lambda_n-\Lambda_0)[h]-\dot Q^{(2)}_{\Lambda_0,F_0}(\hat F_n-F_0)[h]=o_p(n^{-1/2}).
$$

Thus, (B1)-(B4) are satisfied. By Theorem 3.3, we have

$$
-\sqrt{n}\dot Q^{(1)}_{\Lambda_0,\hat F_n}(\hat\Lambda_n-\Lambda_0)[h]=\sqrt{n}\dot Q^{(2)}_{\Lambda_0,F_0}(\hat F_n-F_0)[h]+\sqrt{n}Q_n(\Lambda_0,F_0)[h]+o_p(1).
$$

(ii) To prove the second part, we need to rewrite $\dot Q^{(2)}_{\Lambda_0,F_0}(\hat F_n-F_0)[h]+Q_n(\Lambda_0,F_0)[h]$ and verify (B5). Let

$$
\tilde\varphi_{\Lambda,F}(u,a;X)[h]=(1-F(u))\bar\varphi_{\Lambda,F}(u;X)[h]-\int_u^a\bar\varphi_{\Lambda,F}(s;X)[h]dF(s).
$$

Note that $\hat F_n$ is the KM estimator based on the data $\{(\tilde Y_i,\tilde\Delta_i):i=1,\cdots,n\}$. Setting $G_0(s)=\mathcal{P}1_{\{\tilde Y\ge s\}}$, then for any constant a, by Propositions 3 and 4 of Akritas (2000), we have

$$
\int_0^a\bar\varphi_{\Lambda_0,F_0}(u;X)[h]d(\hat F_n-F_0)(u)=\frac{1}{n}\sum_{i=1}^n\int_0^a\frac{\tilde\varphi_{\Lambda_0,F_0}(u,a;X)[h]}{G_0(u)}d\tilde M_i(u)+o_p(n^{-1/2}),
$$

where

$$\tilde{M}_i(u) = 1_{\{\tilde{Y}_i \le u, \tilde{\Delta}_i = 1\}} - \int_{-\infty}^{u} \frac{1_{\{\tilde{Y}_i \ge s\}}}{1 - F_0(s)} dF_0(s).$$

It follows that

$$\begin{aligned}
&\int_Y^{\infty} \bar{\varphi}_{\Lambda_0, F_0}(u; X)[h] d(\hat{F}_n - F_0)(u) \\
=& \frac{1}{n}\sum_{i=1}^{n} \left\{ \int_0^{\infty} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u, \infty; X)[h]}{G_0(u)} d\tilde{M}_i(u) - \int_0^{Y} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u, Y; X)[h]}{G_0(u)} d\tilde{M}_i(u) \right\} + o_p(n^{-1/2}) \\
=& \frac{1}{n}\sum_{i=1}^{n} \left\{ 1_{\{\tilde{\Delta}_i = 1\}} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(\tilde{Y}_i, \infty; X)[h]}{G_0(\tilde{Y}_i)} - \int_0^{\tilde{Y}_i} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u, \infty; X)[h]}{G_0(u)(1 - F_0(u))} dF_0(u) \right. \\
&\left. + \int_0^{\tilde{Y}_i \wedge Y} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(u, Y; X)[h]}{G_0(u)(1 - F_0(u))} dF_0(u) - 1_{\{\tilde{\Delta}_i = 1, Y \ge \tilde{Y}_i\}} \frac{\tilde{\varphi}_{\Lambda_0,F_0}(\tilde{Y}_i, Y; X)[h]}{G_0(\tilde{Y}_i)} \right\} + o_p(n^{-1/2}).
\end{aligned}$$

Hence, we obtain $\dot{Q}^{(2)}_{\Lambda_0,F_0}(\hat{F}_n - F_0)[h] = \mathbb{P}_n\{\mathcal{P}\varphi(\Lambda_0, F_0; X; \tilde{Y}, \tilde{\Delta})[h]\} + o_p(n^{-1/2})$, where

$$\begin{aligned}
\varphi(\Lambda, F; X; \tilde{Y}, \tilde{\Delta})[h] =& 1_{\{\tilde{\Delta} = 1\}} \frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y}, \infty; X)[h]}{G_0(\tilde{Y})} - \int_0^{\tilde{Y}} \frac{\tilde{\varphi}_{\Lambda,F}(u, \infty; X)[h]}{G_0(u)(1 - F_0(u))} dF(u) \\
&+ \int_0^{\tilde{Y} \wedge Y} \frac{\tilde{\varphi}_{\Lambda,F}(u, Y; X)[h]}{G_0(u)(1 - F_0(u))} dF(u) - 1_{\{\tilde{\Delta} = 1, Y \ge \tilde{Y}\}} \frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y}, Y; X)[h]}{G_0(\tilde{Y})}.
\end{aligned}$$

Noting that $Q_n(\Lambda_0, F_0)[h] = \mathbb{P}_n \psi(\Lambda_0, F_0; X)[h]$, we have

$$-\sqrt{n}\dot{Q}^{(1)}_{\Lambda_0, \hat{F}_n}(\hat{\Lambda}_n - \Lambda_0)[h] \xrightarrow{d} N(0, \sigma_0^2),$$

where $\sigma_0^2 = E[\{\mathcal{P}\varphi(\Lambda_0, F_0; X; \tilde{Y}, \tilde{\Delta})[h] + \psi(\Lambda_0, F_0; X)[h]\}^2]$. □

3.7.7 Proof of Theorem 3.5

Proof. Setting $U_n^{(l)} = \sqrt{n}\mathbb{P}_n \varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h_n]$ for $l = 1, 2$, we have $U_n = U_n^{(1)} - U_n^{(2)}$ and $U_n^{(l)} = U_{1n}^{(l)} + U_{2n}^{(l)} + U_{3n}^{(l)}$, where

$$U_{1n}^{(l)} = \sqrt{n}(\mathbb{P}_n - \mathcal{P})\varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h_n],$$

$$U_{2n}^{(l)} = \sqrt{n}\mathcal{P}\varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h_n - h],$$

and

$$U_{3n}^{(l)} = \sqrt{n}\mathcal{P}\varsigma(\hat{\Lambda}_l, \hat{F}_n; X)[h].$$

For $U_{1n}^{(l)}$, similar to the proof of (3.9), we have

$$\sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\Lambda}_l, \hat{F}_n; X)[h_n] - \psi(\Lambda_0, F_0; X)[h_n]) = o_p(1)$$

and

$$\sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\Lambda_0, \hat{F}_n; X)[h_n] - \psi(\Lambda_0, F_0; X)[h_n]) = o_p(1).$$

Noting that $\varsigma(\Lambda, F; X)[h_n] = \psi(\Lambda_0, F; X)[h_n] - \psi(\Lambda, F; X)[h_n]$, it follows that $U_{1n}^{(l)} = o_p(1)$.

For $U_{2n}^{(l)}$, by Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & U_{2n}^{(l)} \\ \leq & \sqrt{n}\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle\hat{\Lambda}_{l,j}(Y) - \triangle\Lambda_{0,j}(Y)\big)^2 + (1-\Delta)\frac{\int_Y^\infty \big(\triangle\hat{\Lambda}_{l,j}(u) - \triangle\Lambda_{0,j}(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]^{1/2} \\ \times & \mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle h_{n,j}(Y) - \triangle h_j(Y)\big)^2 + (1-\Delta)\frac{\int_Y^\infty \big(\triangle h_{n,j}(u) - \triangle h_j(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right]^{1/2}. \end{aligned}$$

By Lemma 3.1, similar to the proof of the first inequality in Theorem 3.2, we have

$$\begin{aligned} & \mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle\hat{\Lambda}_{l,j}(Y) - \triangle\Lambda_{0,j}(Y)\big)^2 + (1-\Delta)\frac{\int_Y^\infty \big(\triangle\hat{\Lambda}_{l,j}(u) - \triangle\Lambda_{0,j}(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right] \\ \lesssim & d_1(\hat{\Lambda}_l, \Lambda_0)^2 + d_1(\hat{\Lambda}_l, \Lambda_0)d_2(\hat{F}_n, F_0) = O_p(n^{-2r/(1+2r)}) \end{aligned}$$

and

$$\begin{aligned} & \mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle h_{n,j}(Y) - \triangle h_j(Y)\big)^2 + (1-\Delta)\frac{\int_Y^\infty \big(\triangle h_{n,j}(u) - \triangle h_j(u)\big)^2 d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right\}\right] \\ \lesssim & d_1(h_n, h)^2 + d_1(h_n, h)d_2(\hat{F}_n, F_0) = o_p(n^{-1/(1+2r)}). \end{aligned}$$

Hence, $U_{2n}^{(l)} = o_p(1)$.

For $U_{3n}^{(l)}$, by Theorem 3.4, we have

$$U_{3n}^{(l)} = \sqrt{n}\dot{Q}_{\Lambda_0,F_0}^{(2)}(\hat{F}_n - F_0)[h] + \sqrt{n}\mathbb{P}_{n_l}\psi(\Lambda_0, F_0; X)[h] + o_p(1),$$

where $\mathbb{P}_{n_l}$ is the empirical measure based on group l. Thus,

$$\begin{aligned} U_n =& U_{3n}^{(1)} - U_{3n}^{(2)} + o_p(1) \\ =& \sqrt{\frac{n}{n_1}}\sqrt{n_1}\mathbb{P}_{n_1}\psi(\Lambda_0, F_0; X)[h] - \sqrt{\frac{n}{n_2}}\sqrt{n_2}\mathbb{P}_{n_2}\psi(\Lambda_0, F_0; X)[h] + o_p(1). \end{aligned}$$

Note that $\mathbb{P}_{n_1}$ and $\mathbb{P}_{n_2}$ are independent, and $\sqrt{n_l}\mathbb{P}_{n_l}\psi(\Lambda_0, F_0; X)[h] \xrightarrow{d} N(0, \check{\sigma}_0^2)$. Thus, we have $U_n \xrightarrow{d} N(0, (\frac{1}{p} + \frac{1}{1-p})\check{\sigma}_0^2)$.

Finally, we need to prove that $\hat{\sigma}_n^2 - \check{\sigma}_0^2 = o_p(1)$. Note that $\check{\sigma}_0^2 = \mathcal{P}\psi^2(\Lambda_0, F_0; X)[h]$ and $\hat{\sigma}_n^2 = \mathbb{P}_n\psi^2(\hat{\Lambda}_n, \hat{F}_n; X)[h_n]$. Thus, we have

$$\begin{aligned} \hat{\sigma}_n^2 - \check{\sigma}_0^2 =& \mathbb{P}_n\{\psi^2(\hat{\Lambda}_n, \hat{F}_n; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h_n]\} \\ &+ \mathbb{P}_n\{\psi^2(\Lambda_0, F_0; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h]\} + (\mathbb{P}_n - \mathcal{P})\psi^2(\Lambda_0, F_0; X)[h]. \end{aligned}$$

By the consistency of $(\hat{\Lambda}_n, \hat{F}_n)$ and the law of large numbers, we have

$$\mathbb{P}_n\{\psi^2(\hat{\Lambda}_n, \hat{F}_n; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h_n]\} = o_p(1)$$

and

$$(\mathbb{P}_n - \mathcal{P})\psi^2(\Lambda_0, F_0; X)[h] = o_p(1).$$

Then we only need to consider $\mathbb{P}_n\{\psi^2(\Lambda_0, F_0; X)[h_n] - \psi^2(\Lambda_0, F_0; X)[h]\}$. Since Λ_0, F_0, h and h_n are bounded functions, we have

$$\begin{aligned} &|\psi(\Lambda_0, F_0; X)[h_n] + \psi(\Lambda_0, F_0; X)[h]| = |\psi(\Lambda_0, F_0; X)[h_n + h]| \\ \lesssim& \left| \sum_{j=1}^{K} \left[\Delta\{\triangle N_j - \triangle\Lambda_{0,j}(Y)\} + (1-\Delta)\frac{\int_Y^\infty \{(\triangle N_j - \triangle\Lambda_{0,j}(u))\} dF_0(u)}{1 - F_0(Y)} \right] \right| \\ \lesssim& (N(T_K) + \Lambda_0(\tau)) \end{aligned}$$

with probability 1. Thus, by Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\mathcal{P}\left|\psi^2(\Lambda_0,F_0;X)[h_n]-\psi^2(\Lambda_0,F_0;X)[h]\right| \\
\lesssim&\mathcal{P}\left[\left|\psi(\Lambda_0,F_0;X)[h_n]-\psi(\Lambda_0,F_0;X)[h]\right|(N(T_K)+\Lambda_0(\tau))\right] \\
\lesssim&\mathcal{P}\Bigg[(N(T_K)+\Lambda_0(\tau))\sum_{j=1}^{K}\Big\{\Delta|\triangle N_j-\triangle\Lambda_{0,j}(Y)||\triangle h_{n,j}(Y)-\triangle h_j(Y)| \\
&+(1-\Delta)\frac{\int_Y^\infty|\triangle N_j-\triangle\Lambda_{0,j}(u)||\triangle h_{n,j}(u)-\triangle h_j(u)|dF_0(u)}{1-F_0(Y)}\Big\}\Bigg] \\
\lesssim&\mathcal{P}\Bigg[(N(T_K)+\Lambda_0(\tau))^2\sum_{j=1}^{K}\{\Delta|\triangle h_{n,j}(Y)-\triangle h_j(Y)| \\
&+(1-\Delta)\frac{\int_Y^\infty|\triangle h_{n,j}(u)-\triangle h_j(u)|dF_0(u)}{1-F_0(Y)}\Big\}\Bigg] \\
\leq&\mathcal{P}\left[\sum_{j=1}^{K}\left\{\Delta\big(\triangle h_{n,j}(Y)-\triangle h_j(Y)\big)^2+(1-\Delta)\frac{\int_Y^\infty\big(\triangle h_{n,j}(u)-\triangle h_j(u)\big)^2dF_0(u)}{1-F_0(Y)}\right\}\right]^{\frac{1}{2}} \\
&\times\mathcal{P}\left[(N(T_K)+\Lambda_0(\tau))^4\right]^{\frac{1}{2}}\lesssim d_1(h_n,h)=o(1).
\end{aligned}
$$

Therefore, $\mathbb{P}_n\{\psi^2(\Lambda_0,F_0;X)[h_n]-\psi^2(\Lambda_0,F_0;X)[h]\}=o_p(1)$ and $\hat{\sigma}_n^2-\sigma_0^2=o_p(1)$. $\square$

3.7.8 Proof of Theorem 3.6

Proof. (i) First we write $\tilde{U}_n=\sum_{l=1}^2(n_l/\sqrt{n})\mathbb{P}_{n_l}(\varsigma(\hat{\Lambda}_1,\hat{F}_l;X^{(l)})[h_n]-\varsigma(\hat{\Lambda}_2,\hat{F}_l;X^{(l)})[h_n])$ for $l=1,2$. Note that

$$
\begin{aligned}
\mathbb{P}_{n_l}(\varsigma(\Lambda,\hat{F}_l;X^{(l)})[h_n])=&(\mathbb{P}_{n_l}-\mathcal{P}_l)\varsigma(\Lambda,\hat{F}_l;X^{(l)})[h_n]+\mathcal{P}_l\varsigma(\Lambda,\hat{F}_l;X^{(l)})[h_n-h] \\
&+\mathcal{P}_l(\varsigma(\Lambda,\hat{F}_l;X^{(l)})[h]-\varsigma(\Lambda,F_l;X^{(l)})[h])+\mathcal{P}_l\varsigma(\Lambda,F_l;X^{(l)})[h].
\end{aligned}
$$

We next show that

$$
\mathbb{P}_{n_l}(\varsigma(\hat{\Lambda}_1,\hat{F}_l;X^{(l)})[h_n])=\mathcal{P}_l\varsigma(\hat{\Lambda}_1,F_l;X^{(l)})[h]+o_p(n^{-1/2}) \tag{3.10}
$$

and

$$\mathbb{P}_{n_l}(\varsigma(\hat{\Lambda}_2, \hat{F}_l; X^{(l)})[h_n]) = \mathcal{P}_l \varsigma(\hat{\Lambda}_2, F_l; X^{(l)})[h] + o_p(n^{-1/2}). \tag{3.11}$$

According to the proof of Theorem 3.5, we have

$$\sqrt{n_l}(\mathbb{P}_{n_l} - \mathcal{P}_l)\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h_n] = o_p(1) \text{ and } \sqrt{n_l}\mathcal{P}_l\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h_n - h] = o_p(1).$$

Moreover, according to Lemma 3.1, this implies that

$$\begin{aligned}
&\left|\mathcal{P}_l\left(\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h] - \varsigma(\hat{\Lambda}_l, F_l; X^{(l)})[h]\right)\right| \\
\lesssim &\left(\mathcal{P}_l\left[\sum_{j=1}^{K^{(l)}}\left|\triangle\hat{\Lambda}_{l,j}(U^{(l)}) - \triangle\Lambda_{0,j}(U^{(l)})\right|\right] + \mathcal{P}_l\left[\sum_{j=1}^{K^{(l)}}\left|\triangle\hat{\Lambda}'_{l,j}(U^{(l)}) - \triangle\Lambda'_{0,j}(U^{(l)})\right|\right]\right) \\
&\times||\hat{F}_l - F_l||_\infty \\
\lesssim &(||\hat{\Lambda}_l - \Lambda_0||_{L_2(\mu_l)} + ||\hat{\Lambda}'_l - \Lambda'_0||_{L_2(\mu_l)})||\hat{F}_l - F_l||_\infty.
\end{aligned} \tag{3.12}$$

Applying Lemma 3.5 and Corollary 2.1 of He and Shi (1994), we have $||\hat{\Lambda}'_l - \Lambda'_0||_{L_2(\mu_l)} = O_p(n^{-(r-1)/(1+2r)}) = o_p(1)$. This gives $\mathcal{P}_l(\varsigma(\hat{\Lambda}_l, \hat{F}_l; X^{(l)})[h] - \varsigma(\hat{\Lambda}_l, F_l; X^{(l)})[h]) = o_p(n^{-1/2})$. Thus, (3.10) and (3.11) hold.

Moreover, we have

$$\begin{aligned}
\mathcal{P}_l\left[\sum_{j=1}^{K^{(l)}}\triangle h_j(U^{(l)})\right] &= \int E\left[\sum_{j=1}^{K^{(l)}}\triangle h_j(u)\middle| U^{(l)} = u\right] dF_l(u) \\
&= \int E\left[\sum_{j=1}^{K^{(r)}}\triangle h_j(u)\middle| U^{(r)} = u\right] \frac{f_l(u)}{f_r(u)} dF_r(u) = \mathcal{P}_r\left[\sum_{j=1}^{K^{(r)}}\triangle h_j(U^{(r)})\frac{f_l(U^{(r)})}{f_r(U^{(r)})}\right].
\end{aligned} \tag{3.13}$$

Hence, (3.10), (3.11) and (3.13) yield

$$\tilde{U}_n = \sum_{l=1}^{2}\frac{n_l}{\sqrt{n}}\mathcal{P}_l\left(\varsigma(\hat{\Lambda}_1, F_l; X^{(l)})[h] - \varsigma(\hat{\Lambda}_2, F_l; X^{(l)})[h]\right) + o_p(1)$$

$$=\sqrt{n}\sum_{l=1}^{2}\mathcal{P}_l\varsigma(\hat{\Lambda}_1,F_l;X^{(l)})[p_lh]-\sqrt{n}\sum_{l=1}^{2}\mathcal{P}_l\varsigma(\hat{\Lambda}_2,F_l;X^{(l)})[p_lh]+o_p(1)$$

$$=\frac{\sqrt{n_1}}{\sqrt{p_1}}\mathcal{P}_1\varsigma(\hat{\Lambda}_1,F_1;X^{(1)})[w_1]-\frac{\sqrt{n_2}}{\sqrt{p_2}}\mathcal{P}_2\varsigma(\hat{\Lambda}_2,F_2;X^{(2)})[w_2]+o_p(1).$$

Set $G_l(s)=\mathcal{P}_l 1_{\{\tilde{Y}^{(l)}\geq s\}}$ and

$$\varphi_l(\Lambda,F;X;\tilde{Y},\tilde{\Delta})[w]=1_{\{\tilde{\Delta}=1\}}\frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},\infty;X)[w]}{G_l(\tilde{Y})}-\int_0^{\tilde{Y}}\frac{\tilde{\varphi}_{\Lambda,F}(u,\infty;X)[w]}{G_l(u)(1-F(u))}dF(u)$$

$$+\int_0^{\tilde{Y}\wedge Y}\frac{\tilde{\varphi}_{\Lambda,F}(u,Y;X)[w]}{G_l(u)(1-F(u))}dF(u)-1_{\{\tilde{\Delta}=1,Y>\tilde{Y}\}}\frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},Y;X)[w]}{G_l(\tilde{Y})}.$$

By Theorem 3.4 and (3.12), we can get

$$\sqrt{n_l}\mathcal{P}_l\varsigma(\hat{\Lambda}_l,F_l;X^{(l)})[w_l]=\sqrt{n_l}\mathcal{P}_l\varsigma(\hat{\Lambda}_l,\hat{F}_l;X^{(l)})[w_l]+o_p(1)\rightsquigarrow N(0,\sigma_l^2).$$

Since $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ are independent, it then follows that $\tilde{U}_n\rightsquigarrow N(0,(\sigma_1^2/p_1+\sigma_2^2/p_2))$.

(ii) Set

$$\varphi_n^{(l)}(\Lambda,F;X;\tilde{Y},\tilde{\Delta})[w]=1_{\{\tilde{\Delta}=1\}}\frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},\infty;X)[w]}{G_n^{(l)}(\tilde{Y})}-\int_0^{\tilde{Y}}\frac{\tilde{\varphi}_{\Lambda,F}(u,\infty;X)[w]}{G_n^{(l)}(u)(1-F(u))}dF(u)$$

$$+\int_0^{\tilde{Y}\wedge Y}\frac{\tilde{\varphi}_{\Lambda,F}(u,Y;X)[w]}{G_n^{(l)}(u)(1-F(u))}dF(u)-1_{\{\tilde{\Delta}=1,Y>\tilde{Y}\}}\frac{\tilde{\varphi}_{\Lambda,F}(\tilde{Y},Y;X)[w]}{G_n^{(l)}(\tilde{Y})},$$

where $G_n^{(l)}(s)=\mathbb{P}_{n_l}1_{\{\tilde{Y}^{(l)}\geq s\}}$. Then

$$\hat{\sigma}_l^2-\sigma_l^2=\mathbb{P}_{n_l}\left[\left\{\mathbb{P}_n\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}^{(l)},\tilde{\Delta}^{(l)})[w_n^{(l)}]+\psi(\hat{\Lambda}_l,\hat{F}_l;X^{(l)})[w_n^{(l)}]\right\}^2\right.$$

$$\left.-\left\{\mathcal{P}\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}^{(l)},\tilde{\Delta}^{(l)})[w_n^{(l)}]+\psi(\hat{\Lambda}_l,\hat{F}_l;X^{(l)})[w_n^{(l)}]\right\}^2\right]$$

$$+\mathbb{P}_{n_l}\left[\mathcal{P}\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}^{(l)},\tilde{\Delta}^{(l)})[w_n^{(l)}]+\psi(\hat{\Lambda}_l,\hat{F}_l;X^{(l)})[w_n^{(l)}]\right]^2$$

$$-\mathcal{P}\left[\mathcal{P}\varphi_l(\Lambda_0,F_l;X^{(l)};\tilde{Y}^{(l)},\tilde{\Delta}^{(l)})[w_l]+\psi(\Lambda_0,F_l;X^{(l)})[w_l]\right]^2.$$

Note that

$$\mathbb{P}_{n_l}\left[\left\{\mathbb{P}_n\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}^{(l)},\tilde{\Delta}^{(l)})[w_n^{(l)}]+\psi(\hat{\Lambda}_l,\hat{F}_l;X^{(l)})[w_n^{(l)}]\right\}^2\right.$$

$$\left.-\left\{\mathcal{P}\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}^{(l)},\tilde{\Delta}^{(l)})[w_n^{(l)}]+\psi(\hat{\Lambda}_l,\hat{F}_l;X^{(l)})[w_n^{(l)}]\right\}^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left[\left\{\mathbb{P}_n\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]-\mathcal{P}\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]\right\}\right.$$

$$\times\left\{\mathbb{P}_n\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]+\mathcal{P}\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]\right.$$

$$\left.\left.+\ 2\psi(\hat{\Lambda}_l,\hat{F}_l;X_i^{(l)})[w_n^{(l)}]\right\}\right].$$

For each i, we obtain

$$\mathbb{P}_n\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]-\mathcal{P}\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]$$

$$=(\mathbb{P}_n-\mathcal{P})\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]$$

$$+\mathcal{P}\left(\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]-\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]\right).$$

Since $\sup_{s\in[0,\tau]}|G_n^{(l)}(s)-G_l(s)|=O_p(n^{-1/2})$, we get

$$\mathcal{P}\left(\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]-\varphi_l(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]\right)=o_p(1).$$

Similar to the proof of Lemma 3.2, under Conditions (C2), (C4), (C6) and (C10),

$$\{\varphi_n^{(l)}(\Lambda,F;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]:\Lambda\in\Phi,\ F\in\mathcal{F},\ d_2(F,F_l)\le\delta,\ \Lambda \text{ is uniformly bounded}\}$$

is Donsker and it is Glivenko-Cantelli. It follows that

$$(\mathbb{P}_n-\mathcal{P})\varphi_n^{(l)}(\hat{\Lambda}_l,\hat{F}_l;X^{(l)};\tilde{Y}_i^{(l)},\tilde{\Delta}_i^{(l)})[w_n^{(l)}]=o_p(1)$$

and the first term of $\hat{\sigma}_l^2-\sigma_l^2$ is $o_p(1)$. Denote

$$\upsilon_l(\Lambda,F;\tilde{Y},\tilde{\Delta},X)[w]=\mathcal{P}\varphi_l(\Lambda,F;X^{(l)};\tilde{Y},\tilde{\Delta})[w]+\psi(\Lambda,F;X)[w].$$

To verify that the second term of $\hat{\sigma}_l^2 - \sigma_l^2$ is $o_p(1)$, we only need to prove

$$E[|\upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}, X^{(l)})[w_n^{(l)}] - \upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}, X^{(l)})[w_l]|] = o(1).$$

By the definition of $w_n^{(l)}(u,t)$, we obtain

$$\begin{aligned}
&\left|\left\{w_n^{(l)}(u-t_1) - w_n^{(l)}(u-t_2)\right\} - \left\{w_l(u-t_1) - w_l(u-t_2)\right\}\right| \\
=&\left|\{(h_n(u-t_1) - h(u-t_1)) - (h_n(u-t_2) - h(u-t_2))\}\left(\frac{n_l}{n} + \frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)}\right)\right. \\
&\left. + \{h(u-t_1) - h(u-t_2)\}\left(\frac{n_l}{n} + \frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)} - p_r\frac{f_r(u)}{f_l(u)} - p_l\right)\right| \\
\leq&|(h_n(u-t_1) - h(u-t_1)) - (h_n(u-t_2) - h(u-t_2))|\left|\frac{n_l}{n} + \frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)}\right| \\
&+|h(u-t_1) - h(u-t_2)|\left|\frac{n_l}{n} + \frac{n_r}{n}\frac{\hat{f}_r(u)}{\hat{f}_l(u)} - p_r\frac{f_r(u)}{f_l(u)} - p_l\right| \\
\leq& c\left\{|h_n(u-t_1) - h(u-t_1)| + |h_n(u-t_2) - h(u-t_2)| + \left|\frac{\hat{f}_r(u)}{\hat{f}_l(u)} - \frac{f_r(u)}{f_l(u)}\right|\right\}
\end{aligned} \tag{3.14}$$

with probability 1 for some constant c, where $l, r = 1, 2$ and $l \neq r$. According to Theorem 2.2 of Földes, Rejtő, and Winter (1981), we have

$$\sup_u \left|\frac{\hat{f}_r(u)}{\hat{f}_l(u)} - \frac{f_r(u)}{f_l(u)}\right| \xrightarrow{a.s.} 0. \tag{3.15}$$

Moreover, by the definition of $\varphi_l(\Lambda, F; X; Y, \Delta)[w]$, $\bar{\varphi}_{\Lambda,F}(u; X)[w]$ and $\tilde{\varphi}_{\Lambda,F}(u, a; X)[w]$, after some calculations, we have

$$|\upsilon_l(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}, X^{(l)})[w_n^{(l)} + w_l]| \lesssim N^{(l)}(T_{K^{(l)}}^{(l)}) + \Lambda_0(0)$$

and

$$
\begin{aligned}
&|\upsilon_l(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}, X^{(l)})[w_n^{(l)} - w_l]| \\
\lesssim& \Big(N^{(l)}(T_{K^{(l)}}^{(l)}) + \Lambda_0(\tau)\Big) \sum_{j=1}^{K^{(l)}} \Big[\Delta^{(l)} \Big|\triangle w_{n,j}^{(l)}(Y^{(l)}) - \triangle w_{l,j}(Y^{(l)})\Big| \\
&+ (1-\Delta^{(l)}) \frac{\int_{Y^{(l)}}^{\infty} \Big|\triangle w_{n,j}^{(l)}(u) - \triangle w_{l,j}(u)\Big| dF_l(u)}{1 - F_l(Y^{(l)})}\Bigg] \\
&+ |\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)} - w_l]|
\end{aligned} \tag{3.16}
$$

with probability 1. Combining (3.14)–(3.16) and using the Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
&E|\upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}; X^{(l)})[w_n^{(l)}] - \upsilon_l^2(\Lambda_0, F_l; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)}; X^{(l)})[w_l]| \\
\lesssim& E\Bigg[\Big(N^{(l)}(T_K^{(l)}) + \Lambda_0(\tau)\Big)^2 \sum_{j=1}^{K^{(l)}} \Big\{\Delta^{(l)} \Big|\triangle w_{n,j}^{(l)}(Y^{(l)}) - \triangle w_{l,j}(Y^{(l)})\Big| \\
&+ (1-\Delta^{(l)}) \frac{\int_{Y^{(l)}}^{\infty} \Big|\triangle w_{n,j}^{(l)}(u) - \triangle w_{l,j}(u)\Big| dF_l(u)}{1 - F_l(Y^{(l)})}\Bigg\}\Bigg] \\
&+ E\Big[\big(N^{(l)}(\tau) + \Lambda_0(\tau)\big) |\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)} - w_l]|\Big] \\
\lesssim& E\Big[|\mathcal{P}\varphi_l(\Lambda_0, F_l; X^{(l)}; \tilde{Y}^{(l)}, \tilde{\Delta}^{(l)})[w_n^{(l)} - w_l]|\Big] + ||h_n^{(l)} - h_l||_{L_2(\mu_l)} \\
\lesssim& ||h_n^{(l)} - h_l||_{L_2(\mu_l)} \to 0.
\end{aligned}
$$

□

Chapter 4

Semiparametric Statistical Inference for Panel Count Data with Terminal Event

4.1 Introduction

In this chapter, we extend the nonparametric model in Chapter 3 to a semiparametric model. We still use the monotone I-spline functions to approximate the nonparametric function and propose a least squares-based two-stage estimation by treating the distribution of the terminal event as a nuisance functional parameter. We assume that the distribution of the right-censored terminal event satisfies the Cox model (Cox, 1972) with the baseline hazard function estimated by the Breslow estimator (Breslow, 1972) and the coefficient of covariates estimated by the partial likelihood estimator (Cox, 1972). For the asymptotic properties, we prove that the estimator of the mean function is consistent and the convergence rate is between $n^{1/3}$ and $n^{1/2}$, depending on the smoothness of the mean function. We also establish the asymptotic normality for the proposed estimator and provide the asymptotic variance. After demonstrating the finite sample performance of our method by the simulation studies, we use the method to analyze the rate of severe illness for elder people in the dataset of CLHLS.

The rest of this chapter is organized as follows. In Section 4.2, we introduce the semiparametric model and the loss function for our estimation. Under some basic conditions, Section 4.3 verifies the asymptotic consistency, the convergence rate, and the asymptotic normality of the proposed estimation. We demonstrate the performance of our method by the simulation studies in Section 4.4. Section 4.5 is the analysis of CLHLS data. Finally, we show the proofs of the theoretical results in the Appendix.

4.2 Model Setting and Estimation Procedure

We define the notation U, C, Y, Δ, K, $\underline{T}$ and $\underline{N}$ the same way as in Chapter 3. Let the time-independent covariate vector be $\boldsymbol{Z}$. Then we consider a study containing n subjects with the sample of subject i being $X_i = (Y_i, \Delta_i, K_i, \underline{T_i}, \underline{N_i}, \boldsymbol{Z}_i)$ for $i = 1, \cdots, n$, where $\underline{T_i} = (T_{i1}, T_{i2}, \cdots, T_{iK_i})$ and $\underline{N_i} = \{N_i(T_{i1}), \cdots, N_i(T_{iK_i})\}$.

For the occurrence number of recurrent events from time t to the terminal event, $\tilde{N}(t;U)$, our model supposes that given the covariate and terminal event time, the conditional expectation of $\tilde{N}(t;U)$ is

$$E(\tilde{N}(t;U)|U = u, \boldsymbol{Z} = \boldsymbol{z}) = \exp(\boldsymbol{\beta}^T \boldsymbol{z})\Lambda(u - t), 0 \leq t \leq u \leq \tau, \tag{4.1}$$

where Λ is a non-negative and non-decreasing baseline mean function with $\Lambda(0) = 0$. Suppose that the conditional distribution function of U given $\boldsymbol{Z}$ satisfies the Cox model

$$F(u|\boldsymbol{Z} = \boldsymbol{z}) = P(U \leq u|\boldsymbol{Z} = \boldsymbol{z}) = 1 - \exp\{-H(u)\exp(\boldsymbol{\gamma}^T \boldsymbol{z})\}, \tag{4.2}$$

where $H(u)$ is the baseline cumulative hazard function of U. Then the unknown parameters and functions to be estimated under models (4.1) and (4.2) are $(\boldsymbol{\beta}, \Lambda, F_{\boldsymbol{\gamma},H})$. We assume the following basic conditions before the analysis: (i) Given $\boldsymbol{Z}$, C and U are conditional independent; (ii) Given $\boldsymbol{Z}$, C is noninformative to Λ; (iii) Given

$(Y, \Delta, \boldsymbol{Z})$, $(K, \underline{T})$ is noninformative to Λ. Taking $\triangle N_j = N(T_j) - N(T_{j-1})$ and $\triangle\Lambda_j(u) = \Lambda(u - T_{j-1}) - \Lambda(u - T_j)$ for $j = 1, \cdots, K$ with $T_0 = 0$, according to model (4.1), we have

$$E\left[\sum_{j=1}^{K}\left\{\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)\right\}^2 | Y, \Delta, K, \underline{T}, \underline{N}, \boldsymbol{Z}\right]$$

$$=\sum_{j=1}^{K}\Bigg[\Delta\left\{\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(Y)\right\}^2$$

$$+ \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})}\int_Y^\infty \left\{\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u)\right\}^2 dF(u|\boldsymbol{Z})\Bigg].$$

Correspondingly, the least squares-based loss function should be

$$\begin{aligned}&\ell_n(\boldsymbol{\beta}, \Lambda, F; X)\\ =&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_i\left\{\triangle N_{i,j} - \exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\triangle\Lambda_{i,j}(Y_i)\right\}^2\\ +&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\frac{1-\Delta_i}{1-F(Y_i|\boldsymbol{Z}_i)}\int_{Y_i}^\infty \left\{\triangle N_{i,j} - \exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\triangle\Lambda_{i,j}(u)\right\}^2 dF(u|\boldsymbol{Z}_i),\end{aligned} \tag{4.3}$$

where $X = (Y, \Delta, K, \underline{T}, \underline{N}, \boldsymbol{Z})$, $\triangle N_{i,j} = N_i(T_{i,j}) - N_i(T_{i,j-1})$, and $\triangle\Lambda_{i,j}(Y_i) = \Lambda(Y_i - T_{i,j-1}) - \Lambda(Y_i - T_{i,j})$. Replacing $F(u|\boldsymbol{Z}_i)$ by $1 - \exp\{-H(u)\exp(\boldsymbol{\gamma}^T\boldsymbol{Z}_i)\}$, a reasonable estimator is the minimizer of the loss function (4.3) with respect to $(\boldsymbol{\beta}, \Lambda, F_{\boldsymbol{\gamma},H})$. Since it is difficult to minimize (4.3) directly, we consider the two-stage estimation by treating F as the nuisance functional parameter. In stage 1, we calculate the partial likelihood estimator $\hat{\boldsymbol{\gamma}}_n$ and the Breslow estimator $\hat{H}_n$ (Breslow, 1972). Then we have $\hat{F}_n(u|\boldsymbol{z}) = 1 - \exp\{-\hat{H}_n(u)\exp(\hat{\boldsymbol{\gamma}}_n^T\boldsymbol{z})\}$. In stage 2, $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$ is obtained by minimizing

$$\ell_n(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} \Delta_i \big\{\triangle N_{i,j} - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \triangle\Lambda_{i,j}(Y_i)\big\}^2$$

$$+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} \frac{1-\Delta_i}{1-\hat{F}_n(Y_i|\boldsymbol{Z}_i)} \int_{Y_i}^{\infty} \big\{\triangle N_{i,j} - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)\triangle\Lambda_{i,j}(u)\big\}^2 d\hat{F}_n(u|\boldsymbol{Z}_i) \tag{4.4}$$

with respect to $(\boldsymbol{\beta}, \Lambda)$. To distinguish with the sample $(Y, \Delta, \boldsymbol{Z})$ used in stage 2, we denote the sample used in stage 1 by $(\tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}})$.

We still consider the monotone I-spline functions to approximate Λ on $[0, \tau]$. We divide $[0, \tau]$ into $m_n + 1$ subintervals

$$0 = t_1 = \cdots = t_d < t_{d+1} < \cdots < t_{m_n+d} < t_{m_n+d+1} = \cdots = t_{m_n+2d} = \tau$$

with knots $\{t_i : i = 1, \cdots, m_n{+}2d\}$, where d represents the order of I-spline functions. Let the I-spline basis functions be $\{I_l(s), l = 1, \cdots, q_n\}$, where $q_n = m_n + d$. Then we define the functional space of the sieve estimator for Λ to be

$$\Phi_n = \left\{ \sum_{l=1}^{q_n} \alpha_l I_l(s) : \alpha_l \geq 0, l = 1, \cdots, q_n \right\}.$$

Define $\boldsymbol{I}(s) = (I_1(s), \cdots, I_{q_n}(s))^T$ and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{q_n})^T$, and replace $\Lambda(s)$ by $\boldsymbol{I}(s)^T\boldsymbol{\alpha}$ in (4.4). Then we can minimize the loss function (4.4) by the constrained BFGS algorithm (Lange, 2001). Setting the minimizer of (4.4) to be $(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\alpha}}_n)$, the spline estimator of $\Lambda(s)$ is $\hat{\Lambda}_n(s) = \boldsymbol{I}(s)^T\hat{\boldsymbol{\alpha}}_n$.

4.3 Asymptotic Properties of the Estimator

In this section, we establish the asymptotic properties of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$. First, we define the following parametric spaces

$$
\begin{aligned}
\mathcal{H}_r &= \left\{ g : |g^{(r-1)}(s) - g^{(r-1)}(t)| \leq c_0 |s-t| \text{ for all } 0 \leq s, t \leq \tau \right\}, \\
\Phi &= \{\Lambda \in \mathcal{H}_r : \Lambda \text{ is nondecreasing continuous function on } [0,\tau] \text{ with } \Lambda(0) = 0\}, \\
\mathcal{F} &= \{F : F \text{ is a distribution function on } [0,\infty)\},
\end{aligned}
$$

where $g^{(r)}$ is the rth derivative of g for $r \geq 1$. For a bounded and convex set $\mathcal{R} \subset \mathbb{R}^d$, denote the interior of $\mathcal{R}$ by $\mathcal{R}^\circ$. Set $F_{\boldsymbol{Z}}$ to be the distribution function of $\boldsymbol{Z}$ with a bounded support $\mathcal{Z} \subset \mathbb{R}^d$, and $(\boldsymbol{\beta}_0, \Lambda_0, F_0(\cdot|\boldsymbol{z})) \in \mathcal{R}^\circ \times \Phi \times \mathcal{F}$ to be the true value of $(\boldsymbol{\beta}, \Lambda, F)$ for all $\boldsymbol{z} \in \mathcal{Z}$. Rewrite $\ell_n(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) = \mathbb{P}_n m(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X)$ with

$$
\begin{aligned}
m(\boldsymbol{\beta}, \Lambda, F; X) = \sum_{j=1}^{K} \Big[\Delta \big\{ \triangle N_j - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_j(Y) \big\}^2 \\
+ \frac{1-\Delta}{1 - F(Y|\boldsymbol{Z})} \int_Y^\infty \left\{ \triangle N_j - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_j(u) \right\}^2 dF(u|\boldsymbol{Z}) \Big].
\end{aligned}
$$

Let $\mathcal{B}$ and $\mathcal{B}^d$ be the collection of Borel sets in $\mathbb{R}$ and $\mathbb{R}^d$, respectively. Then for $B_1, B_2 \in \mathcal{B}_{[0,\tau]} =: \{B \cap [0,\tau] : B \in \mathcal{B}\}$ and $C \in \mathcal{B}^d$, we define

$$
\begin{aligned}
\mu_1(B_1 \times B_2 \times C) &= \int_C \int \sum_{k=1}^{\infty} P(K=k|U=u, \boldsymbol{Z}=\boldsymbol{z}) \\
&\times \sum_{j=1}^{k} P\big((u-T_j) \in B_1, (u-T_{j-1}) \in B_2 | K=k, U=u, \boldsymbol{Z}=\boldsymbol{z}\big) dF_0(u|\boldsymbol{z}) dF_{\boldsymbol{Z}}(\boldsymbol{z}), \\
\mu_2(B_1 \times B_2) &= \mu_1(B_1 \times B_2 \times \mathbb{R}^d), \\
\mu_3(B_1 \times B_2) &= \int_{\mathbb{R}^d} \int \sum_{k=1}^{\infty} P(K=k|U=u, \boldsymbol{Z}=\boldsymbol{z}) \\
&\times P\big((u-T_K) \in B_1, u \in B_2 | K=k, U=u, \boldsymbol{Z}=\boldsymbol{z}\big) dF_0(u|\boldsymbol{z}) dF_{\boldsymbol{Z}}(\boldsymbol{z}).
\end{aligned}
$$

Setting $\triangle\Lambda(s_1, s_2) = \Lambda(s_2) - \Lambda(s_1)$, for any functions $\Lambda_1, \Lambda_2 \in \Phi$, we define the metric

$$d_1^2(\Lambda_1, \Lambda_2) = ||\triangle\Lambda_1(s_1, s_2) - \triangle\Lambda_2(s_1, s_2)||_{L_2(\mu_2)}^2 = E\left[\sum_{j=1}^{K}\left(\triangle\Lambda_{1,j}(U) - \triangle\Lambda_{2,j}(U)\right)^2\right]$$

$$= E\left[\sum_{j=1}^{K}\left\{\Delta(\triangle\Lambda_{1,j}(Y) - \triangle\Lambda_{2,j}(Y))^2 + (1-\Delta)\frac{\int_Y^\infty(\triangle\Lambda_{1,j}(u) - \triangle\Lambda_{2,j}(u))^2 dF_0(u|\boldsymbol{Z})}{1 - F_0(Y|\boldsymbol{Z})}\right\}\right].$$

For any functions F_1, F_2 such that $F_1(\cdot|\boldsymbol{z}) \in \mathcal{F}$ and $F_2(\cdot|\boldsymbol{z}) \in \mathcal{F}$, we define the metric

$$d_2(F_1, F_2) = \sup_{u,\boldsymbol{z}} |F_1(u|\boldsymbol{z}) - F_2(u|\boldsymbol{z})|.$$

For any $(\boldsymbol{\beta}_1, \Lambda_1)$ and $(\boldsymbol{\beta}_2, \Lambda_2)$ in the space $\mathcal{R} \times \Phi$, we define the metric

$$d_3((\boldsymbol{\beta}_1, \Lambda_1), (\boldsymbol{\beta}_2, \Lambda_2)) = \{||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_2^2 + d_1^2(\Lambda_1, \Lambda_2)\}^{1/2}.$$

We need the following conditions to establish the asymptotic properties.

(C1) $0 < \Lambda_0(\tau) < \infty$.

(C2) $0 < F_0(\tau|\boldsymbol{Z} = \boldsymbol{0}) < 1$. F_0 is absolutely continuous with respect to Lebesgue measure. Furthermore, the density function $f_0(s|\boldsymbol{Z} = \boldsymbol{0})$ has a uniform positive lower bound for all $s \in [M_1, \tau]$, where M_1 is a constant representing the minimum value of the support of F_0.

(C3) $E\big[\sum_{j=1}^{K}\{\triangle N_j - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\}^2\big] < \infty$.

(C4) The probability of censoring $\varrho = P(Y < U)$ satisfies that $0 < \varrho < 1$.

(C5) The measure $\mu_2 \times \mathcal{Z}$ is absolutely continuous with respect to μ_1.

(C6) $P(\boldsymbol{a}^T\boldsymbol{Z} \neq c) > 0$, $\forall \boldsymbol{a} \neq \boldsymbol{0} \in \mathbb{R}^{\mathbf{d}}$ and $\forall c \in \mathbb{R}$.

(C7) There is a constant $M_2 > 0$ such that $P(K \leq M_2) = 1$.

(C8) The number of subinterval in $[0, \tau]$ satisfies $m_n = O(n^\nu)$ for $0 < \nu < 1/2$. Furthermore,

$$\max_{d+1 \leq i \leq m_n+d+1} |t_i - t_{i-1}| = O(n^{-\nu}),$$

and there is a constant $M_3 > 0$ such that

$$\frac{\max_{d+1 \le i \le m_n+d+1} |t_i - t_{i-1}|}{\min_{d+1 \le i \le m_n+d+1} |t_i - t_{i-1}|} \le M_3$$

uniformly for n.

(C9) $P(T_j - T_{j-1} \ge M_4$ for all $j = 1, \cdots, K) = 1$ with some constant $M_4 > 0$.

Remark 1. Condition (C1) is common on the baseline mean function. Condition (C2) holds when the end time of study is smaller than the maximum value of the support of F_0. Condition (C3) assumes a finite second order for the counting processes, which is necessary for the least squares-based estimation. Condition (C4) is common in right-censored data analysis, meaning that the censoring rate is between 0 to 1. According to Wellner and Zhang (2007), Conditions (C5) and (C6) are necessary for the identifiability of the semiparametric model. Condition (C7) holds in many applications for panel count data. Condition (C8) is a regular assumption for the spline approximation by Lu, Zhang, and Huang (2007, 2009). By Wellner and Zhang (2007), Condition (C9) meaning that the adjacent observation times are separable is regular in applications of panel count data.

Theorem 4.1 (Consistency for Two-Stage Estimator)**.** *Suppose that Conditions* (C1)–(C9) *hold. Then for* $0 \le b_1 \le b_2 \le \tau$ *satisfying* $\mu_3([0, b_1] \times [b_2, \tau]) > 0$*, we have*

$$||\triangle\hat{\Lambda}_n(s_1, s_2)1_{\{(s_1,s_2)\in[b_1,b_2]\times[b_1,b_2]\}} - \triangle\Lambda_0(s_1, s_2)1_{\{(s_1,s_2)\in[b_1,b_2]\times[b_1,b_2]\}}||^2_{L_2(\mu_2)}$$
$$+||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_2^2 = o_p(1).$$

In particular, when $\mu_3(\{0\} \times \{\tau\}) > 0$*, we have* $d_3((\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n), (\boldsymbol{\beta}_0, \Lambda_0)) = o_p(1)$.

Besides Conditions (C1)–(C9), we need the following additional conditions to establish the convergence rate and the asymptotic normality.

(C10) $\inf_{\boldsymbol{z}\in\mathcal{Z}} P(U \geq \tau|\boldsymbol{Z} = \boldsymbol{z}) = \omega_1 > 0$ and $P(C \geq \tau) = \omega_2 > 0$ for some constants ω_1 and ω_2.

(C11) μ_2 is absolutely continuous with respect to Lebesgue measure with a derivative $\dot{\mu}_2$, and $\dot{\mu}_2$ has a uniform positive lower bound.

(C12) There is a positive constant M_5 such that $1/M_5 < \Lambda_0'(s) < M_5$ for all $s \in [\tau', \tau]$, where $0 < \tau' \leq \tau$ such that $\Lambda_0(\tau') > 0$.

(C13) $E(e^{cN(t)})$ is uniformly bounded for $t \in [0, \tau]$ and some constant c.

(C14) For all $\boldsymbol{a} \in \mathbb{R}^d$, there is a constant $\eta \in (0, 1)$ such that

$$\boldsymbol{a}^T Var(\boldsymbol{Z}|S_1, S_2)\boldsymbol{a} \geq \eta \boldsymbol{a}^T E(\boldsymbol{Z}\boldsymbol{Z}^T|S_1, S_2)\boldsymbol{a} \text{ a.e.}$$

for $(S_1, S_2, \boldsymbol{Z})$ having the distribution μ_1.

Remark 2. By Kong et al. (2018), Condition (C10) is necessary for the uniform weak convergence rate of $\hat{F}_n$ on a finite interval. According to Wellner and Zhang (2007) and Lu, Zhang, and Huang (2009), Conditions (C11)–(C14) are common in the analysis of panel count data. Condition (C11) supposes that the total observation time of counting processes has a positive density. Condition (C12) assumes that the derivative of the mean function has a uniform positive upper and lower bound, and it may be stronger than necessary. Condition (C13) holds when $N(t)$ is uniformly bounded or from a Poisson-type process. By Remark 3.4 of Wellner and Zhang (2007), Condition (C14) is satisfied in many applications.

Theorem 4.2 (Rate of Convergence)**.** *Suppose that Conditions* (C1)–(C14) *hold, and* $\mu_3(\{0\} \times \{\tau\}) > 0$. *Taking* $\nu = 1/(1 + 2r)$, *we have* $d_3((\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n), (\boldsymbol{\beta}_0, \Lambda_0)) = O_p(n^{-r/(1+2r)})$.

Remark 3. Although the overall convergence rate of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$ is slower than $n^{1/2}$, the convergence rate of $\hat{\boldsymbol{\beta}}_n$ is still $n^{1/2}$, and we can also find a function of $\hat{\Lambda}_n$ having the convergence rate $n^{1/2}$. The following theorem establishes the asymptotic normality of $\hat{\boldsymbol{\beta}}_n$ and a function of $\hat{\Lambda}_n$.

Theorem 4.3 (Asymptotic Normality)**.** *Suppose that Conditions* (C1)–(C14) *hold, and* $\Lambda_0 \in \mathcal{H}_r$ *with* $r \geq 2$*.*

(i) *For all* $\boldsymbol{h}_1 \in \mathcal{R}$ *and* $h_2 \in \mathcal{H}_r$*, the estimators* $\hat{\boldsymbol{\beta}}_n$ *and* $\hat{\Lambda}_n$ *satisfy that*

$$\sqrt{n}R_1(\boldsymbol{h}_1, h_2)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \sqrt{n}R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0) \rightsquigarrow N(0, \sigma_0[\boldsymbol{h}_1, h_2]^2),$$

where $R_1(\boldsymbol{h}_1, h_2)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$*,* $R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0)$ *and* $\sigma_0[\boldsymbol{h}_1, h_2]^2$ *are defined in the Appendix.*

(ii) *Furthermore, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \rightsquigarrow N(0, (A^*)^{-1}B^*((A^*)^{-1})^T),$$

$$\sqrt{n}\mathcal{P}\left[\sum_{j=1}^{K}\left\{\left(\triangle h_{2,j}(U) + \triangle\Lambda_{0,j}(U)R^{**}(h_2)^T\boldsymbol{Z}\right)\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})(\triangle\hat{\Lambda}_{n,j}(U) - \triangle\Lambda_{0,j}(U))\right\}\right]$$

$$\rightsquigarrow N(0, \tilde{\sigma}_0[h_2]^2)$$

for all $h_2 \in \mathcal{H}_r$*, where* $\triangle h_{2,j}(U) = h_2(U - T_{j-1}) - h_2(U - T_j)$*, and* A^**,* B^**,* $R^{**}(h_2)$*, and* $\tilde{\sigma}_0[h_2]$ *are defined in the Appendix.*

4.4 Simulation Studies

In this section, we conducted the simulation studies to demonstrate the performance of our method. We generated the covariate vector $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$ by the following scheme: $Z_{i1} \sim \text{Unif}(0,1)$, $Z_{i2} \sim N(0,1)$ and $Z_{i3} \sim \text{Bernoulli}(0.5)$. Given the covariate vector $\boldsymbol{Z}_i$, the terminal event U_i satisfied model (4.2) with $\boldsymbol{\gamma}_0 = (\gamma_1, \gamma_2, \gamma_3)^T = (-1, 1, 1)^T$ and $H_0(u) = u - 5$ for $u \in [5, \infty)$. The censoring time C_i was from $5 + \kappa\exp(1)$, where κ was selected to yield 20% and 40% censoring rate, respectively. Then we had $Y_i = U_i \wedge C_i$ and $\Delta_i = 1_{\{U_i \leq C_i\}}$. Took the end time of study to be $\tau = 10$, and the total number of observation K_i to be from $\{1, 2, 3, 4, 5, 6\}$ with equal probability. Given Y_i and K_i, the observation time points

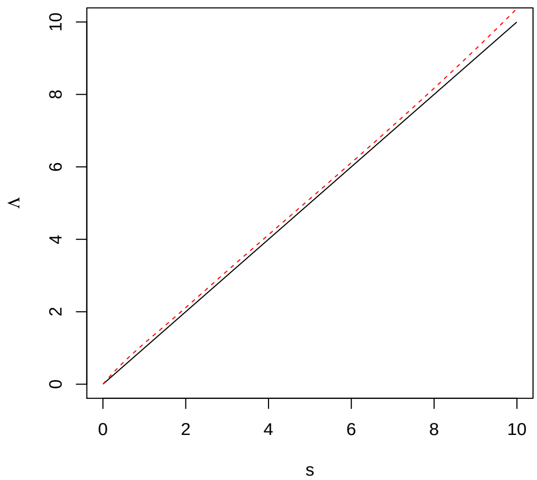
$\underline{T_i} = (T_{i1}, \cdots, T_{iK_i})$ were K_i ordered random variables from $\text{Unif}(0, Y_i)$. Under model (4.1), we considered the following two different cases of Λ_0:

$$\text{Case 1} : \Lambda_0(s) = s, \text{ and Case 2} : \Lambda_0(s) = \frac{10s}{s+1},$$

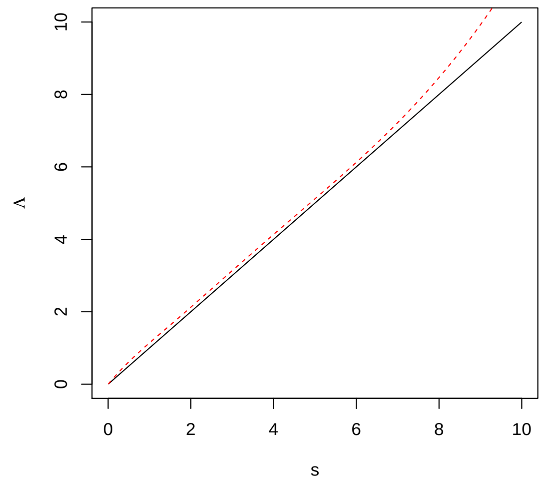
to generate the counting process $\underline{N_i} = \{N_i(T_{i1}), \cdots, N_i(T_{iK_i})\}$ from the Poisson process with $\boldsymbol{\beta}_0 = (\beta_1, \beta_2, \beta_3)^T = (-1, 0.5, 0.5)^T$. That is in Case 1, $N_i(T_{i1})$ was from the Poisson distribution with mean $T_{i1}e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}$, and $N_i(T_{ij}) - N_i(T_{i(j-1)})$ was from the Poisson distribution with mean $(T_{ij} - T_{i(j-1)})e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}$ for $j = 2, \cdots, K_i$; in Case 2, $N_i(T_{i1})$ was from the Poisson distribution with mean $10\{U_i/(U_i+1) - (U_i - T_{i1})/(U_i - T_{i1}+1)\}e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}$, and $N_i(T_{ij}) - N_i(T_{i(j-1)})$ was from the Poisson distribution with mean $10\{(U_i - T_{i(j-1)})/(U_i - T_{i(j-1)} + 1) - (U_i - T_{ij})/(U_i - T_{ij} + 1)\}e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}$ for $j = 2, \cdots, K_i$. For the knots of I-spline, we took t_{d+1}, t_{d+2}, t_{d+3} to be the 25%, 50% and 75% quantiles of $\{Y_i - T_{ij} : j = 1, \cdots, K_i;\ i = 1, \cdots, n\}$ with $d = m_n = 3$. Since it was difficult to obtain σ_0^2 directly, the standard error of our estimator was estimated based on 100 bootstrap samples. The initial value of the BFGS iteration was taken as $\boldsymbol{\alpha} = (1, 1, 1, 1, 1, 1)^T$ and $\boldsymbol{\beta} = \mathbf{0}$. The sample size was $n = 50$ and $n = 100$, respectively. The simulation results were summarized based on 1000 replications.

The dash lines in Figures 4.1–4.2 display the mean of the estimation, and the solid lines in this figure are the true value Λ_0 for comparison. In these Figures, the estimation functions are close to the true value Λ_0, meaning that our estimator $\hat{\Lambda}_n$ is consistent. The simulation results for the regression parameter $\boldsymbol{\beta}$ are summarized in Table 4.1 and Table 4.2. Both of the two tables demonstrate that our estimations are close to the true value of $\boldsymbol{\beta}$, and the biases tend to decrease as the sample size increases and the censoring rate decreases. The tables also show that the sample standard errors (SSE) are close to the corresponding estimation of standard errors (ESE). Both of them are decreasing when the sample size increases and the censoring rate decreases. The 95% empirical coverage probabilities (CP) are close to 0.95.

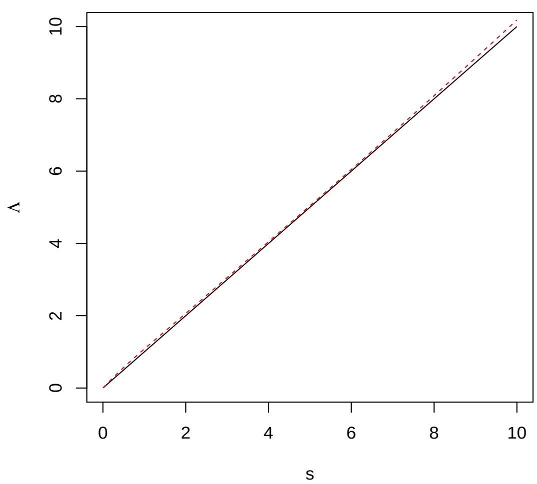
(a) $n = 50$ and censoring rate = 20%



(b) $n = 50$ and censoring rate = 40%



(c) $n = 100$ and censoring rate = 20%



(d) $n = 100$ and censoring rate = 40%

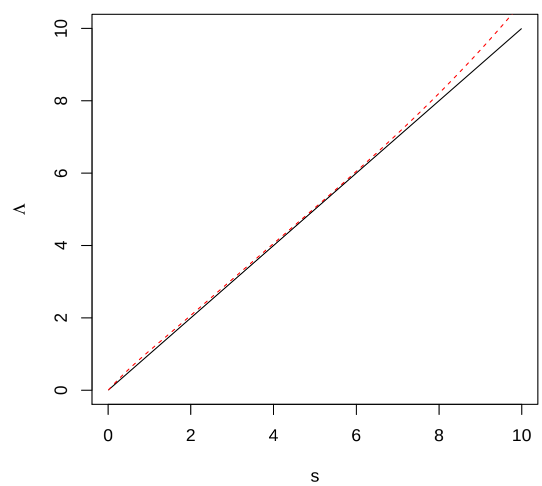
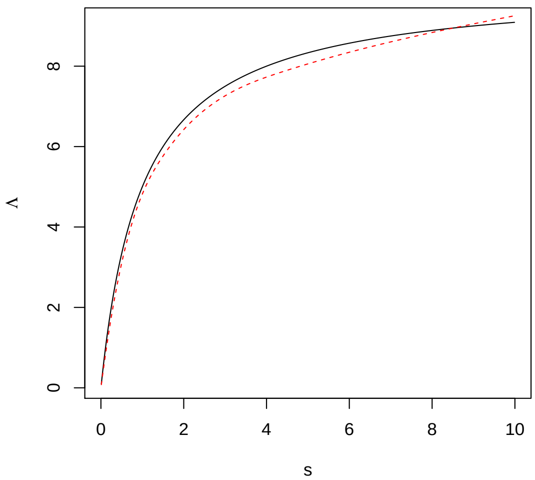
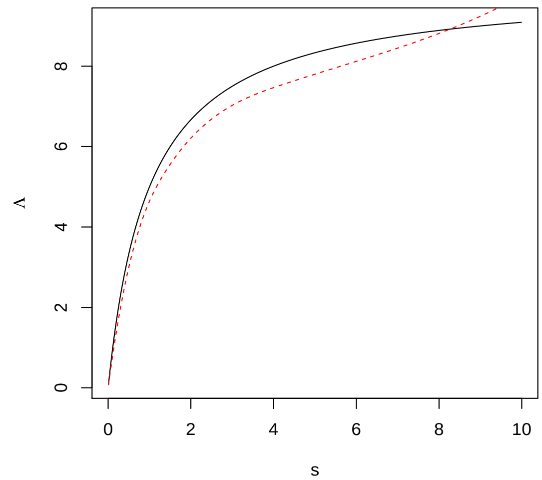


Figure 4.1: Simulation results for the baseline mean function in Case 1. The solid lines are the true functions, and the dash lines are the estimates.

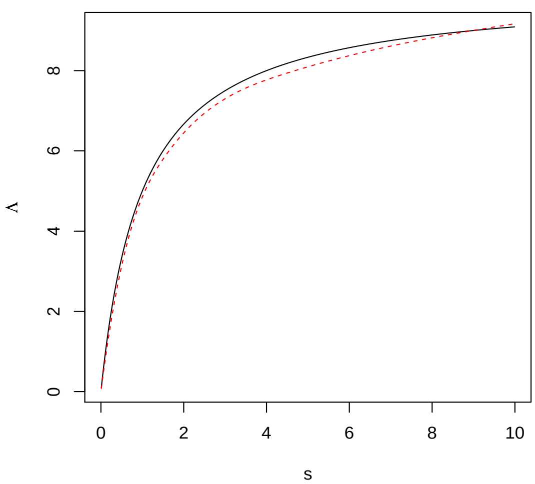
(a) $n = 50$ and censoring rate = 20%



(b) $n = 50$ and censoring rate = 40%



(c) $n = 100$ and censoring rate = 20%



(d) $n = 100$ and censoring rate = 40%

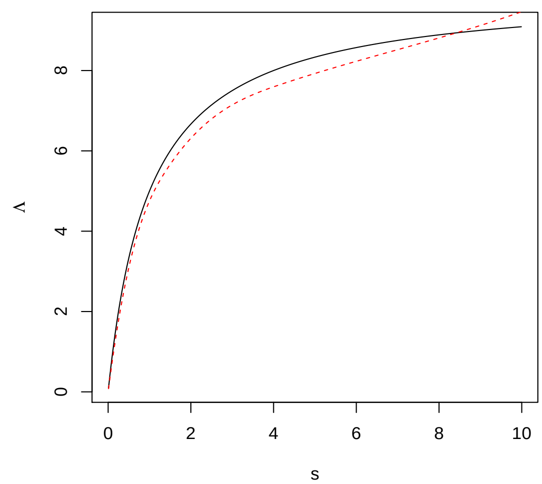


Figure 4.2: Simulation results for the baseline mean function in Case 2. The solid lines are the true functions, and the dash lines are the estimates.

Table 4.1: Simulation results of the estimates of parameter $\boldsymbol{\beta}$ in Case 1.

	Censoring rate = 20%			Censoring rate = 40%		
	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$
			n=50			
Estimates	-1.029	0.511	0.519	-1.032	0.514	0.524
SSE	0.352	0.107	0.194	0.375	0.116	0.204
ESE	0.353	0.112	0.200	0.380	0.125	0.217
CP	0.974	0.948	0.938	0.972	0.964	0.950
			n=100			
Estimates	-1.016	0.508	0.509	-1.016	0.509	0.510
SSE	0.242	0.078	0.142	0.257	0.082	0.148
ESE	0.230	0.071	0.134	0.242	0.076	0.141
CP	0.961	0.909	0.935	0.964	0.922	0.931

Table 4.2: Simulation results of the estimates of parameter $\boldsymbol{\beta}$ in Case 2.

	Censoring rate = 20%			Censoring rate = 40%		
	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$
			n=50			
Estimates	-1.054	0.525	0.536	-1.078	0.541	0.564
SSE	0.456	0.135	0.249	0.491	0.150	0.275
ESE	0.459	0.147	0.257	0.524	0.171	0.291
CP	0.979	0.961	0.952	0.976	0.966	0.944
			n=100			
Estimates	-1.028	0.511	0.517	-1.040	0.525	0.534
SSE	0.287	0.090	0.172	0.307	0.099	0.189
ESE	0.275	0.086	0.161	0.299	0.095	0.176
CP	0.960	0.945	0.928	0.959	0.933	0.927

4.5 Real Data Analysis

In this section, we applied the proposed semiparametric approach to analyze the occurrence rate of serious diseases for elder people in China based on the datasets of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) in the period 1998 to 2014. Similar to Chapter 3, we denoted the terminal event and the censoring event by the death and loss-of-connection, respectively. Then $Y = U \wedge C$ was the follow-up time, and $\Delta = 1_{\{U \leq C\}}$ was the indicator of death. Took the number of months from the baseline survey in 1998 to the jth follow-up wave of survey to be T_j. Then the observation time points were $\underline{T} = \{T_j : j = 1, \cdots, K\}$, where $K \leq 6$ represented the total number of follow-up surveys. Let $\tau = 197$ be the largest number of follow-up months. Set the occurrence numbers of serious diseases before the jth follow-up survey to be $N(T_j)$, and the occurrence numbers of serious diseases from the jth follow-up survey to death to be $\tilde{N}(T_j)$. For the semiparametric model, we chose 6 covariates in our analysis, including the gender ($Z_1 = 1$ for male and $Z_1 = 0$ for female), the living area from 1998 to 2000 ($Z_2 = 1$ for urban and $Z_2 = 0$ for rural), the number of children (Z_3), the systolic blood pressure (Z_4), the heart rates (Z_5), and the lung capacities (Z_6). All the continuous covariates were standardized before the analysis. 4831 individuals were interviewed in both of the surveys in 1998 and 2000. After removing 1099 individuals with missing or typo records and 1160 individuals living in different areas in 1998 and 2000, we focused on 2572 individuals with the censoring rate of 26.36%. We considered the cubic I-spline with order $d = 3$, and we divided $[0, \tau]$ by the knots $t_{d+1} = \tau/4$, $t_{d+2} = \tau/2$ and $t_{d+1} = 3\tau/4$ to calculate the I-spline basis function $\boldsymbol{I}(s)$. We chose the initial value $\boldsymbol{\alpha} = (1, 1, 1, 1, 1, 1)^T$ and $\boldsymbol{\beta} = \mathbf{0}$ to start the BFGS algorithm. Similar to Section 4.4, we applied the bootstrap procedure with 100 replications to estimate the asymptotic variance of the proposed estimator.

Table 4.3: Inference results for the CLHLS data.

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
Estimates	-0.176	0.260	-0.081	-0.101	0.069	0.140
ESE	0.092	0.098	0.041	0.057	0.031	0.053
p-value	0.055*	0.008**	0.047**	0.077*	0.023**	0.009**

* represents significance level of 0.1; ** represents significance level of 0.05.

In Figure 4.3, the solid line represents the estimate of the baseline mean function Λ_0, and the dash lines represent the 2.5- and 97.5- percentiles based on the 100

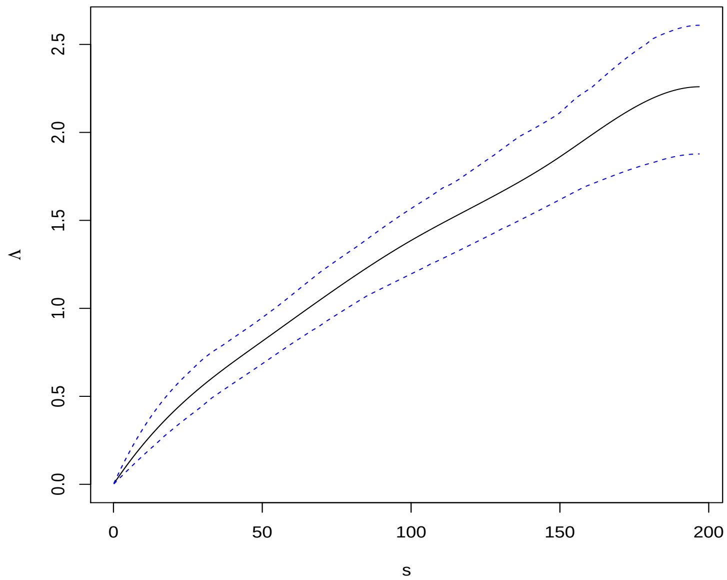


Figure 4.3: Estimate, 2.5-percentile and 97.5-percentile of the baseline mean function for the CLHLS data.

bootstrap samples, respectively. Table 4.3 shows the parameter estimation results, in which all the covariates have significant effects on the occurrence rate of serious diseases. Specifically, Z_1 and Z_4 are significant at the 0.1 level, and Z_2, Z_3, Z_5, and Z_6 are significant at the 0.05 level. Furthermore, Z_2, Z_5, and Z_6 have positive effects on the occurrence rate of serious diseases, meaning that elder people living in urban or having higher heart rates or having larger lung capacities tend to suffer from more serious diseases before death. In contrast, Z_1, Z_3 and Z_4 have negative effects on the occurrence rate of serious diseases, which implies that elder females or elder people having fewer children or elder people having lower systolic blood pressure are likely to suffer from more serious diseases before death.

4.6 Appendix

4.6.1 Lemmas

Lemma 4.1. *Suppose that Conditions* (C1) *and* (C7) *hold. Then we have*

$$E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right]\lesssim d_3^2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0)).$$

In addition, if Condition (C14) *holds, then*

$$E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right]\gtrsim d_3^2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0)).$$

Proof. First, by the Cauchy–Schwarz inequality, under Conditions (C1) and (C7), we have

$$\begin{aligned}
&E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right]\\
=&E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right]\\
+&E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right]\\
+2&E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)\right.\\
\times&\left.\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)\right]\\
\lesssim&E\left[\sum_{j=1}^{K}\big(\triangle\Lambda_j(U)-\triangle\Lambda_{0,j}(U)\big)^2\right]+E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\big)^2\right]\\
+2&\left\{E\left[\sum_{j=1}^{K}\big(\triangle\Lambda_j(U)-\triangle\Lambda_{0,j}(U)\big)^2\right]\right\}^{\frac{1}{2}}\left\{E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\big)^2\right]\right\}^{\frac{1}{2}}
\end{aligned}$$

$$\lesssim E\left[\sum_{j=1}^{K}\big(\triangle\Lambda_j(U)-\triangle\Lambda_{0,j}(U)\big)^2\right]+E\left[\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\big)^2\right].$$

By the mean value theorem, there exists a $\boldsymbol{\beta}_{\xi_1}\in\mathcal{R}$ such that

$$E\left[\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\big)^2\right]=E\left[\exp(2\boldsymbol{\beta}_{\xi_1}^T\boldsymbol{Z})\{\boldsymbol{Z}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}^2\right]\lesssim||\boldsymbol{\beta}-\boldsymbol{\beta}_0||_2^2.$$

It follows that

$$E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right]$$
$$\lesssim||\boldsymbol{\beta}-\boldsymbol{\beta}_0||_2^2+d_1^2(\Lambda,\Lambda_0)=d_3^2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0)).$$

The proof of the second inequality is similar to the proof in Theorem 3.2 of Wellner and Zhang (2007). For $\Lambda\in\Phi$, $\boldsymbol{\beta}\in\mathbb{R}^d$ and $(S_1,S_2,\boldsymbol{Z})\sim\mu_1$, let $g(\xi)=\exp(\boldsymbol{\beta}_\xi^T\boldsymbol{Z})\triangle\Lambda_\xi(S_1,S_2)$, where $\triangle\Lambda_\xi(S_1,S_2)=\xi\triangle\Lambda(S_1,S_2)+(1-\xi)\triangle\Lambda_0(S_1,S_2)$ and $\boldsymbol{\beta}_\xi=\xi\boldsymbol{\beta}+(1-\xi)\boldsymbol{\beta}_0$ with $\xi\in(0,1)$. Then we have $\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda(S_1,S_2)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_0(S_1,S_2)=g(1)-g(0)$. By the mean value theorem, there is a $\xi_2\in(0,1)$ such that

$$\begin{aligned}
&g(1)-g(0)=g'(\xi_2)\\
=&\exp(\boldsymbol{\beta}_{\xi_2}^T\boldsymbol{Z})[(\triangle\Lambda(S_1,S_2)-\triangle\Lambda_0(S_1,S_2))+\triangle\Lambda_{\xi_2}(S_1,S_2)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{Z}]\\
=&\exp(\boldsymbol{\beta}_{\xi_2}^T\boldsymbol{Z})\left[\left\{1+\frac{\xi_2\big(\triangle\Lambda(S_1,S_2)-\triangle\Lambda_0(S_1,S_2)\big)}{\triangle\Lambda_0(S_1,S_2)}\right\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{Z}\triangle\Lambda_0(S_1,S_2)\right.\\
&\left.+(\triangle\Lambda(S_1,S_2)-\triangle\Lambda_0(S_1,S_2))\right],
\end{aligned}$$

where g' is the derivative of g. Setting $g_1=(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{Z}\triangle\Lambda_0(S_1,S_2)$, $g_2=(\triangle\Lambda(S_1,S_2)-\triangle\Lambda_0(S_1,S_2))$ and $g_3=1+\xi_2\big(\triangle\Lambda(S_1,S_2)-\triangle\Lambda_0(S_1,S_2)\big)/\triangle\Lambda_0(S_1,S_2)$, we have

$g(1) - g(0) = \exp(\boldsymbol{\beta}_{\xi_2}^T Z)(g_1 g_3 + g_2)$. This yields that

$$E\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U) - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right] = E_{\mu_1}\left[\big(g(1)-g(0)\big)^2\right]$$

$$= E_{\mu_1}\left[\exp(2\boldsymbol{\beta}_{\xi_2}^T\boldsymbol{Z})\,(g_1g_3+g_2)^2\right] \gtrsim E_{\mu_1}\left[(g_1g_3+g_2)^2\right].$$

Similar to the proof of Theorem 3.2 in Wellner and Zhang (2007), Condition (C14) implies that

$$E_{\mu_1}^2\left[g_1g_2\right] \le (1-\eta)E_{\mu_1}\left[(g_1)^2\right]E_{\mu_1}\left[(g_2)^2\right].$$

According to Lemma 8.8 of van der Vaart (2002), we have

$$E_{\mu_1}\left[(g_1g_3+g_2)^2\right] \gtrsim E_{\mu_1}\left[(g_1)^2\right] + E_{\mu_1}\left[(g_2)^2\right] \gtrsim d_3^2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0)).$$

□

Define

$$s^{(0)}(\boldsymbol{\gamma},t) = E\left[1_{\{\tilde{Y}\ge t\}}\exp(\boldsymbol{\gamma}^T\tilde{\boldsymbol{Z}})\right], \boldsymbol{s}^{(1)}(\boldsymbol{\gamma},t) = E\left[\tilde{\boldsymbol{Z}}1_{\{\tilde{Y}\ge t\}}\exp(\boldsymbol{\gamma}^T\tilde{\boldsymbol{Z}})\right],$$

$$\boldsymbol{s}^{(2)}(\boldsymbol{\gamma},t) = E\left[\tilde{\boldsymbol{Z}}^{\otimes 2}1_{\{\tilde{Y}\ge t\}}\exp(\boldsymbol{\gamma}^T\tilde{\boldsymbol{Z}})\right], \boldsymbol{\varsigma}(\boldsymbol{\gamma},t,\boldsymbol{z}) = H_0(t)\boldsymbol{z} - \int_0^t \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\gamma},u)}{s^{(0)}(\boldsymbol{\gamma},u)}dH_0(u),$$

$$\mathcal{I}(\boldsymbol{\gamma}) = \int_0^\infty\left[\frac{\boldsymbol{s}^{(2)}(\boldsymbol{\gamma},t)}{s^{(0)}(\boldsymbol{\gamma},t)} - \left\{\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\gamma},t)}{s^{(0)}(\boldsymbol{\gamma},t)}\right\}^{\otimes 2}\right]s^{(0)}(\boldsymbol{\gamma},t)dH_0(t),$$

$$\boldsymbol{e}(\boldsymbol{\gamma},t,\boldsymbol{z}) = \mathcal{I}^{-1}(\boldsymbol{\gamma})\left[\boldsymbol{z} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\gamma},t)}{s^{(0)}(\boldsymbol{\gamma},t)}\right],$$

and

$$\tilde{M}_i(t) = 1_{\{\tilde{Y}_i\le t,\tilde{\Delta}_i=1\}} - \exp(\boldsymbol{\gamma}_0^T\tilde{\boldsymbol{Z}}_i)\int_{-\infty}^t 1_{\{\tilde{Y}_i\ge u\}}dH_0(u).$$

Then we have the following lemma.

Lemma 4.2. *Suppose that Condition* (C10) *holds. Then we have*

$$\hat{F}_n(t|\boldsymbol{z}) - F_0(t|\boldsymbol{z}) = \mathbb{P}_n\Omega(t,\boldsymbol{z};\tilde{Y},\tilde{\Delta},\tilde{\boldsymbol{Z}}) + o_p(n^{-1/2}),$$

where

$$\begin{aligned}
&\Omega(t,\boldsymbol{z};\tilde{Y},\tilde{\Delta},\tilde{\boldsymbol{Z}})\\
=&(1-F_0(t|\boldsymbol{z}))e^{\boldsymbol{\gamma}_0^T\boldsymbol{z}}\left[\boldsymbol{\varsigma}(\boldsymbol{\gamma}_0,t,\boldsymbol{z})^T\left\{\tilde{\Delta}\boldsymbol{e}(\boldsymbol{\gamma}_0,\tilde{Y},\tilde{\boldsymbol{Z}}) - \int_0^{\tilde{Y}}\boldsymbol{e}(\boldsymbol{\gamma}_0,u,\tilde{\boldsymbol{Z}})e^{\boldsymbol{\gamma}_0^T\tilde{\boldsymbol{Z}}}dH_0(u)\right\}\right.\\
&\left.+1_{(t\geq\tilde{Y})}\tilde{\Delta}\frac{1}{s^{(0)}(\boldsymbol{\gamma}_0,\tilde{Y})} - \int_0^{\tilde{Y}\wedge t}\frac{e^{\boldsymbol{\gamma}_0^T\tilde{\boldsymbol{Z}}}}{s^{(0)}(\boldsymbol{\gamma}_0,u)}dH_0(u)\right].
\end{aligned}$$

Hence, $d_2(\hat{F}_n,F_0) = O_p(n^{-1/2})$.

Proof. Define the Breslow estimator of $H(t)$ as

$$\hat{H}_n(t) = \int_0^t \sum_{i=1}^n 1_{\{\tilde{Y}_i\geq u\}}\exp(\tilde{\boldsymbol{Z}}_i^T\hat{\boldsymbol{\gamma}}_n)d\left\{\sum_{i=1}^n 1_{\{\tilde{Y}_i\leq u,\tilde{\Delta}_i=1\}}\right\}.$$

According to the proof of Lemma A.3. of Kong et al. (2018), we have

$$\begin{aligned}
&\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0 = \frac{1}{n}\sum_{i=1}^n\int_0^\infty \boldsymbol{e}(\boldsymbol{\gamma}_0,t,\tilde{\boldsymbol{Z}}_i)d\tilde{M}_i(t) + o_p(n^{-1/2}),\\
&\hat{H}_n(t) - H_0(t)\\
=&\left\{-\int_0^t\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\gamma}_0,u)}{s^{(0)}(\boldsymbol{\gamma}_0,u)}dH_0(u)\right\}^T(\hat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0) + \frac{1}{n}\sum_{i=1}^n\int_0^t\frac{1}{s^{(0)}(\boldsymbol{\gamma}_0,u)}d\tilde{M}_i(u) + o_p(n^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
&\hat{F}_n(t|\boldsymbol{z}) - F_0(t|\boldsymbol{z})\\
=&(1-F_0(t|\boldsymbol{z}))\cdot e^{\boldsymbol{\gamma}_0^T\boldsymbol{z}}\left\{H_0(t)\boldsymbol{z}^T(\hat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0) + (\hat{H}_n(t)-H_0(t))\right\} + o_p(n^{-1/2}).
\end{aligned}$$

It follows that

$$\hat{F}_n(t|\boldsymbol{z}) - F_0(t|\boldsymbol{z}) = (1 - F_0(t|\boldsymbol{z}))e^{\boldsymbol{\gamma}_0^T\boldsymbol{z}}\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\varsigma}(\boldsymbol{\gamma}_0, t, \boldsymbol{z})^T\int_0^\infty \boldsymbol{e}(\boldsymbol{\gamma}_0, u, \tilde{\boldsymbol{Z}}_i)d\tilde{M}_i(u)\right.$$

$$\left. + \int_0^t \frac{1}{s^{(0)}(\boldsymbol{\gamma}_0, u)}d\tilde{M}_i(u)\right\} + o_p(n^{-1/2}) = \mathbb{P}_n\Omega(t, \boldsymbol{z}; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}}) + o_p(n^{-1/2}).$$

□

Lemma 4.3. *Suppose that Conditions* (C1)*,* (C4)*,* (C7)*,* (C8) *and* (C11)–(C13) *hold. Define the class*

$$\mathcal{M}_\eta(F) = \{m(\boldsymbol{\beta}, \Lambda, F; X) - m(\boldsymbol{\beta}_0, \Lambda_0, F; X) : \ \boldsymbol{\beta} \in \mathcal{R}, \ \Lambda \in \Phi_n, \ d_3((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) \le \eta\},$$

for a fixed F satisfying $d_2(F, F_0) \le \delta$. For sufficiently small δ and any $\varepsilon < \eta$, we have

$$\log N_{[]}(\varepsilon, \mathcal{M}_\eta(F), ||\cdot||_{P,B}) \lesssim q_n \log(\eta/\varepsilon),$$

where the Bernstein norm is defined as $||f||_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$.

Proof. By the definition of $d_3^2((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0))$, $d_3^2((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) \le \eta^2$ yields that $d_1(\Lambda, \Lambda_0) \le \eta$. By Condition (C8), according to Shen and Wong (1994, page 597), for all $\Lambda \in \Phi_n$, $d_1(\Lambda, \Lambda_0) \le \eta$ and $\varepsilon < \eta$, we can find a set of brackets

$$\{[\triangle\Lambda_i^L, \triangle\Lambda_i^U] : \ ||\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2)||_{L_2(\mu_2)} \le \varepsilon, \ i = 1, \cdots, (\eta/\varepsilon)^{c_1 q_n}\}$$

such that $\triangle\Lambda(s_1, s_2) \in [\triangle\Lambda_i^L(s_1, s_2), \triangle\Lambda_i^U(s_1, s_2)]$ for all $s_1, s_2 \in [0, \tau]$. Similar to the proof of Theorem 3.2 of Wellner and Zhang (2007), under Conditions (C11) and (C12), we have $0 \le \triangle\Lambda_i^L(s_1, s_2) \le \triangle\Lambda_i^U(s_1, s_2) \le \triangle\Lambda_0(s_1, s_2) + \varepsilon^*$ and $\triangle\Lambda_i^U(s_1, s_2) - \triangle\Lambda_i^L(s_1, s_2) \le 2\varepsilon^*$ for all $s_1, s_2 \in [0, \tau]$, where $\varepsilon^* = ((\varepsilon^2 + \eta^2)^{1/2}/c_0)^{2/3}$ for some constant c_0. Furthermore, since $\mathcal{Z}$ is bounded and $\mathcal{R}$ is a compact set in $\mathbb{R}^d$, we can construct an ε-net $\{\boldsymbol{\beta}_s : \ s = 1, \cdots, \lceil (c_2\eta/\varepsilon)^d \rceil\}$ such that for all $\boldsymbol{\beta} \in \mathcal{R}$, there is a

$\boldsymbol{\beta}_s$ satisfying $|\boldsymbol{\beta}^T\boldsymbol{Z} - \boldsymbol{\beta}_s^T\boldsymbol{Z}| \le \varepsilon$ and $|\exp(\boldsymbol{\beta}^T\boldsymbol{Z}) - \exp(\boldsymbol{\beta}_s^T\boldsymbol{Z})| \le c\varepsilon$ for some constant c.

Next we consider the ε-bracket of $m(\boldsymbol{\beta}, \Lambda, F; X) - m(\boldsymbol{\beta}_0, \Lambda_0, F; X)$. After some algebraic calculations,

$$
\begin{aligned}
& m(\boldsymbol{\beta}, \Lambda, F; X) - m(\boldsymbol{\beta}_0, \Lambda_0, F; X) \\
=& \Delta \sum_{j=1}^{K} \left\{ \exp(\boldsymbol{\beta}^T\boldsymbol{Z})^2 \triangle\Lambda_j(Y)^2 - 2\exp(\boldsymbol{\beta}^T\boldsymbol{Z}) \triangle\Lambda_j(Y) \triangle N_j + \triangle N_j^2 \right\} \\
+& \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})} \sum_{j=1}^{K} \int_Y^\infty \Big\{ \exp(\boldsymbol{\beta}^T\boldsymbol{Z})^2 \triangle\Lambda_j(u)^2 - 2\exp(\boldsymbol{\beta}^T\boldsymbol{Z}) \triangle\Lambda_j(u) \triangle N_j \\
+& \triangle N_j^2 \Big\} dF(u|\boldsymbol{Z}) - m(\boldsymbol{\beta}_0, \Lambda_0, F; X).
\end{aligned}
$$

Hence, the ε-brackets can be chosen as $[m_{i,s}^L, m_{i,s}^U]$, where

$$
\begin{aligned}
m_{i,s}^L =& \Delta \sum_{j=1}^{K} \bigg[\Big\{ \big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) - c\varepsilon\big)^2 \big(\triangle\Lambda_{i,j}^L(Y)\big)^2 - 2\big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) + c\varepsilon\big) \triangle\Lambda_{i,j}^U(Y) \triangle N_j \\
& + \triangle N_j^2 \Big\} + \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})} \int_Y^\infty \Big\{ \big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) - c\varepsilon\big)^2 \big(\triangle\Lambda_{i,j}^L(u)\big)^2 \\
& - 2\big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) + c\varepsilon\big) \triangle\Lambda_{i,j}^U(u) \triangle N_j + \triangle N_j^2 \Big\} dF(u|\boldsymbol{Z}) \bigg] - m(\boldsymbol{\beta}_0, \Lambda_0, F; X)
\end{aligned}
$$

and

$$
\begin{aligned}
m_{i,s}^U =& \Delta \sum_{j=1}^{K} \bigg[\Big\{ \big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) + c\varepsilon\big)^2 \big(\triangle\Lambda_{i,j}^U(Y)\big)^2 - 2\big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) - c\varepsilon\big) \triangle\Lambda_{i,j}^L(Y) \triangle N_j \\
& + \triangle N_j^2 \Big\} + \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})} \int_Y^\infty \Big\{ \big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) + c\varepsilon\big)^2 \big(\triangle\Lambda_{i,j}^U(u)\big)^2 \\
& - 2\big(\exp(\boldsymbol{\beta}_s^T\boldsymbol{Z}) - c\varepsilon\big) \triangle\Lambda_{i,j}^L(u) \triangle N_j + \triangle N_j^2 \Big\} dF(u|\boldsymbol{Z}) \bigg] - m(\boldsymbol{\beta}_0, \Lambda_0, F; X).
\end{aligned}
$$

Then for all $\boldsymbol{\beta} \in \mathcal{R}$, $\Lambda \in \Phi_n$ satisfying $d_3^2((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) \le \eta^2$, we have

$$
m(\boldsymbol{\beta}, \Lambda, F; X) - m(\boldsymbol{\beta}_0, \Lambda_0, F; X) \in [m_{i,s}^L, m_{i,s}^U]
$$

for $i = 1, \cdots, (\eta/\varepsilon)^{c_1 q_n}$ and $m = 1, \cdots, \lceil (c_2\eta/\varepsilon)^d \rceil$.

We also need to prove that $||m_{i,s}^L - m_{i,s}^U||_{P,B}^2 \lesssim \varepsilon^2$. Noting that $||f||_{P,B}^2$ is bounded by $2\mathcal{P}(|f|^2 \exp(|f|))$, we need to consider $\mathcal{P}(|m_{i,s}^L - m_{i,s}^U|^2 \exp(|m_{i,s}^L - m_{i,s}^U|))$. After some algebraic calculations,

$$
\begin{aligned}
& m_{i,s}^U - m_{i,s}^L \\
=& \sum_{j=1}^K \Delta \Big[\big\{ \exp(\boldsymbol{\beta}_s^T \boldsymbol{Z})^2 + c^2\varepsilon^2 \big\} \big\{ \big(\triangle\Lambda_{i,j}^U(Y)\big)^2 - \big(\triangle\Lambda_{i,j}^L(Y)\big)^2 \big\} \\
&+2c\varepsilon \exp(\boldsymbol{\beta}_s^T \boldsymbol{Z}) \big\{ \big(\triangle\Lambda_{i,j}^U(Y)\big)^2 + \big(\triangle\Lambda_{i,j}^L(Y)\big)^2 \big\} + 2\exp(\boldsymbol{\beta}_s^T \boldsymbol{Z}) \big(\triangle\Lambda_{i,j}^U(Y) - \triangle\Lambda_{i,j}^L(Y)\big) \triangle N_j \\
&+2c\varepsilon \big(\triangle\Lambda_{i,j}^U(Y) + \triangle\Lambda_{i,j}^L(Y)\big) \triangle N_j \Big] \\
&+ \sum_{j=1}^K \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})} \int_Y^\infty \Big[\big\{ \exp(\boldsymbol{\beta}_s^T \boldsymbol{Z})^2 + c^2\varepsilon^2 \big\} \big\{ \big(\triangle\Lambda_{i,j}^U(u)\big)^2 - \big(\triangle\Lambda_{i,j}^L(u)\big)^2 \big\} \\
&+2c\varepsilon \exp(\boldsymbol{\beta}_s^T \boldsymbol{Z}) \big\{ \big(\triangle\Lambda_{i,j}^U(u)\big)^2 + \big(\triangle\Lambda_{i,j}^L(u)\big)^2 \big\} + 2\exp(\boldsymbol{\beta}_s^T \boldsymbol{Z}) \big(\triangle\Lambda_{i,j}^U(u) - \triangle\Lambda_{i,j}^L(u)\big) \triangle N_j \\
&+2c\varepsilon \big(\triangle\Lambda_{i,j}^U(u) + \triangle\Lambda_{i,j}^L(u)\big) \triangle N_j \Big] dF(u|\boldsymbol{Z}).
\end{aligned}
$$

Since $\triangle\Lambda_i^L(s)$, $\triangle\Lambda_i^U(s)$ and $1-F(Y)$ are uniformly bounded functions, and $\exp(\boldsymbol{\beta}_s^T \boldsymbol{Z})$ is also bounded, we have $\exp(|m_{i,s}^L - m_{i,s}^U|) \lesssim \exp(cN(T_K))$. Note that $d_2(F, F_0) \le \delta$ and δ is sufficiently small. Then Conditions (C7) and (C13) imply that

$$
\begin{aligned}
& \mathcal{P}(|m_{i,s}^L - m_{i,s}^U|^2 \exp(|m_{i,s}^L - m_{i,s}^U|)) \lesssim \mathcal{P}\left[|m_{i,s}^L - m_{i,s}^U|^2 \exp\left(cN(T_K)\right) \right] \\
\lesssim& \mathcal{P}\Bigg[\Bigg| \sum_{j=1}^K \Delta \Big[\big\{ \big(\triangle\Lambda_{i,j}^U(Y)\big)^2 - \big(\triangle\Lambda_{i,j}^L(Y)\big)^2 \big\} + \varepsilon + \big(\triangle\Lambda_{i,j}^U(Y) - \triangle\Lambda_{i,j}^L(Y)\big) \triangle N_j \\
&+ \varepsilon \triangle N_j \Big] + \sum_{j=1}^K \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})} \int_Y^\infty \big\{ \big(\triangle\Lambda_{i,j}^U(u)\big)^2 - \big(\triangle\Lambda_{i,j}^L(u)\big)^2 + \varepsilon \\
&+ \big(\triangle\Lambda_i^U(u) - \triangle\Lambda_i^L(u)\big) \triangle N_j + \varepsilon \triangle N_j \big\} dF(u|\boldsymbol{Z}) \Bigg| \Bigg]
\end{aligned}
$$

$$\lesssim \mathcal{P}\left[\left|\sum_{j=1}^{K}\Delta\Big\{\big(\triangle\Lambda_{i,j}^{U}(Y)-\triangle\Lambda_{i,j}^{L}(Y)+\varepsilon\big)(\triangle N_j+1)\Big\}\right.\right.$$

$$\left.\left.+\sum_{j=1}^{K}\frac{1-\Delta}{1-F(Y|\boldsymbol{Z})}\int_{Y}^{\infty}\Big\{\big(\triangle\Lambda_{i,j}^{U}(u)-\triangle\Lambda_{i,j}^{L}(u)+\varepsilon\big)(\triangle N_j+1)\Big\}dF(u|\boldsymbol{Z})\right|^{2}\right]$$

$$\lesssim \mathcal{P}\left[\Delta\sum_{j=1}^{K}\big(\triangle\Lambda_{i,j}^{U}(Y)-\triangle\Lambda_{i,j}^{L}(Y)\big)^{2}\right.$$

$$\left.+\frac{1-\Delta}{1-F(Y|\boldsymbol{Z})}\sum_{j=1}^{K}\int_{Y}^{\infty}\big(\triangle\Lambda_{i,j}^{U}(u)-\triangle\Lambda_{i,j}^{L}(u)\big)^{2}dF(u|\boldsymbol{Z})\right]+\varepsilon^{2}$$

$$\lesssim ||\triangle\Lambda_{i}^{U}(s_1,s_2)-\triangle\Lambda_{i}^{L}(s_1,s_2)||_{L_2(\mu_2)}^{2}+d_2(F,F_0)+\varepsilon^{2}\lesssim\varepsilon^{2}.$$

Thus, $||m_{i,s}^{L}-m_{i,s}^{U}||_{P,B}^{2}\lesssim\varepsilon^{2}$ and $N_{[]}(\varepsilon,\mathcal{M}_{\eta}(F),||\cdot||_{P,B})\leq c_2^{d}(\eta/\varepsilon)^{c_1 q_n+d}$. Noting that q_n goes to infinity as n goes to infinity, we obtain $\log N_{[]}(\varepsilon,\mathcal{M}_{\eta}(F),||\cdot||_{P,B})\lesssim q_n\log(\eta/\varepsilon)$. □

Lemma 4.4. *Suppose that Condition* (C2) *holds. For sufficiently small* δ*, any* F *satisfying* $d_2(F,F_0)\leq\delta$*, and any differentiable function* g*, we have*

$$\mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\left|\frac{\int_{Y}^{\infty}g(u-T_j)dF(u|\boldsymbol{Z})}{1-F(Y|\boldsymbol{Z})}-\frac{\int_{Y}^{\infty}g(u-T_j)dF_0(u|\boldsymbol{Z})}{1-F_0(Y|\boldsymbol{Z})}\right|\right]$$

$$\lesssim\left(E\left[\sum_{j=1}^{K}|g'(U-T_j)|\right]+E\left[\sum_{j=1}^{K}|g(U-T_j)|\right]\right)d_2(F,F_0).$$

Lemma 4.5. *Suppose that Conditions* (C2) *and* (C4) *hold. Then for sufficiently small* δ*,* $\{m(\boldsymbol{\beta},\Lambda,F;X):\boldsymbol{\beta}\in\mathcal{R},\ \Lambda\in\Phi,\ \Lambda$ *is uniformly bounded,* $d_2(F,F_0)\leq\delta\}$ *is Donsker.*

Lemma 4.6. *Suppose that for every* $\boldsymbol{\beta}\in\mathcal{R}$*,* $\Lambda\in\Phi_n$*, sufficiently large* n *and suffi-*

ciently small η,

$$\begin{aligned}&\mathcal{P}(m(\boldsymbol{\beta}_0,\Lambda_0,\hat{F}_n;X)-m(\boldsymbol{\beta},\Lambda,\hat{F}_n;X))\\&\lesssim -d_3^2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))+d_3((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))d_2(\hat{F}_n,F_0)+d_2^2(\hat{F}_n,F_0)\end{aligned}$$

and

$$E\sup_{\{\Lambda\in\Phi_n:d_3((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))<\eta\}}|(\mathbb{P}_n-\mathcal{P})(m(\boldsymbol{\beta}_0,\Lambda_0,\hat{F}_n;X)-m(\boldsymbol{\beta},\Lambda,\hat{F}_n;X))|\lesssim\frac{\phi_n(\eta)}{\sqrt{n}}$$

hold, where $\phi_n(\eta)$ satisfies that $\eta\mapsto\phi_n(\eta)/\eta^\alpha$ is decreasing for some $\alpha<2$. Let $r_n>0$ satisfy $\phi_n(r_n)\lesssim\sqrt{n}r_n^2$. If the sequence $\hat{\Lambda}_n$ satisfies

$$\mathbb{P}_n m(\boldsymbol{\beta}_0,\Lambda_0,\hat{F}_n;X)\geq\mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n,\hat{F}_n;X)-O_p(r_n^2)$$

and converges in outer probability to Λ_0, then $d_1(\hat{\Lambda}_n,\Lambda_0)=O_p(r_n+d_2(\hat{F}_n,F_0))$.

4.6.2 A General Theorem for the Asymptotic Normality of Semiparametric M-estimation with Nuisance Parameter

In this section, we establish a general theorem for the asymptotic normality of semiparametric M-estimator with nuisance parameter overcoming the difficulty that the overall convergence rate is slower than $n^{1/2}$. Consider the two-stage semiparametric M-estimator $(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n)$, which is obtained by minimizing the objective function $\mathbb{P}_n m(\boldsymbol{\beta},\Lambda,\hat{F}_n;X)$ with respect to $(\boldsymbol{\beta},\Lambda)\in\mathcal{R}\times\Phi_n$. For a parameter path $\Lambda_\eta\in\Phi$ satisfying $\Lambda_\eta|_{\eta=0}=\Lambda$, define $\mathcal{H}=\{h:h=\frac{\partial\Lambda_\eta}{\partial\eta}|_{\eta=0}\}$, $\tilde{\mathcal{H}}=\{(\boldsymbol{h}_1,h_2):\boldsymbol{h}_1\in\mathcal{R},h_2\in\mathcal{H}_r\}$ and

$$\psi(\boldsymbol{\beta},\Lambda,F;X)[\boldsymbol{h}_1,h_2]=\frac{\partial}{\partial\eta}m(\boldsymbol{\beta}+\eta\boldsymbol{h}_1,\Lambda+\eta h_2,F;X)|_{\eta=0}$$

for $(\boldsymbol{h}_1, h_2) \in \tilde{\mathcal{H}}$. In particular, in the case of semiparametric estimation for panel count data with an informative terminal event, ignoring constant factor, we have

$$\psi(\boldsymbol{\beta}, \Lambda, F; X)[\boldsymbol{h}_1, h_2] = \sum_{j=1}^{K} \Big[\Delta\big\{\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(Y)\big\}$$

$$\times \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\big\{\triangle\Lambda_j(Y)\boldsymbol{h}_1^T\boldsymbol{Z} + \triangle h_{2,j}(Y)\big\} + \frac{1-\Delta}{1-F(Y|Z)}$$

$$\times \int_Y^\infty \big\{\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u)\big\} \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\big\{\triangle\Lambda_j(u)\boldsymbol{h}_1^T\boldsymbol{Z} + \triangle h_{2,j}(u)\big\} dF(u|Z)\Big].$$

Set

$$Q_n(\boldsymbol{\beta}, \Lambda, F)[\boldsymbol{h}_1, h_2] = \mathbb{P}_n\psi(\boldsymbol{\beta}, \Lambda, F; X)[\boldsymbol{h}_1, h_2]$$

and

$$Q(\boldsymbol{\beta}, \Lambda, F)[\boldsymbol{h}_1, h_2] = \mathcal{P}\psi(\boldsymbol{\beta}, \Lambda, F; X)[\boldsymbol{h}_1, h_2].$$

According to Theorem 1 of Zhao and Zhang (2017), we need the following conditions to establish the asymptotic normality.

(B1) $Q(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] = 0$ and $Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] = o_p(n^{-1/2})$.

(B2) $\sqrt{n}(Q_n - Q)(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - \sqrt{n}(Q_n - Q)(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] = o_p(1)$.

(B3) $Q(\boldsymbol{\beta}, \Lambda, F)[\boldsymbol{h}_1, h_2]$ is Fréchet-differentiable with respect to $(\boldsymbol{\beta}, \Lambda)$ at $(\boldsymbol{\beta}_0, \Lambda_0, F_0)$ with a continuous derivative $\dot{Q}_{1,\boldsymbol{\beta}_0,\Lambda_0,F_0}[\boldsymbol{h}_1, h_2]$; $Q(\boldsymbol{\beta}, \Lambda, F)[\boldsymbol{h}_1, h_2]$ is Fréchet-differentiable with respect to F at $(\boldsymbol{\beta}_0, \Lambda_0, F_0)$ with a continuous derivative $\dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}[\boldsymbol{h}_1, h_2]$.

(B4) $Q(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - Q(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] - \dot{Q}_{1,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\boldsymbol{h}_1, h_2] - \dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2] = o_p(n^{-1/2})$.

(B5) $\sqrt{n}Q_n(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] + \sqrt{n}\dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2]$ converges in distribution to a tight Gaussian progress.

Lemma 4.7 (General Theorem for the Asymptotic Normality)**.** *Suppose that Con-*

ditions (B1)–(B5) *hold. Then we have*

$$
\begin{aligned}
&-\sqrt{n}\dot{Q}_{1,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0,\hat{\Lambda}_n-\Lambda_0)[\boldsymbol{h}_1,h_2]\\
=&\sqrt{n}Q_n(\boldsymbol{\beta}_0,\Lambda_0,F_0)[\boldsymbol{h}_1,h_2]+\sqrt{n}\dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n-F_0)[\boldsymbol{h}_1,h_2]+o_p(1)
\end{aligned}
$$

converges in distribution to a tight Gaussian process.

Proof. By the relations in (B1), (B3) and (B4),

$$
\begin{aligned}
&Q(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n,\hat{F}_n)[\boldsymbol{h}_1,h_2]-\dot{Q}_{1,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0,\hat{\Lambda}_n-\Lambda_0)[\boldsymbol{h}_1,h_2]\\
&-\dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n-F_0)[\boldsymbol{h}_1,h_2]=o_p(n^{-1/2}).
\end{aligned}
$$

According to (B1) and (B2), we have $-Q(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n,\hat{F}_n)[\boldsymbol{h}_1,h_2]=Q_n(\boldsymbol{\beta}_0,\Lambda_0,F_0)[\boldsymbol{h}_1,h_2]+o_p(n^{-1/2})$. Combining the above two equations, it follows that

$$
\begin{aligned}
&-\dot{Q}_{1,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0,\hat{\Lambda}_n-\Lambda_0)[\boldsymbol{h}_1,h_2]\\
=&Q_n(\boldsymbol{\beta}_0,\Lambda_0,F_0)[\boldsymbol{h}_1,h_2]+\dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n-F_0)[\boldsymbol{h}_1,h_2]+o_p(n^{-1/2}).
\end{aligned}
$$

□

4.6.3 Proof of Theorem 4.1

Proof. We first prove that $\hat{\Lambda}_n$ is uniformly bounded. Since $(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n)$ minimizes $\mathbb{P}_n m(\boldsymbol{\beta},\Lambda,\hat{F}_n;X)$ with respect to $(\boldsymbol{\beta},\Lambda)\in\mathcal{R}\times\Phi_n$, for any direction function $h\in\Phi_n$, we have

$$
\begin{aligned}
0=&\lim_{\varepsilon\to 0}\frac{\mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n+\varepsilon h,\hat{F}_n;X)-\mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n,\hat{F}_n;X)}{\varepsilon}\\
=&-2\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\Delta\big((\triangle N_j-\exp(\hat{\boldsymbol{\beta}}_n^T\boldsymbol{Z})\triangle\hat{\Lambda}_{n,j}(Y))\exp(\hat{\boldsymbol{\beta}}_n^T\boldsymbol{Z})\triangle h_j(Y)\big)+\frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})}\right.\right.\\
&\left.\left.\times\int_Y^\infty\big((\triangle N_j-\exp(\hat{\boldsymbol{\beta}}_n^T\boldsymbol{Z})\triangle\hat{\Lambda}_{n,j}(u))\exp(\hat{\boldsymbol{\beta}}_n^T\boldsymbol{Z})\triangle h_j(u)\big)d\hat{F}_n(u|\boldsymbol{Z})\right\}\right].
\end{aligned}
\tag{4.5}
$$

Taking $h(s) = s$, by Condition (C8), it follows that

$$\mathbb{P}_n \left[\exp(2\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}) \sum_{j=1}^{K} \Big\{ (T_j - T_{j-1}) \Big(\Delta \triangle \hat{\Lambda}_{n,j}(Y) \right.$$

$$\left. + \ \frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})} \int_Y^\infty \triangle \hat{\Lambda}_{n,j}(u) d\hat{F}_n(u|\boldsymbol{Z}) \Big) \Big\} \right]$$

$$= \mathbb{P}_n \left[\exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}) \sum_{j=1}^{K} \{ \triangle N_j (T_j - T_{j-1}) \} \right] \leq c_1 \mathbb{P}_n \left[\sum_{j=1}^{K} \{ \triangle N_j (T_j - T_{j-1}) \} \right]$$

$$\xrightarrow{a.s.} c_1 E \left[\exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \sum_{j=1}^{K} \triangle \Lambda_{0,j}(U)(T_j - T_{j-1}) \right] \leq c_2,$$

for some constant c_1 and c_2. Furthermore, by Condition (C9),

$$\limsup_{n \to \infty} \mathbb{P}_n \left[\exp(2\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}) \sum_{j=1}^{K} \Big\{ (T_j - T_{j-1}) \Big(\Delta \triangle \hat{\Lambda}_{n,j}(Y) \right.$$

$$\left. + \ \frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})} \int_Y^\infty \triangle \hat{\Lambda}_{n,j}(u) d\hat{F}_n(u|\boldsymbol{Z}) \Big) \Big\} \right]$$

$$\gtrsim \limsup_{n \to \infty} \mathbb{P}_n \left[\sum_{j=1}^{K} \Big\{ (T_j - T_{j-1}) \Big(\Delta \triangle \hat{\Lambda}_{n,j}(Y) \right.$$

$$\left. + \ \frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})} \int_Y^\infty \triangle \hat{\Lambda}_{n,j}(u) d\hat{F}_n(u|\boldsymbol{Z}) \Big) \Big\} \right]$$

$$\gtrsim \limsup_{n \to \infty} \mathbb{P}_n \left[\Delta \big\{ \hat{\Lambda}_n(Y) - \hat{\Lambda}_n(Y - T_K) \big\} \right.$$

$$\left. + \frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})} \int_Y^\infty \big\{ \hat{\Lambda}_n(u) - \hat{\Lambda}_n(u - T_K) \big\} d\hat{F}_n(u|\boldsymbol{Z}) \right]$$

$$\geq \limsup_{n \to \infty} \triangle \hat{\Lambda}_n(b_1, b_2) \mathbb{P}_n \Big[\Delta 1_{\{Y - T_K \in [0, b_1], Y \in [b_2, \tau]\}}$$

$$+ \frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})} \int_Y^\infty 1_{\{u - T_K \in [0, b_1], u \in [b_2, \tau]\}} d\hat{F}_n(u|\boldsymbol{Z}) \Big]$$

$$= \limsup_{n\to\infty} \triangle\hat{\Lambda}_n(b_1, b_2) E\left[1_{\{U-T_K\in[0,b_1], U\in[b_2,\tau]\}}\right]$$

$$= \limsup_{n\to\infty} \triangle\hat{\Lambda}_n(b_1, b_2)\mu_3([0, b_1] \times [b_2, \tau]).$$

According to the above two inequalities, we have $\limsup_{n\to\infty} \triangle\hat{\Lambda}_n(b_1, b_2)\mu_3([0, b_1] \times [b_2, \tau]) \le c_2$. Then for all $0 \le b_1 \le s_1 \le s_2 \le b_2 \le \tau$ with $\mu_3([0, b_1]\times[b_2, \tau]) > 0$, since $\triangle\hat{\Lambda}_n(s_1, s_2) = \hat{\Lambda}_n(s_2)-\hat{\Lambda}_n(s_1) \le \hat{\Lambda}_n(b_2)-\hat{\Lambda}_n(b_1) = \triangle\hat{\Lambda}_n(b_1, b_2)$, we have $\triangle\hat{\Lambda}_n(s_1, s_2)$ is uniformly bounded. In particular, when $\mu_3(\{0\} \times \{\tau\}) > 0$, $\hat{\Lambda}_n(s)$ is uniformly bounded on $[0, \tau]$.

We next consider the minimal point of $\mathcal{P}m(\boldsymbol{\beta}, \Lambda, F_0; X)$. After some algebraic calculations, we have

$$\begin{aligned}
&\mathcal{P}m(\boldsymbol{\beta}, \Lambda, F_0; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)\\
=&\mathcal{P}\left[\Delta\sum_{j=1}^{K}\left\{\left(\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(Y)\right)^2 - \left(\triangle N_j - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(Y)\right)^2\right\}\right.\\
&+\frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})}\sum_{j=1}^{K}\int_Y^\infty\left\{\left(\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u)\right)^2\right.\\
&-\left.\left.\left(\triangle N_j - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\right)^2\right\}dF_0(u|\boldsymbol{Z})\right]\\
=&\mathcal{P}\left[\sum_{j=1}^{K}\Delta\left\{2\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(Y) - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(Y)\right\}\right.\\
&\times\left\{\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(Y) - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(Y)\right\}\\
&+\sum_{j=1}^{K}\frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})}\int_Y^\infty\left\{2\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u) - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\right\}\\
&\left.\times\left\{\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u) - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u)\right\}dF_0(u|\boldsymbol{Z})\right]\\
=&\mathcal{P}\left[\sum_{j=1}^{K}\left(\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U) - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)\right)^2\right] \ge 0.
\end{aligned}$$

It follows that

$$\mathcal{P}m(\boldsymbol{\beta}, \Lambda, F_0; X) \geq \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X),$$

and $\mathcal{P}m(\boldsymbol{\beta}, \Lambda, F_0; X) = \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)$ if and only if $\triangle\Lambda(s_1, s_2)\exp(\boldsymbol{\beta}^T\boldsymbol{z}) = \triangle\Lambda_0(s_1, s_2)\exp(\boldsymbol{\beta}_0^T\boldsymbol{z})$ a.e. with respect to μ_1. By conditions (C5) and (C6), similar to the proof in Theorem 3.1 of Wellner and Zhang (2007), we have $\mathcal{P}m(\boldsymbol{\beta}, \Lambda, F_0; X) = \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)$ if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\triangle\Lambda(s_1, s_2) = \triangle\Lambda_0(s_1, s_2)$ a.e. with respect to μ_2. Hence, for every $\delta > 0$, we have

$$\sup_{d_3((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))>\delta} \mathcal{P}m(\boldsymbol{\beta}, \Lambda, F_0; X) > \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X),$$

which yields that

$$\{d_3((\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n), (\boldsymbol{\beta}_0, \Lambda_0)) > \delta\} \subset \{\mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0; X) > \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)\}.$$

By Lemma A1 of Lu, Zhang, and Huang (2007), under Condition (C8), there is a $\Lambda_n^* \in \Phi_n$ such that $||\Lambda_n^* - \Lambda_0||_\infty = O(n^{-\nu r})$. Note that

$$\begin{aligned}
0 \leq & \mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X) \\
= & \mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) + \mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) \\
- & \mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) + \mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) \\
+ & \mathbb{P}_n m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) + \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) \\
- & \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, F_0; X) + \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, F_0; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X).
\end{aligned} \tag{4.6}$$

According to Conditions (C2) and (C7),

$$0 \leq \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, F_0; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X) \lesssim ||\Lambda_n^* - \Lambda_0||_\infty^2 = o(1).$$

The definition of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$ yields that $\mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) \leq \mathbb{P}_n m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X)$. By Lemma 4.4, we have

$$\mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) - \mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0; X) = o_p(1)$$

and

$$\mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, F_0; X) = o_p(1).$$

By Lemma 4.5, $\{m(\boldsymbol{\beta}, \Lambda, F; X) : \boldsymbol{\beta} \in \mathcal{R},\ \Lambda \in \Phi,\ F \in \mathcal{F}\}$ is Donsker, meaning that it is Glivenko-Cantelli, and we have $(\mathbb{P}_n - \mathcal{P})m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) = o_p(1)$ and $(\mathbb{P}_n - \mathcal{P})m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) = o_p(1)$. Combining them with (4.6), we have

$$0 \leq \mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X) \leq o_p(1). \tag{4.7}$$

According to (4.7), $\{\mathcal{P}m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0; X) > \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)\}$ goes into a null set as $n \to \infty$. According to the argument in Theorem A.2.3. of Kong el al. (2018) and Theorem 5.8 in van der Vaart (2002), it follows that $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n) \to (\boldsymbol{\beta}_0, \Lambda_0)$ almost surely. Thus, $d_3((\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n), (\boldsymbol{\beta}_0, \Lambda_0)) = o_p(1)$. □

4.6.4 Proof of Theorem 4.2

Proof. We use Lemma 4.6 to prove the rate of convergence.

According to the second part of Lemma 4.1, we have

$$\begin{aligned}
&\mathcal{P}\big(m(\boldsymbol{\beta}_0, \Lambda_0, F; X) - m(\boldsymbol{\beta}, \Lambda, F; X)\big) \\
&= - \mathcal{P}\left[\sum_{j=1}^{K} \Delta\big(\exp(\boldsymbol{\beta}^T \boldsymbol{Z})\triangle\Lambda_j(Y) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\triangle\Lambda_{0,j}(Y)\big)^2\right. \\
&\left. + \sum_{j=1}^{K} \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})} \int_Y^\infty \big(\exp(\boldsymbol{\beta}^T \boldsymbol{Z})\triangle\Lambda_j(u) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\triangle\Lambda_{0,j}(u)\big)^2 dF(u|\boldsymbol{Z})\right] \\
&\lesssim - d_3^2((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) \\
&+ \mathcal{P}\left[\sum_{j=1}^{K} \frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})} \int_Y^\infty \big(\exp(\boldsymbol{\beta}^T \boldsymbol{Z})\triangle\Lambda_j(u) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\triangle\Lambda_{0,j}(u)\big)^2 dF_0(u|\boldsymbol{Z})\right. \\
&\left. - \sum_{j=1}^{K} \frac{1-\Delta}{1-F(Y|\boldsymbol{Z})} \int_Y^\infty \big(\exp(\boldsymbol{\beta}^T \boldsymbol{Z})\triangle\Lambda_j(u) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\triangle\Lambda_{0,j}(u)\big)^2 dF(u|\boldsymbol{Z})\right].
\end{aligned}$$

By Conditions (C1), (C2) and (C7), Cauchy–Schwarz inequality and Lemma 4.4,

$$\Bigg|\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left\{\frac{\int_Y^\infty\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\big)^2dF_0(u|\boldsymbol{Z})}{1-F_0(Y|\boldsymbol{Z})}\right.\right.$$

$$\left.\left.-\ \frac{\int_Y^\infty\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\big)^2dF(u|\boldsymbol{Z})}{1-F(Y|\boldsymbol{Z})}\right\}\right]\Bigg|$$

$$\lesssim\mathcal{P}\left[\sum_{j=1}^{K}\big(\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big)^2\right]d_2(F,F_0)$$

$$+\mathcal{P}\left[\sum_{j=1}^{K}2\big|\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)\big|\right.$$

$$\left.\times\big|\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j'(U)-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}'(U)\big|\right]d_2(F,F_0)$$

$$\lesssim d_3((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))d_2(F,F_0)+d_3^2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))d_2(F,F_0).$$

This yields that

$$\mathcal{P}\big(m(\boldsymbol{\beta}_0,\Lambda_0,\hat{F}_n;X)-m(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n,\hat{F}_n;X)\big)\lesssim-d_3^2((\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n),(\boldsymbol{\beta}_0,\Lambda_0))$$

$$+d_3((\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n),(\boldsymbol{\beta}_0,\Lambda_0))d_2(\hat{F}_n,F_0)+d_3^2((\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n),(\boldsymbol{\beta}_0,\Lambda_0))d_2(\hat{F}_n,F_0).$$

Second, we need to find a $\phi_n(\eta)$ such that

$$E\sup_{\{(\boldsymbol{\beta},\Lambda)\in\mathcal{R}\times\Phi_n:d_3((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))<\eta\}}|(\mathbb{P}_n-\mathcal{P})(m(\boldsymbol{\beta},\Lambda,\hat{F}_n;X)-m(\boldsymbol{\beta}_0,\Lambda_0,\hat{F}_n;X))|\lesssim\frac{\phi_n(\eta)}{\sqrt{n}}.$$

By Lemma 4.3, for sufficiently large n, we have $\log N_{[]}(\varepsilon,\mathcal{L}_\eta(\hat{F}_n),||\cdot||_{P,B})\lesssim q_n\log(\eta/\varepsilon)$, where

$$\mathcal{M}_\eta(\hat{F}_n)=\{m(\boldsymbol{\beta},\Lambda,\hat{F}_n;X)-m(\boldsymbol{\beta}_0,\Lambda_0,\hat{F}_n;X):$$

$$\boldsymbol{\beta}\in\mathcal{R},\ \Lambda\in\Phi_n,\ d_3^2((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))\le\eta^2\}.$$

For $(\boldsymbol{\beta},\Lambda)\in\mathcal{R}\times\Phi_n$ satisfying $d_3((\boldsymbol{\beta},\Lambda),(\boldsymbol{\beta}_0,\Lambda_0))<\eta$, similar to the proof of Lemma

4.3, we have

$$
\begin{aligned}
&|m(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)| \\
\lesssim & \sum_{j=1}^{K} \Big[(\triangle N_j + 1) \Big\{ \Delta \big| \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_j(Y) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(Y) \big| \\
+ & \frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})} \sum_{j=1}^{K} \int_Y^\infty \big| \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_j(u) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(u) \big| d\hat{F}_n(u|\boldsymbol{Z}) \Big\} \Big].
\end{aligned}
$$

Furthermore, since $\exp(\boldsymbol{\beta}^T \boldsymbol{Z})$, $\triangle \Lambda_j$, $\exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})$ and $\triangle \Lambda_{0,j}$ are bounded and $d_2(\hat{F}_n, F_0) = o_p(1)$, we have $e^{|m(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)|} \lesssim e^{CN(T_K)}$. The above two inequalities yield that

$$
\begin{aligned}
&\mathcal{P}\left[e^{|m(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)|} |m(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)|^2 \right] \\
\lesssim & \mathcal{P}\Bigg[\sum_{j=1}^{K} \Delta \big(\exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_j(Y) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(Y) \big)^2 \\
+ & \sum_{j=1}^{K} \frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})} \int_Y^\infty \big(\exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_j(u) - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(u) \big)^2 d\hat{F}_n(u|\boldsymbol{Z}) \Bigg] \\
\lesssim & d_3^2((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) + d_3((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) d_2(\hat{F}_n, F_0).
\end{aligned}
$$

That means for sufficiently large n, $||m(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)||_{P,B}^2 \lesssim \eta^2$. By Lemma 3.4.3 of van der Vaart and Wellner (1996),

$$
E||n^{1/2}(\mathbb{P}_n - \mathcal{P})||_{\mathcal{M}_\eta(\hat{F}_n)} \lesssim J_{[]}(\eta, \mathcal{M}_\eta(\hat{F}_n), ||\cdot||_{P,B}) \left\{ 1 + \frac{J_{[]}(\eta, \mathcal{M}_\eta(\hat{F}_n), ||\cdot||_{P,B})}{\eta^2 n^{1/2}} \right\},
$$

where $J_{[]}(\eta, \mathcal{M}_\eta(\hat{F}_n), ||\cdot||_{P,B}) := \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{M}_\eta(\hat{F}_n), ||\cdot||_{P,B})\}^{1/2} d\varepsilon \lesssim q_n^{1/2}\eta$. It follows that

$$
\begin{aligned}
&E \sup_{\{(\boldsymbol{\beta}, \Lambda) \in \mathcal{R} \times \Phi_n : d_3((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) < \eta\}} \sqrt{n} |(\mathbb{P}_n - \mathcal{P})(m(\boldsymbol{\beta}, \Lambda, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X))| \\
\lesssim & q_n^{1/2}\eta + q_n n^{-1/2}.
\end{aligned}
$$

Setting $\phi_n(\eta) = q_n^{1/2}\eta + q_n n^{-1/2}$ such that $\phi_n(\eta)/\eta$ decreases about η, for a sequence $r_n = O(n^a)$, we have $r_n^2\phi(\frac{1}{r_n}) = q_n^{\frac{1}{2}} r_n + n^{-\frac{1}{2}} q_n r_n^2$. Note that $q_n = O(n^\nu), 0 < \nu < 1/2$. This yields that

$$r_n^2\phi(\frac{1}{r_n}) = O(n^{a+\frac{\nu}{2}} + n^{2a+\nu-\frac{1}{2}}).$$

Since $a \leq (1-\nu)/2$ ensures $r_n^2\phi(1/r_n) \lesssim n^{1/2}$, we choose $r_n = O(n^{(1-\nu)/2})$.

Finally, we calculate ν satisfying $\mathbb{P}_n\big(m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)\big) \leq O_p(r_n^{-2})$. Note that for $\Lambda_0 \in \mathcal{H}_r$, there is a $\Lambda_n^* \in \Phi_n$ such that $||\Lambda_n^* - \Lambda_0||_\infty = O(n^{-\nu r})$. By the definition of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$ and $0 \leq \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X) \lesssim ||\Lambda_n^* - \Lambda_0||_\infty^2$, we have

$$\begin{aligned}
&\mathbb{P}_n\big(m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)\big) \\
=&\mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) - \mathbb{P}_n m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) + \mathbb{P}_n m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) \\
+&\mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X) + \mathcal{P}m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X) - \mathbb{P}_n m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X) \\
\leq& n^{-\nu r+\varepsilon}(\mathbb{P}_n - \mathcal{P})\left(\frac{m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)}{n^{-\nu r+\varepsilon}}\right) + O_p(n^{-2\nu r}).
\end{aligned}$$

According to Lemma 4.3, $\mathcal{M}_\eta(F)$ is Donsker. After some algebraic calculations,

$$\mathcal{P}(m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X))^2 \lesssim ||\Lambda_n^* - \Lambda_0||_\infty^2.$$

Hence for any $f \in \mathcal{M}_\eta(F)$, we have $P(f/n^{-\nu r+\varepsilon})^2 \to 0$ as $n \to 0$ for any $\varepsilon > 0$. Using Corollary 2.3.12 of van der Vaart and Wellner (1996), we have

$$(\mathbb{P}_n - \mathcal{P})(m(\boldsymbol{\beta}_0, \Lambda_n^*, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)) = o_p(n^{-\nu r+\varepsilon-1/2}).$$

Taking $0 < \varepsilon \leq 1/2 - r\nu$, we have $\mathbb{P}_n\big(m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)\big) \leq O_p(n^{-2\nu r})$, meaning that $\nu \geq 1/(1+2r)$ ensures $\mathbb{P}_n\big(m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X) - m(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n; X)\big) \leq O_p(r_n^{-2})$. Thus, taking $\nu = 1/(1+2r)$, we have $d_3((\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n), (\boldsymbol{\beta}_0, \Lambda_0)) = O_p(n^{-\frac{r}{1+2r}})$.

□

4.6.5 Proof of Theorem 4.3

Proof. (i) We verify Conditions (B1)–(B5) presented in Lemma 4.7 to prove this theorem.

For (B1), under model (4.1), we have $Q(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] = 0$. By the definition of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n)$, for all $(\boldsymbol{h}_1, h_2) \in \mathcal{R} \times \Phi_n$, we obtain

$$\lim_{\eta \to 0} \frac{\mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n + \eta \boldsymbol{h}_1, \hat{\Lambda}_n + \eta h_2, F; X) - \mathbb{P}_n m(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X)}{\eta} = 0.$$

This implies that $Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] = 0$ for all $(\boldsymbol{h}_1, h_2) \in \mathcal{R} \times \Phi_n$. By Lemma A1 of Lu, Zhang, and Huang (2007) and the properties of spline functions (Schumaker, 2007), for any $h_2 \in \mathcal{H}_r$, we can find an $h_{2,n} \in \Phi_n$ satisfying $||h_{2,n} - h_2||_\infty = O(n^{-r/(1+2r)})$ and $||h'_{2,n} - h'_2||_\infty = o(1)$, where h'_2 is the derivative of h_2. Thus, for each $h_2 \in \mathcal{H}_r$, we need to prove $Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\mathbf{0}, h_2 - h_{2,n}] = \mathbb{P}_n \psi(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\mathbf{0}, h_2 - h_{2,n}] = o_p(n^{-1/2})$ to verify $Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] = o_p(n^{-1/2})$. Note that

$$\begin{aligned}
&Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\mathbf{0}, h_2 - h_{2,n}] \\
=&\big\{Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\mathbf{0}, h_2 - h_{2,n}] - Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0)[\mathbf{0}, h_2 - h_{2,n}]\big\} \\
&+\big\{Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0)[\mathbf{0}, h_2 - h_{2,n}] - Q_n(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\mathbf{0}, h_2 - h_{2,n}]\big\} \\
&+Q_n(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\mathbf{0}, h_2 - h_{2,n}] =: I_{1n} + I_{2n} + I_{3n}.
\end{aligned}$$

For the first term I_{1n}, Lemma 4.4 yields that

$$\begin{aligned}
\mathcal{P}|I_{1n}| =& \mathcal{P}\left|Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\mathbf{0}, h_2 - h_{2,n}] - Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0)[\mathbf{0}, h_2 - h_{2,n}]\right| \\
\leq&\mathcal{P}\left[\sum_{j=1}^{K}(1-\Delta)\left|\frac{\int_Y^\infty \{\triangle N_j - \exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z})\triangle\hat{\Lambda}_{n,j}(u)\}(\triangle h_{2,j}(u) - \triangle h_{2,n,j}(u))d\hat{F}_n(u)}{1 - \hat{F}_n(Y)}\right.\right. \\
&- \left.\left.\frac{\int_Y^\infty \{\triangle N_j - \exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z})\triangle\hat{\Lambda}_{n,j}(u)\}(\triangle h_{2,j}(u) - \triangle h_{2,n,j}(u))dF_0(u)}{1 - F_0(Y)}\right| \exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z})\right] \\
\lesssim& d_2(\hat{F}_n, F_0)(||h_2 - h_{2,n}||_\infty + ||h'_2 - h'_{2,n}||_\infty) = o_p(n^{-1/2}).
\end{aligned}$$

For the second term I_{2n}, after some algebraic calculations, we have

$$\mathcal{P}|I_{2n}| = \mathcal{P}\left|Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0)[\mathbf{0}, h_2 - h_{2,n}] - Q_n(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\mathbf{0}, h_2 - h_{2,n}]\right|$$

$$\leq \mathcal{P}\left[\sum_{k=1}^{K} \Delta \left| \left\{ \triangle N_j - \exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}) \triangle \hat{\Lambda}_{n_j}(Y) \right\} \exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}) \right.\right.$$

$$\left. - \left\{ \triangle N_j - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(Y) \right\} \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \right|$$

$$+ \frac{1-\Delta}{1-F_0(Y)} \int_Y^\infty \left| \left\{ \triangle N_j - \exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}) \triangle \hat{\Lambda}_{n,j}(u) \right\} \exp(\hat{\boldsymbol{\beta}}_n^T \boldsymbol{Z}) \right.$$

$$\left.\left. - \left\{ \triangle N_j - \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(u) \right\} \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}) \right| dF_0(u) \right] ||h_2 - h_{2,n}||_\infty$$

$$\lesssim \left\{ ||\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0||_2 + d_3((2\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n), (2\boldsymbol{\beta}_0, \Lambda_0)) \right\} ||h - h_n||_\infty = o_p(n^{-1/2}).$$

For the third term I_{3n}, note that $Q(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\mathbf{0}, h_2 - h_{2,n}] = 0$. By the independence of X_i and X_j, it follows that

$$\mathcal{P} I_{3n}^2 = \mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n} \psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X_i)[\mathbf{0}, h_2 - h_{2,n}]\right)^2$$

$$= n^{-1}\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n} \psi^2(\boldsymbol{\beta}_0, \Lambda_0, F_0; X_i)[\mathbf{0}, h_2 - h_{2,n}]\right)$$

$$\lesssim n^{-1}\mathcal{P}\left[\sum_{j=1}^{K} \left\{ \Delta \left| \triangle N_j - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(Y) \right| \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \right.\right.$$

$$\left.\left. + \frac{1-\Delta}{1-F_0(Y)} \int_Y^\infty \left| \triangle N_j - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \triangle \Lambda_{0,j}(Y) \right| \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) dF_0(u) \right\}\right]^2 ||h_2 - h_{2,n}||_\infty^2$$

$$\lesssim n^{-1} ||h_2 - h_{2,n}||_\infty^2.$$

Then we have $Q_n(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\mathbf{0}, h_2 - h_{2,n}] = o_p(n^{-1/2})$, and (B1) holds.

For (B2), after some algebraic calculations, we have

$$\sqrt{n}(Q_n - Q)(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\boldsymbol{h}_1, h_2] - \sqrt{n}(Q_n - Q)(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2]$$
$$= \sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X)[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1, h_2]).$$

For each fixed bounded $(\boldsymbol{h}_1, h_2) \in \tilde{\mathcal{H}}$, set

$$\bar{\Psi}_\eta(\boldsymbol{h}_1, h_2) = \{\psi(\boldsymbol{\beta}, \Lambda, F; X)[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1, h_2] : \ \boldsymbol{\beta} \in \mathcal{R}, \ \Lambda \in \Phi_n, \ F \in \mathcal{F}$$
$$d_3((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0)) < \eta, \ d_2(F, F_0) < \eta, \ \Lambda \text{ is uniformly bounded}\}.$$

Similar to Lemma 4.5, it follows that $\bar{\Psi}_\eta(\boldsymbol{h}_1, h_2)$ is Donsker. By Condition (C6) and Lemma 4.4, after some algebraic calculations, we obtain

$$\mathcal{P}(\psi(\boldsymbol{\beta}, \Lambda, F; X)[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1, h_2])^2 \lesssim d_3((\boldsymbol{\beta}, \Lambda), (\boldsymbol{\beta}_0, \Lambda_0))^2 + d_2(F, F_0)^2.$$

Then Corollary 2.3.12 of van der Vaart and Wellner (1996) implies that

$$\sqrt{n}(\mathbb{P}_n - \mathcal{P})(\psi(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n; X)[\boldsymbol{h}_1, h_2] - \psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1, h_2]) = o_p(1),$$

and (B2) holds.

For (B3), since $Q(\boldsymbol{\beta}, \Lambda, F)[\boldsymbol{h}_1, h_2]$ is a smooth function with respect to $(\boldsymbol{\beta}, \Lambda, F)$, $Q(\boldsymbol{\beta}, \Lambda, F)[\boldsymbol{h}_1, h_2]$ is Fréchet-differentiable with respect to $(\boldsymbol{\beta}, \Lambda)$ at $(\boldsymbol{\beta}_0, \Lambda_0, F_0)$. Similarly, $Q(\boldsymbol{\beta}, \Lambda, F)[\boldsymbol{h}_1, h_2]$ is Fréchet-differentiable with respect to F at $(\boldsymbol{\beta}_0, \Lambda_0, F_0)$. Setting

$$R_1(\boldsymbol{h}_1, h_2)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = -\mathcal{P}\left[\sum_{j=1}^{K} \left\{\Delta\Big(\triangle N_j - 2\exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\triangle\Lambda_{0,j}(Y)\Big)\exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\right.\right.$$
$$\times\Big(\triangle h_{2,j}(Y) + \triangle\Lambda_{0,j}(Y)\boldsymbol{h}_1^T \boldsymbol{Z}\Big) + \frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})}\int_Y^\infty \Big(\triangle N_j - 2\exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\triangle\Lambda_{0,j}(u)\Big)$$
$$\left.\left.\times \ \exp(\boldsymbol{\beta}_0^T \boldsymbol{Z})\Big(\triangle h_{2,j}(u) + \triangle\Lambda_{0,j}(u)\boldsymbol{h}_1^T \boldsymbol{Z}\Big)dF_0(u|\boldsymbol{Z})\right\}\boldsymbol{Z}^T\right](\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \tag{4.8}$$

and

$$R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0) = -\mathcal{P}\left[\sum_{j=1}^{K}\Big\{\Delta \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})(\triangle\hat{\Lambda}_{n,j}(Y) - \triangle\Lambda_{0,j}(Y))\right.$$

$$\times\big(\triangle N_j \boldsymbol{h}_1^T\boldsymbol{Z} - 2\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(Y)\boldsymbol{h}_1^T\boldsymbol{Z} - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle h_{2,j}(Y)\big)$$

$$+\frac{1-\Delta}{1-\hat{F}_n(Y|\boldsymbol{Z})}\int_Y^{\infty}\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})(\triangle\hat{\Lambda}_{n,j}(u) - \Lambda_{0,j}(u))$$

$$\left.\times\ \big(\triangle N_j \boldsymbol{h}_1^T\boldsymbol{Z} - 2\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\boldsymbol{h}_1^T\boldsymbol{Z} - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle h_{2,j}(u)\big)d\hat{F}_n(u|\boldsymbol{Z})\Big\}\right], \tag{4.9}$$

we obtain

$$\dot{Q}_{1,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\boldsymbol{h}_1, h_2]$$

$$=\frac{d}{d\varepsilon}\left\{\mathcal{P}\left[\sum_{j=1}^{K}\Big\{\Delta\Big(\triangle h_{2,j}(Y) + \big(\triangle\Lambda_{0,j}(Y) + \varepsilon(\triangle\hat{\Lambda}_{n,j}(Y) - \triangle\Lambda_{0,j}(Y))\big)h_1^T\boldsymbol{Z}\Big)\right.\right.$$

$$\times\Big(\triangle N_j - \big(\triangle\Lambda_{0,j}(Y) + \varepsilon(\triangle\hat{\Lambda}_{n,j}(Y) - \triangle\Lambda_{0,j}(Y))\big)\exp\big((\boldsymbol{\beta}_0 + \varepsilon(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0))^T\boldsymbol{Z}\big)\Big)$$

$$\times\exp\big((\boldsymbol{\beta}_0 + \varepsilon(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0))^T\boldsymbol{Z}\big) + \frac{1-\Delta}{1-F_0(Y)}\int_Y^{\infty}\exp\big((\boldsymbol{\beta}_0 + \varepsilon(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0))^T\boldsymbol{Z}\big)$$

$$\times\Big(\triangle N_j - \big(\triangle\Lambda_{0,j}(u) + \varepsilon(\triangle\hat{\Lambda}_{n,j}(u) - \triangle\Lambda_{0,j}(u))\big)\exp\big((\boldsymbol{\beta}_0 + \varepsilon(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0))^T\boldsymbol{Z}\big)\Big)$$

$$\left.\left.\times\ \Big(\triangle h_{2,j}(u) + \big(\triangle\Lambda_{0,j}(u) + \varepsilon(\triangle\hat{\Lambda}_{n,j}(u) - \triangle\Lambda_{0,j}(u))\big)h_1^T\boldsymbol{Z}\Big)dF_0(u)\Big\}\right]\right\}\Big|_{\varepsilon=0}$$

$$= - R_1(\boldsymbol{h}_1, h_2)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) - R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0).$$

Since the equation

$$d\frac{\int_Y^{\infty} g(u-T_j)d(F_0 + \varepsilon(\hat{F}_n - F_0))(u|\boldsymbol{Z})}{1 - F_0(Y|\boldsymbol{Z}) - \varepsilon(\hat{F}_n - F_0)(Y|\boldsymbol{Z})}/d\varepsilon\Big|_{\varepsilon=0}$$

$$=\frac{\int_Y^{\infty} g(u-T_j)d(\hat{F}_n - F_0)(u|\boldsymbol{Z})}{1-F_0(Y|\boldsymbol{Z})} + \frac{(\hat{F}_n - F_0)(Y|\boldsymbol{Z})\int_Y^{\infty} g(u-T_j)dF_0(u|\boldsymbol{Z})}{(1-F_0(Y|\boldsymbol{Z}))^2}$$

$$=\frac{1}{1-F_0(Y|\boldsymbol{Z})}\int_Y^\infty \left\{g(u-T_j)-\int_Y^\infty \frac{g(s-T_j)}{1-F_0(Y|\boldsymbol{Z})}dF_0(s|\boldsymbol{Z})\right\}d(\hat{F}_n-F_0)(u|\boldsymbol{Z})$$

holds for any differentiable function g, we obtain

$$\dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n-F_0)[\boldsymbol{h}_1,h_2]=\frac{d}{d\varepsilon}\left\{Q(\boldsymbol{\beta}_0,\Lambda_0,F_0+\varepsilon(\hat{F}_n-F_0))[\boldsymbol{h}_1,h_2]\right\}\Big|_{\varepsilon=0}$$

$$=\frac{d}{d\varepsilon}\left\{\mathcal{P}\left[\sum_{j=1}^K \frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})-\varepsilon(\hat{F}_n-F_0)(Y|\boldsymbol{Z})}\int_Y^\infty \left\{\triangle N_j-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\right\}\right.\right.$$

$$\left.\left.\times\ \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\big\{\triangle h_{2,j}(u)+\triangle\Lambda_{0,j}(u)\boldsymbol{h}_1^T\boldsymbol{Z}\big\}d(F_0+\varepsilon(\hat{F}_n-F_0))(u|\boldsymbol{Z})\right]\right\}\Big|_{\varepsilon=0}$$

$$=\mathcal{P}\left[\int_Y^\infty \bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]d(\hat{F}_n-F_0)(u|\boldsymbol{Z})\right],$$

where

$$\bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]$$
$$=\frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})}\sum_{j=1}^K\left\{\tilde{\varphi}_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]-\int_Y^\infty \frac{\tilde{\varphi}_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(s;X)[\boldsymbol{h}_1,h_2]}{1-F_0(Y|\boldsymbol{Z})}dF_0(s|Z)\right\}$$

and

$$\tilde{\varphi}_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]$$
$$=\big\{\triangle N_j-\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\big\}\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\big\{\triangle h_{2,j}(u)+\triangle\Lambda_{0,j}(u)\boldsymbol{h}_1^T\boldsymbol{Z}\big\}.$$

Then (B3) is verified.

For (B4), since $||\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0||_2=O_p(n^{-1/2})$ and by the Taylor expansion, we have

$$\exp(\hat{\boldsymbol{\beta}}_n^T\boldsymbol{Z})=\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})+\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})Z^T(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)+o_p(n^{-1/2}).$$

By the above equation and Lemma 4.4, after some algebraic calculations, we obtain

$$Q(\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}_n,F_0)[\boldsymbol{h}_1,h_2]-Q(\boldsymbol{\beta}_0,\hat{\Lambda}_n,F_0)[\boldsymbol{h}_1,h_2]=R_1(\boldsymbol{h}_1,h_2)(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)+o_p(n^{-1/2}).\tag{4.10}$$

Noting that

$$
\begin{aligned}
&Q(\boldsymbol{\beta}_0, \hat{\Lambda}_n, F_0)[\boldsymbol{h}_1, h_2] - Q(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] \\
=&\mathcal{P}\left[\sum_{j=1}^{K}\Big\{\Delta\Big(\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\big(\triangle\hat{\Lambda}_{n,j}(Y)^2 - \triangle\Lambda_{0,j}(Y)^2\big)h_1^T\boldsymbol{Z}\right. \\
+&\big(\triangle\hat{\Lambda}_{n,j}(Y) - \triangle\Lambda_{0,j}(Y)\big)\big(\triangle N_j\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})h_1^T\boldsymbol{Z} - \exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle h_{2,j}(Y)\big)\Big) \\
+&\frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})}\int_Y^\infty\Big(\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\big(\triangle\hat{\Lambda}_{n,j}(u)^2 - \triangle\Lambda_{0,j}(u)^2\big)h_1^T\boldsymbol{Z} \\
+&\left.\big(\triangle\hat{\Lambda}_{n,j}(u) - \triangle\Lambda_{0,j}(u)\big)\big(\triangle N_j\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})h_1^T\boldsymbol{Z} - \exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle h_{2,j}(u)\big)\Big)dF_0(u|\boldsymbol{Z})\Big\}\right] \\
=&R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0) - \mathcal{P}\left[\sum_{j=1}^{K}\Big\{\Delta\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})(\triangle\hat{\Lambda}_{n,j}(Y) - \triangle\Lambda_{0,j}(Y))^2h_1^T\boldsymbol{Z}\right. \\
+&\left.\frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})}\int_Y^\infty\big(\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})(\triangle\hat{\Lambda}_{n,j}(u) - \triangle\Lambda_{0,j}(u))^2h_1^T\boldsymbol{Z}\big)dF_0(u|\boldsymbol{Z})\Big\}\right] \\
=&R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0) + o_p(n^{-1/2}),
\end{aligned}
$$

and by (4.10), it follows that

$$
\begin{aligned}
&Q(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0)[\boldsymbol{h}_1, h_2] - Q(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] \\
=&R_1(\boldsymbol{h}_1, h_2)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0) + o_p(n^{-1/2}) \qquad (4.11) \\
=&\dot{Q}_{1,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\Lambda}_n - \Lambda_0)[\boldsymbol{h}_1, h_2] + o_p(n^{-1/2}).
\end{aligned}
$$

By Lemma 4.4,

$$
\begin{aligned}
&|Q(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n)[\boldsymbol{h}_1, h_2] - Q(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] - \dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2]| \\
=&\left|\mathcal{P}\left[(1-\Delta)\frac{\hat{F}_n(Y) - F_0(Y)}{1-F_0(Y)}\sum_{j=1}^{K}\left\{\frac{\int_Y^\infty\tilde{\varphi}_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1, h_2]d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right.\right.\right. \\
-&\left.\left.\left.\frac{\int_Y^\infty\tilde{\varphi}_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1, h_2]dF_0(u)}{1-F_0(Y)}\right\}\right]\right|
\end{aligned}
$$

$$\lesssim ||\hat{F}_n - F_0||_\infty \mathcal{P}\left[(1-\Delta)\sum_{j=1}^{K}\left|\frac{\int_Y^\infty \tilde{\varphi}_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]d\hat{F}_n(u)}{1-\hat{F}_n(Y)}\right.\right.$$
$$\left.\left. - \frac{\int_Y^\infty \tilde{\varphi}_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]dF_0(u)}{1-F_0(Y)}\right|\right] \lesssim ||\hat{F}_n - F_0||_\infty^2 = o_p(n^{-1/2}). \tag{4.12}$$

Furthermore, since $\Lambda_0 \in \mathcal{H}_r$ with $r \geq 2$, we have $d_1(\hat{\Lambda}_n', \Lambda_0') = o_p(1)$. By Lemma 4.4,

$$\Big|\big(Q(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{F}_n)[\boldsymbol{h}_1,h_2] - Q(\boldsymbol{\beta}_0, \Lambda_0, \hat{F}_n)[\boldsymbol{h}_1,h_2]\big)$$
$$-\big(Q(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, F_0)[\boldsymbol{h}_1,h_2] - Q(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1,h_2]\big)\Big|$$
$$\lesssim (d_1(\hat{\Lambda}_n, \Lambda_0) + d_1(\hat{\Lambda}_n', \Lambda_0') + ||\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0||_2) d_2(\hat{F}_n, F_0) = o_p(n^{-1/2}).$$

Thus, according to (4.11), (4.12) and the above inequality, (B4) holds.

Finally, we consider (B5). Note that

$$\sqrt{n}Q_n(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1,h_2] + \sqrt{n}\dot{Q}_{2,\boldsymbol{\beta}_0,\Lambda_0,F_0}(\hat{F}_n - F_0)[\boldsymbol{h}_1,h_2]$$
$$= \sqrt{n}\mathbb{P}_n\psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1,h_2] + \sqrt{n}\mathcal{P}\left[\int_Y^\infty \bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]d(\hat{F}_n - F_0)(u|\boldsymbol{Z})\right].$$

According to Lemma 4.2, we have

$$\mathcal{P}\left[\int_Y^\infty \bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]d(\hat{F}_n(u|\boldsymbol{Z}) - F_0(u|\boldsymbol{Z}))\right]$$
$$= \mathcal{P}\left[\int_Y^\infty \frac{\partial\bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]}{\partial u}(\hat{F}_n(u|\boldsymbol{Z}) - F_0(u|\boldsymbol{Z}))du\right.$$
$$\left. - \bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(Y;X)[\boldsymbol{h}_1,h_2](\hat{F}_n(Y|\boldsymbol{Z}) - F_0(Y|\boldsymbol{Z}))\right]$$
$$= \mathcal{P}\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\int_Y^\infty \frac{\partial\bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)[\boldsymbol{h}_1,h_2]}{\partial u}\Omega(u,\boldsymbol{Z};\tilde{Y}_i,\tilde{\Delta}_i,\tilde{\boldsymbol{Z}}_i)du\right.\right.$$
$$\left.\left. - \bar{\varphi}_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(Y;X)[\boldsymbol{h}_1,h_2]\Omega(Y,\boldsymbol{Z};\tilde{Y}_i,\tilde{\Delta}_i,\tilde{\boldsymbol{Z}}_i)\right\}\right]$$
$$=: \mathbb{P}_n\varphi(\boldsymbol{\beta}_0, \Lambda_0, F_0; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}})[\boldsymbol{h}_1,h_2],$$

where

$$\varphi(\boldsymbol{\beta}_0, \Lambda_0, F_0; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}})[\boldsymbol{h}_1, h_2] = \mathcal{P}_X \left[\left\{ \int_Y^\infty \frac{\partial \bar{\varphi}_{\boldsymbol{\beta}_0, \Lambda_0, F_0}(u; X)[\boldsymbol{h}_1, h_2]}{\partial u} \Omega(u, \boldsymbol{Z}; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}}) du \right.\right.$$
$$\left.\left. - \ \bar{\varphi}_{\boldsymbol{\beta}_0, \Lambda_0, F_0}(Y; X)[\boldsymbol{h}_1, h_2] \Omega(Y, \boldsymbol{Z}; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}}) \right\} \right].$$

By the central limit theorem, it follows that

$$\sqrt{n} Q_n(\boldsymbol{\beta}_0, \Lambda_0, F_0)[\boldsymbol{h}_1, h_2] + \sqrt{n} \dot{Q}_{2, \boldsymbol{\beta}_0, \Lambda_0, F_0}(\hat{F}_n - F_0)[\boldsymbol{h}_1, h_2] \rightsquigarrow N(0, \sigma_0[\boldsymbol{h}_1, h_2]^2),$$

where

$$\sigma_0[\boldsymbol{h}_1, h_2]^2 = E\left[\left\{ \psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1, h_2] + \varphi(\boldsymbol{\beta}_0, \Lambda_0, F_0; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}})[\boldsymbol{h}_1, h_2] \right\}^2 \right]. \tag{4.13}$$

By Lemma 4.7, $\sqrt{n} R_1(\boldsymbol{h}_1, h_2)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \sqrt{n} R_2(\boldsymbol{h}_1, h_2)(\hat{\Lambda}_n - \Lambda_0) \rightsquigarrow N(0, \sigma_0[\boldsymbol{h}_1, h_2]^2)$.

(ii) To prove the asymptotic normality of $\hat{\boldsymbol{\beta}}_n$, we need to find an $(\boldsymbol{h}_1^*, h_2^*)$ such that $R_2(\boldsymbol{h}_1^*, h_2^*)(\hat{\Lambda}_n - \Lambda_0) = 0$. After some algebraic calculations, we obtain

$$R_2(\boldsymbol{h}_1^*, h_2^*)(\hat{\Lambda}_n - \Lambda_0) = \mathcal{P}\left[\sum_{j=1}^K \left\{ (\triangle \hat{\Lambda}_{n,j}(U) - \triangle \Lambda_{0,j}(U)) \right.\right.$$
$$\left.\left. \times E\left[\left(\triangle h_{2,j}^*(U) + \triangle \Lambda_{0,j}(U) \boldsymbol{h}_1^{*T} \boldsymbol{Z} \right) \exp(2\boldsymbol{\beta}_0^T \boldsymbol{Z}) | U, K, \underline{T} \right] \right\} \right].$$

This implies that

$$\triangle h_{2,j}^*(U) = \frac{-\boldsymbol{h}_1^{*T} E[\exp(2\boldsymbol{\beta}_0^T \boldsymbol{Z}) \boldsymbol{Z} | U, K, \underline{T}]}{E[\exp(2\boldsymbol{\beta}_0^T \boldsymbol{Z}) | U, K, \underline{T}]} \triangle \Lambda_{0,j}(U) = -\boldsymbol{h}_1^{*T} \boldsymbol{R}^*(U, K, \underline{T}) \triangle \Lambda_{0,j}(U),$$

where $\boldsymbol{R}^*(U, K, \underline{T}) = E[\exp(2\boldsymbol{\beta}_0^T \boldsymbol{Z}) \boldsymbol{Z} | U, K, \underline{T}] / E[\exp(2\boldsymbol{\beta}_0^T \boldsymbol{Z}) | U, K, \underline{T}]$. Then we have

$$\triangle \Lambda_j(U) \boldsymbol{h}_1^{*T} \boldsymbol{Z} + \triangle h_{2,j}^*(U) = \triangle \Lambda_j(U) \boldsymbol{h}_1^{*T} (\boldsymbol{Z} - \boldsymbol{R}^*(U, K, \underline{T})). \tag{4.14}$$

It follows that

$$
\begin{aligned}
& R_1(\boldsymbol{h}_1^*, h_2^*)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
=& \boldsymbol{h}_1^{*T}\mathcal{P}\left[\sum_{j=1}^{K}\left\{\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)^2\big(\boldsymbol{Z}-\boldsymbol{R}^*(U,K,\underline{T})\big)\right\}\boldsymbol{Z}^T\right](\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
=& \boldsymbol{h}_1^{*T}\boldsymbol{A}^*(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0),
\end{aligned}
$$

where $\boldsymbol{A}^* = \mathcal{P}\big[\sum_{j=1}^{K}\{\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(U)^2(\boldsymbol{Z}-\boldsymbol{R}^*(U,K,\underline{T}))^{\otimes 2}\}\big]$. Furthermore, by (4.14), we obtain

$$
\begin{aligned}
& \psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1^*, h_2^*] \\
=& \boldsymbol{h}_1^{*T}\sum_{j=1}^{K}\Big[\Delta\big\{\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(Y)\big\}\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(Y)(\boldsymbol{Z}-\boldsymbol{R}^*(Y,K,\underline{T})) \\
+& \frac{1-\Delta}{1-F(Y|Z)}\int_Y^\infty\big\{\triangle N_j - \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u)\big\}\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\triangle\Lambda_j(u) \\
\times& (\boldsymbol{Z}-\boldsymbol{R}^*(u,K,\underline{T}))dF(u|Z)\Big] =: \boldsymbol{h}_1^{*T}\psi^*(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)
\end{aligned}
$$

and

$$
\begin{aligned}
& \varphi(\boldsymbol{\beta}_0, \Lambda_0, F_0; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}})[\boldsymbol{h}_1^*, h_2^*] \\
=& \boldsymbol{h}_1^{*T}\mathcal{P}_X\left[\left\{\int_Y^\infty \frac{\partial\bar{\varphi}^*_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)}{\partial u}\Omega(u,\boldsymbol{Z};\tilde{Y},\tilde{\Delta},\tilde{\boldsymbol{Z}})du \right.\right. \\
-& \left.\left. \bar{\varphi}^*_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(Y;X)\Omega(Y,\boldsymbol{Z};\tilde{Y},\tilde{\Delta},\tilde{\boldsymbol{Z}})\right\}\right] \\
=:& \boldsymbol{h}_1^{*T}\varphi^*(\boldsymbol{\beta}_0, \Lambda_0, F_0; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}}),
\end{aligned}
$$

where

$$
\bar{\varphi}^*_{\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X)\frac{1-\Delta}{1-F_0(Y|\boldsymbol{Z})}\sum_{j=1}^{K}\left\{\tilde{\varphi}^*_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X) - \int_Y^\infty \frac{\tilde{\varphi}^*_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(s;X)}{1-F_0(Y|\boldsymbol{Z})}dF_0(s|Z)\right\}
$$

and $\tilde{\varphi}^*_{j,\boldsymbol{\beta}_0,\Lambda_0,F_0}(u;X) = \{\triangle N_j - \exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_{0,j}(u)\}\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z})\triangle\Lambda_j(u)(\boldsymbol{Z}-\boldsymbol{R}^*(u,K,\underline{T}))$.

After some algebraic calculations, we have

$$\sigma_0[\boldsymbol{h}_1^*, h_2^*]^2 = E\left[\left\{\psi(\boldsymbol{\beta}_0, \Lambda_0, F_0; X)[\boldsymbol{h}_1^*, h_2^*] + \varphi(\boldsymbol{\beta}_0, \Lambda_0, F_0; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}})[\boldsymbol{h}_1^*, h_2^*]\right\}^2\right]$$

$$= \boldsymbol{h}_1^{*T} E\left[\left\{\psi^*(\boldsymbol{\beta}_0, \Lambda_0, F_0; X) + \varphi^*(\boldsymbol{\beta}_0, \Lambda_0, F_0; \tilde{Y}, \tilde{\Delta}, \tilde{\boldsymbol{Z}})\right\}^2\right] \boldsymbol{h}_1^* =: \boldsymbol{h}_1^{*T} \boldsymbol{B}^* \boldsymbol{h}_1^*.$$

It follows that $\sqrt{n}\boldsymbol{h}_1^{*T} \boldsymbol{A}^*(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \rightsquigarrow N(0, \boldsymbol{h}_1^{*T}\boldsymbol{B}^*\boldsymbol{h}_1^*)$ for all $\boldsymbol{h}_1^* \in \mathcal{R}$. Then we obtain $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \rightsquigarrow N(0, (\boldsymbol{A}^*)^{-1}\boldsymbol{B}^*((\boldsymbol{A}^*)^{-1})^T)$.

Finally, we turn to consider $(\boldsymbol{h}_1^{**}, h_2^{**})$ such that $R_1(\boldsymbol{h}_1^{**}, h_2^{**})(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = 0$, which implies $\boldsymbol{h}_1^{**} = \boldsymbol{R}^{**}(h_2^{**})$ with

$$\boldsymbol{R}^{**}(h_2) = -\left\{E\left[\sum_{j=1}^K \left\{\triangle\Lambda_{0,j}(U)^2 \exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\boldsymbol{Z}^{\otimes 2}\right\}\right]\right\}^{-1}$$

$$\times E\left[\sum_{j=1}^K \left\{\triangle h_{2,j}(U)\triangle\Lambda_{0,j}(U)\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})\boldsymbol{Z}\right\}\right].$$

Then we have

$$R_2(\boldsymbol{h}_1^{**}, h_2^{**})(\hat{\Lambda}_n - \Lambda_0) = \mathcal{P}\left[\sum_{j=1}^K \left\{\left(\triangle h_{2,j}^{**}(U) + \triangle\Lambda_{0,j}(U)\boldsymbol{R}^{**}(h_2^{**})^T\boldsymbol{Z}\right)\right.\right.$$

$$\left.\left.\times \exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})(\triangle\hat{\Lambda}_{n,j}(U) - \triangle\Lambda_{0,j}(U))\right\}\right].$$

Taking $\tilde{\sigma}_0[h_2]^2 = \sigma_0[\boldsymbol{R}^{**}(h_2), h_2]^2$ for all $h_2 \in \mathcal{H}_r$, it follows that

$$\sqrt{n}\mathcal{P}\left[\sum_{j=1}^K \left\{\left(\triangle h_{2,j}(U) + \triangle\Lambda_{0,j}(U)\boldsymbol{R}^{**}(h_2)^T\boldsymbol{Z}\right)\exp(2\boldsymbol{\beta}_0^T\boldsymbol{Z})(\triangle\hat{\Lambda}_{n,j}(U) - \triangle\Lambda_{0,j}(U))\right\}\right]$$

converges in distribution to $N(0, \tilde{\sigma}_0[h_2]^2)$. □

Chapter 5

Concluding Remarks

In this thesis, we study two topics in the area of survival analysis.

In Chapter 2, we conduct the subgroup analysis for the heterogenous Cox model using the concave fusion penalized partial likelihood approach. The proposed approach can identify the grouping structure and estimate the heterogeneous covariate effects involved in the model simultaneously and automatically. To obtain an efficient solution to the objective function, we apply the majorized ADMM algorithm which not only converges faster but also calculates more accurately than the local quadratic approximated ADMM algorithm suggested by Ma et al. (2019). Our simulation and real data analysis demonstrate that the proposed method performs well. We expect that the proposed approach can be extensively used for subgroup analysis with survival data.

Based on the residual time between observation and the terminal event, Chapter 3 builds a conditional nonparametric mean functional model to study the explicit effects of the terminal events on the occurrence rate of panel count data. We propose a two-stage estimation procedure and obtain the consistency and the convergence rate of the nonparametric estimator by extending modern empirical process theories. In addition, we establish the asymptotic normality of the nonparametric estimator and conduct the two-sample hypothesis test based on the theoretical results. At last, we use the proposed method to analyze the CLHLS data and draw some reasonable and practical conclusions.

Chapter 4 considers the effect of the covariates and studies the intricate interaction between the terminal event and the recurrent event with panel count data more efficiently and accurately. We propose a semiparametric two-stage estimation for the baseline conditional mean function and the parameter of covariates, and we also establish the asymptotic properties for the estimator. Our simulation study demonstrates that the proposed estimation has satisfactory performance with the finite sample size. Finally, we use the approach to analyze the CLHLS data and obtain some practical conclusions.

There still exist some interesting questions for future researches.

For the subgroup analysis, after the identification of the subgroup structure, we can utilize this result for the prediction purpose. Considering the statistical methods for classification, such as the support vector machines, we can divide a new individual into existing subgroups. Based on the estimate for the coefficient of this subgroup, we can predict the treatment effect. Next, we can consider the algorithm based on second-order optimization techniques instead of the majorized ADMM algorithm because the ADMM-based algorithm converges slowly. Since the performances of the proposed estimation highly depend on the choice of the initial value in the algorithm, we can improve our estimation by taking a better initial value. Further, the proposed method can be extended to handling the case where the unknown number of subgroups and the dimension of covariates can increase with sample size in the proposed heterogenous Cox model. For this situation, we propose to use the criterion function

$$Q_n(\eta, \boldsymbol{\beta}) = \ell_n(\eta, \boldsymbol{\beta}) + \sum_{i<j} p_\gamma^{(1)}(||\beta_i - \beta_j||, \lambda_1) + \sum_{j=1}^{q} p_\gamma^{(2)}(\eta_j, \lambda_2).$$

With the penalty functions $p_\gamma^{(1)}(\cdot,\lambda_1)$ and $p_\gamma^{(2)}(\cdot,\lambda_2)$, we can conduct subgroup analysis and variable selection simultaneously. In subgroup analysis, a common problem of interest is to test the existence of a subgroup with an enhanced treatment effect. We can use our subgroup analysis method to study the signs of coefficients rather than the values of coefficients to solve the problem.

For the panel count data with an informative terminal event, we may first consider the test for goodness-of-fit by using the coefficient of determination or the Pearson's chi-squared test. Similar to the techniques using in the nonparametric model, we can conduct the hypothesis test in the semiparametric model to identify the differences between two groups. Instead of using the least squares-based two-stage estimation, it is natural to adapt the maximum-likelihood-based two-stage approach to enhance the efficiency of the estimation. To study the maximum-likelihood-based two-stage approach, we need to overcome the challenges brought by the difficulties of theoretical proofs. We can also extend our model to

$$E(\tilde{N}(t;U)|U=u)=\Lambda(u,t).$$

Using a two-dimensional function, this model has a more widespread application. Another straightforward improvement is that we can consider the variable selection problem by introducing a penalized function such as SCAD (Fan and Li, 2001) or MCP (Zhang, 2010). Furthermore, in this thesis, the observation progress of the recurrent event $\underline{T}$ is independent of the covariates, and the covariates are also time-independent. As these assumptions are usually not satisfied in many applications, our model will be more accurate when we consider the effect of $\underline{T}$ and the time-dependent covariates. Finally, subgroup analysis is also applicable to the semiparametric regression model for panel count data when the treatment heterogeneity exists.

Bibliography

[1] Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* **6,** 701–726.

[2] Aalen, O. O., and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine* **10,** 1227–1240.

[3] Aalen, O. O., and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* **5,** 141–150.

[4] Akritas, M. G. (2000). The central limit theorem under censoring. *Bernoulli* **6,** 1109–1120.

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

[6] Aranda-Ordaz, F. J. (1983). An extension of the proportional-hazards model for grouped data. *Biometrics* **39,** 109–117.

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

[8] Banfield, J. D., and Raftery, A. E. (1993). Model-based gaussian and non-gaussian clustering. *Biometrics* **49,** 803—821.

[9] Bradic, J., Fan, J., and Jiang, J. (2011). Regularization for Cox's proportional hazards model with NP-dimensionality. *The Annals of Statistics* **39,** 3092–3120.

[10] Breslow, N. (1970). A generalized Kruskal–Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* **57,** 579—594.

[11] Breslow, N. (1972). Discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society: Series B* **34,** 216–217.

[12] Breslow, N., and Crowley, J. A. (1974). Large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics* **2,** 437–453.

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

[14] Buckley, J., and James, I. (1979). Linear regression with censored data. *Biometrika* **66,** 429—436.

[15] Chan, K. C. G., and Wang, M.-C. (2010). Backward estimation of stochastic processes with failure events as time origins. *The Annals of Applied Statistics* **4,** 1602–1620.

[16] Chen, K., Chen, K., Muller, H. G., and Wang, J. L. (2011). Stringing high-dimensional data for function analysis. *Journal of the American Statistical Association* **106,** 275–284.

[17] Christensen, E., Schlichting, P., Andersen, P. K., et al. (1986). Updating prognosis and therapeutic effect evaluation in cirrhosis using Cox's multiple regression model for time-dependent variables. *Scandinavian Journal of Gastroenterology* **21,** 163—174.

[18] Cook, R. J., and Lawless, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* **16,** 911–924.

[19] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B* **34,** 187–202.

[20] Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

[21] De Gruttola, V., and Tu, X. M. (1994). Modeling progression of CD4+ lymphocyte count and its relationship to survival time. *Biometrics* **50,** 1003-–1014.

[22] Diamond, I. D., McDonald, J. W., and Shah, I. H. (1986). Proportional hazard models for current status data: Applications to the study of differentials in age at weaning in Pakistan. *Demography* **28,** 607-–620.

[23] Everitt, B., and Hand, D. J. (1981). *Finite mixture distributions.* New York: Chapman & Hall.

[24] Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96,** 1348–1360.

[25] Fan, J., and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Annals of Statistics* **30,** 74–99.

[26] Fan, J., and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B* **70,** 849–911.

[27] Fang, E. X., Ning, Y., and Liu, H. (2017). Testing and confidence intervals for high dimensional proportional hazards models. *Journal of the Royal Statistical Society: Series B* **79,** 1415–1437.

[28] Fleming, T. R., and Harrington, D. P. (1991). *Counting processes and survival analysis.* John Wiley & Sons.

[29] Földes, A., Rejtő, L., and Winter, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate. *Periodica Mathematica Hungarica* **12,** 15–29.

[30] Gehan, E. A. (1965). A generalized wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika* **52,** 203-–223.

[31] Gill, R. (1980). Nonparametric estimation based on censored observations of a markov renewal process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* textbf53, 97—116.

[32] Gill, R. (1981). Testing with replacement and the product-limit estimator. *The Annals of Statistics* **9,** 853-–860.

[33] Ghosh, D., and Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12,** 663-–688.

[34] Ghosh, D., and Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59,** 877-–885.

[35] Harrington, D. P., and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69,** 553–566.

[36] Hastie, T., and Tibshirani, R. (1996). Discriminant analysis by gaussian mixtures. *Journal of the Royal Statistical Society: Series B* **58,** 155-–176.

[37] He, X., and Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics* **3,** 299–308.

[38] Hu, C., and Lin, D. Y. (2002). Cox regression with covariate measurement error. *Scandinavian Journal of Statistics* **29,** 637–655.

[39] Hu, X. J., Sun, J., and Wei, L. J. (2003). Regression analysis of panel count data. *Scandinavian Journal of Statistics* **25,** 25–43.

[40] Hu, P., Tsiatis, A. A., and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54,** 1407–1419.

[41] Huang, C. Y., and Wang, M. C. (2004). Joint modeling and estimation of recurrent event processes and failure time. *Journal of the American Statistical Association* **99,** 1153-–1165.

[42] Huang, J., Sun, T., Ying, Z., Yu, Y., and Zhang, C. (2013). Oracle inequalities for the lasso in the Cox model. *The Annals of Statistics* **41,** 1142–1165.

[43] 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.

[44] Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53,** 457–481.

[45] Klein, J. P. (1991). Small sample moments of some estimators of the variance of the Kaplan-Meier and Nelson-Aalen estimators. *Scandinavian Journal of Statistics* **18,** 333–340.

[46] Klein, J. P., and Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data.* Springer Science & Business Media.

[47] Kong, D., Ibrahim, J. G., Lee, E., and Zhu, H. (2018). FLCRM: Functional linear cox regression model. *Biometrics* **74,** 109–117.

[48] Kong, S., Nan, B., Kalbfleisch, J. D., Saran, R., and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *Journal of the American Statistical Association* **113,** 357–368.

[49] Kravitz, R. L., Duan, N., and Braslow, J. (2004). Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly* **82,** 661–687.

[50] Lagakos, S. W. (2006). The challenge of subgroup analyses-reporting without distorting. *New England Journal of Medicine* **354,** 1667–1669.

[51] Lancaster, A., and Intrator, O. (1998). Panel data with survival: Hospitalization of HIV-positive patients. *Journal of the American Statistical Association* **93,** 46—53.

[52] Lange, K. (2001). *Numerical Analysis for Statisticians.* Springer.

[53] Lee, E. R., Noh, H., and Park, B. U. (2014). Model selection via Bayesian Information Criterion for quantile regression models. *Journal of the American Statistical Association* **109,** 216–229.

[54] Li, Q., and Lagakos, S. (1997). Use of the Wei–Lin–Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine* **16,** 925-–940.

[55] Li, M., Sun, D., and Toh, K. C. (2016). A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM Journal on Optimization* **26,** 922–950.

[56] Lin, D., Wei, L., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85,** 605-–618.

[57] Lin, D. Y., and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81,** 61–71.

[58] Little, R. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90,** 1112-–1121.

[59] Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60,** 747–756.

[60] Lu, M., Zhang, Y., and Huang, J. (2007). Estimaton of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94,** 705–718.

[61] Lu, M., Zhang, Y., and Huang, J. (2009). Semiparametric estiamtion methods for panel count data using monotone B-splines. *Journal of the American Statistical Association* **104,** 1060–1070.

[62] Ma, S., and Huang, J. (2017). A concave pairwise fusion approach to subgroup analysis. *Journal of the American Statistical Association* **112,** 410–423.

[63] Ma, S., Huang, J., Zhang, Z., and Liu, M. (2019). Exploration of heterogeneous treatment effects via concave fusion. *The International Journal of Biostatistics* **16**.

[64] McNichols, D. T., and Padgett, W. J. (1986). Mean and variance of a kernel density estimator under the Koziol–Green model of random censorship. *Sankhyā: The Indian Journal of Statistics* **48,** 150–168.

[65] McNicholas, P. D. (2010). Model-based classification using latent gaussian mixture models. *Journal of Statistical Planning and Inference* **140,** 1175-–1181.

[66] Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data. *The Annals of Statistics* **14,** 766—773.

[67] Muthén, B., and Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society: Series A* **172,** 639—657.

[68] Muthén, B., and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55,** 463—469.

[69] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14,** 945–966.

[70] Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69,** 331–342.

[71] Qu, S., Wang, J. L., and Wang, X. (2016). Optimal estimation for the functional cox model. *The Annals of Statistics* **44,** 1708–1738.

[72] Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical science* **3,** 425–441.

[73] Rothwell, P. M. (2005). Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet* **365,** 176–186.

[74] Sellke, T. (1988) Weak convergence of the Aalen estimator for a censored renewal process. *Statistical Decision Theory and Related Topics IV* **2,** 183–194.

[75] Shen, J., and He, X. (2015). Inference for subgroup analysis with a structured logistic-normal mixture model. *Journal of the American Statistical Association* **110,** 303–312.

[76] Shen, X., and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics* **14,** 590–606.

[77] Sorensen, T. (1996). Which patients may be harmed by good treatments? *Lancet* **348,** 351–352.

[78] Sun, J., and Fang, H. B. (2003). A nonparametric test for panel count data. *Biometrika* **90,** 199–208.

[79] Sun, J., and Kalbfleisch, J. D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association* **88,** 1449-–1454.

[80] Sun, J., and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* **5,** 279–290.

[81] Sun, J., and Song, P. (2001). Statistical analysis of repeated measurements with informative censoring times. *Statistics in Medicine* **20,** 63-–73.

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

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

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

[85] Tarone, R. E., and Ware, J. H. (1997). On distribution-free tests for equality for survival distributions. *Biometrika* **64,** 156-–160.

[86] Tong, X., He, X., Sun, L., and Sun, J. (2009). Variable selection for panel count data via non-concave penalized estimating function. *Scandinavian Journal of Statistics* **36,** 620–635.

[87] van de Vijver, M. J., He, Y., Van't Veer, L. J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* **347,** 1999–2009.

[88] van der Vaart, A. W. (1996). New donsker classes. *The Annals of Probability* **24,** 862–878.

[89] van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge.

[90] van der Vaart, A. W. (2002). Semiparametric statistics. *Lectures on Probability Theory and Statistics. École d'Eté de Probabilités de Saint-Flour XXIX—1999. Lecture Notes in Math.* **1781** 331—457. Springer, New York.

[91] van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

[92] van't Veer, L. J., Dai, H., van de Vijver, M. J., et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536.

[93] Wang, M. C., and Chang, S. H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association* **94,** 146—153.

[94] Wang, M., Qin, J., and Chiang, C. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96,** 1057—1065.

[95] Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the cox regression model in survival analysis. *Statistics in Medicine* **11,** 1871—1879.

[96] Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065—1073.

[97] Wellner, J. A., and Zhang, Y. (2000). Two estimator of the mean of a counting process with panel count data. *The Annals of Statistics* **28,** 779–814.

[98] Wellner, J. A., and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics* **35,** 2106–2142.

[99] Wong, C. S., and Li, W. K. (2001). On a logistic mixture autoregressive model. *Biometrika* **88,** 833-–846.

[100] Wu, R., Zheng, M., and Yu, W. (2016). Subgroup analysis with time-to-event data under a logistic-Cox mixture model. *Scandinavian Journal of Statistics* **43,** 863–878.

[101] Yan, X., Yin, G., and Zhao, X. (2020). Subgroup analysis in censored linear regression. *Statistica Sinica Perprint*

[102] Ye, Y., Kalbfleisch, J. D., and Schaubel, D. E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63,** 78—87.

[103] Zeng, D., and Cai, J. W. (2010). A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika* **97,** 699—712.

[104] Zeng, Y., Vaupel, J., Xiao, Z., Liu, Y., and Zhang, C. Chinese Longitudinal Healthy Longevity Survey (CLHLS), 1998–2014. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2017-04-11. https://doi.org/10.3886/ICPSR36692.v1

[105] Zhang, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38,** 894–942.

[106] Zhang, H., Sun, J., and Wang, D. (2013). Variable selection and estimation for multivariate panel count data via the seamless-L0 penalty. *Canadian Journal of Statistics* **41,** 368–385.

[107] Zhang, H. H., and Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika* **94,** 691–703.

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

[109] Zhang, Y. (2006). Nonparametric k-sample tests for panel count data. *Biometrika* **93,** 777–790.

[110] Zhang, Y., Wang, J. H., and Zhu, Z. (2019). Robust subgroup identification. *Statistica Sinica* **29,** 1873–1889.

[111] Zhao, H., Li, Y., and Sun, J. (2013a). Analyzing panel count data with a dependent observation process and a terminal event. *Canadian Journal of Statistics* **41,** 174–191.

[112] Zhao, H., Li, Y., and Sun, J. (2013b). Semiparametric analysis of multivariate panel count data with dependent observation processes and a terminal event. *Journal of Nonparametric Statistics* **25,** 379–394.

[113] Zhao, H., Wu, Q., Li, G., and Sun, J. (2020). Simultaneous estimation and variable selection for interval-censored data with broken adaptive ridge regression. *Journal of the American Statistical Association* **115,** 204–216.

[114] Zhao, X., and Sun, J. (2011). Nonparametric comparison for panel count data with unequal observation processes. *Biometrics* **67,** 770–779.

[115] Zhao, X., and Zhang, Y. (2017). Asymptotic normality of nonparametrix M-estimators with aplications to hypothesis testing for panel count data. *Statistica Sinica* **27,** 931–950.

[116] Zhao, X., Zhou, J., and Sun, L. (2011). Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. *Biometrics* **67,** 404–414.

[117] Zhou, J., Zhang, H., Sun, L., and Sun, J. (2017). Joint analysis of panel count data with an informative observation process and a dependent terminal event. *Lifetime Data Analysis* **23,** 560–584.