GENERALIZED CHANGE-POINT HAZARD MODELS WITH CENSORED DATA

Yu Deng

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2016

Approved by:

Jianwen Cai

Donglin Zeng

David Couper

Quefeng Li

Jinying Zhao

ABSTRACT

Yu Deng: Generalized Change-Point Hazard Models with Censored Data
(Under the direction of Jianwen Cai and Donglin Zeng)

In many epidemiology studies, the biomarkers and survival endpoints of diseases are collected to investigate the association between risk factors and disease incidence. The risk of the disease may change when a certain risk factor exceeds a threshold. Finding this threshold value is important for individual risk prediction and disease prevention. In this dissertation, we develop semiparametric statistical approaches for the change point in both univariate and clustered survival data.

Family studies are very popular in public health studies due to its cost-efficiency in collecting the data. In Chapter 3, we propose a Cox-type marginal hazards model with an unknown change point for clustered event data. The marginal pseudo-partial likelihood functions are maximized to estimate the change point. We develop a supremum test and m out of n bootstrap to make inferences of the change point. We establish the consistency and asymptotic distributions of the proposed estimators. We evaluate the small sample performance of the proposed methods via extensive simulation studies. Finally, the Strong Heart Family Study dataset is analyzed to illustrate the methods.

To improve the performance in identifying the high-risk individuals, we propose a change hyperplane model, which is an extended change point model based on a linear combination of multiple risk factors. We develop the Cox proportional hazards model and Cox-type marginal hazards model with a change hyperplane for univariate and clustered event data in Chapter 4 and 5, respectively. To ensure identifiability, two different sets of criterion are shown to be equivalent and proved to guarantee the identifiability. The two-step procedure with application of the genetic algorithm is applied to maximize the partial or pseudo-partial likelihood function. We introduce the m out of n bootstrap to generate confidence intervals for the change hyperplane parameters. The supremum tests with score and robust score statistics are conducted to verify the existence of the change hyperplane in univariate and clustered data, respectively. The asymptotic properties of the proposed estimators are derived in both cases. The performance of the proposed approach is demonstrated via simulation studies and analyses in the Cardiovascular Health Study and the Strong Heart Family Study.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisers Dr. Jianwen Cai and Dr. Donglin Zeng. They gave me great guidance, support, and encouragement throughout my graduate studies. I have learned a great deal from their unique perspective on research, their sharp insight on statistical problems, and their passions. All of these will inspire me to strive for excellence in my future career.

I would like to give sincere thanks to my committee members: Drs. David Couper, Quefeng Li, and Jinying Zhao. I appreciate them for their time and insightful comments to improve my research.

I am deeply grateful to my supervisors Dr. Jianwen Cai and Dr. Daniela Sotres-Alvarez at the Collaborative Studies Coordinating Center, and Dr. Russell Reeve at Quintiles. They have offered me tremendous opportunities to develop and improve my collaborative and interdisciplinary research skills. All these working experiences enhanced my understanding of statistics in the applied field. In addition, their generous financial supports have helped me through my entire graduate study. My thanks also go to Dr. Kelly R Evenson for giving me invaluable opportunities to work with her. I have learned a lot from her passion in epidemiology and disciplined focus on research.

Last but not the least, I want to extend my gratitude to my parents, my husband, and all my friends for their unconditional love, support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: LITERATURE REVIEW 4

2.1 Semiparametric Regression Models for Censored Data 4

2.1.1 Cox Proportional Hazards Model 4

2.1.2 Proportional Hazards Model for Clustered Survival Data 5

2.2 Semiparametric single-index model 8

2.3 Change-Point Models with Censored Data 12

2.3.1 Models 12

2.3.2 M out of n Bootstrap 13

2.3.3 Tests 14

2.3.4 Optimization Methods 15

CHAPTER 3: PROPORTIONAL HAZARDS MODEL WITH A CHANGE POINT FOR CLUSTERED EVENT DATA 17

3.1 Introduction 17

3.2 Methods 19

3.2.1 Model and Parameter Estimation 19

3.2.2 Inference for Parameter Estimators 21

3.2.3 Hypothesis Testing for the Change Point 23

3.3 Asymptotic Results ... 24

3.4 Simulation Studies ... 27

3.5 Application to the Strong Heart Family Study ... 30

3.6 Discussion ... 33

3.7 Proof of Lemma and Theorems ... 34

CHAPTER 4: STATISTICAL INFERENCE FOR COX PROPORTIONAL HAZARDS MODEL WITH A CHANGE HYPERPLANE ... 49

4.1 Introduction ... 49

4.2 Methods ... 51

4.2.1 Model and Parameter Estimation ... 51

4.2.2 Inference ... 53

4.2.3 Hypothesis Testing for the Change Hyperplane ... 54

4.3 Asymptotic Properties ... 55

4.4 Simulation Studies ... 58

4.5 Application to the Cardiovascular Health Study ... 61

4.6 Conclusion ... 63

4.7 Proof of Lemma and Theorems ... 64

CHAPTER 5: CHANGE HYPERPLANE HAZARDS MODELS FOR CLUSTERED SURVIVAL DATA ... 77

5.1 Introduction ... 77

5.2 Methods ... 78

5.2.1 Model and Parameter Estimation ... 78

5.2.2 Inference ... 79

5.2.3 Hypothesis Testing for the Change Hyperplane ... 80

5.3 Asymptotic Properties ... 81

5.4 Simulation Studies ... 83

5.5 Application to the Strong Heart Family Study 86
5.6 Conclusion 88
5.7 Proof of Lemma and Theorems 90
CHAPTER 6: SUMMARY AND FUTURE RESEARCH 97
REFERENCES 100

LIST OF TABLES

3.1 Simulation Results for the Change Point and Regression Parameters. 29

3.2 Type I Error and Power for SUP Tests for the Existence of the Change Point 30

3.3 Analysis Results Based on the Strong Heart Family Study: Model 1 ($\widehat{\zeta} = 0.870[0.834, 0.907]$) and Model 2 ($\widehat{\zeta}_{ad-hoc} = 0.872$). 32

4.4 Simulation Results for the Change Hyperplane Parameters. 59

4.5 Simulation Results for the Regression Parameters. 60

4.6 Type I Error and Power for SUP Tests for the Existence of the Change Hyperplane 61

4.7 Change Hyperplane Estimates Based on the CHS 62

4.8 Regression Coefficients Estimates Based on the CHS 62

5.9 Simulation Results for the Change Hyperplane Parameters. 85

5.10 Simulation Results for the Regression Parameters. 85

5.11 Type I Error and Power for SUP Tests for the Existence of the Change Hyperplane 86

5.12 Change Hyperplane Estimates Based on the Strong Heart Family Study: Model 1 (full model) and Model 2 (final reduced model). 88

5.13 Regression Coefficients Estimates Based on Model 2 (final reduced model) in the Strong Heart Family Study. 88

LIST OF FIGURES

3.1 Diagnostic Plots. The log of negative log of survival functions versus time are plotted for leukocyte telomere length, gender, and BMI. Schoenfeld residuals are plotted for age, fasting glucose, and total triglycerides. 32

4.2 The Kaplan-Meier Plot of the Change Hyperplane (the logrank test with p-value< 0.001). 63

5.3 Diagnostic Plots. The log of negative log of survival functions versus time are plotted for leukocyte telomere length. Schoenfeld residuals are plotted for age, BMI. 89

CHAPTER 1: INTRODUCTION

Change-point effects have been observed in many epidemiology studies. The identified change points of certain biomakers are applied to predict disease risks. Such risk scores may influence the decision of early intervention. For example, in the Strong Heart Family Study, Zhao et al. (2014) found that participants with leukocyte telomere length (LTL) less than the first quartile have a significantly higher risk for developing incident diabetes compared to those with LTL longer than the first quartile. In addition, the change point based on LTL has also been observed to depend on some other biomarkers, such as triglycerides, HDL, etc. Thus, finding the threshold value based on a single or multiple covariates is important for identifying at-risk individuals. Furthermore, one important characteristic of such studies is that the collected samples may be correlated, e.g. family data. Consequently, it is of utmost interest to develop a rigorous and comprehensive framework to conduct the change point analysis for both univariate and clustered survival data subject to censoring.

Change Point Analysis for Clustered Event Data

The Cox proportional hazards model (Cox 1972) has been widely applied to estimate the association between potential risk factors and disease incidence under the assumption of independent observations. The change point analysis has been extensively studied in the univariate Cox proportional hazards model (Liang et al. 1990; Luo 1996; Pons 2002; Luo 1996; Gandy et al. 2005; Gandy and Jensen 2005; Jensen and Lütkebohmert 2008; Luo and Boyett 1997, Pons 2003, Kosorok and Song 2007). Among them, Luo and Boyett (1997), Pons (2003), Kosorok and Song (2007) assumed a non-smooth “jump" effect at an unknown threshold of a covariate. A two-step procedure is applied to maximize the partial likelihood to obtain the change point and regression coefficients. The consistency and asymptotic distribution of the proposed estimator are proved by Luo and Boyett (1997) and Pons (2003), respectively. Kosorok and Song (2007) generalized the asymptotic properties in the Cox model to the class of transformation models. However, the developed asymptotic properties and inference procedures for the univariate case cannot be applied directly to clustered survival data. The major difficulty for the change point analysis in the Cox marginal hazards model is the complicated asymptotic distribution of the change point estimator for clustered data. In Chapter 3, we focus on developing a Cox-type marginal hazards model (Lee et al. 1992, Cai and Prentice 1997) with an unknown change point in a covariate for clustered survival data.

Change Hyperplane Model for Univariate Survival Data

To improve the performance in identifying the rules to classify the high-risk individuals, we extend the change point model to the change hyperplane model, where the change hyperplane is based on a linear combination of multiple risk factors. The change hyperplane in the Cox proportional hazards model can be viewed as a single index. The single index has been widely applied in the Cox proportional hazards model as a remedy of the violation of the proportional hazards assumption. Wang (2004) and Huang and Liu (2006) assumed unspecified smooth link functions of the single index. However, the estimation approach in the single index Cox model cannot be directly applied to the change hyperplane model, because the unknown smooth function in the single index model is replaced by a discontinuous indicator function in the change hyperplane model. The change hyperplane function equals one or zero depending on whether the single index is larger or smaller than the change point. In addition, we prove that the convergence rate of the change hyperplane estimators is $1/n$ as opposed to the convergence rate $1/\sqrt{n}$ in the single index model. Consequently, the asymptotic properties for the single index Cox proportional hazards model cannot be extended to the Cox proportional hazards model with a change hyperplane. In Chapter 4, we propose a partially linear Cox proportional hazards model with a change hyperplane.

Change Hyperplane Model for Clustered Event Data

The strong heart family study is a family-based longitudinal study. The observed change point for incident diabetes may depend on more than one covariate. Finding the change hyperplane which has a significant association with the incidence of diabetes helps gain the understanding of the disease and evaluate the risk of each individual. However, the model proposed in Chapter 4 can only handle univariate data. Considering the varying number of cluster sizes, the asymptotic distributions of the change hyperplane estimators become more complicated. Aiming at gaining a wider application of the change hyperplane model, we consider a Cox-type marginal hazards model (Lee et al. 1992) with a change hyperplane based on multiple covariates for clustered survival data in Chapter 5.

Outlines of the Dissertation

In this dissertation, we conduct a literature review in Chapter 2. In Chapter 3, we develop a Cox-type marginal hazards model with an unknown change point depending on a single covariate for clustered survival data. In Chapter 4, we propose a partially linear Cox proportional hazards model with a change hyperplane. In Chapter 5, we investigate the Cox-type marginal hazards model with a change hyperplane for the Strong Heart Family Study. In Chapter 6, we discuss possible extensions and future work.

CHAPTER 2: LITERATURE REVIEW

2.1 Semiparametric Regression Models for Censored Data

2.1.1 Cox Proportional Hazards Model

The Cox proportional hazards model (Cox 1972) has been widely used to study the effect of covariates on failure times. Cox (1972) assumed that the hazard function for the failure time $\tilde{T}$ given the vector of covariates $\boldsymbol{Z}(t)$ follows

$$\lambda(t|\boldsymbol{Z}_i) = \lambda_0(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}, \tag{2.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $i = 1, ..., n$ is the subject indicator. The Cox proportional hazards model is a semiparametric model with the linear combination of the covariates as the parametric part and the unspecified positive baseline hazard function as the nonparametric part.

Cox (1975) proposed the partial likelihood to estimate the parameters of interest. We assume that the observed failure time $T_i = \min(\widetilde{T}_i, C_i)$ is the minimum of the failure time $\widetilde{T}_i$ and the censoring time C_i, $\Delta_i = I(\widetilde{T}_i \leq C_i)$ is the failure indicator, where $I(.)$ is the indicator function, and T_i and C_i are independent conditional on Z_i. Thus, the partial likelihood function is written as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left[\frac{\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(T_i)\}}{\sum_{l=1}^{n} I(T_l \geq T_i)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_l(T_i)\}}\right]^{\Delta_i}.$$

Obviously, the partial likelihood function depends only on the parameters of interest $\boldsymbol{\beta}$, and does not depend on the infinite dimensional nuisance function $\lambda_0(t)$. In order to obtain the estimates of $\boldsymbol{\beta}$, we maximize the partial likelihood using the Newton-Raphson method, which is equivalent to solving the score equation given by

$$U(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \Delta_i \left\{ \boldsymbol{Z}_i(T_i) - \frac{S_n^{(1)}(T_i;\boldsymbol{\beta})}{S_n^{(0)}(T_i;\boldsymbol{\beta})} \right\} = 0, \tag{2.2}$$

where $\boldsymbol{S}_n^{(r)}(t;\boldsymbol{\beta}) = \frac{1}{n}\left[\sum_{i=1}^{n} Y_i(t)\boldsymbol{Z}_i^{\otimes r}(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(T_i)\}\right]$, $Y_i(t) = I(T_i \geq t)$, and $r = 0, 1$.

For the proposed maximum partial likelihood estimators, Tsiatis (1981) and Andersen and Gill (1982) established the large sample properties by showing the strong consistency and asymptotic normality. Andersen and Gill (1982) generalized the proof by using counting process. Under asymptotic regularity conditions, $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ converges in distribution to a normal random variable $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \rightarrow_d N\left(\mathbf{0}, I_1(\boldsymbol{\beta}_0)^{-1}\right)$, where $I_1(\boldsymbol{\beta}) = \int_0^{\tau} v(t;\boldsymbol{\beta})s^{(0)}(t;\boldsymbol{\beta})\lambda_0(t)dt$, $s^{(r)}(t;\boldsymbol{\beta}) = \mathrm{E}\left(\boldsymbol{Z}_i^{\otimes r}(t)Y_i(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)\}\right)$, $v(t;\boldsymbol{\beta}) = \frac{s^{(2)}(t;\boldsymbol{\beta})}{s^{(0)}(t;\boldsymbol{\beta})} - \left\{\frac{s^{(1)}(t;\boldsymbol{\beta})}{s^{(0)}(t;\boldsymbol{\beta})}\right\}^{\otimes 2}$, and $r = 0, 1, 2$.

2.1.2 Proportional Hazards Model for Clustered Survival Data

The clustered failure time data are commonly seen in the observational studies and clinical trials. The correlated data can be multiple types of events within one subject or a small cluster of subjects. When failure time data are correlated, the original Cox proportional hazards model is no longer applicable, because the model depends on the assumption that the observations are independent of each other. Later, Andersen and Gill (1982) extended the Cox proportional hazards model to a more generalized version, whose intensity function can be applied to clustered events. The model assumed that the later occurring events would not be unaffected by the earlier events within the same individual. In addition to this model, there are two most commonly used approaches: marginal models and frailty models. The difference between these two models is whether to explicitly specify the dependence structure for the intra-cluster association or not. If the correlation among the observations is of interest, then the frailty model is preferred because it explicitly specifies the correlation structure for the observations within the same cluster. However, the marginal models are desirable when we focus on making inferences on the population average effect of risk factors on failure times. We discuss these two types of models in the following sections.

Marginal Model

The marginal models assume that the multivariate failure times follow the proportional hazards assumption in their marginal distributions. Wei et al. (1989) proposed the marginal model with failure-specific baseline hazard function and failure-specific regression parameters. Later, Lee et al. (1992) proposed a marginal model with common baseline hazard function for the situations that were consisted of correlated failure times from small groups. Combing the models in Wei et al. (1989) and Lee et al. (1992), Spiekerman and Lin (1998) and Clegg et al. (1999) proposed a general Cox-type regression model. The proposed model included both models as special cases. It allowed for both the failure-specific baseline hazard functions for different types of failures and the common baseline hazard functions for the cluster within the same type of failure. Here, we focus on clustered event data, instead of multiple event data. We assume that $i = 1, ..., n$ is the cluster indicator, and $j = 1, ..., K_i$ is the subject indicator within the ith cluster. For the jth subject in the ith cluster, the marginal hazard function is given by

$$\lambda(t|\boldsymbol{Z}_{ij}) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(t)\right\}, \tag{2.3}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and K_i is bounded. Let the observed failure time $T_{ij} = \min(\widetilde{T}_{ij}, C_{ij})$ be the minimum of the failure time $\widetilde{T}_{ij}$ and the censoring time C_{ij}, $\Delta_{ij} = I(\widetilde{T}_{ij} \leq C_{ij})$ be the failure indicator, and T_{ij} and C_{ij} be independent conditional on Z_{ij}. The corresponding pseudo-partial likelihood function is written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i}\left(\frac{\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(T_{ij})\}}{\sum_{l=1}^{n}\sum_{k=1}^{K_i} I(T_{lk} \geq T_{ij})\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_{lk}(T_{ij})\}}\right)^{\Delta_{ij}}. \tag{2.4}$$

The score equation is generated based on this pseudo-partial likelihood function (2.4). Lee et al. (1992) proved that the estimators based on the estimating equations, which ignored dependencies among the failure times, would still be consistent. Even though the model does not specify the dependence structure in the model, the correlation is adjusted for inference. The estimators are asymptotically normally distributed with a sandwich-type variance estimator.

In addition to this mentioned estimating method, several other approaches were developed for estimating the parameters of interest. Liang et al. (1990) proposed a different estimating equations for $\boldsymbol{\beta}$, which replaced the elements of the pseudo-partial likelihood function by combinations of all possible pairs of failure times in a cluster. Cai and Prentice (1995) and Cai and Prentice (1997) proposed weighted partial likelihood estimation equations for failure-specific and common baseline hazard functions, respectively. This weighted method could improve the efficiency when pairwise dependencies are strong and censoring rates are moderate.

Frailty Model

Unlike the marginal models, the frailty model, which is first studied by Clayton and Cuzick (1985), incorporates a frailty term which induces dependence among correlated data. The frailty model for clustered data is specified as

$$\lambda(t|Q_i) = Q_i\lambda_0(t)\exp\left\{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(t)\right\}, \tag{2.5}$$

where Q_i is assumed to be independent and identically distributed under the frailty distribution. Conditional on the frailty Q_i, the proportional hazards assumption still holds for each subject. In Clayton and Cuzick (1985), it assumed that the frailty Q_i followed a gamma distribution with mean one and an unknown variance. The other popular frailty distributions are proposed, for example, positive stable distribution (Hougaard 1986b), inverse Gaussian distribution (Hougaard 1986a), log Normal distribution (McGilchrist and Aisbett 1991), and uniform distribution (Lee and Klein 1988).

EM algorithm has been proposed to estimate the regression parameters in the frailty model with the assumption of gamma distribution (Klein 1992, Nielsen et al. 1992). The main step is to take a numerical integration for conditional expectations, and then take the maximization. If the Laplace transform of the frailty distribution is known, then EM algorithm is applicable in estimation for that class of distributions (Parner et al. 1997). The consistency and asymptotic distributions of the estimators in the frailty model with the assumption of the gamma distribution were proved in Murphy (1994) and Murphy (1995), respectively.

2.2 Semiparametric single-index model

The single-index model is becoming more and more popular in medical or economic studies. The reason is that it reduces the dimensionality of the covariates $\boldsymbol{X}$ by introducing the single index $\boldsymbol{\beta}^T\boldsymbol{X}$, which is a linear combination of all the covariates $\boldsymbol{X}$. The model is first proposed in the linear regression, which is given by

$$Y = g(\boldsymbol{\beta}^T\boldsymbol{X}) + \epsilon, \tag{2.6}$$

where Y is the response variable, $g(.)$ is the unknown univariate link function, $\boldsymbol{\beta}$ is an p-dimensional unknown index vector, and ϵ is a random variable with $\mathrm{E}(\epsilon|\boldsymbol{X}) = 0$. In this model, the index vector $\boldsymbol{\beta}$ is only identifiable under certain criterion. There are two criterion which are most commonly seen. One way is to assume that $\|\boldsymbol{\beta}\| = 1$ with its first non-zero element positive, the other way is to assume one of its non-zero elements to be one.

In the single-index model, one of the most challenging problem is to estimate the unknown link function $g(.)$ and the index vector $\boldsymbol{\beta}$. There are two most commonly applied estimators: average derivative estimator and semiparametric least square estimator. We introduce these two estimators and their asymptotic properties in the following sections.

Average Derivative Estimation

Härdle and Stoker (1989) applied a nonparametric estimating approach to obtain the marginal density $f(\boldsymbol{X})$ such that $\widehat{f}_{a_n} = n^{-1}a_n^{-p}\sum_{j=1}^{n} K\left(\frac{\boldsymbol{X}-\boldsymbol{X}_j}{a_n}\right)$, where $K(.)$ is a kernel function, $a_n > 0$ is the bandwidth, $a_n \to 0$, and p indicates that $K(.)$ is a function of p arguments. Under general conditions, $\widehat{f}(\boldsymbol{X})$, $\widehat{f}(\boldsymbol{X})'$, and $\widehat{l}(\boldsymbol{X}) = -\frac{\widehat{f}(\boldsymbol{X})'}{\widehat{f}(\boldsymbol{X})}$ are consistent estimators of the corresponding counterpart in δ. Thus, the average derivative estimator $\widehat{\delta}$ is defined as

$$\widehat{\delta} = \frac{1}{n}\sum_{i=1}^{n}\widehat{l}(\boldsymbol{X}_i)y_i I(\widehat{f}(\boldsymbol{X}_i) > b_n),$$

where b_n is a trimming bound such that $b_n \to 0$. This method leads to a $1/\sqrt{n}$ consistent estimator and the proposed estimator is proved to be asymptotically normally distributed.

Semiparametric Least Squares Estimator

The objective function in the single-index model is expressed as

$$J(\boldsymbol{\beta}) = \mathrm{E}\left(\left[y - \mathrm{E}\left\{y|g(\boldsymbol{\beta}^T\boldsymbol{X})\right\}\right]^2\right).$$

Since $\mathrm{E}\{y|g(\boldsymbol{\beta}^T\boldsymbol{X})\}$ is unknown, Ichimura (1993) proposed to a kernel estimator, which is

$$\widehat{\mathrm{E}}\left(\boldsymbol{\beta}^T\boldsymbol{X}_i\right) = \frac{\sum_{j\neq i} y_j K\left\{\frac{g(\boldsymbol{\beta}^T\boldsymbol{X}_i)-g(\boldsymbol{\beta}^T\boldsymbol{X}_j)}{a_n}\right\}}{\sum_{j\neq i} K\left\{\frac{g(\boldsymbol{\beta}^T\boldsymbol{X}_i)-g(\boldsymbol{\beta}^T\boldsymbol{X}_j)}{a_n}\right\}},$$

where $K(.)$ is a kernel function, the bandwidth $a_n > 0$, and $a_n \to 0$. Then $\boldsymbol{\beta}$ is estimated by minimizing the least square function $J(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - \widehat{\mathrm{E}}\left(\boldsymbol{\beta}^T\boldsymbol{X}_i\right)\right\}^2$. In addition, Ichimura (1993) established consistency and asymptotic normality of the proposed estimator. The selection of the bandwidth was later studied in Hardle et al. (1993).

These two estimators were originally proposed to estimate the link function and the index vector in Model (2.6). They can also be adapted to other forms of the single-index models. Carroll et al. (1997) extended the single-index model (2.6) to the generalized partially linear single-index model, which is written as $Y = \alpha^T\boldsymbol{Z} + g(\boldsymbol{\beta}^T\boldsymbol{X}) + \epsilon$. Obviously, when $\boldsymbol{\alpha} \equiv 0$, it is reduced to Model (2.6). The other form of the extended model is to include more than one single index in an unknown link function $g(.)$, which is given by $Y = g(\boldsymbol{\beta}_1^T\boldsymbol{X}, ..., \boldsymbol{\beta}_D^T\boldsymbol{X}) + \epsilon$ (Cook and Li 2002, Xia et al. 2002, Yin and Cook 2002).

In censored data, the Cox model assumes that the covariates has a log-linear relationship with the hazard function. However, this assumption may not be valid under some situations. Wang (2004) generalized the exponential link function to an unspecified form $\lambda(t|\boldsymbol{Z}) = \lambda_0(t)g\{\boldsymbol{\beta}^T\boldsymbol{Z}(t)\}$, where $g(.)$ is the unknown link function. They presented a two-step iterative algorithm to estimate the link function and regression parameters $\boldsymbol{\beta}$. The iterative algorithm is based on the proposed kernel estimated local log partial likelihood and global partial likelihood. The unknown link function $g(.)$ is estimated by the local likelihood approach. Let $r(v) = \{g^{(1)}(v), ..., g^{(p)}(v)/p!\}^T$, where $g^{(p)}(v)$ represents the pth order of derivatives of the function $g(.)$ at v. Then $r(v)$ could be obtained through the following local log partial likelihood

$$\sum_{i=1}^{n}\Big(K_a\left\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(T_i) - v\right\}\Big[\left\{\boldsymbol{\beta}^T\boldsymbol{Z}_i(T_i)\right\}^T r(v) \\ -\log\left(\sum_{j=1}^{n}\exp\left[\{\boldsymbol{\beta}^T\boldsymbol{Z}_j(T_i)\}^T r(v)\right]K_a\left\{\boldsymbol{\beta}^T\boldsymbol{Z}_j(T_i) - v\right\}Y_j(T_i)\right)\Big]\Big)^{\Delta_i},$$

where $K_a(.)$ is the kernel function with the bandwidth a, and $Y_i(t) = I(T_i \geq t)$. Based on the estimated $r(v)$, we can obtain the function $g(.)$ through integration, since $g(\boldsymbol{Z}(t)) \approx$

$g(v) + \{\boldsymbol{\beta}^T \boldsymbol{Z}(t)\}^T r(v)$. Then we can estimate $\boldsymbol{\beta}$ through the global partial likelihood $l(\boldsymbol{\beta}, \widehat{g}) = \sum_{i=1}^n \left[\widehat{g}\{Z_i(T_i)\} - \log\left\{\sum_{j=1}^n \exp[\widehat{g}\{Z_j(T_i)\}]Y_j(T_i)\right\}\right]^{\Delta_i}$. The asymptotic properties are established for the estimated link function.

The advantage of the model proposed by Wang (2004) is that it does not have any assumption for the relationship between the hazard function and risk factors. However, the interpretation based on the estimated link function is not straightforward. Thus, the model, which combines the Cox proportional hazards model and the single-index model, is of great interest. For example, Huang and Liu (2006) extended the Cox proportional hazards model as $\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{g(\boldsymbol{\beta}^T \boldsymbol{X})\}$, where $g(.)$ is the unknown smooth function. This model keeps the log linear relationship between the hazard function and $g(\boldsymbol{\beta}^T \boldsymbol{X})$. This model is applied when there is a violation of the linear relationship between the risk factors and the log of hazard function. Thus, it provides a more straightforward interpretation for the regression parameters based on the Cox proportional hazards model. In the paper, they adopted a spline smoothing approach to approximate the unknown link function $g(.)$, and then maximized the partial likelihood to estimate $\boldsymbol{\beta}$. Define B_j, $j = 1, ..., k$, are the B-spline basis functions (De Boor 1978). Then the unknown link function is expressed by $g(\boldsymbol{\beta}^T \boldsymbol{x}) = \sum_{j=1}^k \gamma_j \bar{B}_j(\boldsymbol{\beta}^T \boldsymbol{x})$, where $\bar{B}_j(u) = \int_0^u B_j(s)ds$, and γ_j, $j = 1, ..., k$, are unknown parameters. Both the unknown $\boldsymbol{\beta}$ and γ_j, $j = 1, ..., k$, are estimated by maximizing the following partial likelihood function

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[\frac{\exp\left\{\sum_{j=1}^k \gamma_j \bar{B}_j(\boldsymbol{\beta}^T \boldsymbol{X}_i(T_i))\right\}}{\sum_{l=1}^n I(T_l \geq T_i)\exp\left\{\sum_{j=1}^k \gamma_j \bar{B}_j(\boldsymbol{\beta}^T \boldsymbol{X}_l(T_i))\right\}}\right]^{\Delta_i}.$$

Similar to the linear model, the single index Cox model is also extended to the partially linear single index Cox model $\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp\{\alpha^T \boldsymbol{Z} + g(\boldsymbol{\beta}^T \boldsymbol{X})\}$, which is proposed by Sasieni (1992b). The estimation is based on maximizing a spline smoothed partial likelihood. Sasieni (1992a) provided the efficient score and information bound for $\boldsymbol{\beta}$.

2.3 Change-Point Models with Censored Data

2.3.1 Models

The change point analysis has been studied in the univariate Cox proportional hazards model. Therneau et al. (1990) first extended this model to include a continuous covariate dichotomized by a threshold value, which is defined as the change point. The model is written as

$$\lambda(t|\boldsymbol{Z}_i) = \lambda_0(t)\exp\left\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_i(t) + \beta_2 I(X_i > \zeta)\right\}, \tag{2.7}$$

where $\lambda_0(t)$ is an unknown baseline function, ζ is the unknown change point for which the covariate X_i has different effects for $X_i \le \zeta$ and $X_i > \zeta$, and $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1^T, \beta_2)^T$ is a vector of $J+1$ unknown parameters with $J = \dim(\boldsymbol{Z}_i(t))$. Therneau et al. (1990) applied the martingale residual plots to identify the existence of the change point. The partial likelihood function for n subjects with right censoring can be formulated as

$$L(\zeta,\boldsymbol{\beta}) = \prod_{i=1}^{n}\left(\frac{\exp\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_i(T_i) + \beta_2 I(X_i > \zeta)\}}{\sum_{l=1}^{n} I(T_l \ge T_i)\exp\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_l(T_i) + \beta_2 I(X_i > \zeta)\}}\right)^{\Delta_i},$$

where $T_i = \min(\widetilde{T}_i, C_i)$ with C_i being the censoring time assumed to be independent of $\widetilde{T}_i$ given the covariates, and $\Delta_i = I(\widetilde{T}_i \le C_i)$ is the failure indicator. To estimate the change point and regression coefficients, Luo and Boyett (1997) applied a two-step procedure to maximize the logarithm of partial likelihood function, which is defined as $l_n(\zeta,\boldsymbol{\beta}) \equiv \log\{L(\zeta,\boldsymbol{\beta})\}$. Since it is possible to have multiple ζ reaching the same maximum value, they choose the smallest one as the estimator for ζ, denoted by $\widehat{\zeta}$. The corresponding estimator for $\boldsymbol{\beta}$ is denoted as $\widehat{\boldsymbol{\beta}}$. Thus, $(\widehat{\zeta},\widehat{\boldsymbol{\beta}}) = \arg\max_{\zeta\in[\zeta_1,\zeta_2],\boldsymbol{\beta}} l_n(\zeta,\boldsymbol{\beta})$.

Luo and Boyett (1997) further proved the consistency of a resulting change-point estimator and its convergence rate of $1/n$. Later, Pons (2003) proved that the estimators of the regression parameters are asymptotically independent of the change-point estimator, and thus, $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$ still weakly converges to a multivariate normal distribution. In addition, they proved that the asymptotic distribution of the change-point estimator is a composite Poisson process. Kosorok and Song (2007) generalized this estimator to transformation models and established the asymptotic properties of this class of models, which includes the Cox model as a special case.

Except for Model (2.7), there are many other forms of change-point models based on Cox proportional hazards model in censored data. Liang et al. (1990), Luo (1996), Luo et al. (1997), and Pons (2002) discussed the change point at an unknown time for the lag effect of the covariates. The model is specified as

$$\lambda(t|\boldsymbol{Z}, \boldsymbol{X}) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}^T \boldsymbol{Z} + \boldsymbol{\theta}^T I(t \leq \tau)\boldsymbol{Z} + \boldsymbol{\gamma}^T \boldsymbol{X}\right\},$$

where τ is an unknown change point in time. There is a change in the effect of $\boldsymbol{Z}$ before and after the time τ from $\boldsymbol{\beta} + \boldsymbol{\theta}$ to $\boldsymbol{\beta}$. Gandy et al. (2005), Gandy and Jensen (2005), and Jensen and Lütkebohmert (2008) considered the Cox model with a smooth change in the regression coefficient. To address the cases, they assumed that the slopes are different for the covariates above and below the change point, which is given by

$$\lambda(t|\boldsymbol{Z}, X) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(t) + \beta_2 X_i + \beta_3 (X_i - \eta)^+\right\},$$

where there is a change of influence of the covariate X_i at η from β_2 to $\beta_2 + \beta_3$.

2.3.2 M out of n Bootstrap

Given the dataset of size n, the m out of n bootstrap approach is defined as sampling with replacement of size m, where $m \to \infty$, and $m/n \to 0$. The key step is how to chose the optimal m, which determines the consistency of the bootstrap estimators. Hall et al. (1995)

showed that the choices of m are varying in different contexts. For example, the optimal choice of m is $n^{1/3}$ for the setting of estimation of bias. However, the optimal choice of m becomes $n^{1/5}$ in the setting of constructing two-sided confidence intervals for an unknown parameters. Several data-driven approaches for choosing m have been proposed. Datta and McCormick (1995) proposed a jackknife scheme by drawing m samples with replacement from the dataset, which leaves out one sample in each sampling. The choice of m is the one which minimizes the function $L(m) = \sum_{i=1}^{n}(\widetilde{q}_{-i} - \hat{q})^2$, where $\widetilde{q}_{-i}$ is estimate of the ith draw from the dataset without the ith sample, and $\hat{q}$ is the estimate based on the full dataset. Bickel and Sakov (2005) and Bickel and Sakov (2008) proposed approaches to chose m for confidences bounds for extreme percentiles and extrema functions. Based on their approach, the desired m is selected as the maximum value that achieves the stable empirical distributions of the extrema function.

2.3.3 Tests

In practice, one important question is whether the change point exists. The null hypothesis is $H_0 : \beta_2 = 0$ in Model (2.7). Since the estimation of the change point relies on β_2 unequal to zero, the change point is not identifiable given $\beta_2 = 0$ under the null hypothesis. To handle it, in general, there are two testing methods in the change-point method literatures.

The maximum efficiency robust tests (MERT) are proposed by Gastwirth (1966) and Gastwirth (1985). For the set $\Gamma = \{\tau_1, \tau_2, ..., \tau_K\}$ within a certain range $[\tau_{\min}, \tau_{\max}]$, the maximum efficiency robust test is defined as $Q(\boldsymbol{\alpha})$ which $\sup_Q \inf_{\tau\epsilon\Gamma} \rho\left(Q(\boldsymbol{\alpha}), U(\tau)\right)$, where $Q(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \alpha_k U(\tau_k)$, $U(\tau_k)$ is the score statistics, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ...\alpha_K)$ is a vector of positive coefficients, and $\rho(.,.)$ is the correlation function. The optimal $\boldsymbol{\alpha}$ can be obtained by the quadratic programming (Gastwirth 1966). The supremum (SUP) tests are proposed by Davies (1977) and Davies (1987). The test statistics for the Supremum tests is defined as $\text{SUP} = \sup_{\tau\epsilon\Gamma} |U(\tau)|$. Later, Kosorok and Song (2007) proposed the Supremum tests with a weighted bootstrap to obtain critical values. Zucker et al. (2013) conducted extensive simulations to compare these two approaches. Based on their simulation results, the SUP tests are more powerful under different scenarios.

2.3.4 Optimization Methods

The optimization methods can be classified into two groups depending on whether the objective functions have derivatives or not. For the most commonly used Newton-Raphson and quasi-Newton methods, they completely rely on the derivatives of the objective function. They are not applicable if the derivatives do not exist (Gill et al. 1981). However, the genetic algorithm still works when the derivative information is not available (Sekhon and Mebane 1998, Mebane Jr et al. 2011).

Newton-Raphson Method: The Newton-Raphson method is proposed to solve the equations $\frac{\partial f(X)}{\partial X} = 0$. The formula is $X_{j+1} = X_j - \left(\frac{\partial^2 f(X_j)}{\partial X_j^2}\right)^{-1} \frac{\partial f(X_j)}{\partial X_j}$. The algorithm starts with an initial value X_0. If the initial guess of X_0 is close enough to the true value, it ensures the convergence of the algorithm. The local linear and quadratic convergences of the algorithm are proved in Kantorovitch (1939) and Kantorovich (1948), respectively.

quasi-Newton Method: The quasi-Newton method entirely depends on the derivatives. In order to avoid the computation of Hessian matrix, quasi-Newton method proposes to use an approximation for the Hessian matrix $B_j \approx \left(\frac{\partial^2 f(X_j)}{\partial X_j^2}\right)^{-1}$ to speed up the computation. Thus the formula is $X_{j+1} = X_j - B_j \frac{\partial f(X_j)}{\partial X_j}$. There are several iterative methods of choosing B_j. Wells (1965) and Fletcher and Powell (1963) proposed Davidon-Fletcher-Powell inverse-Hessian approximation as

$$B_{j+1}^* = B_j + \frac{(X_{j+1} - X_j)(X_{j+1} - X_j)^T}{(X_{j+1} - X_j)^T \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}}\right)} - \frac{B_j \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}}\right)\left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}}\right)^T B_j}{\left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}}\right)^T B_j \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}}\right)}.$$

Battiti (1990) updated the above algorithm and named as Broyden-Fletcher-Goldfarb-Shanno algorithm,which is specified as.

$$\begin{aligned}\widetilde{B}_{j+1} =& B^{*}_{j+1} + \phi_j \left\{ \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right)^T B_j \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right) \right\} \\ &\times \left(\frac{X_{j+1} - X_j}{(X_{j+1} - X_j)^T \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right)} - \frac{B_j \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right)}{\left\{ \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right)^T B_j \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right) \right\}} \right) \\ &\times \left(\frac{X_{j+1} - X_j}{(X_{j+1} - X_j)^T \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right)} - \frac{B_j \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right)}{\left\{ \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right)^T B_j \left(\frac{\partial f(X_j)}{\partial X_j} - \frac{\partial f(X_{j+1})}{\partial X_{j+1}} \right) \right\}} \right)^T\end{aligned}$$

Genetic Algorithm: The Genetic algorithm is a random search algorithm based on a group of heuristic rules. Holland (1975) first proposed this algorithm to mimic the natural process of evolution. The purpose of the algorithm is to search the best solution which optimize the target objective function. In each generation, there are a group of trial values which will result in different values of the target objective function. The genetic algorithm is to ensure the average of trial values in each generation is better than their predecessors. The basic heuristic rules include reproduction, mutation, and crossover to guarantee the evolution process (Goldberg 1989). Even though there might exists decreasing average trial values, Holland (1975) showed that the above average predecessor trial values are sampled at an exponential rate for inclusion in the successive generation. This algorithm is especially helpful when the derivatives do not exist. Mebane Jr et al. (2011) showed that this algorithm worked for the discontinuous step function proposed in Yao et al. (1999).

CHAPTER 3: PROPORTIONAL HAZARDS MODEL WITH A CHANGE POINT FOR CLUSTERED EVENT DATA

3.1 Introduction

The change point models have been widely applied in clinical research to decide the subgroup of participants who have a much higher risk for specific diseases. Change point effects have been observed in many medical studies for different traits, such as fasting plasma glucose in the Australian Diabetes Obesity and Lifestyle Study (AusDiab) (Tapp et al. 2006), midthigh muscle cross-sectional area in the COPD Study (Marquis et al. 2002), and leukocyte telomere length in the Strong Heart Family Study (SHFS) (Zhao et al. 2014). In these studies, risk of disease changes when a continuous risk factor passes a threshold value. For example, Zhao et al. (2014) investigated the association between leukocyte telomere length (LTL) and diabetes incidence in the SHFS. SHFS is a longitudinal family-based cohort study of cardiovascular disease, type 2 diabetes and their risk factors among American Indians residing in Oklahoma, Arizona and South/North Dakota. The authors found that participants with shorter LTL (lower quartile) have nearly two-fold increased risk for developing incident diabetes compared to those with longer LTL. Such a change point for LTL and diabetes incidence was also observed by Willeit et al. (2014). It is well-known that telomere length shortens progressively with each cell division until it reaches a threshold value beyond which cells enter into senescence or die, a phenomenon called “Hayflick limit". Even though the change point observed in these studies is consistent with the theory of “Hayflick limit", the precise change point location in LTL remains to be determined. Finding this threshold value is helpful to identify at-risk individuals and risk prediction. Thus, it is of great interest to develop a rigorous and comprehensive framework to conduct the change point analysis for survival data subject to censoring.

The change point analysis has been studied in the univariate Cox proportional hazards model. The Cox proportional hazards model (Cox 1972) was widely used to estimate the association between disease incidence and potential risk factors. Different change point models in the Cox proportional hazards model are proposed for various purposes. Liang et al. (1990), Luo (1996), Luo et al. (1997), and Pons (2002) discussed the change point at an unknown time for the lag effect of the covariates. Gandy et al. (2005), Gandy and Jensen (2005), and Jensen and Lütkebohmert (2008) considered the Cox model with a smooth change in the regression coefficient. They assumed that the slopes are different for the covariates above and below the change points. Another class of models assumes a non-smooth "jump" effect at an unknown threshold of a covariate (Luo and Boyett 1997, Pons 2003, Kosorok and Song 2007). Here, we focus on the change point analysis based on a non-smooth "jump" effect of a covariate. Maximum partial likelihood methods were proposed to estimate the change point and regression coefficients in this type of models. Luo and Boyett (1997) applied a two-step procedure to estimate the change point and proved the consistency of a resulting estimator. Later, Pons (2003) proved that this estimator asymptotically follows a composite Poisson process. Kosorok and Song (2007) generalized this estimator to transformation models and established the asymptotic properties of this class of models, which includes the Cox model as a special case. The change point analysis proposed for the univariate case cannot be applied directly to clustered survival data, because the proposed methods did not take into account the correlation between subjects within the same cluster.

In this paper, we focus on developing a Cox-type marginal hazards model (Lee et al. 1992) with a change point in a covariate for clustered survival data. The Cox marginal hazards model uses a pseudo-likelihood approach with a working independence assumption, while adjusting for the correlation by sandwich estimate in estimating the covariance matrix. The marginal hazards model is useful when the focus is on making inferences on the population average effect of risk factors on failure time. One major difficulty for the change point analysis in the Cox marginal hazards model is the complicated asymptotic distribution of the change point estimator for clustered data. With univariate survival data, Pons (2003) proved that the change point estimator asymptotically follows a composite Poisson process which depends on the change point locations across all the subjects. However, the existing theory for the univariate Cox model cannot be applied directly to the change point analysis in the Cox marginal hazards model. The asymptotic distribution of the change point estimator for the clustered data depends on the correlation structure of all the clusters. Considering the varying cluster size and all the possible situations of the covariate passing the true threshold across every member within each cluster, we prove that the asymptotic distribution of the proposed change point estimator follows a more complicated composite Poisson process.

The structure of this Section is as follows. In Section 3.2, we describe the estimation method based on a two-step procedure. We then provide an inference method based on m out of n bootstrap, and a testing procedure for the existence of a change point. In Section 3.3, we establish the consistency, convergence rates and asymptotic distributions of the proposed estimators. Simulation studies evaluating the small sample performance of the method are presented in Section 3.4. In Section 3.5, data from the Strong Heart Family Study are analyzed using our approach. The details of the proofs are given in the Proof of Lemma and Theorems.

3.2 Methods

3.2.1 Model and Parameter Estimation

Consider n independent and identically distributed (i.i.d) clusters with the ith cluster containing K_i subjects ($i = 1, ..., n$). For the jth subject in the ith cluster, $j = 1, ..., K_i$, let $\widetilde{T}_{ij}$ be the survival time, X_{ij} denote a one-dimensional continuous covariate whose effect on the response may have a change point, and $\boldsymbol{Z}_{ij}(t)$ denote other potentially time-dependent covariates whose effects could be different before or after X_{ij} passes the change point. In other words, the proportional hazards model with a change point assumes that the hazard rate function for $\widetilde{T}_{ij}$ given $\boldsymbol{W}_{ij}(t) \equiv \left(X_{ij}, \boldsymbol{Z}_{ij}^T(t)\right)^T$ takes a form

$$\lambda(t|\boldsymbol{W}_{ij}) = \lambda_0(t)\exp\left\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t) + \beta_2 I(X_{ij} > \zeta) + \boldsymbol{\beta}_3^T\boldsymbol{Z}_{ij}(t)I(X_{ij} > \zeta)\right\},$$

where $\lambda_0(t)$ is an unknown baseline function, ζ is the unknown change point for which the covariate X_{ij} has different effects for $X_{ij} \le \zeta$ and $X_{ij} > \zeta$, and $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1^T, \beta_2, \boldsymbol{\beta}_3^T)^T$ is a vector of $2J+1$ unknown parameters with $J = \dim(\boldsymbol{Z}_{ij}(t))$. Therefore, the proposed model implies that the effect of $\boldsymbol{Z}_{ij}$ is $\boldsymbol{\beta}_1$ when $X_{ij} \le \zeta$, and it becomes $(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_3)$ when $X_{ij} > \zeta$. Furthermore, the hazard ratio between $X_{ij} > \zeta$ and $X_{ij} \le \zeta$ is $\exp\{\beta_2 + \boldsymbol{\beta}_3^T\boldsymbol{Z}_{ij}(t)\}$ for given $\boldsymbol{Z}_{ij}(t)$.

If we define $r_\theta\{\boldsymbol{W}_{ij}(t)\} \equiv \boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t) + \beta_2 I(X_{ij} > \zeta) + \boldsymbol{\beta}_3^T\boldsymbol{Z}_{ij}(t)I(X_{ij} > \zeta)$ and $\boldsymbol{\theta} \equiv (\zeta, \boldsymbol{\beta}^T)^T$, then a marginal pseudo-partial likelihood function for n clusters with right censoring can be formulated as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i}\left(\frac{\exp\left[r_\theta\left\{\boldsymbol{W}_{ij}(T_{ij})\right\}\right]}{\sum_{l=1}^{n}\sum_{k=1}^{K_i} I(T_{lk} \ge T_{ij})\exp\left[r_\theta\left\{\boldsymbol{W}_{\boldsymbol{lk}}(T_{ij})\right\}\right]}\right)^{\Delta_{ij}},$$

where $T_{ij} = \min(\widetilde{T}_{ij}, C_{ij})$ with C_{ij} being the censoring time assumed to be independent of $\widetilde{T}_{ij}$ given the covariates W_{ij}, and $\Delta_{ij} = I(\widetilde{T}_{ij} \le C_{ij})$ is the failure indicator.

To estimate the model parameters, we propose to maximize the logarithm of the pseudo-likelihood function, which is defined as $l_n(\zeta, \boldsymbol{\beta}) \equiv \log\{L(\boldsymbol{\theta})\}$. Computationally, we adopt the following two-step procedure for maximization. For any fixed value of ζ in a pre-specified range $[\zeta_1, \zeta_2]$, we maximize the logarithm of the pseudo-likelihood function via the Newton-Raphson method, which yields the global maximum due to the strict concavity of $l_n(\zeta, \boldsymbol{\beta})$ for the given ζ. We thus obtain the profile function for ζ. In the second step, we apply a grid-search algorithm to find the optimal estimator for ζ. Since it is possible to have multiple

ζ reaching the same maximum value, we choose the smallest one as our final estimator for ζ, denoted by $\widehat{\zeta}$. The corresponding estimator for $\boldsymbol{\beta}$ is denoted as $\widehat{\boldsymbol{\beta}}$. Thus, $(\widehat{\zeta}, \widehat{\boldsymbol{\beta}}) = \arg\max_{\zeta \in [\zeta_1, \zeta_2], \boldsymbol{\beta}} l_n(\zeta, \boldsymbol{\beta})$.

3.2.2 Inference for Parameter Estimators

To make inference for ζ and $\boldsymbol{\beta}$, we utilize the asymptotic results which will be given in Section 3.3. In that section, we show that $\widehat{\zeta}$ and $\widehat{\boldsymbol{\beta}}$ are asymptotically independent and the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ remains the same regardless whether ζ is known or not. Thus, the inference for $\widehat{\boldsymbol{\beta}}$ can be carried out in a similar manner as the marginal proportional hazard model for clustered survival data, treating $\zeta = \widehat{\zeta}$ as fixed (c.f. Lee et al. 1992). However, the inference for $\widehat{\zeta}$ is challenging due to the intractable asymptotic distribution shown in Section 3.3. The bootstrap approach is commonly applied to generate the empirical distributions of the estimators with complicate asymptotic distributions (Efron and Tibshirani 1994). The usual bootstrap approach is to draw a sample of n with replacement from the dataset of n samples. Efron and Tibshirani (1986) demonstrated its performance in generating standard errors and confidence intervals under regular conditions. However, the usual bootstrap approach produces inconsistent estimators in some non-standard problems. Dümbgen (1993) and Shao (1994) demonstrated the failure of the usual bootstrap in non-differentiable objective functions or non-smooth statistics. In addition, Shao (1994) proposed a remedy of such situation by sampling a ratio of the size of the original dataset. Given the dataset of size n, the m out of n bootstrap approach is defined as sampling with replacement of size m, where $m \to \infty$, and $m/n \to 0$. Similar concepts are also proposed by Bickel et al. (2012) and Politis and Romano (1999). Such method is widely used in non-standard problems, such as non-differentiable objective functions (Huang et al. 1996, Chakraborty et al. 2013) and non-$n^{-1/2}$ asymptotics (Abrevaya and Huang 2005, Sen et al. 2010). In addition, Xu et al. (2014) proved the consistency of the m out of n bootstrap in the case of the Cox proportional hazards model with a change point.

For the m out of n bootstrap, several data-driven approaches for choosing m have been proposed (Hall et al. 1995; Lee 1999; Cheung et al. 2005; Bickel and Sakov 2005; Bickel and Sakov 2008). Among them, Bickel and Sakov (2008) proposed a method to select m for extrema functions. Based on their approach, the desired m is selected from a sequence of possible re-sampling sample sizes. The rule is to select the maximum sample size that achieves the minimum distance defined on supremum norm between two empirical distributions, which are based on any two adjacent re-sampling sample sizes. Thus, the selected m can achieve the stable empirical distributions of the proposed estimator. Hence, we adapt this algorithm to select m in the following way.

(1) Construct a sequence of the re-sampling sample sizes $m_j = [j \times \frac{n}{q}]$, where $j = q, q-1, ..., 1$, n/q is the interval between two adjacent re-sampling sample sizes, and $[a]$ is the largest integer no larger than a.

(2) For the m_j out of n bootstrap, the empirical cumulative distribution function for the change point estimator is constructed as follows:

$$R_{m_j}(x, \widehat{\zeta}) = \frac{1}{B} \sum_{b=1}^{B} I\left\{m_j\left(\widehat{\zeta}_{m_j}^{(b)} - \widehat{\zeta}\right) \leq x\right\},$$

where $\widehat{\zeta}$ is the change point estimator based on the full dataset, $\widehat{\zeta}_{m_j}^{(b)}$ is the change point estimator based on the dataset with m_j samples in the bth replication, $b = 1, 2..., B$, and B is the total number of bootstrap replications.

(3) The m will be selected as the maximum value which minimizes the supremum difference between two adjacent empirical cumulative distributions in the m_j sequence.

$$m = \max \arg \min_{m_j} \sup_{x} \left| R_{m_j}(x, \widehat{\zeta}) - R_{m_{j+1}}(x, \widehat{\zeta}) \right|$$

Based on the selected m, the m out of n bootstrap is to draw m samples with replacement out of the overall n samples. The standard error of the proposed estimator is estimated by the sample standard error based on B replicates divided by n/m. In addition, the equal-tailed 95% confidence intervals are generated as $\left[\widehat{\zeta} - \frac{Q_{\widehat{\zeta},0.95}}{n/m}, \widehat{\zeta} + \frac{Q_{\widehat{\zeta},0.95}}{n/m}\right]$, where $Q_{\widehat{\zeta},0.95}$ is the 95th quantile of the absolute value $\left|\widehat{\zeta} - \widehat{\zeta}_m^{(b)}\right|$ for the replicate $b = 1, 2..., B$. Both the standard error estimator and the confidence interval are adjusted by n/m, which corrects the over-estimated variance and wide confidence intervals based on the m out of n bootstrap.

3.2.3 Hypothesis Testing for the Change Point

In practice, one important question is whether the change point exists. The null hypothesis is specified as $H_0 : \beta_2 = 0, \boldsymbol{\beta}_3^T = \mathbf{0}$ in our proposed model. However, the change point is not identifiable given both β_2 and $\boldsymbol{\beta}_3$ are zero, because the estimation of the change point relies on either β_2 or $\boldsymbol{\beta}_3$ unequal to zero. To handle it, in general, there are two testing methods in the change point method literatures, which are the maximum efficiency robust tests (MERT) (Gastwirth 1966, Gastwirth 1985) and the supremum (SUP) tests (Davies 1977, Davies 1987, Kosorok and Song 2007). Zucker et al. (2013) conducted extensive simulations to compare these two approaches. Based on their simulation results, the SUP tests are more powerful under different scenarios. Here, we adopt the SUP type of test but rely on robust score statistics for the clustered survival time. Specifically, our test statistic is

$$\mathrm{SUP}_k = \sup_{\zeta \in [\zeta_1, \ldots, \zeta_k]} \boldsymbol{U}(\zeta)^T \boldsymbol{\Sigma}(\zeta)^{-1} \boldsymbol{U}(\zeta),$$

where $\boldsymbol{U}(\zeta) = \frac{\partial l_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}(\zeta) = \sum_{i=1}^n \sum_{j=1}^{K_i} \sum_{l=1}^{K_i} \boldsymbol{H_{ij}}(\zeta) \boldsymbol{H_{il}}(\zeta)^T$,

$$\begin{aligned} \boldsymbol{H_{ij}}(\zeta) &= \left\{ \widetilde{\boldsymbol{Z}}_{ij}(T_{ij}) - \frac{\boldsymbol{S}_n^{(1)}(T_{ij}; \zeta, \boldsymbol{\beta})}{S_n^{(0)}(T_{ij}; \zeta, \boldsymbol{\beta})} \right\} \\ &\quad - \sum_{s=1}^{n} \sum_{l=1}^{K_i} \frac{\Delta_{sl} \exp\{\boldsymbol{\beta}^T \widetilde{\boldsymbol{Z}}_{ij}(T_{sl})\}}{n S_n^{(0)}(T_{sl}; \zeta, \boldsymbol{\beta})} \left\{ \widetilde{\boldsymbol{Z}}_{ij}(T_{sl}) - \frac{\boldsymbol{S}_n^{(1)}(T_{sl}; \zeta, \boldsymbol{\beta})}{S_n^{(0)}(T_{sl}; \zeta, \boldsymbol{\beta})} \right\}, \end{aligned}$$

$\boldsymbol{S}_n^{(r)}(t;\zeta,\boldsymbol{\beta}) = \frac{1}{n}\left(\sum_{i=1}^{n}\sum_{j=1}^{K_i} Y_{ij}(t)\widetilde{\boldsymbol{Z}}_{ij}^{\otimes r}(t;\zeta)\exp\left[r_\theta\left\{\boldsymbol{W}_{ij}(t)\right\}\right]\right)$, $Y_{ij}(t) = I(T_{ij} \geq t)$, and $\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta) = (\boldsymbol{Z}_{ij}(t), I(X_{ij} > \zeta), \boldsymbol{Z}_{ij}(t)I(X_{ij} > \zeta))$ for $r = 0,1$. We propose to use permutation to generate the null distribution of the proposed test statistic. Under the null hypothesis, there is no change point effect on the response. Thus, we randomly shuffle the covariate X_{ij} for sufficient times. Then, we obtain the permutation distribution of the proposed test statistics. We reject the null hypothesis at a significance level of α if SUP_k is larger than the upper α-quantile of the permutation distribution.

3.3 Asymptotic Results

In this section, we establish the consistency and asymptotic distributions of the estimators for both the change point and the regression parameters. The following conditions are needed to establish the asymptotic properties of the estimators.

(C.1) The density of X_{ij} is assumed to be strictly positive, bounded and continuous in a neighborhood of ζ_0, denoted by $\mathscr{V}_0$.

(C.2) For any ζ in $\mathscr{V}_0$, the information matrix $I(\boldsymbol{\theta}) = \int_0^\tau v(t;\zeta,\boldsymbol{\beta})s^{(0)}(t;\zeta,\boldsymbol{\beta})\lambda_0(t)dt$ is positive definite, where $\boldsymbol{v}(t;\zeta,\boldsymbol{\beta}) = \boldsymbol{s}^{(2)}(t;\zeta,\boldsymbol{\beta})/s^{(0)}(t;\zeta,\boldsymbol{\beta}) - \left[\boldsymbol{s}^{(1)}(t;\zeta,\boldsymbol{\beta})/s^{(0)}(t;\zeta,\boldsymbol{\beta})\right]^{\otimes 2}$, $\boldsymbol{s}^{(r)}(t;\zeta,\boldsymbol{\beta}) = \mathrm{E}\left(\sum_{j=1}^{K_i} Y_{ij}(t)\widetilde{\boldsymbol{Z}}_{ij}^{\otimes r}(t;\zeta)\exp\left[r_\theta\left\{\boldsymbol{W}_{ij}(t)\right\}\right]\right)$, and $r = 0,1,2$. In addition, $\lambda_{\min}\left(\int_0^\tau \mathrm{E}\left[Y_{ij}(t)\left\{1,\boldsymbol{Z}_{ij}(t)\right\}^{\otimes 2}\middle| X_{ij} = \zeta_0\right]d\Lambda_0(t)\right) > 0$, where $\lambda_{\min}(\boldsymbol{A})$ is the smallest eigenvalue of any square matrix $\boldsymbol{A}$.

(C.3) There exists a convex and bounded neighborhood $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$ such that for $k = 0,1,2$, and $r = 1,2$, $\sup_{\zeta\in[\zeta_1,\zeta_2]}\mathrm{E}\left\{\sup_{t\in[0,\tau]}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left(\|\boldsymbol{Z}_{ij}(t)\|^k\exp\left[r_\theta\left\{\boldsymbol{W}_{ij}(t)\right\}\right]\right)^r\middle| X_{ij} = \zeta\right\} < \infty$.

(C.4) The random process $\sup_{t\in[0,\tau]}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{S}_n^{(r)}(t;\zeta,\boldsymbol{\beta}) - \boldsymbol{s}^{(r)}(t;\zeta,\boldsymbol{\beta})\right\|$ converges almost surely to zero, where $\boldsymbol{s}^{(r)}(t;\zeta,\boldsymbol{\beta}) < \infty$, and $r = 0,1,2$. When $r = 0$, $\boldsymbol{s}^{(0)}(t;\zeta,\boldsymbol{\beta})$ is bounded away from zero.

(C.5) $\sup_{t\in[0,\tau]}\lambda_0(t) < \infty$, and $P(Y_{ij}(t) = 1) > 0$ for all $t \in [0,\tau]$.

(C.6) $P(K_i \leq k_0) = 1$, where $1 \leq k_0 < \infty$.

(C.1) and (C.2) are needed for the identifiability of the change point and regression coefficients. (C.3) is required to guarantee that $\boldsymbol{s}^{(r)}(t;\zeta,\boldsymbol{\beta})$ are uniformly continuous on $\boldsymbol{\Theta}$ and $t \in [0,\tau]$. As $\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|$ tends to zero, $\boldsymbol{s}^{(r)}(t;\zeta',\boldsymbol{\beta}') - \boldsymbol{s}^{(r)}(t;\zeta,\boldsymbol{\beta}) = O(|\zeta-\zeta'| + \|\boldsymbol{\beta}-\boldsymbol{\beta}'\|)$ uniformly on $t \in [0,\tau]$. (C.4) and (C.5) are required to guarantee that $n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{K_i} M_{ij}(t)$ converges in distribution to a Gaussian process with mean zero, where $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s)\exp\left[r_{\theta_0}\{\boldsymbol{W}_{ij}(s)\}\right] d\Lambda_0(s)$ and $N_{ij}(t) = \Delta_{ij} I(T_{ij} \le t)$. (C.6) is the standard condition for clustered failure time data to ensure the finite cluster size.

Our first two theorems establish the consistency and convergence rates of the estimators.

Theorem 3.3.1. *Under conditions (C.1)-(C.6), $\widehat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}_0$ in the neighborhood $\boldsymbol{\Theta}$ as $n \to \infty$.*

Theorem 3.3.2. *Under conditions (C.1)-(C.6),*

$$\lim_{A\to\infty}\lim_{n\to\infty} P_0\left(n|\widehat{\zeta}-\zeta_0| > A\right) = 0,$$

$$\lim_{A\to\infty}\lim_{n\to\infty} P_0\left(n^{1/2}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\| > A\right) = 0.$$

Let $\boldsymbol{\theta}_{n,u} = (\zeta_{n,u}, \boldsymbol{\beta}_{n,u}^T)^T$, $\zeta_{n,u} = \zeta_0 + n^{-1}u_1$, and $\boldsymbol{\beta}_{n,u} = \boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{u}_2$, where u_1 and $\boldsymbol{u}_2$ satisfy that $\left(|u_1| + \|\boldsymbol{u}_2\|^2\right)^{1/2} \le n^{1/2}\epsilon$. To obtain the asymptotic distributions of the estimators, we first need the expansions of $\{l_n(\boldsymbol{\theta}_{n,u}) - l_n(\boldsymbol{\theta}_0)\}$. In Theorem 3.3.3, we prove that $l_n(\boldsymbol{\theta}_{n,u}) - l_n(\boldsymbol{\theta}_0) = Q_n(u_1) + \boldsymbol{u}_2^T\widetilde{\boldsymbol{l}}_n - \frac{1}{2}\boldsymbol{u}_2^T\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{u}_2 + o_p(1)$, where $Q_n(u_1)$ and $\widetilde{\boldsymbol{l}}_n$ are defined as

$$\begin{aligned} Q_n(u_1) =& \sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_{ij}\Big[\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(T_{ij})\right\}\left\{I(\zeta_0 \ge X_{ij} > \zeta_{n,u}) - I(\zeta_{n,u} \ge X_{ij} > \zeta_0)\right\} \\ & - \frac{S_n^{(0)}(T_{ij};\zeta_{n,u},\boldsymbol{\beta}_0) - S_n^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)}\Big], \\ \widetilde{\boldsymbol{l}}_n =& n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\int_0^{\tau}\left(\widetilde{\boldsymbol{Z}}_{ij}(t,\zeta_0) - \frac{\boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right)dM_{ij}(t), \end{aligned}$$

where $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s)\exp\left[r_{\theta_0}\left\{\boldsymbol{W}_{ij}(s)\right\}\right]d\Lambda_0(s)$ and $N_{ij}(t) = \Delta_{ij} I(T_{ij} \le t)$.

For the cluster with m subjects, we define the set $A^{+}_{m1} = \{K_i = m, \text{only one } X_{ij} > \zeta_0, \text{all the other } X_{i1}, ..., X_{ij-1}, X_{ij+1}, ..., X_{im} \le \zeta_0\}$, where $m = 1, ..., K$, and $K = \max(K_i)$ is the maximum cluster size. We further define the element of A^{+}_{m1} as $A^{k+}_{m1} = \{X_{ik} > \zeta_0, X_{i1}, ..., X_{ik-1}, X_{ik+1}, ..., X_{im} \le \zeta_0\}$, where $k = 1, 2, ..., m$. Similarly, A^{-}_{m1} and A^{k-}_{m1} are defined for the situations when only one $X_{ij} \le \zeta_0$. Let $V^{+}_{mk,l}$ and $V^{-}_{mk,l}$ be independent sequences of identically and independently distributed random variables with the characteristic functions

$$\mathrm{E}\left\{\exp\left(itV^{+}_{mk,l}\right)\right\} = \mathrm{E}\left\{\exp\left(itq_s\eta^{(1)}_{ns,ik}\right)\Big|K_i = m, A^{k+}_{m1}, X_{ik} = \zeta_0^{+}\right\},$$
$$\mathrm{E}\left\{\exp\left(itV^{-}_{mk,l}\right)\right\} = \mathrm{E}\left\{\exp\left(itq_s\eta^{(1)}_{ns,ik}\right)\Big|K_i = m, A^{k-}_{m1}, X_{ik} = \zeta_0^{-}\right\},$$

where $l \ge 1$, q_s is an arbitrary constant, and $\eta^{(1)}_{ns,ik} = -\Delta_{ik}\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ik}(T_{ik})\right\}$ $-\int_0^\tau Y_{ik}(t)\exp\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{ik}(t)\}\left[1-\exp\{\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ik}(t)\}\right]d\Lambda_0(t)$. We further denote v^{+}_{mk} and v^{-}_{mk} to be the real jump processes such that $v^{+}_{mk} = 0$ on R^{-} and $v^{-}_{mk} = 0$ on R^{+}. We further denote $v^{+}_{mk}(s)$ to be a Poisson variable with the parameter $sp(m)f_{X_{ik}}(\zeta_0^{+})P(A^{k+}_{m1}|K_i = m, A^{+}_{m1})$, and $v^{-}_{mk}(s)$ to be a Poisson variable with parameter $sp(m)f_{X_{ik}}(\zeta_0^{-})P(A^{k-}_{m1}|K_i = m, A^{-}_{m1})$, where $p(m)$ is probability of the cluster with m subjects, $P(A^{k+}_{m1}|K_i = m, A^{+}_{m1})$ is the conditional probability of A^{k+}_{m1} given A^{+}_{m1}, and $f_{X_{ik}}(\zeta_0^{+})$ is marginal density function of X_{ij} at ζ_0^{+}. Similarly, we define $P(A^{k-}_{m1}|K_i = m, A^{-}_{m1})$ and $f_{X_{ik}}(\zeta_0^{-})$ for $X_{ik} \le \zeta_0$. Let $Q(s) \equiv Q^{+}(s) - Q^{-}(s)$, where

$$Q^{+}(s) \equiv \sum_{m=1}^{K}\sum_{k=1}^{m}\sum_{0\le l\le v^{+}_{mk}(s)} V^{+}_{mk,l} \text{ and } Q^{-}(s) \equiv \sum_{m=1}^{K}\sum_{k=1}^{m}\sum_{0\le l\le v^{-}_{mk}(s)} V^{-}_{mk,l}.$$

Thus, we further establish the following Theorem 3.3.3.

Theorem 3.3.3. *Under conditions (C.1)-(C.6),* $n(\widehat{\zeta}-\zeta_0)$ *and* $n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ *are asymptotically independent. Furthermore,* $n(\widehat{\zeta}-\zeta_0)$ *converges in distribution to* $\arg\max Q(s)$. $n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ *converges weakly to a Gaussian variable* $N(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1})$.

3.4 Simulation Studies

We conducted simulation studies to evaluate the performance of our proposed method. Our first set of studies was designed to assess the bias of the estimators and the coverage rate of the confidence interval. We considered one covariate $Z \sim$ N(1,4) and one change point variable $X \sim$ Uniform(0, 2) with the true change point at 1. We generated the marginal survival times $\widetilde{T}_{ij}$ under the proportional hazards model $\Lambda(t|X,Z) = t\exp\{\beta_1 Z + \beta_2 I(X > \zeta) + \beta_3 Z I(X > \zeta)\}$, where $(\beta_1, \beta_2, \beta_3) = (-1, -0.5, 0.6)$. The censoring time follows Uniform(0, 800). The correlated failure times were generated in the same way as in Cai and Shen (2000), which is a multivariate extension of the Clayton and Cuzick (1985) method. The conditional cumulative density function of the survival time for the jth subject in the ith cluster is

$$F_i(\widetilde{T}_{ij}|\widetilde{T}_{i1},...,\widetilde{T}_{i(j-1)}) = 1 - \left\{\sum_{h=1}^{j} S_{ih}(\widetilde{T}_{ih})^{-1/\gamma} - (j-1)\right\}\left\{\sum_{h=1}^{j} S_{ih}(\widetilde{T}_{ih})^{1/\gamma} - (j-2)\right\}^{\gamma+j-1},$$

where $S_{ij}(t) = P(\widetilde{T}_{ij} > t)$ is the marginal survival function, γ indicates the degree of dependence between $\widetilde{T}_{ij}$ and $\widetilde{T}_{ih} (h = 1, ..., j-1)$. The Kendall's tau coefficient can be expressed as $\tau_K = \frac{1}{2\gamma+1}$, where $\gamma = 0.25$ or 1.5 indicates strong or moderate positive dependence within each cluster. We considered both the small cluster sizes with 2 or 2–5 subjects and the large cluster size with 20 subjects. The number of clusters is 100 or 200. The searching range of the change point is $[0.5, 1.5]$. To select m for each simulation, we considered q to be 5 or 10. The number of grids is 500 for the small cluster size, and 1000 for the large cluster size. All results are based on 500 replications and each m out of n bootstrap consists of 150 replicates.

In Table 3.1, the proposed method provides approximately unbiased estimates for the change point ζ, and the m out of n bootstrap generates proper coverage rates. When the cluster size and/or the number of clusters increase, the bias of the change point estimate and the variance estimates decrease. For the m out of n bootstrap (results not shown), the choices of m are not influenced by the dependence (moderate vs high dependence) within the clusters. However, the choice of m increases as the number of clusters increases. The results also show that the estimates for the regression coefficients ($\boldsymbol{\beta}$) are approximately unbiased and the confidence intervals using normal approximation generate proper coverage rates for both highly and moderately correlated clusters.

Our second set of simulation studies were aimed at comparing type I error and power of the SUP_1, SUP_3, and SUP_{11} tests under varying scenarios. We examine the performance of these tests with the highly/moderately correlated clusters of size 2 or 2 to 5 with 100 clusters, and clusters of size 20 with 50 clusters. We set the true change point to be 1 or 0.75, the grid for the SUP_1 test to be 1, the grids for the SUP_3 test to be $\{0.5, 1, 1.5\}$, and the grids for the SUP_{11} test to be $\{0.5, 0.6, 0.7, ..., 1.4, 1.5\}$. Thus, the SUP_1 test is the optimal test if the true change point is the same as the pre-assumed change point 1. The regression coefficients (β_{20}, β_{30}) are set to $(0, 0)$ for type I error, and $(0.2, -0.3)$, $(0.2, -0.23)$, or $(0.2, -0.12)$ for power under the cluster size 2, 2 to 5 and 20, respectively. The results for type I error and power are based on 10000 and 1000 replicates, respectively. All the other specifications are the same as the first set of simulations.

Table 3.2 shows that type I errors of all three tests are close to 0.05, regardless of where the true change point is. For the power, the performance of the supremum tests is determined by the number of grids and the minimum distance between the grids and the true change point. The minimum distance is calculated as the smallest absolute difference between the true change point and the grids. For example, when the true change point is 1, the minimum distances for all three tests are 0. In this case, the SUP_1 test is the optimal test with the highest power, while the SUP_{11} test has the lowest power. This finding is expected because the SUP_1 test is only evaluated once, while the SUP_{11} test is evaluated on more grids. When the true value is 0.75, these tests have different minimum distances. In this case, the SUP_{11} test is the most powerful test among the three tests because it has the smallest minimum

Table 3.1: Simulation Results for the Change Point and Regression Parameters.

Correlation	Cluster Size	# of Clusters	Bias ($\widehat{\zeta}$) ($\times 10^{-2}$)	SSD($\widehat{\zeta}$) ($\times 10^{-2}$)	95% CI($\widehat{\zeta}$)	Length($\widehat{\zeta}$) ($\times 10^{-2}$)	Parameters	Bias	SSD	SEE	95% CI
High	2	100	-8.85	1.65	0.93	8.27	$\widehat{\beta}_1$	-0.020	0.098	0.089	0.920
							$\widehat{\beta}_2$	-0.028	0.199	0.192	0.940
							$\widehat{\beta}_3$	0.039	0.172	0.160	0.926
		200	-3.14	0.82	0.95	4.21	$\widehat{\beta}_1$	-0.005	0.068	0.062	0.930
							$\widehat{\beta}_2$	-0.005	0.138	0.135	0.948
							$\widehat{\beta}_3$	0.011	0.127	0.113	0.920
	2–5	100	-4.20	0.89	0.94	4.75	$\widehat{\beta}_1$	-0.013	0.088	0.079	0.914
							$\widehat{\beta}_2$	-0.021	0.173	0.156	0.924
							$\widehat{\beta}_3$	0.028	0.169	0.147	0.916
		200	-1.25	0.50	0.93	2.35	$\widehat{\beta}_1$	-0.009	0.064	0.058	0.936
							$\widehat{\beta}_2$	-0.013	0.119	0.113	0.940
							$\widehat{\beta}_3$	0.016	0.116	0.107	0.924
	20	100	-0.49	0.17	0.93	0.82	$\widehat{\beta}_1$	-0.013	0.072	0.066	0.926
							$\widehat{\beta}_2$	-0.019	0.111	0.107	0.934
							$\widehat{\beta}_3$	0.025	0.140	0.130	0.926
		200	-0.12	0.09	0.92	0.40	$\widehat{\beta}_1$	-0.003	0.051	0.047	0.944
							$\widehat{\beta}_2$	-0.005	0.084	0.077	0.944
							$\widehat{\beta}_3$	0.006	0.101	0.093	0.932
Moderate	2	100	-8.72	1.82	0.93	8.38	$\widehat{\beta}_1$	-0.016	0.089	0.081	0.914
							$\widehat{\beta}_2$	-0.017	0.192	0.184	0.936
							$\widehat{\beta}_3$	0.030	0.152	0.141	0.924
		200	-3.03	0.81	0.94	4.22	$\widehat{\beta}_1$	-0.007	0.061	0.057	0.926
							$\widehat{\beta}_2$	-0.004	0.129	0.130	0.938
							$\widehat{\beta}_3$	0.011	0.112	0.100	0.932
	2–5	100	-4.38	0.88	0.94	4.87	$\widehat{\beta}_1$	-0.011	0.071	0.064	0.892
							$\widehat{\beta}_2$	-0.017	0.153	0.141	0.936
							$\widehat{\beta}_3$	0.023	0.131	0.114	0.900
		200	-1.19	0.50	0.94	2.38	$\widehat{\beta}_1$	-0.005	0.050	0.047	0.952
							$\widehat{\beta}_2$	-0.005	0.104	0.101	0.944
							$\widehat{\beta}_3$	0.008	0.085	0.083	0.938
	20	100	-0.52	0.18	0.92	0.82	$\widehat{\beta}_1$	-0.008	0.046	0.041	0.892
							$\widehat{\beta}_2$	-0.011	0.077	0.074	0.922
							$\widehat{\beta}_3$	0.015	0.086	0.077	0.904
		200	-0.09	0.09	0.92	0.40	$\widehat{\beta}_1$	0.001	0.033	0.030	0.924
							$\widehat{\beta}_2$	0.001	0.061	0.054	0.922
							$\widehat{\beta}_3$	-0.001	0.063	0.056	0.930

NOTE: SSD and SEE stand for sample standard deviation and standard error estimate, respectively.

distance. The SUP_1 test has a slightly higher power than SUP_3, since both tests have the same minimum distance and the SUP_3 test is evaluated on a larger set. Consequently, the power of the supremum test increases if the minimum distance decreases. Given the same minimum distance, the tests based on a smaller set of grids have a slightly higher power.

Table 3.2: Type I Error and Power for SUP Tests for the Existence of the Change Point

Correlation	ζ_0	Cluster Size	Number of Clusters	(β_{20}, β_{30})	SUP_1(Optimal)	SUP_3	SUP_{11}
High	1	2	100	$(0, 0)$	0.051	0.051	0.050
				$(0.2, -0.3)$	0.892	0.818	0.812
		2–5	100	$(0, 0)$	0.056	0.055	0.055
				$(0.2, -0.23)$	0.928	0.854	0.840
		20	50	$(0, 0)$	0.050	0.047	0.047
				$(0.2, -0.12)$	0.904	0.816	0.800
	0.75	2	100	$(0, 0)$	0.051	0.051	0.050
				$(0.2, -0.3)$	0.678	0.630	0.760
		2–5	100	$(0, 0)$	0.056	0.055	0.055
				$(0.2, -0.23)$	0.688	0.652	0.796
		20	50	$(0, 0)$	0.050	0.047	0.047
				$(0.2, -0.12)$	0.666	0.650	0.742
Moderate	1	2	100	$(0, 0)$	0.049	0.050	0.051
				$(0.2, -0.3)$	0.908	0.830	0.822
		2–5	100	$(0, 0)$	0.049	0.050	0.049
				$(0.2, -0.23)$	0.916	0.844	0.826
		20	50	$(0, 0)$	0.049	0.049	0.050
				$(0.2, -0.12)$	0.896	0.836	0.842
	0.75	2	100	$(0, 0)$	0.049	0.050	0.051
				$(0.2, -0.3)$	0.658	0.632	0.774
		2–5	100	$(0, 0)$	0.049	0.050	0.049
				$(0.2, -0.23)$	0.704	0.664	0.794
		20	50	$(0, 0)$	0.049	0.049	0.050
				$(0.2, -0.12)$	0.724	0.676	0.754

3.5 Application to the Strong Heart Family Study

The SHFS recruited 3665 American Indians (aged 15 and older) from 94 extended families in three geographic areas: Arizona, Oklahoma, and Dakota. Each participant attended clinical and physical examinations at baseline (2001-2003) and 5-year follow-up (2006-2009). There are 2315 participants free of diabetes at baseline, among whom 292 developed incident diabetes by the end of 5-year follow-up (median survival time=5.4 years). Zhao et al. (2014) used a trial-and-error approach and observed that those individuals with LTL less than the 25th percentile had a significantly higher risk of developing new diabetes than the other individuals. Here, we took a more systematic approach to identify the change point in LTL for diabetes incidence.

We included LTL with an unknown change point to be estimated, gender, age, body mass index (BMI) (<25 kg/m^2, [25, 30) kg/m^2, and ≥ 30 kg/m^2), fasting glucose, total triglycerides, and their interactions with the dichotomized LTL (long vs. short) as predictors in the Cox marginal hazards model. First of all, we applied the proposed supremum test with the robust score statistics to verify the existence of the change point. We set the grids for the supremum test to be $\{0.5, 0.9, 1.3\}$, which correspond to the lower 5% quantile, median, and upper 5% quantile of LTL, respectively. The p-value is 0.002, which is highly significant. This indicates the existence of a change point in LTL for diabetes incidence. We next applied the two-step procedure to estimate the change point and the m out of n bootstrap with 500 replicates to generate the 95% confidence interval of the change point. The estimated change point is 0.870 and its 95% confidence interval is $[0.834, 0.907]$. Only the interaction between the change point of LTL and total triglycerides is statistically significant (p=0.036). We removed the non-significant interaction terms and presented the final model as Model 1 in Table 3.3. The marginal test for the effect of total triglycerides among the participants with LTL larger than the change point is highly significant with p-value < 0.001. For this group of participants, the increase in the level of total triglycerides results in an increase in the risk of developing incident diabetes. In contrast, the marginal effect of total triglycerides among the participants with LTL less than the change point is not significant (p=0.583). The hazard ratio of diabetes for shorter LTL ($< \zeta$) compared to longer LTL given the mean total triglycerides (147 mg/dL) is 2.476 [1.866, 3.285]. We verified proportional hazard assumptions for all covariates in Figure 3.1. For categorical variables (leukocyte telomere length, gender, and BMI), we generated plots of log of negative log of survival functions versus time, which show parallel trends between different levels for each covariate. For continuous variables (age, fasting glucose, and total triglycerides), the scattered plots show that the Schoenfeld residuals based on Model 1 in Table 3.3 are evenly distributed on both sides of the reference line, suggesting that the proportional hazards assumptions are satisfied for all predictors.

As mentioned before, Zhao et al. (2014) used a trial-and-error approach to find the change point. After trying different cutpoints, they located the change point somewhere near the

Table 3.3: Analysis Results Based on the Strong Heart Family Study: Model 1 ($\widehat{\zeta}$ = 0.870[0.834, 0.907]) and Model 2 ($\widehat{\zeta}_{ad-hoc}$ = 0.872).

	Model 1			Model 2		
Parameter	Estimate	SE	p-value	Estimate	SE	p-value
TOTAL TRIGLYCERIDES (mg/dL)	-0.001	0.001	0.583	0.001	0.001	0.136
GENDER	-0.333	0.115	0.004	-0.348	0.121	0.004
AGE	-0.002	0.005	0.723	-0.001	0.005	0.838
BMI [25, 30)	0.329	0.335	0.326	0.341	0.334	0.308
BMI ($\geq$ 30)	1.100	0.342	0.001	1.126	0.343	0.001
FASTING GLUCOSE (mg/dL)	0.068	0.006	< 0.001	0.066	0.006	< 0.001
TELOMERE LENGTH($> \zeta$)	-1.334	0.270	< 0.001	-0.768	0.146	< 0.001
TELOMERE LENGTH($> \zeta$)$\times$TOTAL TRIGLYCERIDES	0.003	0.001	0.036			

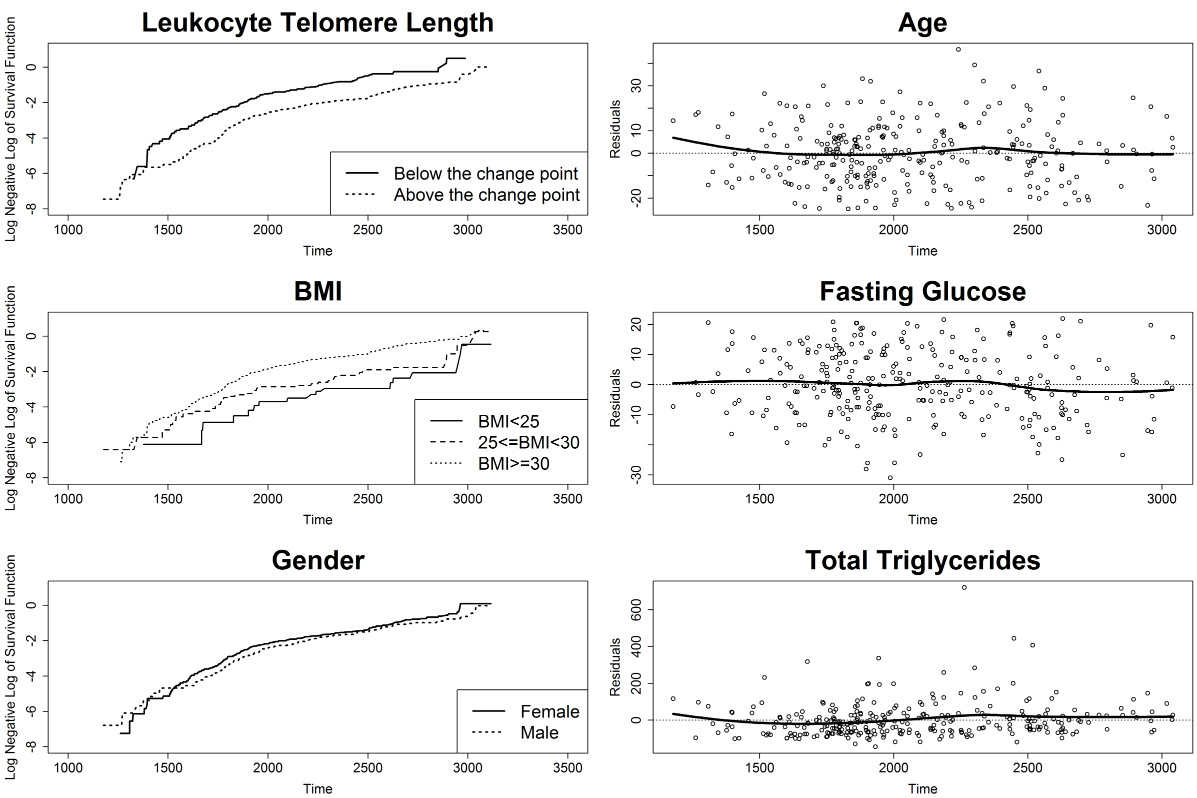

Figure 3.1: Diagnostic Plots. The log of negative log of survival functions versus time are plotted for leukocyte telomere length, gender, and BMI. Schoenfeld residuals are plotted for age, fasting glucose, and total triglycerides.

first quartile (0.872). Their results are presented under Model 2 in Table 3.3. Although the ad-hoc estimate of the change point is very close to our estimate, their approach was not able to detect the interaction between the change point and total triglycerides. Thus, it could not differentiate the effect of total triglycerides on developing incident diabetes among the short and long LTL participants. Based on this ad-hoc estimate, total triglycerides did not have a significant effect on developing incident diabetes for both short and long LTL participants (p=0.136). In addition, the ad-hoc method cannot provide a confidence interval for the change point estimate. In contrast, our approach can estimate the change point and corresponding 95% CI.

3.6 Discussion

Change point effects are commonly seen in regression problems. Although a number of approaches have been developed to estimate the change point in linear regression and the univariate Cox model, no research has been done for clustered survival data. In this paper, we developed for the first time a two-step approach to estimate the change point and a testing procedure to verify the existence of a change point for clustered survival data. We developed an adaptive m out of n bootstrap to construct the confidence interval and provide an easy way to determine the appropriate m. We proved the asymptotic properties of the proposed change point estimator. As shown in our simulation studies, the estimator is approximately unbiased and its confidence interval has a good coverage rate.

We applied our methods to estimate the change point of LTL for diabetes incidence in the SHFS. Because telomere length is genetically determined (Zhu et al. 2013), it is likely that the change point is racial or ethnic specific. Thus, it will be of interest to investigate the change point of LTL in other ethical groups. In addition, the change point of LTL is disease-specific. The estimated change point for LTL may be different for diabetes from that for other diseases, such as carotid atherosclerosis. We can apply our methods to identify the change point of LTL for other diseases in future studies.

Here, we consider the situation that the change point exists in one continuous variable. In reality, the change point may depend on multiple covariates, such as the single index $\boldsymbol{\beta}^T X$. In the change point $I(\boldsymbol{\beta}^T X > \zeta)$, the parameters $\boldsymbol{\beta}$ and ζ are identifiable only under certain conditions. The current estimation and inference procedures cannot be directly extended to the single index model. The computation time for grid search increases geometrically as the number of covariates increases. Thus, it is essential to devise more efficient algorithms to estimate the change point for such single index model in future work.

3.7 Proof of Lemma and Theorems

For convenience, we define

$$\boldsymbol{s}^{(r)+}(t;\zeta,\boldsymbol{\beta}) = \mathrm{E}\left[\sum_{j=1}^{K_i} Y_{ij}(t)I(X_{ij} > \zeta)\boldsymbol{Z}_{ij}^{\otimes r}(t)\exp\left\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t) + \beta_2 + \boldsymbol{\beta}_3^T\boldsymbol{Z}_{ij}(t)\right\}\right],$$
$$\boldsymbol{s}^{(r)-}(t;\zeta,\boldsymbol{\beta}) = \mathrm{E}\left[\sum_{j=1}^{K_i} Y_{ij}(t)I(X_{ij} \le \zeta)\boldsymbol{Z}_{ij}^{\otimes r}(t)\exp\left\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t)\right\}\right],$$
$$\boldsymbol{s}^{(r)}(t;\zeta,\boldsymbol{\beta}) = \mathrm{E}\left(\sum_{j=1}^{K_i} Y_{ij}(t)\widetilde{\boldsymbol{Z}}_{ij}^{\otimes r}(t;\zeta)\exp\left[r_\theta\{\boldsymbol{W}_{ij}(t)\}\right]\right),$$
$$\boldsymbol{S}_n^{(r)}(t;\zeta,\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i} Y_{ij}(t)\widetilde{\boldsymbol{Z}}_{ij}^{\otimes r}(t;\zeta)\exp\left[r_\theta\{\boldsymbol{W}_{ij}(t)\}\right],$$

where $\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta) = \left(\boldsymbol{Z}_{ij}^T(t), I(X_{ij} > \zeta), \boldsymbol{Z}_{ij}^T(t)I(X_{ij} > \zeta)\right)^T$, and $r = 0, 1$.

Proof of Theorem 3.3.1

We first show that $G_n(\boldsymbol{\theta}) = n^{-1}\{l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta}_0)\}$ converges uniformly to $G(\boldsymbol{\theta})$ in probability, where $G(\boldsymbol{\theta})$ is defined as

$$\int_0^\tau (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T\boldsymbol{s}^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0) - s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)\log\frac{s^{(0)}(t;\zeta,\boldsymbol{\beta}_0)}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}d\Lambda_0(t) + g(\boldsymbol{\theta})$$

with $g(\boldsymbol{\theta}) = \int_0^\tau \beta_2\left\{s^{(0)-}(t;\zeta_0,\boldsymbol{\beta}_0) - s^{(0)-}(t;\zeta,\boldsymbol{\beta}_0)\right\} + \boldsymbol{\beta}_3^T\left\{\boldsymbol{s}^{(1)-}(t;\zeta_0,\boldsymbol{\beta}_0) - \boldsymbol{s}^{(1)-}(t;\zeta,\boldsymbol{\beta}_0)\right\}$
$- \beta_2\left\{s^{(0)+}(t;\zeta_0,\boldsymbol{\beta}_0) - s^{(0)+}(t;\zeta,\boldsymbol{\beta}_0)\right\} - \boldsymbol{\beta}_3^T\left\{\boldsymbol{s}^{(1)+}(t;\zeta_0,\boldsymbol{\beta}_0) - \boldsymbol{s}^{(1)+}(t;\zeta,\boldsymbol{\beta}_0)\right\}d\Lambda_0(t)$.

To this end, we write $G_n(\boldsymbol{\theta})$ as

$$
\begin{aligned}
G_n(\boldsymbol{\theta}) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_{ij}\left[r_\theta\left\{\boldsymbol{W}_{ij}(T_{ij})\right\} - r_{\theta_0}\left\{\boldsymbol{W}_{ij}(T_{ij})\right\} - \log\frac{S_n^{(0)}(T_{ij};\zeta,\boldsymbol{\beta})}{S_n^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)}\right] \\
&= (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\left\{\boldsymbol{M}_n^{(1)}(\tau) + \int_0^\tau \boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)d\Lambda_0(t)\right\} - n^{-1}\int_0^\tau \log\frac{S_n^{(0)}(t;\zeta,\boldsymbol{\beta})}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}d\bar{N}_n(t) \\
&+ n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_{ij}\{\beta_2+\boldsymbol{\beta}_3^T\boldsymbol{Z}_{ij}(T_{ij})\}\{I(\zeta_0\ge X_{ij}>\zeta)-I(\zeta\ge X_{ij}>\zeta_0)\},
\end{aligned} \tag{3.8}
$$

where $\boldsymbol{M}_n^{(1)}(\tau) = n^{-1}\left\{\sum_{i=1}^n\sum_{j=1}^{K_i}\int_0^\tau \widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_0)dN_{ij}(t)\right\} - \int_0^\tau \boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)d\Lambda_0(t)$, and $\bar{N}_n(t) = \sum_{i=1}^n\sum_{j=1}^{K_i}N_{ij}(t)$. Since $\mathrm{E}\left\{\boldsymbol{M}_n^{(1)}(\tau)\right\}=0$ and $\mathrm{E}\left[\left\{\boldsymbol{M}_n^{(1)}(\tau)\right\}^2\right]$ are bounded, $\boldsymbol{M}_n^{(1)}(\tau)$ converges in probability to zero by the law of large numbers. From (C.4), $\int_0^\tau \boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)d\Lambda_0(t)$ converges to $\int_0^\tau \boldsymbol{s}^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)d\Lambda_0(t)$ in probability. Therefore, the first term in (3.8) converges uniformly to $\int_0^\tau(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{s}^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)d\Lambda_0(t)$ in probability. For the second term in (3.8), $S_n^{(0)}(t;\zeta,\boldsymbol{\beta})$ converges uniformly to $s^{(0)}(t;\zeta,\boldsymbol{\beta})$ and $\int_0^t \frac{1}{nS_n^{(0)}(s;\zeta_0,\boldsymbol{\beta}_0)}$ $d\bar{N}_n(s)$ converges uniformly to $\Lambda_0(t)$ in the $BV[0,\tau]$, where $BV[0,\tau]$ is bounded variation over $[0,\tau]$. Thus,

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|n^{-1}\int_0^\tau \log\frac{S_n^{(0)}(t;\zeta,\boldsymbol{\beta})}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}d\bar{N}_n(t) - \int_0^\tau s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)\log\frac{s^{(0)}(t;\zeta,\boldsymbol{\beta})}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}d\Lambda_0(t)\right| \to_p 0.
$$

In the last term of (3.8), the indicator functions $\{I(\zeta_0\ge X_{ij}>\zeta)-I(\zeta\ge X_{ij}>\zeta_0)\}$ belong to a Donsker class. Thus, $n^{-1}\sum_{i=1}^n\sum_{j=1}^{K_i}\Delta_{ij}\beta_2 I(\zeta_0\ge X_{ij}>\zeta)$ converges to $E(\sum_{j=1}^{K_i}\Delta_{ij}\beta_2 I(\zeta_0\ge X_{ij}>\zeta)) = \int_0^\tau \beta_2\left\{s^{(0)-}(t;\zeta_0,\boldsymbol{\beta}_0)-s^{(0)-}(t;\zeta,\boldsymbol{\beta}_0)\right\}d\Lambda_0$. Similarly, we could prove that the other three terms converge to the remaining terms in $g(\boldsymbol{\theta})$. Therefore, this term converges uniformly to its expectation $g(\boldsymbol{\theta})$. Combining these results, we conclude that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|G_n(\boldsymbol{\theta})-G(\boldsymbol{\theta})|$ converges in probability to zero.

Next, we verify that $G(\boldsymbol{\theta})$ is a strictly concave function in a neighborhood of $\boldsymbol{\theta}_0$. We define $\dot{G}_\zeta^-(\zeta,\boldsymbol{\beta})$ and $\dot{G}_\zeta^+(\zeta,\boldsymbol{\beta})$ to be the left partial derivative for $\zeta<\zeta_0$ and the right partial derivative for $\zeta>\zeta_0$, respectively. In a neighborhood of $\boldsymbol{\theta}_0$, if $\boldsymbol{\theta}$ tends to $\boldsymbol{\theta}_0$ with $\zeta<\zeta_0$, we have

$$
\begin{aligned}
&\dot{G}_{\zeta}^{-}(\zeta,\boldsymbol{\beta}) \\
&= \lim_{\zeta^- \to \zeta} \frac{G(\zeta,\boldsymbol{\beta}) - G(\zeta^-,\boldsymbol{\beta})}{\zeta - \zeta^-} \\
&= \lim_{\zeta^- \to \zeta} \frac{1}{\zeta - \zeta^-} \int_0^\tau -\frac{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}{s^{(0)}(t;\zeta^-,\boldsymbol{\beta})} \left\{ s^{(0)}(t;\zeta,\boldsymbol{\beta}) - s^{(0)}(t;\zeta^-,\boldsymbol{\beta}) \right\} \\
&\quad - \beta_2 \left\{ s^{(0)-}(t;\zeta,\boldsymbol{\beta}) - s^{(0)-}(t;\zeta^-,\boldsymbol{\beta}) \right\} - \boldsymbol{\beta}_3^T \left\{ s^{(1)-}(t;\zeta,\boldsymbol{\beta}) - s^{(1)-}(t;\zeta^-,\boldsymbol{\beta}) \right\} d\Lambda_0(t),
\end{aligned}
$$

which is close to

$$
\begin{aligned}
&\int_0^\tau \sum_{j=1}^{K_i} \mathrm{E}\left(-\exp\left\{ \boldsymbol{\beta}_{10}^T \boldsymbol{Z}_{ij}(t) \right\} \left[\beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(t) + 1 - \exp\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(t) \right\} \right] \middle| X_{ij} = \zeta_0 \right) \\
&f_{X_{ij}}(\zeta_0) d\Lambda_0(t) \\
&= \int_0^\tau \sum_{j=1}^{K_i} \mathrm{E}\left[\exp\left\{ \boldsymbol{\beta}_{10}^T \boldsymbol{Z}_{ij}(t) \right\} \frac{\left\{ \beta_{2*} + \boldsymbol{\beta}_{3*} \boldsymbol{Z}_{ij}(t) \right\}^2}{2} \middle| X_{ij} = \zeta_0 \right] f_{X_{ij}}(\zeta_0) d\Lambda_0(t),
\end{aligned}
$$

where β_{2*} is between 0 and β_{20}, $\boldsymbol{\beta}_{3*}$ is between 0 and $\boldsymbol{\beta}_{30}$, and $f_{X_{ij}}(x)$ is the marginal density function for X_{ij}. Based on the above equation, $\dot{G}_{\zeta}^{-}(\zeta,\boldsymbol{\beta})$ is strictly positive in the neighborhood of $\boldsymbol{\theta}_0$. Similarly, we can verify that $\dot{G}_{\zeta}^{+}(\zeta,\boldsymbol{\beta})$ is strictly negative in the neighborhood of $\boldsymbol{\theta}_0$. This implies that $G(\boldsymbol{\theta})$ is concave for ζ in the neighborhood of $\boldsymbol{\theta}_0$. Furthermore, we have $G(\zeta_0,\boldsymbol{\beta}_0) - G(\zeta,\boldsymbol{\beta}) = G(\zeta_0,\boldsymbol{\beta}_0) - G(\zeta_0,\boldsymbol{\beta}) + G(\zeta_0,\boldsymbol{\beta}) - G(\zeta,\boldsymbol{\beta})$. Since the first derivative of the function $G(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ are zero at fixed ζ_0 and the second derivative of the function $G(\boldsymbol{\theta})$ is negative definite based on Condition (C.2), we have $G(\zeta_0,\boldsymbol{\beta}_0) - G(\zeta_0,\boldsymbol{\beta}) \geq 0$. In addition, $G(\zeta_0,\boldsymbol{\beta}) - G(\zeta,\boldsymbol{\beta}) \geq 0$ when $\boldsymbol{\theta}$ belongs to the neighborhood of $\boldsymbol{\theta}_0$. Thus, $G(\zeta_0,\boldsymbol{\beta}_0) - G(\zeta,\boldsymbol{\beta}) \geq 0$ in a neighborhood of $\boldsymbol{\theta}_0$, denoted by $\mathscr{V}_0$.

Consequently, if $\widehat{\boldsymbol{\theta}} \in \mathscr{V}_0$, then $G_n(\widehat{\boldsymbol{\theta}}) \geq G_n(\boldsymbol{\theta}_0)$. From the uniform convergence of $G_n(\boldsymbol{\theta})$ to $G(\boldsymbol{\theta})$, it gives $\liminf G(\widehat{\boldsymbol{\theta}}) \geq G(\boldsymbol{\theta}_0)$ with probability one. Since $G(\boldsymbol{\theta})$ has the unique maximum $\boldsymbol{\theta}_0$ in $\mathscr{V}_0$, we conclude that $\widehat{\boldsymbol{\theta}}$ should converge to $\boldsymbol{\theta}_0$ in probability. Thus, Theorem 3.3.1 holds.

Proof of Theorem 3.3.2

It suffices to show that the limit of $P_0\left\{n^{1/2}\left(|\widehat{\zeta}-\zeta_0|+\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2\right)^{1/2}>A\right\}$ can be arbitrarily small if A is large enough. For a given ϵ, let $U_\epsilon(\boldsymbol{\theta}_0)=\{\boldsymbol{\theta}\in\boldsymbol{\Theta}:A<n^{1/2}\left(|\zeta-\zeta_0|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2\right)^{1/2}\leq n^{1/2}\epsilon\}$ and $V_\epsilon(\boldsymbol{\theta}_0)=\{\boldsymbol{\theta}\in\Theta:\left(|\zeta-\zeta_0|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2\right)^{1/2}<\epsilon\}$. From Theorem 3.3.1, $P_0\left\{\widehat{\boldsymbol{\theta}}\in V_\epsilon(\boldsymbol{\theta}_0)\right\}$ $>1-\eta$ for any $\eta>0$, when n is large enough. Hence,

$$P_0\left\{n^{1/2}\left(|\widehat{\zeta}-\zeta_0|+\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2\right)^{1/2}>A\right\}=P_0\left\{\widehat{\boldsymbol{\theta}}\in U_\epsilon(\boldsymbol{\theta}_0)\right\}+P_0\left\{\widehat{\boldsymbol{\theta}}\in V_\epsilon^C(\boldsymbol{\theta}_0)\right\}$$
$$\leq P_0\left\{\sup_{\boldsymbol{\theta}\in U_\epsilon(\boldsymbol{\theta}_0)}L_n(\boldsymbol{\theta})\geq L_n(\boldsymbol{\theta}_0)\right\}+\eta=P_0\left\{\sup_{\boldsymbol{\theta}\in U_\epsilon(\boldsymbol{\theta}_0)}G_n(\boldsymbol{\theta})\geq 0\right\}+\eta.$$

The Taylor expression for $\zeta<\zeta_0$ gives

$$G(\boldsymbol{\theta})=-|\zeta-\zeta_0|\dot{G}_\zeta^-(\boldsymbol{\theta}_0)-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{I}(\boldsymbol{\theta}^*)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o_p(|\zeta-\zeta_0|),$$

where $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Similarly, for $\zeta>\zeta_0$,

$$G(\boldsymbol{\theta})=|\zeta-\zeta_0|\dot{G}_\zeta^+(\boldsymbol{\theta}_0)-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{I}(\boldsymbol{\theta}^*)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o_p(|\zeta-\zeta_0|).$$

In the proof of Theorem 3.3.1, we show that $\dot{G}_\zeta^-(\boldsymbol{\theta}_0)$ is strictly positive if $\zeta<\zeta_0$, and $\dot{G}_\zeta^+(\boldsymbol{\theta}_0)$ is strictly negative if $\zeta>\zeta_0$. The matrix $I(\boldsymbol{\theta}^*)$ is positive definite by (C.2). Therefore, there exists a positive constant k_0 which ensures $G(\boldsymbol{\theta})\leq -k_0(|\zeta-\zeta_0|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2)$. We split $U_\epsilon(\boldsymbol{\theta}_0)$ into subsets $H_{n,j}=\{(\zeta,\boldsymbol{\beta}):g(j)<n^{1/2}(|\zeta-\zeta_0|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2)^{1/2}<g(j+1)\}$, where $g(j)=2^j$, and $j=1,2,....$ In the subset $H_{n,j}$, we have $G(\boldsymbol{\theta})\leq -n^{-1}k_0g(j)^2$ and

$$P_0\left\{\sup_{\boldsymbol{\theta}\in U_\epsilon(\boldsymbol{\theta}_0)}G_n(\boldsymbol{\theta})\geq 0\right\}\leq\sum_{j:g(j)>A}P_0\left[\sup_{H_{n,j}}n^{1/2}\left\{G_n(\boldsymbol{\theta})-G(\boldsymbol{\theta})\right\}\geq n^{-1/2}g^2(j)k_0\right].$$

Using Lemma A.1 whose proof is given at the end of this proof, we obtain

$$\begin{aligned}
&\limsup_{n} \sum_{j;g(j)>A} P_0\left[\sup_{H_{n,j}} n^{1/2}\left\{G_n(\boldsymbol{\theta}) - G(\boldsymbol{\theta})\right\} \geq n^{-1/2} g^2(j) k_0\right] \\
&\leq \limsup_{n} \sum_{j;g(j)>A} \frac{\mathrm{E}\left[\sup_{H_{n,j}} n\left\{G_n(\boldsymbol{\theta}) - G(\boldsymbol{\theta})\right\}\right]^2}{g^4(j) k_0^2} \leq \sum_{j;g(j)>A} \frac{k^2 g^2(j+1)}{k_0^2 g^4(j)} \to 0,
\end{aligned}$$

as A goes to infinity.

Hence, it gives

$$\lim_{A\to\infty} \limsup_{n\to\infty} P_0\left\{n^{1/2}\left(|\zeta - \zeta_0| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2\right)^{1/2} > A\right\} = 0.$$

In other words, the convergence rate of $\widehat{\zeta}$ is n^{-1} and the convergence rate of $\widehat{\boldsymbol{\beta}}$ is $n^{-1/2}$. Theorem 3.3.2 has been proved.

Lemma A.1 Under conditions (C.1)-(C.6), for every $\epsilon > 0$, there exists a constant $k > 0$ such that $\mathrm{E}\sup_{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0)} |n^{1/2}\left\{G_n(\boldsymbol{\theta}) - G(\boldsymbol{\theta})\right\}| \leq k\epsilon$ as $n \to \infty$.

Proof of Lemma A.1 The process $n^{1/2}\left\{G_n(\boldsymbol{\theta}) - G(\boldsymbol{\theta})\right\}$ is written as $B_1(\boldsymbol{\theta}) - B_2(\boldsymbol{\theta})$, where

$$\begin{aligned}
B_1(\boldsymbol{\theta}) &= n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Delta_{ij}\left(r_\theta\left\{\boldsymbol{W}_{ij}(T_{ij})\right\} - r_{\theta_0}\left\{\boldsymbol{W}_{ij}(T_{ij})\right\}\right. \\
&\quad \left. - \mathrm{E}\left[r_\theta\left\{\boldsymbol{W}_{ij}(T_{ij})\right\} - r_{\theta_0}\left\{\boldsymbol{W}_{ij}(T_{ij})\right\}\right]\right), \\
B_2(\boldsymbol{\theta}) &= n^{-1/2} \sum_{i=1}^{n}\left[\sum_{j=1}^{K_i} \Delta_{ij} \log\left\{\frac{S_n^{(0)}(T_{ij};\zeta,\boldsymbol{\beta})}{S_n^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)}\right\}\right. \\
&\quad \left. - \int_0^\tau \log\left\{\frac{s^{(0)}(t;\zeta,\boldsymbol{\beta})}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right\} s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0) d\Lambda_0(t)\right].
\end{aligned}$$

We define

$$(I) = \int_0^\tau \beta_2 \left\{s^{(0)-}(t;\zeta_0,\boldsymbol{\beta}_0) - s^{(0)-}(t;\zeta,\boldsymbol{\beta}_0)\right\} d\Lambda_0(t),$$
$$(II) = \int_0^\tau \boldsymbol{\beta}_3^T \left\{\boldsymbol{s}^{(1)-}(t;\zeta_0,\boldsymbol{\beta}_0) - \boldsymbol{s}^{(1)-}(t;\zeta,\boldsymbol{\beta}_0)\right\} d\Lambda_0(t),$$
$$(III) = \int_0^\tau \beta_2 \left\{s^{(0)+}(t;\zeta_0,\boldsymbol{\beta}_0) - s^{(0)+}(t;\zeta,\boldsymbol{\beta}_0)\right\} d\Lambda_0(t),$$
$$(IV) = \int_0^\tau \boldsymbol{\beta}_3^T \left\{\boldsymbol{s}^{(1)+}(t;\zeta_0,\boldsymbol{\beta}_0) - \boldsymbol{s}^{(1)+}(t;\zeta,\boldsymbol{\beta}_0)\right\} d\Lambda_0(t).$$

First, we prove the supremum of $B_1(\boldsymbol{\theta})$ is bounded by $O(\epsilon)$ in a neighborhood of $\boldsymbol{\theta}_0$. We rewrite $B_1(\boldsymbol{\theta})$ as

$$\begin{aligned}
B_1(\boldsymbol{\theta}) = & n^{-1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \sum_{i=1}^n \left\{\int_0^\tau \widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_0) dN_{i\cdot} - \int_0^\tau \boldsymbol{s}^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0) d\Lambda_0(t)\right\} \\
& + n^{-1/2}\sum_{i=1}^n \left[\int_0^\tau \left\{\beta_2 + \boldsymbol{\beta}_3^T \boldsymbol{Z}_{ij}(T_{ij})\right\} I(\zeta_0 \geq X_{ij} > \zeta) dN_{i\cdot} - (I) - (II)\right] \\
& - n^{-1/2}\sum_{i=1}^n \left[\int_0^\tau \left\{\beta_2 + \boldsymbol{\beta}_3^T \boldsymbol{Z}_{ij}(T_{ij})\right\} I(\zeta \geq X_{ij} > \zeta_0) dN_{i\cdot} - (III) - (IV)\right],
\end{aligned}$$

where $N_{i\cdot} = \sum_{j=1}^{K_i} N_{ij}$. Clearly,

$$\begin{aligned}
& \mathrm{E}\left\| n^{-1/2}\sum_{i=1}^n \left\{\int_0^\tau \widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_0) dN_{i\cdot} - \int_0^\tau \boldsymbol{s}^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0) d\Lambda_0(t)\right\}\right\|^2 \\
\leq & \mathrm{E}\left\|\boldsymbol{M}_n^{(1)}(\tau)\right\|^2 + \mathrm{E}\left\|\int_0^\tau n^{1/2}\left\{\boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0) - \boldsymbol{s}^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)\right\} d\Lambda_0(t)\right\|^2 \leq O(\epsilon).
\end{aligned}$$

Additionally,

$$\begin{aligned}
& \mathrm{E}\sup_{\zeta \in V_{\epsilon^2}(\zeta_0)} \left\| n^{-1/2}\sum_{i=1}^n \left[\int_0^\tau \{\beta_2 + \boldsymbol{\beta}_3^T \boldsymbol{Z}_{ij}(T_{ij})\} I(\zeta_0 \geq X_{ij} > \zeta) dN_{i\cdot} - (I) - (II)\right]\right\| \\
\leq & n^{1/2}\mathrm{E}\sup_{f \in \mathscr{F}_1^-} \left| n^{-1}\sum_{i=1}^n \left\{\sum_{j=1}^{K_i} \Delta_{ij}\beta_2 I(\zeta_0 \geq X_{ij} > \zeta) - (I)\right\}\right| \\
& + n^{1/2}\mathrm{E}\sup_{f \in \mathscr{F}_2^-} \left| n^{-1}\sum_{i=1}^n \left\{\sum_{j=1}^{K_i} \Delta_{ij}\boldsymbol{\beta}_3^T \boldsymbol{Z}_{ij}(T_{ij}) I(\zeta_0 \geq X_{ij} > \zeta) - (II)\right\}\right|,
\end{aligned}$$

where $\mathscr{F}_1^- = \{\Delta_{ij} I(\zeta_0 \geq X_{ij} > \zeta) : \zeta_0 - \epsilon^2 < \zeta \leq \zeta_0\}$ and $\mathscr{F}_2^- = \{\Delta_{ij} \boldsymbol{Z}_{ij} I(\zeta_0 \geq X_{ij} > \zeta) : \zeta_0 - \epsilon^2 < \zeta \leq \zeta_0\}$. The expectations of the right-hand side are bounded by $O(\epsilon)$ as a consequence of Theorem 2.14.1 in van der Vaart and Wellner (1996). Similarly, we can prove that the expectation of the last term of $B_1(\boldsymbol{\theta})$ is bounded by $O(\epsilon)$.

For $B_2(\boldsymbol{\theta})$, it can be rewritten as

$$\begin{aligned} B_2(\boldsymbol{\theta}) = n^{-1/2} \sum_{i=1}^{n} &\left[\sum_{j=1}^{K_i} \Delta_{ij} \log \left\{ \frac{s^{(0)}(T_{ij}; \zeta, \boldsymbol{\beta})}{s^{(0)}(T_{ij}; \zeta_0, \boldsymbol{\beta}_0)} \right\} \right. \\ &\left. - \int_0^\tau \log \left\{ \frac{s^{(0)}(t; \zeta, \boldsymbol{\beta})}{s^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0)} \right\} s^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0) d\Lambda_0(t) \right] \\ &+ n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Delta_{ij} \left[\log \left\{ \frac{S_n^{(0)}(T_{ij}; \zeta, \boldsymbol{\beta})}{s^{(0)}(T_{ij}; \zeta, \boldsymbol{\beta})} \right\} - \log \left\{ \frac{S_n^{(0)}(T_{ij}; \zeta_0, \boldsymbol{\beta}_0)}{s^{(0)}(T_{ij}; \zeta_0, \boldsymbol{\beta}_0)} \right\} \right]. \end{aligned} \tag{3.9}$$

The first term of (3.9) is the empirical process of $\Delta_{ij} \log \left\{ \frac{s^{(0)}(t;\zeta,\boldsymbol{\beta})}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)} \right\}$. Following the proof in Pons (2003), its bracketing integral is finite and it is bounded by $O(\epsilon)$ as a consequence of Theorem 2.14.2 (van der Vaart and Wellner 1996). Using the approximation $\log(1+x) \approx x$ as x goes to zero, the second term of (3.9) is uniformly approximated by $n^{-3/2} \sum_{i,i'=1}^{n} \sum_{j,j'=1}^{K_i} \Delta_{ij} \left(\frac{Y_{i'j'}(T_{ij}) \exp[r_\theta\{\boldsymbol{W}_{i'j'}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta,\boldsymbol{\beta})} - \frac{Y_{i'j'}(T_{ij}) \exp[r_{\theta_0}\{\boldsymbol{W}_{i'j'}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)} \right) \{1 + o_{a.s.}(1)\}$, which is bounded by $n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Delta_{ij} \left(\frac{Y_{ij}(T_{ij}) \exp[r_\theta\{\boldsymbol{W}_{ij}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta,\boldsymbol{\beta})} - \frac{Y_{ij}(T_{ij}) \exp[r_{\theta_0}\{\boldsymbol{W}_{ij}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)} \right) \{1 + o_{a.s.}(1)\} + n^{-3/2} \sum_{i \neq i'} \sum_{j \neq j'} \Delta_{ij} \left(\frac{Y_{i'j'}(T_{ij}) \exp[r_\theta\{\boldsymbol{W}_{i'j'}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta,\boldsymbol{\beta})} - \frac{Y_{i'j'}(T_{ij}) \exp[r_{\theta_0}\{\boldsymbol{W}_{i'j'}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)} \right) \{1 + o_{a.s.}(1)\}$.

Note that $E\left\{ \sup_{\boldsymbol{\theta} \in V_\epsilon(\boldsymbol{\theta}_0)} n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Delta_{ij} \left(\frac{\exp[r_\theta\{\boldsymbol{W}_{ij}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta,\boldsymbol{\beta})} - \frac{\exp[r_{\theta_0}\{\boldsymbol{W}_{ij}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)} \right) \right\} = o_p(1)$, and $E\left\{ \sup_{\boldsymbol{\theta} \in V_\epsilon(\boldsymbol{\theta}_0)} n^{-3/2} \left| \sum_{i \neq i'} \sum_{j \neq j'} \Delta_{ij} \left(\frac{Y_{i'j'}(T_{ij}) \exp[r_\theta\{\boldsymbol{W}_{i'j'}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta,\boldsymbol{\beta})} - \frac{Y_{i'j'}(T_{ij}) \exp[r_{\theta_0}\{\boldsymbol{W}_{i'j'}(T_{ij})\}]}{s^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)} \right) \right| \right\}$ is bounded by

$$\begin{aligned} &\mathrm{E} \int_0^\tau \sup_{\boldsymbol{\theta} \in V_\epsilon(\boldsymbol{\theta}_0)} n^{-1/2} \left| \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left(\frac{Y_{ij}(t) \exp[r_\theta\{\boldsymbol{W}_{ij}(t)\}]}{s^{(0)}(t;\zeta,\boldsymbol{\beta})} - \frac{Y_{ij}(t) \exp[r_{\theta_0}\{\boldsymbol{W}_{ij}(t)\}]}{s^{(0)}(T_{ij};\zeta_0,\boldsymbol{\beta}_0)} \right) \right| \\ &\exp[r_\theta\{\boldsymbol{W}_{ij}(t)\}] d\Lambda_0(t). \end{aligned}$$

According to the location of the variable X_{ij} with respect to ζ and ζ_0, we split the integrand

$\sum_{j=1}^{K_i}\left(\frac{Y_{ij}(t)\exp[r_\theta\{\boldsymbol{W}_{ij}(t)\}]}{s^{(0)}(t;\zeta,\boldsymbol{\beta})} - \frac{Y_{ij}(t)\exp[r_{\theta_0}\{\boldsymbol{W}_{ij}(t)\}]}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right)$ into four terms,

$$
\begin{aligned}
\phi_{1,t,\theta} &= \sum_{j=1}^{K_i}\left[\frac{\exp\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta,\boldsymbol{\beta})} - \frac{\exp\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right] I(X_{ij}\le\zeta_0),\\
\phi_{2,t,\theta} &= \sum_{j=1}^{K_i}\left[\frac{\exp\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t)+\beta_2+\boldsymbol{\beta}_3^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta,\boldsymbol{\beta})} - \frac{\exp\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{ij}(t)+\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right]\\
&\quad I(X_{ij}>\zeta_0),\\
\phi_{3,t,\theta} &= \sum_{j=1}^{K_i}\left[\frac{\exp\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t)+\beta_2+\boldsymbol{\beta}_3^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta,\boldsymbol{\beta})} - \frac{\exp\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right] I(\zeta<X_{ij}\le\zeta_0),\\
\phi_{4,t,\theta} &= \sum_{j=1}^{K_i}\left[\frac{\exp\{\boldsymbol{\beta}_1^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta,\boldsymbol{\beta})} - \frac{\exp\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{ij}(t)+\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(t)\}}{s^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right] I(\zeta_0<X_{ij}\le\zeta),
\end{aligned}
$$

Both $\phi_{1,t,\theta}$ and $\phi_{2,t,\theta}$ are continuously differentiable with respect to $\boldsymbol{\theta}$. In addition, their derivatives are uniformly square integrable on $[0,\tau]\times V_\epsilon(\boldsymbol{\theta}_0)$. Thus, $\sup_{t\in[0,\tau]}\mathrm{E}\sup_{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0)}$ $|n^{-1/2}\sum_{i=1}^n\{\phi_{1,t,\theta} - \mathrm{E}(\phi_{1,t,\theta})\}|$ and $\sup_{t\in[0,\tau]}\mathrm{E}\sup_{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0)}|n^{-1/2}\sum_{i=1}^n\{\phi_{2,t,\theta}-\mathrm{E}(\phi_{2,t,\theta})\}|$ are bounded by $O(\epsilon)$. For $\phi_{3,t,\theta}$ and $\phi_{4,t,\theta}$, they are the products of the indicator function $I(\zeta_0<X_{ij}\le\zeta)$ or $I(\zeta<X_{ij}\le\zeta_0)$ and a continuously differentiable function with respect to $\boldsymbol{\theta}$ having uniformly square integrable derivatives on $[0,\tau]\times V_\epsilon(\boldsymbol{\theta}_0)$. Since their finite L_2 bracketing integral which does not depend on t, $\sup_{t\in[0,\tau]}\mathrm{E}\sup_{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0)}|n^{-1/2}\sum_{i=1}^n\{\phi_{3,t,\theta}-\mathrm{E}(\phi_{3,t,\theta})\}|$ and $\sup_{t\in[0,\tau]}\mathrm{E}\sup_{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0)}|n^{-1/2}\sum_{i=1}^n\{\phi_{4,t,\theta}-\mathrm{E}(\phi_{4,t,\theta})\}|$ are bounded by $O(\epsilon)$. Consequently, $\mathrm{E}\int_0^\tau\sup_{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0)} n^{-1/2}|\sum_{i=1}^n\phi_{k,t,\theta}\exp[r_\theta\{\boldsymbol{W}_{ij}(t)\}]\,d\Lambda_0(t)$ is bounded by $O(\epsilon)$ for $k=1,2,3,4$. Therefore, both $B_1(\boldsymbol{\theta})$ and $B_2(\boldsymbol{\theta})$ are bounded by $O(\epsilon)$ and the lemma is proved.

Proof of Theorem 3.3.3

The whole proof can be divided into the following steps. First, recall the definitions of $\boldsymbol{\theta}_{n,u}=(\zeta_{n,u},\boldsymbol{\beta}_{n,u})$, $\zeta_{n,u}=\zeta_0+n^{-1}u_1$, and $\boldsymbol{\beta}_{n,u}=\boldsymbol{\beta}_0+n^{-1/2}\boldsymbol{u}_2$. We partition the $l_n(\boldsymbol{\theta}_{n,u})-l_n(\boldsymbol{\theta}_0)$ into three terms and obtain their asymptotic expansions. In the second step, we derive the asymptotic distributions based on the first two terms in the partition. Finally, we obtain Theorem 3.3.3 using the argmax continuous mapping theorem.

In the first step, similar to the partition in Theorem 3.3.1, $l_n(\boldsymbol{\theta}_{n,u}) - l_n(\boldsymbol{\theta}_0)$ is partitioned as

$$
\begin{aligned}
& l_n(\boldsymbol{\theta}_{n,u}) - l_n(\boldsymbol{\theta}_0) \\
&= (\boldsymbol{\beta}_{n,u} - \boldsymbol{\beta}_0)^T \left\{ \sum_{i=1}^{n} \sum_{j=1}^{K_i} \int_0^\tau \widetilde{\boldsymbol{Z}}_{ij}(\zeta_{n,u}) dN_{ij} \right\} - \int_0^\tau \log \left\{ \frac{S_n^{(0)}(t; \zeta_{n,u}, \boldsymbol{\beta}_{n,u})}{S_n^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0)} \right\} d\bar{N}_n(t) \\
&+ \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Delta_{ij} \{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(T_{ij}) \} \{ I(\zeta_0 \geq X_{ij} > \zeta_{n,u}) - I(\zeta_{n,u} \geq X_{ij} > \zeta_0) \}.
\end{aligned}
$$

By the Taylor expansion for $\boldsymbol{\beta}_{n,u}$ at $\boldsymbol{\beta}_0$,

$$
\begin{aligned}
& \log \left\{ \frac{S_n^{(0)}(t; \zeta_{n,u}, \boldsymbol{\beta}_{n,u})}{S_n^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0)} \right\} \\
&= \log \left\{ \frac{S_n^{(0)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0)}{S_n^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0)} \right\} + n^{-1/2} \boldsymbol{u}_2^T \frac{\boldsymbol{S}_n^{(1)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0)}{S_n^{(0)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0)} \\
&+ \frac{n^{-1}}{2} \boldsymbol{u}_2^T \frac{\boldsymbol{S}_n^{(2)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0) S_n^{(0)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0) - \boldsymbol{S}_n^{(1)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0)^{\otimes 2}}{S_n^{(0)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0)^{\otimes 2}} \boldsymbol{u}_2 + o_p(n^{-1}) \\
&= \frac{S_n^{(0)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0) - S_n^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0)}{S_n^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0)} + n^{-1/2} \boldsymbol{u}_2^T \frac{\boldsymbol{S}_n^{(1)}(t; \zeta_{n,u}, \boldsymbol{\beta}_0)}{S_n^{(0)}(t; \zeta_0, \boldsymbol{\beta}_0)} \\
&+ \frac{n^{-1}}{2} \boldsymbol{u}_2^T \boldsymbol{V}_n(t; \zeta_{n,u}, \boldsymbol{\beta}_0) \boldsymbol{u}_2 + o_p(n^{-1}),
\end{aligned}
$$

where $\boldsymbol{V}_n(t; \zeta, \boldsymbol{\beta}) = \boldsymbol{S}_n^{(2)}(t; \zeta, \boldsymbol{\beta}) / S_n^{(0)}(t; \zeta, \boldsymbol{\beta}) - \left\{ \boldsymbol{S}_n^{(1)}(t; \zeta, \boldsymbol{\beta}) / S_n^{(0)}(t; \zeta, \boldsymbol{\beta}) \right\}^{\otimes 2}$. Thus, we have

$$
l_n(\boldsymbol{\theta}_{n,u}) - l_n(\boldsymbol{\theta}_0) = Q_n(u_1) + \boldsymbol{u}_2^T \boldsymbol{C}_n(u_1) + \frac{1}{2} \boldsymbol{u}_2^T \boldsymbol{V}_n(t; \zeta_{n,u}, \boldsymbol{\beta}_0) \boldsymbol{u}_2 + o_p(n^{-1}),
$$

where

$$
\begin{aligned}
Q_n(u_1) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Delta_{ij} \Big[& \left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(T_{ij}) \right\} \{ I(\zeta_0 \geq X_{ij} > \zeta_{n,u}) - I(\zeta_{n,u} \geq X_{ij} > \zeta_0) \} \\
& - \frac{S_n^{(0)}(T_{ij}; \zeta_{n,u}, \boldsymbol{\beta}_0) - S_n^{(0)}(T_{ij}; \zeta_0, \boldsymbol{\beta}_0)}{S_n^{(0)}(T_{ij}; \zeta_0, \boldsymbol{\beta}_0)} \Big],
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{C}_n(u_1) &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\int_0^{\tau}\left\{\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_{n,u}) - \frac{\boldsymbol{S}_n^{(1)}(t;\zeta_{n,u},\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right\}dN_{ij}(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\int_0^{\tau}\left\{\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_{n,u}) - \frac{\boldsymbol{S}_n^{(1)}(t;\zeta_{n,u},\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right\}dM_{ij}(t)\\
&+ n^{-1/2}\int_0^{\tau}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_{n,u})\left(\exp\left[r_{\theta_0}\left\{\boldsymbol{W}_{ij}(t)\right\}\right] - \exp\left[r_{\zeta_{n,u},\beta_0}\left\{\boldsymbol{W}_{ij}(t)\right\}\right]\right)d\Lambda_0(t).
\end{aligned}
$$

It is easy to verify that

$$
\begin{aligned}
&\mathrm{E}\left(\int_0^{\tau}\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_{n,u}) - \widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_0)dM_{ij}(t)\right) = 0,\\
&\mathrm{E}\left(\int_0^{\tau}\frac{\boldsymbol{S}_n^{(1)}(t;\zeta_{n,u},\boldsymbol{\beta}_0) - \boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}dM_{ij}(t)\right) = 0,\\
&\mathrm{E}\left\{\int_0^{\tau}\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_{n,u})\left(\exp\left[r_{\theta_0}\left\{\boldsymbol{W}_{ij}(t)\right\}\right] - \exp\left[r_{\zeta_{n,u},\beta_0}\left\{\boldsymbol{W}_{ij}(t)\right\}\right]\right)d\Lambda_0(t)\right\} = 0.
\end{aligned}
$$

In addition, we can show that

$$
\begin{aligned}
&E\sup_{|u_1|\le A}\left\|\int_0^{\tau}\left\{\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_{n,u}) - \widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_0)\right\}dM_{ij}(t)\right\|^2 = O(n^{-1}),\\
&E\sup_{|u_1|\le A}\left\|\int_0^{\tau}\left\{\frac{\boldsymbol{S}_n^{(1)}(t;\zeta_{n,u},\boldsymbol{\beta}_0) - \boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right\}dM_{ij}(t)\right\|^2 = O(n^{-1}),\\
&E\sup_{|u_1|\le A}\left\|\int_0^{\tau}\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_{n,u})\left(\exp\left[r_{\theta_0}\left\{\boldsymbol{W}_{ij}(t)\right\}\right] - \exp\left[r_{\zeta_{n,u},\beta_0}\left\{\boldsymbol{W}_{ij}(t)\right\}\right]\right)d\Lambda_0(t)\right\|^2 = O(n^{-1}).
\end{aligned}
$$

Thus, $\boldsymbol{C}_n(u_1)$ uniformly converges to $\widetilde{\boldsymbol{l}}_n = n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\int_0^{\tau}\left\{\widetilde{\boldsymbol{Z}}_{ij}(t;\zeta_0) - \frac{\boldsymbol{S}_n^{(1)}(t;\zeta_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}\right\}dM_{ij}(t)$ in probability, as $|u_1| \le A$. Then we have

$$
l_n(\boldsymbol{\theta}_{n,u}) - l_n(\boldsymbol{\theta}_0) = Q_n(u_1) + \boldsymbol{u}_2^T\widetilde{\boldsymbol{l}}_n - \frac{1}{2}\boldsymbol{u}_2^T\boldsymbol{I}(\boldsymbol{\theta}_0)\boldsymbol{u}_2 + o_p(1). \tag{3.10}
$$

In the second step, we derive the asymptotic distributions of $Q_n(u_1)$ and $\widetilde{\boldsymbol{l}}_n$ in (3.10). The variable $\widetilde{\boldsymbol{l}}_n$ converges weakly to a Gaussian variable following the normal distribution

$N(\mathbf{0}, \mathbf{\Sigma}(\boldsymbol{\theta}_0))$. The process Q_n is written as $Q_n = Q_n^+ - Q_n^-$, where

$$Q_n^+(u_1) = I(u_1 > 0)\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_{ij}\Big(-\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(T_{ij})\right\}I(\zeta_0 + n^{-1}u_1 \geq X_{ij} > \zeta_0)$$
$$-\frac{\sum_{l=1}^{n}\sum_{k=1}^{K_l}Y_{lk}(T_{ij})\exp\left\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{lk}(T_{ij})\right\}\left[1-\exp\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{lk}(T_{ij})\right\}\right]I(\zeta_0 + n^{-1}u_1 \geq X_{lk} > \zeta_0)}{nS_n^{(0)}(T_{ij};\boldsymbol{\theta}_0)}\Bigg),$$
$$Q_n^-(u_1) = I(u_1 < 0)\sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_{ij}\Big(-\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(T_{ij})\right\}I(\zeta_0 + n^{-1}u_1 < X_{ij} \leq \zeta_0)$$
$$-\frac{\sum_{l=1}^{n}\sum_{k=1}^{K_l}Y_{lk}(T_{ij})\exp\left\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{lk}(T_{ij})\right\}\left[1-\exp\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{lk}(T_{ij})\right\}\right]I(\zeta_0 + n^{-1}u_1 < X_{lk} \leq \zeta_0)}{nS_n^{(0)}(T_{ij};\boldsymbol{\theta}_0)}\Bigg).$$

Note that Q_n is defined as a random variable on the space D of right-continuous functions with left-hand limits on R equipped with the Skorohod topology. Now we prove that the process Q_n^+ converges weakly to Q^+ on the space $D[0,\infty)$, where Q^+ is defined in the Section 3. Let $0 = v_0 < v_1 < v_2 < ... < v_S \leq A$ be an increasing sequence, $I_{ns} = [\zeta_0 + n^{-1}v_{s-1}, \zeta_0 + n^{-1}v_s]$, and $q_1, q_2, ..., q_S$ be constants. Consider the variable $\Sigma_n \equiv \sum_{s=1}^{S} q_s\left\{Q_n^+(v_s) - Q_n^+(v_{s-1})\right\} \equiv \sum_{i=1}^{n}\sum_{s=1}^{S} q_s\eta_{ns,i}$, where $\eta_{ns,i} \equiv \eta_{ns,i}^{(1)} - \eta_{ns,i}^{(2)}$,

$$\eta_{ns,i}^{(1)} = \sum_{j=1}^{K_i} I(X_{ij} \in I_{ns})\left[-\Delta_{ij}\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(T_{ij})\right\} - \int_0^\tau \phi_{ij}(t)d\Lambda_0(t)\right],$$

$$\eta_{ns,i}^{(2)} = \sum_{j=1}^{K_i} I(X_{ij} \in I_{ns})\int_0^\tau \phi_{ij}(t)\left\{\frac{1}{nS_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}d\bar{N}_n(t) - d\Lambda_0(t)\right\},$$

with $\phi_{ij}(t) = Y_{ij}(t)\exp\left\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{ij}(t)\right\}\left[1-\exp\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(t)\right\}\right]$. We define $\Sigma_{1n} \equiv \sum_{i=1}^{n}\sum_{s=1}^{S} q_s$ $\eta_{ns,i}^{(1)}$ and $\Sigma_{2n} \equiv \sum_{i=1}^{n}\sum_{s=1}^{S} q_s\eta_{ns,i}^{(2)}$. Spiekerman and Lin (1998) shows that $\int_0^s \frac{1}{nS_n^{(0)}(t;\zeta_0,\boldsymbol{\beta}_0)}d\bar{N}_n(t)$ converges in probability to $\Lambda_0(s)$ uniformly for $t \in [0,\tau]$. Since $|\phi_{ij}(t)|$ is bounded, $\eta_{ns,i}^{(2)} = o_p(1/n)$ and $\Sigma_{2n} = o_p(1)$. Thus, Σ_{2n} converges to zero in probability and the asymptotic distribution of $\widehat{\zeta}$ only depends on Σ_{1n}. In addition, the sequence of the distributions of Σ_n is tight, since $\mathrm{E}\left(\Sigma_{1n}^2\right)$ is bounded.

Next, we show that the characteristic function of Σ_{1n} converges to the characteristic function of Q^{+}. The characteristic function for Σ_{1n} is expressed as

$$\phi_n^{+}(t) = \left\{ \mathrm{E}\left(\exp \sum_{s \le S} itq_s \eta_{ns,i}^{(1)} \right) \right\}^n = \left\{ \sum_{m=0}^{K} P(K_i = m) \mathrm{E}\left(\exp \sum_{s \le S} itq_s \eta_{ns,i}^{(1)} \middle| K_i = m \right) \right\}^n ,$$

where $K = \max(K_i)$ is the maximum cluster size. We define the set

$$A_{m0}^{+} = \{K_i = m, \text{no one } X_{ij} > \zeta_0\},$$

$$A_{m1}^{+} = \{K_i = m, \text{only one } X_{ij} > \zeta_0, \text{all the other } X_{i1}, ..., X_{ij-1}, X_{ij+1}, ..., X_{im} \le \zeta_0\},$$

and $A_{mu}^{+} = \{K_i = m, \text{only u subjects } X_{ij_1}, ..., X_{ij_u} > \zeta_0, \text{all the others } \le \zeta_0\}, \text{for u} \ge 2.$

Thus, we have

$$\mathrm{E}\left\{ \exp\left(\sum_{s \le S} itq_s \eta_{ns,i}^{(1)} \right) \middle| K_i = m \right\} = \sum_{u=0}^{m} \mathrm{E}\left\{ I(A_{mu}^{+}) \exp\left(\sum_{s \le S} itq_s \eta_{ns,i}^{(1)} \right) \middle| K_i = m \right\},$$

where $\mathrm{E}\left\{ I(A_{m0}^{+}) \exp\left(\sum_{s \le S} itq_s \eta_{ns,i}^{(1)} \right) | K_i = m \right\} = P(A_{m0}^{+} | K_i = m)$. Define $A_{m1}^{k+} = \{X_{ik} > \zeta_0, X_{i1}, ..., X_{ik-1}, X_{ik+1}, ..., X_{im} \le \zeta_0\}$, and $\eta_{ns,ij}^{(1)} = -\Delta_{ij} \{\beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(T_{ij})\} - \int_0^{\tau} \phi_{ij}(t) d\Lambda_0(t)$. For the subset A_{m1}^{+},

$$\begin{aligned}
&\mathrm{E}\left\{ I(A_{m1}^{+}) \exp\left(\sum_{s \le S} itq_s \eta_{ns,i}^{(1)} \right) \middle| K_i = m \right\} \\
&= \sum_{s \le S} \sum_{k=1}^{m} P(A_{m1}^{k+} | K_i = m, A_{m1}^{+}) \int_{X_{ik} \in I_{ns}} \mathrm{E}\left\{ \exp\left(itq_s \eta_{ns,ik}^{(1)} \right) \middle| K_i = m, A_{m1}^{k+} \right\} f(X_{ik}) dX_{ik} \\
&= \sum_{s \le S} \sum_{k=1}^{m} P(A_{m1}^{k+} | K_i = m, A_{m1}^{+}) \left(\frac{v_s - v_{s-1}}{n} \right) \\
&\quad f_{X_{ik}}(\zeta_0^{+}) \mathrm{E}\left\{ \exp\left(itq_s \eta_{ns,ik}^{(1)} \right) \middle| K_i = m, A_{m1}^{k+}, X_{ik} = \zeta_0^{+} \right\} + o_p(1/n).
\end{aligned}$$

Similarly, for the subset A^{+}_{mu} with $u \geq 2$,

$$\sum_{u=2}^{m} \mathrm{E}\left\{ I(A^{+}_{mu}) \exp\left(\sum_{s \leq S} itq_s \eta^{(1)}_{ns,i} \right) \middle| K_i = m \right\}$$
$$= \sum_{u=2}^{m} \sum_{s_1,\ldots,s_u \leq S} \sum_{k=1}^{|A^{+}_{mu}|} f_{X_{ik_1},\ldots,X_{ik_u}}(\zeta^{+}_0) P(A^{k+}_{mu}|K_i = m, A^{+}_{mu}) \left(\prod_{l=1}^{u} \frac{v_{s_l} - v_{s_l-1}}{n} \right)$$
$$\mathrm{E}\left\{ \exp\left(\sum_{l=1}^{u} itq_{s_l} \eta^{(1)}_{ns_l,ik_l} \right) \middle| K_i = m, A^{k+}_{mu}, X_{ik_1} = \ldots = X_{ik_u} = \zeta^{+}_0 \right\} + o_p(1/n^u)$$
$$= o_p(1/n),$$

where $A^{k+}_{mu} = \{X_{ik_1}, \ldots, X_{ik_u} > \zeta_0, \text{all the others } \leq \zeta_0\}$, and $|A^{+}_{mu}|$ is the total number of all the possible sets. In addition, $P(A^{+}_{m0}|K_i = m) = 1 - \sum_{u=1}^{m} P(A^{+}_{mu}|K_i = m)$, where

$$P(A^{+}_{m1}|K_i = m) = \sum_{s \leq S} \sum_{k=1}^{m} P(A^{k+}_{m1}|K_i = m, A^{+}_{m1}) \int_{X_{ik} \in I_{ns}} f(X_{ik}) dX_{ik}$$
$$= \sum_{s \leq S} \sum_{k=1}^{m} P(A^{k+}_{m1}|K_i = m, A^{+}_{m1}) \left(\frac{v_s - v_{s-1}}{n} \right) f_{X_{ik}}(\zeta^{+}_0),$$
$$P(A^{+}_{mu}|K_i = m) = \sum_{s_1,\ldots,s_u \leq S} \sum_{k=1}^{|A^{+}_{mu}|} P(A^{k+}_{mu}|K_i = m, A^{+}_{mu})$$
$$\int_{X_{ik_1} \in I_{ns_1}} \int_{X_{ik_2} \in I_{ns_2}} \cdots \int_{X_{ik_u} \in I_{ns_u}} f(X_{ik_1}, \ldots, X_{ik_u}) dX_{ik_1} \ldots dX_{ik_u}$$
$$= \sum_{s_1,\ldots,s_u \leq S} \sum_{k=1}^{|A^{+}_{mu}|} P(A^{k+}_{mu}|K_i = m, A^{+}_{mu}) \left(\prod_{l=1}^{u} \frac{v_{s_l} - v_{s_l-1}}{n} \right) f_{X_{ik_1},\ldots,X_{ik_u}}(\zeta^{+}_0),$$

where $u \geq 2$. Thus, $P(A^{+}_{m0}|K_i = m) = 1 - \sum_{s \leq S} \sum_{k=1}^{m} P(A^{k+}_{m1}|K_i = m, A^{+}_{m1}) \left(\frac{v_s - v_{s-1}}{n} \right) f_{X_{ik}}(\zeta^{+}_0) + o_p(1/n)$.

Then, we have

$$
\begin{aligned}
&\mathrm{E}\left\{\exp\left(\sum_{s\le S} itq_s\eta^{(1)}_{ns,i}\right)\middle| K_i = m\right\}\\
=&P(A^{+}_{m0}|K_i = m) + \sum_{s\le S}\sum_{k=1}^{m} P(A^{k+}_{m1}|K_i = m, A^{+}_{m1})\left(\frac{v_s - v_{s-1}}{n}\right)\\
&f_{X_{ik}}(\zeta^{+}_0)\mathrm{E}\left\{\exp\left(itq_s\eta^{(1)}_{ns,ik}\right)\middle| K_i = m, A^{k+}_{m1}, X_{ik} = \zeta^{+}_0\right\} + o_p(1/n)\\
=&1 + \sum_{s\le S}\frac{v_s - v_{s-1}}{n}\phi_m(t;q_s) + o_p(1/n),
\end{aligned}
$$

where

$$
\begin{aligned}
&\phi_m(t;q_s)\\
&=\sum_{k=1}^{m} f_{X_{ik}}(\zeta^{+}_0)P(A^{k+}_{m1}|K_i = m, A^{+}_{m1})\left[\mathrm{E}\left\{\exp\left(itq_s\eta^{(1)}_{ns,ik}\right)\middle| K_i = m, A^{k+}_{m1}, X_{ik} = \zeta^{+}_0\right\} - 1\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\phi^{+}_n(t) &= \left[\sum_{m=0}^{K} P(K_i = m)\left\{1 + \sum_{s\le S}\left(\frac{v_s - v_{s-1}}{n}\right)\phi_m(t;q_s) + o_p(1/n)\right\}\right]^n\\
&= 1 + \sum_{m=0}^{K} P(K_i = m)\sum_{s\le S}(v_s - v_{s-1})\phi_m(t;q_s) + o_p(1)\\
&\to \exp\left\{\sum_{s\le S}(v_s - v_{s-1})\sum_{m=0}^{K} p(m)\phi_m(t;q_s)\right\} = \phi^{+}(t),
\end{aligned}
$$

where $p(m)$ is the probability of the cluster with m subjects.

In the Section 3, we define $Q(s) \equiv Q^{+}(s) - Q^{-}(s)$, where

$$
Q^{+}(s) \equiv \sum_{m=0}^{K}\sum_{k=1}^{m}\sum_{0\le l\le v^{+}_{mk}(s)} V^{+}_{mk,l}, Q^{-}(s) \equiv \sum_{m=0}^{K}\sum_{k=1}^{m}\sum_{0\le l\le v^{-}_{mk}(s)} V^{-}_{mk,l}.
$$

Recall that $V^+_{mk,l}$ and $V^-_{mk,l}$ are independent sequences of identically and independently distributed random variables with the characteristic functions

$$\mathrm{E}\left\{\exp\left(itV^+_{mk,l}\right)\right\} = \mathrm{E}\left\{\exp\left(itq_s\eta^{(1)}_{ns,ik}\right)\middle|K_i = m, A^{k+}_{m1}, X_{ik} = \zeta^+_0\right\},$$
$$\mathrm{E}\left\{\exp\left(itV^-_{mk,l}\right)\right\} = \mathrm{E}\left\{\exp\left(itq_s\eta^{(1)}_{ns,ik}\right)\middle|K_i = m, A^{k-}_{m1}, X_{ik} = \zeta^-_0\right\},$$

where $m = 1,...,K$, $k = 1,...,m$, and $l \geq 1$. We denote v^+_{mk} and v^-_{mk} to be the real jump processes such that $v^+_{mk} = 0$ on R^- and $v^-_{mk} = 0$ on R^+. We further denote $v^+_{mk}(s)$ to be a Poisson variable with the parameter $sp(m)f_{X_{ik}}(\zeta^+_0)P(A^{k+}_{m1}|K_i = m, A^+_{m1})$, and $v^-_{mk}(s)$ to be a Poisson variable with parameter $sp(m)f_{X_{ik}}(\zeta^-_0)P(A^{k-}_{m1}|K_i = m, A^-_{m1})$.

The characteristic function of Q^+ is as follows.

$$\begin{aligned}
\phi^+(t) =& E\exp\left\{itQ^+(s)\right\} \\
=& \prod_{m=0}^{K}\prod_{k=1}^{m}\exp\left\{-sp(m)f_{X_{ik}}(\zeta^+_0)P(A^{k+}_{m1}|K_i = m, A^+_{m1})\right\} \\
&\times\sum_{j\geq 0}\frac{\left[sp(m)f_{X_{ik}}(\zeta^+_0)P(A^{k+}_{m1}|K_i = m, A^+_{m1})\mathrm{E}\left\{\exp\left(itV^+_{mk,l}\right)\right\}\right]^j}{j!} \\
=& \exp\left(\sum_{m=0}^{K} sp(m)\sum_{k=1}^{m} f_{X_{ik}}(\zeta^+_0)P(A^{k+}_{m1}|K_i = m, A^+_{m1})\left[\mathrm{E}\left\{\exp\left(itV^+_{mk,l}\right)\right\} - 1\right]\right) \\
=& \exp\left\{\sum_{m=0}^{K} sp(m)\phi_m(t;q_s)\right\}.
\end{aligned}$$

Let $0 \leq v_1 \leq v \leq v_2 \leq A$, $I_{n1} =]\zeta_0 + n^{-1}v_1, \zeta_0 + n^{-1}v]$ and $I_{n2} =]\zeta_0 + n^{-1}v, \zeta_0 + n^{-1}v_2]$. To prove Q^+_n is tight, we have $E|Q^+_n(v) - Q^+_n(v_1)||Q^+_n(v_2) - Q^+_n(v)|$ is bounded by $(v_2 - v_1)^2$ times a constant. Thus, the processes Q^+_n converge weakly to Q^+, using the D-tightness criterion. (Billingsley 2009) Similarly, we can prove that Q^-_n converges weakly to Q^-. Finally, because of argmax continuous mapping theorem, $n\left(\widehat{\zeta} - \zeta_0\right) = \arg\max_{u_1} Q_n(u_1) + o_p(1)$ and it converges weakly to $\arg\max Q(s)$.

CHAPTER 4: STATISTICAL INFERENCE FOR COX PROPORTIONAL HAZARDS MODEL WITH A CHANGE HYPERPLANE

4.1 Introduction

Change point analysis has been widely used in epidemiology studies and clinical trials to identify high-risk subjects whose hazard rates are substantially different from the others (Tapp et al. 2006; Marquis et al. 2002; Zhao et al. 2014). For example, Zhao et al. (2014) and Chapter 3 investigated the change point of leukocyte telomere length (LTL) for diabetes incidence in the Strong Heart Family Study (SHFS). In the same study, however, the change point based on LTL has been observed to depend on some other biomarkers, such as triglycerides, HDL, and etc. In other words, the incidence of diabetes can change dramatically depending on a combination of LTL and other biomarkers. To improve the performance in identifying the high-risk individuals, we aim to introduce an extended change point model based on a linear combination of multiple risk factors. This model is defined as a change hyperplane model, where the hyperplane is the affine hyperplane in geometry. Specifically, the change hyperplane is a single linear function larger or less than a threshold.

In the univariate survival data, the Cox proportional hazards model (Cox 1972) has been widely used for estimating the association between potential risk factors and disease incidence. In this type of model, the change point analysis has been extensively studied (Liang et al. 1990; Luo 1996; Pons 2002; Luo 1996; Gandy et al. 2005; Gandy and Jensen 2005; Jensen and Lütkebohmert 2008; Luo and Boyett 1997, Pons 2003, Kosorok and Song 2007). Among them, Pons (2003) proved that the change point estimator based on a single covariate follows a composite Poisson process which depends on the change point locations across all the subjects. However, all these methods assume that the change point depends on only one covariate. The current estimation and inference procedures cannot be directly extended to the Cox model with a change hyperplane based on multiple covariates, due to the fact that the linear combination of covariates is unknown in addition to the threshold. Thus, it is of great interest to develop a rigorous Cox proportional hazards model with a change hyperplane based on multiple covariates for survival data subject to censoring.

The change hyperplane in the Cox proportional hazards model can be viewed as a function of the single index. The single index model reduces the dimensionality of the covariates by introducing an unspecified function of the single index. The single index has been widely applied in the Cox proportional hazards model. Wang (2004) generalized the exponential link function in the Cox proportional hazards model to an unspecified link function of the single index. The proportional hazards assumption is no longer needed in this model. Later, Huang and Liu (2006) proposed a model which assumed the log linear relationship between the hazard function and the unknown link function of the single index. All methods assume the link function to be a smooth function. However, in a change hyperplane model, the unknown smooth function is replaced by a discontinuous indicator function, where the indicator function equals one or zero depending on whether the single index is larger or smaller than the change point. Indeed, we will prove that the convergence rate of the coefficients in the change hyperplane is $1/n$ in contrast to the convergence rate $1/\sqrt{n}$ in the single index model. Consequently, the existing methods for the Cox proportional hazards model with an unknown link function of the single index cannot be applied to the Cox proportional hazards model with a change hyperplane.

In this paper, we propose a partially linear Cox proportional hazards model with a change hyperplane. The model includes both the covariates which have a linear relationship with the log-hazard rate of the disease incidence and those defining the change hyperplane. The change hyperplane parameters are identifiable under certain conditions. We propose a genetic optimization algorithm (Sekhon and Mebane 1998) for estimating parameters. One major challenge for the Cox proportional hazards model with a change hyperplane is the complicated asymptotic distribution of the change hyperplane estimators. The existing theory of the change point estimator for a single covariate cannot be applied to the proof in the Cox proportional hazards model with a change hyperplane. The asymptotic distribution of the change hyperplane estimators for multiple covariates depends on the joint distribution of the all the covariates in the change hyperplane. We will show that the asymptotic distribution of the proposed change hyperplane estimators is an integrated composite Poisson process.

The structure of this paper is as follows. In Section 4.2, we describe the estimation method based on a two-step procedure. We then provide an inference method based on the m out of n bootstrap and a testing procedure for the existence of a change hyperplane. In Section 4.3, we provide the asymptotic properties of the proposed estimators. Simulation studies evaluating the small sample performance of the method are presented in Section 4.4. The Cardiovascular Health Study is analyzed in Section 4.5.

4.2 Methods

4.2.1 Model and Parameter Estimation

For $i = 1, ..., n$, let $\widetilde{T}_i$ and C_i be the failure times and censoring times with respect to the ith subject. The Cox proportional hazards model with a change hyperplane assumes that the hazard rate function for $\widetilde{T}_i$ given $\boldsymbol{W}_i(t) \equiv (\boldsymbol{X}_i^T, \boldsymbol{Z}_i^T(t))^T$ takes a form

$$\lambda(t|\boldsymbol{W}_i) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(t) + \beta_2 I(\boldsymbol{\eta}^T \boldsymbol{X}_i > 0) + \boldsymbol{\beta}_3^T \boldsymbol{Z}_i(t) I(\boldsymbol{\eta}^T \boldsymbol{X}_i > 0)\right\},$$

where $\lambda_0(t)$ is an unknown baseline function, $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1^T, \beta_2, \boldsymbol{\beta}_3^T)^T$ is a vector of $2p_2 + 1$ unknown parameters, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_{p_1}, \eta_0)^T$ is a vector of $p_1 + 1$ unknown change hyperplane parameters, $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip_1}, -1)^T$ is a vector of the change hyperplane covariates, and $\boldsymbol{Z}_i(t) = (Z_{i1}(t), ..., Z_{ip_2}(t))^T$ is a vector of other potentially time-dependent covariates. Then the change hyperplane is denoted as $I(\boldsymbol{\eta}^T\boldsymbol{X}_i > 0) = I(\eta_1 X_{i1} + \eta_2 X_{i2} + ... + \eta_{p_1} X_{ip_1} - \eta_0 > 0)$. The covariate $\boldsymbol{Z}_i(t)$ has different effects for $\boldsymbol{\eta}^T\boldsymbol{X}_i \leq 0$ and $\boldsymbol{\eta}^T\boldsymbol{X}_i > 0$. Therefore, the proposed model implies that the effect of $\boldsymbol{Z}_i(t)$ is $\boldsymbol{\beta}_1$ when $\boldsymbol{\eta}^T\boldsymbol{X}_i \leq 0$, and it becomes $(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_3)$ when $\boldsymbol{\eta}^T\boldsymbol{X}_i > 0$. Furthermore, the hazard ratio between two groups $\boldsymbol{\eta}^T\boldsymbol{X}_i > 0$ and $\boldsymbol{\eta}^T\boldsymbol{X}_i \leq 0$ is $\exp\{\beta_2 + \boldsymbol{\beta}_3^T\boldsymbol{Z}_i(t)\}$, given the fixed $\boldsymbol{Z}_i(t)$. When $p_1 = 1$, it reduces to the change point model in Pons (2003).

We assume that C_i is independent of $\widetilde{T}_i$ given the covariates $\boldsymbol{W}_i$. If we define $r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\boldsymbol{W}_i(t)\} \equiv \boldsymbol{\beta}_1^T\boldsymbol{Z}_i(t) + \beta_2 I(\boldsymbol{\eta}^T\boldsymbol{X}_i > 0) + \boldsymbol{\beta}_3^T\boldsymbol{Z}_i(t) I(\boldsymbol{\eta}^T\boldsymbol{X}_i > 0)$, then the partial likelihood function for n subjects with right censoring can be formulated as

$$L_n(\boldsymbol{\eta}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp\left[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\boldsymbol{W}_i(T_i)\}\right]}{\sum_{l=1}^{n} I(T_l \geq T_i) \exp\left[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\boldsymbol{W}_l(T_i)\}\right]} \right)^{\Delta_i},$$

where $T_i = \min(\widetilde{T}_i, C_i)$ and $\Delta_i = I(\widetilde{T}_i \leq C_i)$.

To estimate the model parameters, we propose to maximize the logarithm of the partial likelihood function, which is defined as $l_n(\boldsymbol{\eta}, \boldsymbol{\beta}) \equiv \log\{L_n(\boldsymbol{\eta}, \boldsymbol{\beta})\}$. For the purpose of identifiability of the model with a change hyperplane, we impose the constraints that $\eta_1^2 + \eta_2^2 + ... + \eta_{p_1}^2 = 1$ and the coefficient of the first continuous covariate η_1 is positive. The first condition restricts the scale of $\boldsymbol{\eta}$ and the second condition determines the uniqueness of sign for each parameter in η. The proof of the identifiability under these constraints is shown in Theorem 4.3.1 in Proof of Lemma and Theorems. The estimation of the change point for a single covariate is based on maximizing the partial likelihood with a two-step procedure (Luo and Boyett 1997). We propose a similar procedure to obtain the change hyperplane estimates. In the first step, for any fixed value of $\boldsymbol{\eta}$ in a pre-specified range

$S_0 = \{(\eta_1, ..., \eta_{p_1}, \eta_0) : \eta_1^2 + \eta_2^2 + ... + \eta_{p_1}^2 = 1, \eta_1 > 0, \eta_0 \in [\eta_a, \eta_b]\}$, we obtain the estimates of $\boldsymbol{\beta}$ by applying the Newton-Raphson method to maximize the logarithm of the partial likelihood function. In the second step, we apply an evolutionary algorithm with a derivative-based, quasi-Newton method to maximize the profile function for $\boldsymbol{\eta}$ in the first step (Sekhon and Mebane 1998). This evolutionary algorithm has been widely applied to optimize the function when the objective function is not a continuous function of the parameter of interest. Thus, we obtain $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) = \underset{\boldsymbol{\eta} \in S_0, \boldsymbol{\beta}}{\arg\max} l_n(\boldsymbol{\eta}, \boldsymbol{\beta})$.

4.2.2 Inference

We will prove that $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\beta}}$ are asymptotically independent. In addition, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ remains to be normal no matter whether $\boldsymbol{\eta}$ is known or not. The theorems are shown in Section 4.3 and the details of proofs are in Proof of Lemma and Theorems. Consequently, the inference of $\boldsymbol{\beta}$ can be based on the exact covariance matrix in the Cox proportional hazards model and the corresponding confidence intervals are generated by normal approximation. One challenge question is how to make inference for $\boldsymbol{\eta}$ because it is impossible to derive the covariance matrix based on the asymptotic distribution shown in Section 4.3. Shao (1994), Bickel et al. (2012), Politis and Romano (1999), and Xu et al. (2014) proposed the m out of n bootstrap to generate the 95% confidence intervals under this situation. Additionally, Xu et al. (2014) proved the consistency of the m out of n bootstrap in the Cox proportional hazards model with a change point. For the m out of n bootstrap, m is usually determined by the data-driven approaches (Hall et al. 1995; Lee 1999; Cheung et al. 2005; Bickel and Sakov 2005; Bickel and Sakov 2008). We choose to adapt the algorithm proposed by Bickel and Sakov (2008) to select m. In this algorithm, m_j is selected from a group of possible values as the maximum sample size that achieves the stable empirical distribution for each η_j, where $j = 0, 1, ..., p_1$. The common m is the minimum among all the m_j, $j = 0, 1, ..., p_1$. Here is the algorithm about how to select m.

(1) Construct a sequence of $m_{js} = [s \times \frac{n}{q}]$, where $s = q, q-1, ..., 1$, q is a reasonable interval between two evaluation points, and $[a]$ is the smallest integer larger than a.

(2) For the m_{js} out of n bootstrap, the empirical cumulative distribution function for the change hyperplane estimator is constructed as follows:

$$R_{m_{js}}(x, \widehat{\eta}_j) = \frac{1}{B} \sum_{b=1}^{B} I \left\{ m_{js} \left(\widehat{\eta}_{j(m_{js})}^{(b)} - \widehat{\eta}_j \right) \leq x \right\},$$

where $\widehat{\eta}_j$ is the change hyperplane estimator based on the full dataset, $\widehat{\eta}_{j(m_{js})}^{(b)}$ is the change hyperplane estimator based on the dataset with m_{js} samples in the bth replication, $b = 1, 2..., B$, and B is the total number of bootstrap replications.

(3) The m_j will be selected as the maximum value which minimizes the supremum difference between two adjacent empirical cumulative distributions in the m_{js} sequence.

$$\widehat{m}_j = \max \arg \min_{m_{js}} \sup_x \left| R_{m_{js}}(x, \widehat{\eta}_j) - R_{m_{js+1}}(x, \widehat{\eta}_j) \right|$$

(4) Finally, $\widehat{m} = \min(\widehat{m}_0, \widehat{m}_1, ..., \widehat{m}_{p_1})$.

Both the standard error estimator for $\widehat{\boldsymbol{\eta}}$ and the confidence interval for $\boldsymbol{\eta}$ will be adjusted by n/m based on the convergence rate $1/n$ of $\widehat{\boldsymbol{\eta}}$ (Theorem 4.3.3). Particularly, the equal-tailed 95% confidence intervals are generated as $\left[\widehat{\boldsymbol{\eta}} - \frac{Q_{\widehat{\boldsymbol{\eta}},0.95}}{n/m}, \widehat{\boldsymbol{\eta}} + \frac{Q_{\widehat{\boldsymbol{\eta}},0.95}}{n/m} \right]$, where $Q_{\widehat{\boldsymbol{\eta}},0.95}$ is the 95th quantile of the absolute value $\left| \widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_m^{(b)} \right|$ for $b = 1, 2..., B$.

4.2.3 Hypothesis Testing for the Change Hyperplane

In practice, one important question is whether the change hyperplane exists. Equivalently, we wish to test the null hypothesis $H_0 : \beta_2 = 0, \boldsymbol{\beta}_3^T = \mathbf{0}$ in our proposed model. Since the estimation of the change hyperplane relies on either β_2 or $\boldsymbol{\beta}_3$ unequal to zero, the model is not identifiable given the fact that both β_2 and $\boldsymbol{\beta}_3$ are zero under the null hypothesis. The supremum (SUP) tests is proposed to verify the existence of the change point based on single covariate (Davies 1977, Davies 1987, Kosorok and Song 2007). Here, we extend this SUP test with score statistics to multi-dimensional covariates. Specifically, our test statistic is

$$\mathrm{SUP}_{k^{p_1}} = \sup_{\eta_j \in [\eta_{j1},...,\eta_{jk}], j=2,...,p_1,0} \boldsymbol{U}(\boldsymbol{\eta})^T \boldsymbol{\Sigma}(\boldsymbol{\eta})^{-1} \boldsymbol{U}(\boldsymbol{\eta}),$$

where $\boldsymbol{U}(\boldsymbol{\eta}) = \frac{\partial l_n(\boldsymbol{\eta},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$, $\boldsymbol{\Sigma}(\boldsymbol{\eta}) = -\frac{\partial^2 l_n(\boldsymbol{\eta},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}$, p_1 is the dimension of the covariates in the change hyperplane, and k is the number of grids selected for each η_j, $j = 2, ..., p_1, 0$. We use permutation to generate the null distribution of the proposed test statistic. Under the null hypothesis, there is no change hyperplane effect on the response. Thus, we randomly shuffle the covariate $\boldsymbol{X}_i$ for sufficient times. Then, we obtain the permutation distribution of the proposed test statistics. We reject the null hypothesis at a significance level of α if $\mathrm{SUP}_{k^{p_1}}$ is larger than the upper α-quantile of the permutation distribution.

4.3 Asymptotic Properties

The consistency and asymptotic distributions of the estimators for both the change hyperplane and the regression parameters are established in this section. We prove these asymptotic properties based on the following conditions. The details of proofs are shown in Proof of Lemma and Theorems.

(C.1) The joint density of $(X_{i1}, X_{i2}, ..., X_{ip_1})$ is assumed to be strictly positive, bounded and continuous in a neighborhood $V_0 = \{\boldsymbol{x} : |\boldsymbol{\eta}_0^T \boldsymbol{x}| < \epsilon\}$. In addition, the joint density of $(Z_{i1}(t), ..., Z_{ip_2}(t))$ is assumed to be strictly positive and bounded.

(C.2) For any $V_\delta(\boldsymbol{\eta}_0) = \{\boldsymbol{\eta} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \delta\}$, the covariance matrix

$$I(\boldsymbol{\eta},\boldsymbol{\beta}) = \int_0^\tau v(t;\boldsymbol{\eta},\boldsymbol{\beta}) s^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta}) \lambda_0(t) dt$$

is positive definite, where $\boldsymbol{\eta}_0$ is the true value of $\boldsymbol{\eta}$, $\boldsymbol{v}(t;\boldsymbol{\eta},\boldsymbol{\beta}) = \boldsymbol{s}^{(2)}(t;\boldsymbol{\eta},\boldsymbol{\beta})/s^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta}) - \left[\boldsymbol{s}^{(1)}(t;\boldsymbol{\eta},\boldsymbol{\beta})/s^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})\right]^{\otimes 2}$, $\boldsymbol{s}^{(r)}(t;\boldsymbol{\eta},\boldsymbol{\beta}) = \mathrm{E}\left(Y_i(t) \widetilde{\boldsymbol{Z}}_i^{\otimes r}(t;\boldsymbol{\eta}) \exp\left[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\left\{\boldsymbol{W}_i(t)\right\}\right]\right)$, $\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}) =$

$(\boldsymbol{Z}_i^T(t), I(\boldsymbol{\eta}^T\boldsymbol{X}_i > 0), \boldsymbol{Z}_i^T(t)I(\boldsymbol{\eta}^T\boldsymbol{X}_i > 0))^T$, and $r = 0, 1, 2$. In addition,

$$\lambda_{\min}\left(\int_0^\tau \mathrm{E}\left[Y_i(t)\left\{1, \boldsymbol{Z}_i(t)\right\}^{\otimes 2} \middle| \boldsymbol{\eta}_0^T\boldsymbol{X}_i = 0\right] d\Lambda_0(t)\right) > 0,$$

where $\lambda_{\min}(\boldsymbol{A})$ is the smallest eigenvalue of any square matrix $\boldsymbol{A}$.

(C.3) The random process $\sup_{t\in[0,\tau]} \sup_{\|\boldsymbol{\eta}-\boldsymbol{\eta}_0\|<\delta} \left\|\boldsymbol{S}_n^{(r)}(t;\boldsymbol{\eta},\boldsymbol{\beta}) - \boldsymbol{s}^{(r)}(t;\boldsymbol{\eta},\boldsymbol{\beta})\right\|$ converges almost surely to zero, where $\boldsymbol{s}^{(r)}(t;\boldsymbol{\eta},\boldsymbol{\beta}) < \infty$, and $r = 0, 1, 2$. When $r = 0$, $\boldsymbol{s}^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})$ is bounded away from zero.

(C.4) $\sup_{t\in[0,\tau]} \lambda_0(t) < \infty$, and $P(Y_i(t) = 1) > 0$ for all $t \in [0, \tau]$.

(C.1) and (C.2) are needed for the identifiability of the change hyperplane and regression coefficients. (C.2) holds if $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, ..., \boldsymbol{Z}_n)$ has a full rank given $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)^T$. (C.3) guarantees that $\boldsymbol{S}_n^{(r)}(t;\boldsymbol{\eta},\boldsymbol{\beta})$ converges almost surely to $\boldsymbol{s}^{(r)}(t;\boldsymbol{\eta},\boldsymbol{\beta})$. (C.4) requires that $\lambda_0(t)$ is bounded and the at risk probability is non-zero for $t \in [0, \tau]$.

Our first theorem establishes the identifiability of the change hyperplane parameters and regression coefficient parameters.

Theorem 4.3.1. *Under the condition that* $\|\boldsymbol{\eta}\| = 1$ *with the first continuous covariate* $\eta_1 > 0$ *and at least one of the elements in* β_2 *or* $\boldsymbol{\beta}_3$ *is nonzero, the change hyperplane parameters* $\boldsymbol{\eta}$ *and regression parameters* $\boldsymbol{\beta}$ *are identifiable.*

Theorem 4.3.2 and Theorem 4.3.3 show the consistency and convergence rates of the change hyperplane estimators and regression coefficients estimators. Theorem 4.3.3 implies that the convergence rates for $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\beta}}$ are $1/n$ and $1/\sqrt{n}$, respectively. These rates will be applied to establish the asymptotic distributions of the estimators in Theorem 4.3.4.

Theorem 4.3.2. *Under conditions (C.1)-(C.4), there exists a neighborhood of* $\boldsymbol{\eta}_0$ *and* $\boldsymbol{\beta}_0$*, where* $\widehat{\boldsymbol{\eta}}$ *and* $\widehat{\boldsymbol{\beta}}$ *converge in probability to* $\boldsymbol{\eta}_0$ *and* $\boldsymbol{\beta}_0$ *as* $n \to \infty$*, respectively.*

Theorem 4.3.3. *Under conditions (C.1)-(C.4),*

$$\lim_{A\to\infty}\lim_{n\to\infty} P_0\left(n\|\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\|>A\right)=0,$$
$$\lim_{A\to\infty}\lim_{n\to\infty} P_0\left(n^{1/2}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|>A\right)=0.$$

Thus, $\|\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\|=o_p(1/n)$ *and* $\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|=o_p(1/\sqrt{n})$.

Let $\boldsymbol{\eta}_{n,u_1}=\boldsymbol{\eta}_0+n^{-1}\boldsymbol{u}_1$, $\boldsymbol{\beta}_{n,u_2}=\boldsymbol{\beta}_0+n^{-1/2}\boldsymbol{u}_2$, and $W_{i0}=\boldsymbol{\eta}_0^T\boldsymbol{X}_i$, where $\boldsymbol{u}_1=(a_1,a_2,...,a_{p_1})^T$ and $\boldsymbol{u}_2=(b_1,b_2,...,b_{2p_2+1})^T$. To derive the asymptotic distributions of the change hyperplane estimations, we first partition the $l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2})-l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)$ into three terms and obtain their asymptotic expansions. In Theorem 4.3.4, we show that

$$l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2})-l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)=Q_n(\boldsymbol{u}_1)+\boldsymbol{u}_2^T\widetilde{\boldsymbol{l}}_n-\frac{1}{2}\boldsymbol{u}_2^T\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\boldsymbol{u}_2+o_p(1),$$

where

$$\begin{aligned}Q_n(\boldsymbol{u}_1)=\sum_{i=1}^n\Delta_i\Big[&\left\{\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_i(T_i)\right\}\\&\left[I\left\{0\geq W_{i0}>n^{-1}\boldsymbol{X}_i^T\boldsymbol{u}_1,\boldsymbol{X}_i^T\boldsymbol{u}_1<0\right\}-I\left\{0<W_{i0}\leq n^{-1}\boldsymbol{X}_i^T\boldsymbol{u}_1,\boldsymbol{X}_i^T\boldsymbol{u}_1\geq 0\right\}\right]\\&-\frac{S_n^{(0)}(T_i;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)-S_n^{(0)}(T_i;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(T_i;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\Bigg],\\\widetilde{\boldsymbol{l}}_n=n^{-1/2}&\sum_{i=1}^n\int_0^\tau\left\{\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_0)-\frac{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right\}dM_i(t),\end{aligned}$$

where $M_i(t)=N_i(t)-\int_0^\tau Y_i(s)\exp\{r_{\boldsymbol{\eta}_0,\boldsymbol{\beta}_0}(\boldsymbol{W}_i(s))d\Lambda_0(s)\}$. The asymptotic distributions are based on the first two terms in the partition. Then, we obtain Theorem 4.3.4 using the argmax continuous mapping theorem.

We define that $V_l^+(\boldsymbol{x})$ and $V_l^-(\boldsymbol{x})$ are random variables with the conditional distributions of η^+ and η^- given $W = 0$ and $\boldsymbol{X} = \boldsymbol{x}$, respectively, where

$$\eta^- = -\Delta\left(\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}(T)\right\} - \int_0^\tau \phi(t) d\Lambda_0(t)\right),$$
$$\eta^+ = -\Delta\left(\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}(T)\right\} + \int_0^\tau \phi(t) d\Lambda_0(t)\right),$$
$$\phi(t) = Y(t)\exp\left\{\boldsymbol{\beta}_{10}^T \boldsymbol{Z}(t)\right\}\left[1 - \exp\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}(t)\right\}\right].$$

In addition, $V_l^+(\boldsymbol{x})$ and $V_l^-(\boldsymbol{x})$ are assumed to be independent for any $\boldsymbol{x}$. We further define that $v^+(\boldsymbol{x}, t)$ and $v^-(\boldsymbol{x}, t)$ are multivariate Poisson processes with Poisson intensity $E[v^+(d\boldsymbol{x}, dt)] = f_{\boldsymbol{X}}(\boldsymbol{x}) f_W(0+) d\boldsymbol{x} dt$ for $t > 0$ and $E[v^-(d\boldsymbol{x}, dt)] = f_{\boldsymbol{X}}(\boldsymbol{x}) f_W(0-) d\boldsymbol{x} dt$ for $t < 0$, respectively, where $f_{\boldsymbol{X}}(\boldsymbol{x})$ is the joint density of $\boldsymbol{X}$, $f_W(0+)$ is the density function of $\boldsymbol{\eta}_0^T \boldsymbol{X}$ at zero given $\boldsymbol{\eta}_0^T \boldsymbol{X} \leq 0$, and $f_W(0-)$ is the density function of $\boldsymbol{\eta}_0^T \boldsymbol{X}$ at zero given $\boldsymbol{\eta}_0^T \boldsymbol{X} > 0$. Thus, the asymptotic distribution is defined as $Q(\boldsymbol{u}_1) \equiv Q^+(\boldsymbol{u}_1) - Q^-(\boldsymbol{u}_1)$, where

$$Q^+(\boldsymbol{u}_1) \equiv \sum_{\boldsymbol{x}} \sum_{0 \leq l \leq v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{u}_1)} V_l^+(\boldsymbol{x}),$$
$$Q^-(\boldsymbol{u}_1) \equiv \sum_{\boldsymbol{x}} \sum_{0 \leq l \leq v^-(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{u}_1)} V_l^-(\boldsymbol{x}),$$

Theorem 4.3.4. *Under conditions (C.1)-(C.4), $n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ and $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ are asymptotically independent. Furthermore, $n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges weakly to* $\inf\{\boldsymbol{u}_1 : \arg\max Q(\boldsymbol{u}_1)\}$, *and $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges weakly to $N(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)^{-1})$.*

4.4 Simulation Studies

We conducted simulation studies to evaluate the performance of our proposed method. Our first set of studies was designed to assess the performance of the estimators and the coverage rate of the confidence interval. We considered one covariate $Z \sim$ Uniform(-1,1) and the change hyperplane with two covariates $X_1 \sim N(2, 1.5^2)$ and $X_2 \sim N(0, 1)$. We generated the survival times $\widetilde{T}_i$ under the proportional hazards model $\Lambda(t|X_1, X_2, Z) = t\exp\{\beta_1 Z +$

$\beta_2 I(\eta_1 X_1 + \eta_2 X_2 - \eta_0 > 0) + \beta_3 Z I(\eta_1 X_1 + \eta_2 X_2 - \eta_0 > 0)\}$, where $(\beta_1, \beta_2, \beta_3) = (-1, 1.8, 0.5)$, $(\eta_1, \eta_2, \eta_0) = (0.8, -0.6, 1.7)$, and $\eta_1^2 + \eta_2^2 = 1$. In order to obtain the censoring rates of 10%, 30%, and 50%, we generated the censoring time from Uniform (0,680), Uniform(0,220), and Uniform(0,118), respectively. The number of subjects is 200 or 300. To select m for each simulation, we considered q to be 10. All results are based on 500 replications and each m out of n bootstrap consists of 100 replicates.

In Table 4.4, the proposed method provides approximately unbiased estimates for the change hyperplane parameters η_2 and η_0. Here, we only presented the results for η_2 and η_0 because that η_1 and η_2 satisfy $\eta_1^2 + \eta_2^2 = 1$. In addition, the m out of n bootstrap generates proper coverage rates. When the number of subjects increases or the censoring rate decreases, the bias of the change point estimate and the variance estimates decrease. In Table 4.5, the results show that the estimates for the regression coefficients $\boldsymbol{\beta}$ are also approximately unbiased and the confidence intervals using normal approximation have proper coverage rates.

Table 4.4: Simulation Results for the Change Hyperplane Parameters.

Censoring Rate	Sample Size	Parameters	Bias	SSD	95% CI	Length
50%	200	$\widehat{\eta}_2$	0.0017	0.083	0.960	0.427
		$\widehat{\eta}_0$	0.0111	0.157	0.952	0.739
	300	$\widehat{\eta}_2$	0.0006	0.052	0.956	0.283
		$\widehat{\eta}_0$	0.0083	0.095	0.946	0.506
30%	200	$\widehat{\eta}_2$	0.0040	0.064	0.964	0.325
		$\widehat{\eta}_0$	0.0076	0.117	0.966	0.569
	300	$\widehat{\eta}_2$	-0.0013	0.040	0.962	0.210
		$\widehat{\eta}_0$	0.0123	0.077	0.950	0.373
10%	200	$\widehat{\eta}_2$	-0.0023	0.050	0.972	0.269
		$\widehat{\eta}_0$	0.0181	0.098	0.958	0.468
	300	$\widehat{\eta}_2$	0.0020	0.040	0.954	0.179
		$\widehat{\eta}_0$	0.0086	0.071	0.956	0.317

NOTE: SSD stands for sample standard deviation. Length is the length of the 95% CI.

Our second set of simulation studies were aimed at comparing type I error and power of the SUP_{5^2}, SUP_{10^2}, and SUP_{20^2} tests under various scenarios. Since our test is based

Table 4.5: Simulation Results for the Regression Parameters.

Censoring Rate	Sample Size	Parameters	Bias	SSD	SSE	95% CI
50%	200	$\widehat{\beta}_1$	-0.0469	0.332	0.344	0.944
		$\widehat{\beta}_2$	0.1163	0.255	0.244	0.944
		$\widehat{\beta}_3$	0.0392	0.399	0.407	0.950
	300	$\widehat{\beta}_1$	-0.0254	0.267	0.270	0.954
		$\widehat{\beta}_2$	0.0657	0.204	0.205	0.950
		$\widehat{\beta}_3$	0.0051	0.321	0.324	0.952
30%	200	$\widehat{\beta}_1$	-0.0346	0.250	0.245	0.962
		$\widehat{\beta}_2$	0.0854	0.218	0.209	0.952
		$\widehat{\beta}_3$	0.0189	0.320	0.314	0.956
	300	$\widehat{\beta}_1$	-0.0240	0.202	0.206	0.948
		$\widehat{\beta}_2$	0.0576	0.175	0.171	0.950
		$\widehat{\beta}_3$	0.0035	0.259	0.264	0.956
10%	200	$\widehat{\beta}_1$	-0.0228	0.210	0.207	0.950
		$\widehat{\beta}_2$	0.0692	0.197	0.191	0.952
		$\widehat{\beta}_3$	0.0112	0.281	0.273	0.962
	300	$\widehat{\beta}_1$	-0.0153	0.170	0.181	0.942
		$\widehat{\beta}_2$	0.0457	0.160	0.169	0.926
		$\widehat{\beta}_3$	0.0031	0.228	0.229	0.952

NOTE: SSD and SEE stand for sample standard deviation and average standard error estimate, respectively.

on two change hyperplane parameters, the SUP test will be evaluated on the set with k^2 points, where k is the number of grids in the pre-specified range $[-1, 1]$ for η_2 and $[-10, 10]$ for η_0. The range for η_2 is determined by the conditions in Theorem 4.3.1. The range of η_0 is determined by the range of each covariate as well as the value of η_2. For example, the test SUP_{5^2} stands for the test which is evaluated on the grids $[-1, -0.5, 0, 0.5, 1] \times [-10, -5, 0, 5, 10]$. We examine the performance of these tests with the sample sizes 200, 300, and 400. The results for type I error and power are based on 10000 and 1000 replicates, respectively. All the other specifications are the same as the first set of simulations.

Table 4.6 shows that type I errors of all three tests are generally closer to 0.05. As the sample sizes increase and the censoring rates decrease, the type I errors get close to 0.05. For the power, the performance of the supremum tests is determined by the numbers of grids, sample sizes, and censoring rates. Given the same sample size and censoring rate, the power gets stabilized after the number of grids reaches 10 for each parameter. Given the tests with the same number of grids, the power increases as the sample size increases and the censoring rate decreases.

Table 4.6: Type I Error and Power for SUP Tests for the Existence of the Change Hyperplane

			Sample Size		
(β_{20}, β_{30})	Censoring Rate	Test	200	300	400
$\beta_{20} = \beta_{30} = 0$	10%	$SUP5^2$	0.056	0.050	0.051
		$SUP10^2$	0.051	0.053	0.052
		$SUP20^2$	0.049	0.053	0.051
	30%	$SUP5^2$	0.054	0.048	0.053
		$SUP10^2$	0.052	0.051	0.054
		$SUP20^2$	0.049	0.058	0.054
	50%	$SUP5^2$	0.054	0.049	0.051
		$SUP10^2$	0.055	0.050	0.052
		$SUP20^2$	0.051	0.055	0.052
$\beta_{20} = 0.8, \beta_{30} = -0.4$	10%	$SUP5^2$	0.144	0.260	0.294
		$SUP10^2$	0.718	0.846	0.972
		$SUP20^2$	0.748	0.940	0.996
	30%	$SUP5^2$	0.110	0.280	0.294
		$SUP10^2$	0.700	0.852	0.958
		$SUP20^2$	0.744	0.948	0.988
	50%	$SUP5^2$	0.094	0.196	0.232
		$SUP10^2$	0.600	0.772	0.904
		$SUP20^2$	0.602	0.876	0.960

4.5 Application to the Cardiovascular Health Study

We applied the proposed method to the Cardiovascular Health Study (CHS). The CHS recruited 5888 participants aged 65 years and older from four U.S. communities to study the development and progression of CHD and stroke. We applied our approach to the cohort of male participants, who were free of CHD at baseline. It resulted in 995 subjects after excluding the ones with missing responses and covariates. Among them, there are 851 subjects developed CHD before the end of the study. We included the linear combination of HDL, systolic blood pressure, and cholesterol level to form the risk categories (high vs.

low). We investigated the association between these risk categories and the risk of CHD in a Cox proportional hazards model, which is adjusted by the confounding covariates age, hypertension, diabetes, and smoking status.

The analysis was conducted in the following two steps. First, we applied the SUP_{10^3} test to verify the existence of these risk categories. The test is significant with p-value less than 0.01. Second, we obtained the parameter estimates to form the risk categories by applying the two-step estimating procedures. The corresponding 95% confidence intervals was generated by the m out of n bootstrap. The results were summarized in Table 4.7. All the estimates are significant and included in the final model. Based on these risk categories, the regression coefficient estimates were summarized in Table 4.8. Except for hypertension, all the other covariates have significant effects. The hazard ratio of CHD for the low risk group $I(\boldsymbol{\eta}^T\boldsymbol{X} > 0)$ vs. the high risk group $I(\boldsymbol{\eta}^T\boldsymbol{X} < 0)$ is 0.652. To show the survival functions of these two risk groups, we generated the Kaplan Meier curves in Figure 4.2.

Table 4.7: Change Hyperplane Estimates Based on the CHS

Change Hyperplane Covariate	Estimate	95% CI
HDL	0.671	[0.338, 1.003]
SBP	-0.604	[-0.796, -0.412]
CHOL	-0.431	[-0.811, -0.051]
Change Point	-0.209	-

Table 4.8: Regression Coefficients Estimates Based on the CHS

	Estimate	exp(Est)	p-value
Age	0.071	1.073	< 0.01
Change Hyperplane	-0.428	0.652	< 0.01
Diabetes	0.385	1.469	< 0.01
Smoke	0.315	1.370	< 0.01
Hypertension	0.027	1.028	0.707

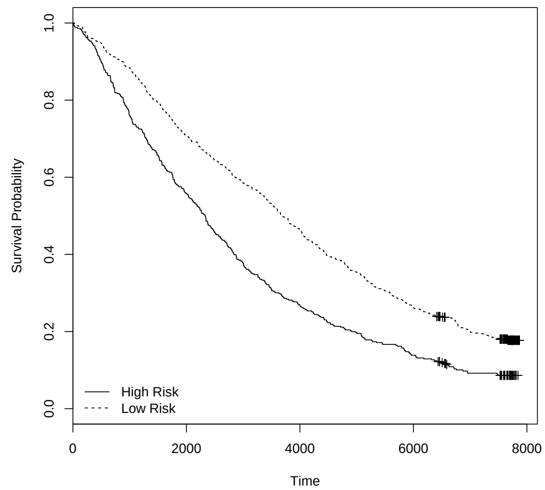

Figure 4.2: The Kaplan-Meier Plot of the Change Hyperplane (the logrank test with p-value< 0.001).

4.6 Conclusion

Change point effects are commonly observed in regression problems. Although a number of approaches have been developed to estimate the change point that is based on a single covariate, no research has been done for the change hyperplane that is based on multiple covariates. In this paper, we developed for the first time a two-step approach to estimate the change hyperplane parameters and a testing procedure to verify the existence of a change hyperplane for univariate survival data. We developed an adaptive m out of n bootstrap to construct the confidence interval and provided an easy way to determine the appropriate m. We proved the asymptotic properties of the proposed change hyperplane estimators. As shown in our simulation studies, the estimator is approximately unbiased and its confidence interval has a good coverage rate.

We applied our methods to estimate the change hyperplane parameters based on LTL and the lipid profiles for diabetes incidence in the SHFS. In the final model, the change hyperplane consists of LTL and total triglycerides. This finding is consistent with the conclusion in Chapter 3 and improved the understanding of the relationship between LTL and total triglycerides.

In this paper, we consider the situation that the linear combination of the multiple risk factors has only one change point. In reality, the change hyperplane may have multiple change points. Instead of categorizing the participants into low and high risk groups, we may further define a moderate risk group. In this situation, the inference procedures and the asymptotic properties cannot be directly extended to the change hyperplane with multiple thresholds. Thus, it is essential to devise valid and efficient inference procedures for general change hyperplane models in the future. Moreover, when the proportional hazards assumption is violated, we could extend the change hyperplane model to other survival models, e.g. additive hazard models and accelerated failure time model. Such extension will have a wide application in the univariate and clustered survival data.

4.7 Proof of Lemma and Theorems

To simplify the proofs, we write the proposed change hyperplane $I(\eta_1^* X_{i1} + \eta_2^* X_{i2} + ... + \eta_{p_1}^* X_{ip_1} - \eta_0^* > 0)$ in an equivalent form $I(X_{i1} + \eta_2 X_{i2} + ... + \eta_{p_1} X_{ip_1} - \eta_0 > 0)$, where η_1^* is the first positive nonzero component of $\boldsymbol{\eta}^*$, $\eta_1^* = \frac{1}{\sqrt{1+\eta_2^2+...+\eta_{p_1}^2}}$, $\eta_j^* = \frac{\eta_j}{\sqrt{1+\eta_2^2+...+\eta_{p_1}^2}}$ for $j = 2, ..., p_1$, and $\eta_0^* = \frac{\eta_0}{\sqrt{1+\eta_2^2+...+\eta_{p_1}^2}}$. Thus, we define $\boldsymbol{\eta} = (1, \eta_2, ..., \eta_{p_1}, \eta_0)^T$ to be a vector of $p_1 + 1$ elements and $\boldsymbol{\eta}^T \boldsymbol{X}_i = X_{i1} + \eta_2 X_{i2} + ... + \eta_{p_1} X_{ip_1} - \eta_0$. The following proofs are carried out based on this equivalent form.

For convenience, we define $V_\delta(\boldsymbol{\eta}_0) = \{\boldsymbol{\eta} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \delta\}$, $V_\epsilon(\boldsymbol{\beta}_0) = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon\}$,

$$\boldsymbol{s}^{(r)+}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \mathrm{E}\left[Y_i(t) I(\boldsymbol{\eta}^T \boldsymbol{X}_i > 0) \boldsymbol{Z}_i^{\otimes r}(t) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(t) + \beta_2 + \boldsymbol{\beta}_3^T \boldsymbol{Z}_i(t)\right\} \middle| \boldsymbol{X}_i\right],$$

$$\boldsymbol{s}^{(r)-}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \mathrm{E}\left[Y_i(t) I(\boldsymbol{\eta}^T \boldsymbol{X}_i \le 0) \boldsymbol{Z}_i^{\otimes r}(t) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(t)\right\} \middle| \boldsymbol{X}_i\right],$$

where $r = 0, 1$.

Proof of Theorem 4.3.1

The proof of the identifiability of $\boldsymbol{\eta}^*$ is equivalent to the proof of the identifiability of $\boldsymbol{\eta}$. To prove the parameters are identifiable, we need to show that $\boldsymbol{\eta}$ can be uniquely determined from the distribution function. In the Cox model, we assume that the failure times and censoring times are independent. Thus, the full likelihood function for subject i is written as

$$P(\boldsymbol{\eta}, \boldsymbol{\beta}) = \{f(T_i)S_c(T_i)\}^{\Delta_i}\{S(T_i)f_c(T_i)\}^{1-\Delta_i} \propto S(T_i)\lambda(T_i)^{\Delta_i},$$

where $f(T_i)$ and $S(T_i)$ are the density function and survival function of the failure time, respectively. Similarly, $f_c(T_i)$ and $S_c(T_i)$ are the density function and survival function of the censoring time, respectively. Based on this full likelihood function, we apply the mathematical induction to prove the identifiability.

First, we prove the identifiability for the case with only one factor in the change hyperplane, i.e., $I(X_i > \eta_0)$, where X_i is continuous and η_0 is in the support of X_i. Assume that $(\eta_0, \boldsymbol{\beta})$ and $(\widetilde{\eta}_0, \widetilde{\boldsymbol{\beta}})$ are two sets of parametrization for the model and satisfy the conditions specified in the theorem. Let

$$\begin{aligned} m(\eta_0, \boldsymbol{\beta}|t) = &\left[\lambda_0(t)\exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(t) + \beta_2 I(X_i > \eta_0) + \boldsymbol{\beta}_3^T \boldsymbol{Z}_i(t) I(X_i > \eta_0)\right\}\right]^{\Delta_i} \\ &\times \exp\left[-\int_0^t \lambda_0(s)\exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(s) + \beta_2 I(X_i > \eta_0) + \boldsymbol{\beta}_3^T \boldsymbol{Z}_i(s) I(X_i > \eta_0)\right\} ds\right], \\ h\left(\widetilde{\eta}_0, \widetilde{\boldsymbol{\beta}}|t\right) = &\left[\widetilde{\lambda}_0(t)\exp\left\{\widetilde{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_i(t) + \widetilde{\beta}_2 I(X_i > \widetilde{\eta}_0) + \widetilde{\boldsymbol{\beta}}_3^T \boldsymbol{Z}_i(t) I(X_i > \widetilde{\eta}_0)\right\}\right]^{\Delta_i} \\ &\times \exp\left[-\int_0^t \widetilde{\lambda}_0(s)\exp\left\{\widetilde{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_i(s) + \widetilde{\beta}_2 I(X_i > \widetilde{\eta}_0) + \widetilde{\boldsymbol{\beta}}_3^T \boldsymbol{Z}_i(s) I(X_i > \widetilde{\eta}_0)\right\} ds\right]. \end{aligned}$$

If we take log on both sides of $m(\eta_0, \boldsymbol{\beta}|t) = h(\widetilde{\eta}_0, \widetilde{\boldsymbol{\beta}}|t)$, then we obtain

$$\begin{aligned}
&\Delta_i \left[\log\{\lambda_0(t)\} + \boldsymbol{\beta}_1^T \boldsymbol{Z}_i(t) + \beta_2 I(X_i > \eta_0) + \boldsymbol{\beta}_3^T \boldsymbol{Z}_i(t) I(X_i > \eta_0)\right] \\
&- \int_0^t \lambda_0(s) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(s) + \beta_2 I(X_i > \eta_0) + \boldsymbol{\beta}_3^T \boldsymbol{Z}_i(s) I(X_i > \eta_0)\right\} ds \\
=&\Delta_i \left[\log\{\widetilde{\lambda}_0(t)\} + \widetilde{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_i(t) + \widetilde{\beta}_2 I(X_i > \widetilde{\eta}_0) + \widetilde{\boldsymbol{\beta}}_3^T \boldsymbol{Z}_i(t) I(X_i > \widetilde{\eta}_0)\right] \\
&- \int_0^t \widetilde{\lambda}_0(s) \exp\left\{\widetilde{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_i(s) + \widetilde{\beta}_2 I(X_i > \widetilde{\eta}_0) + \widetilde{\boldsymbol{\beta}}_3^T \boldsymbol{Z}_i(s) I(X_i > \widetilde{\eta}_0)\right\} ds
\end{aligned}$$

We will prove that $\eta_0 = \widetilde{\eta}_0$ and $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$ by contradiction. Assume that $\eta_0 \neq \widetilde{\eta}_0$. Without loss of generality, we assume that $\eta_0 > \widetilde{\eta}_0$. Then, if $X_i < \widetilde{\eta}_0$ and $\boldsymbol{Z}_i(t) = \mathbf{0}$, then we have

$$\begin{aligned}
\Delta_i \log\{\lambda_0(t)\} - \int_0^t \lambda_0(s) ds &= \Delta_i \log\{\widetilde{\lambda}_0(t)\} - \int_0^t \widetilde{\lambda}_0(s) ds \\
\Delta_i \log\{\frac{\lambda_0(t)}{\widetilde{\lambda}_0(t)}\} &= \int_0^t \{\lambda_0(s) - \widetilde{\lambda}_0(s)\} ds.
\end{aligned}$$

Since it should hold for any t, we have $\lambda_0(t) = \widetilde{\lambda}_0(t)$. In addition, if $\eta_0 > X_i > \widetilde{\eta}_0$ and $\boldsymbol{Z}_i(t) = \mathbf{0}$, then we have

$$- \int_0^t \lambda_0(s) ds = \Delta_i \widetilde{\beta}_2 - \int_0^t \lambda_0(s) \exp\left(\widetilde{\beta}_2\right) ds.$$

Since it holds for any t, $\widetilde{\beta}_2$ has to be zero if $t \to 0$. In addition, if $\eta_0 > X_i > \widetilde{\eta}_0$ and $\boldsymbol{Z}_i(t) \neq \mathbf{0}$, then we have

$$\begin{aligned}
&\Delta_i \left[\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(t)\right] - \int_0^t \lambda_0(s) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_i(s)\right\} ds \\
=&\Delta_i \left[\widetilde{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_i(t) + \widetilde{\boldsymbol{\beta}}_3^T \boldsymbol{Z}_i(t)\right] - \int_0^t \lambda_0(s) \exp\left\{\widetilde{\boldsymbol{\beta}}_1^T \boldsymbol{Z}_i(s) + \widetilde{\boldsymbol{\beta}}_3^T \boldsymbol{Z}_i(s)\right\} ds
\end{aligned}$$

The equation holds if $\boldsymbol{\beta}_1 = \widetilde{\boldsymbol{\beta}}_1 + \widetilde{\boldsymbol{\beta}}_3$. If $\widetilde{\eta}_0 > X_i$ and $\boldsymbol{Z}_i(t) \neq \mathbf{0}$, then we can simply derive that $\boldsymbol{\beta}_1 = \widetilde{\boldsymbol{\beta}}_1$. Thus, $\widetilde{\boldsymbol{\beta}}_3 = \mathbf{0}$. Therefore, $\widetilde{\beta}_2$ and $\widetilde{\boldsymbol{\beta}}_3$ equal zeros at the same time. This is conflicted with the assumption that at least one of $\widetilde{\beta}_2$ and $\widetilde{\boldsymbol{\beta}}_3$ is nonzero. Thus, $\eta_0 = \widetilde{\eta}_0$. Given that, it is straightforward to prove that $(\boldsymbol{\beta}_1^T, \beta_2, \boldsymbol{\beta}_3^T)^T = (\widetilde{\boldsymbol{\beta}}_1^T, \widetilde{\beta}_2, \widetilde{\boldsymbol{\beta}}_3^T)^T$. Hence, the identifiability of η_0 and $\boldsymbol{\beta}$ is proved for the one factor case.

Second, we extend the proof to the case with two factors $I(X_{i1} + \eta_2 X_{i2} - \eta_0 > 0)$, where $\eta_2 \neq 0$ and the conditional distribution of X_{i1} given X_{i2} is continuous. The indicator function can be written as $I(X_{i1} > \eta_0 - \eta_2 X_{i2})$. It is easy to verify the identifiability of $\eta_0 - \eta_2 X_{i2}$ based on the similar arguments in the first step. Additionally, η_0 is identifiable given $X_{i2} = 0$. Since η_0 is identifiable, η_2 is also identifiable. Consequently, we have established the identifiability of η_0 and η_2.

In order to apply the mathematical induction, we assume that $(\eta_2, ..., \eta_k, \eta_0)$ are identifiable in $I(X_{i1} + \eta_2 X_{i2} + ... + \eta_k X_{ik} > \eta_0)$, where $k > 2$. The conditional distribution of X_{i1} given all other covariates $(X_{i2}, ..., X_{ik})$ is continuous. Based on that, we further prove the identifiability for the case $I(X_{i1} + \eta_2 X_{i2} + ... + \eta_k X_{ik} + \eta_{k+1} X_{ik+1} > \eta_0)$, where $\eta_{k+1} \neq 0$. Assuming that $X_{i2} = ... = X_{ik} = 0$, it is reduced to $I(X_{i1} + \eta_{k+1} X_{ik+1} > \eta_0)$. Similar to the proof in the second step, η_{k+1} is also identifiable. Thus, we can obtain the identifiability of $(\eta_2, ..., \eta_{k+1}, \eta_0)$.

From the mathematical induction, we conclude that the identifiability of the parameters in the single index are verified given the proposed conditions. Thus, Theorem 4.3.1 is proved.

Proof of Theorem 4.3.2

We first show that $G_n(\boldsymbol{\eta},\boldsymbol{\beta}) = n^{-1}\{l_n(\boldsymbol{\eta},\boldsymbol{\beta}) - l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\}$, which is expressed as

$$G_n(\boldsymbol{\eta},\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\Delta_i\left[r_{\boldsymbol{\eta},\boldsymbol{\beta}}\{\boldsymbol{W}_i(T_i)\} - r_{\boldsymbol{\eta}_0,\boldsymbol{\beta}_0}\{\boldsymbol{W}_i(T_i)\} - \log\frac{S_n^{(0)}(T_i;\boldsymbol{\eta},\boldsymbol{\beta})}{S_n^{(0)}(T_i;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right]$$
$$= (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\left\{\boldsymbol{M}_n^{(1)}(\tau) + \int_0^\tau \boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)d\Lambda_0(t)\right\} - n^{-1}\int_0^\tau \log\frac{S_n^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}d\bar{N}_n(t)$$
$$+ n^{-1}\sum_{i=1}^{n}\Delta_i\{\beta_2 + \boldsymbol{\beta}_3^T\boldsymbol{Z}_i(T_i)\}\left\{I(\boldsymbol{\eta}^T\boldsymbol{X}_i > 0, \boldsymbol{\eta}_0^T\boldsymbol{X}_i \le 0) - I(\boldsymbol{\eta}^T\boldsymbol{X}_i \le 0, \boldsymbol{\eta}_0^T\boldsymbol{X}_i > 0)\right\},$$

where $\boldsymbol{M}_n^{(1)}(\tau) = n^{-1}\left\{\sum_{i=1}^n\int_0^\tau \widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_0)dN_i(t)\right\} - \int_0^\tau \boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)d\Lambda_0(t)$, and $\bar{N}_n(t) = \sum_{i=1}^n N_i(t)$. Let $Y_{i1} = \boldsymbol{\eta}^T\boldsymbol{X}_i$, $Y_{i2} = X_{i2}$,..., $Y_{ip_1} = X_{ip_1}$, and $\widetilde{Y}_{i1} = \boldsymbol{\eta}_0^T\boldsymbol{X}_i$. Then,

$$G_n(\boldsymbol{\eta},\boldsymbol{\beta})$$
$$= (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\left\{\boldsymbol{M}_n^{(1)}(\tau) + \int_0^\tau \boldsymbol{S}_n^{(1)}(t;\widetilde{Y}_{i1},\boldsymbol{\beta}_0)d\Lambda_0(t)\right\} - n^{-1}\int_0^\tau \log\frac{S_n^{(0)}(t;Y_{i1},\boldsymbol{\beta})}{S_n^{(0)}(t;\widetilde{Y}_{i1},\boldsymbol{\beta}_0)}d\bar{N}_n(t)$$
$$+ n^{-1}\sum_{i=1}^{n}\Delta_i\{\beta_2 + \boldsymbol{\beta}_3^T\boldsymbol{Z}_i(T_i)\}\left\{I\left(Y_{i1} > 0 \ge \widetilde{Y}_{i1}\right) - I\left(\widetilde{Y}_{i1} > 0 \ge Y_{i1}\right)\right\}.$$

We define $G(\boldsymbol{\eta},\boldsymbol{\beta}) = \mathrm{E}(G_n(\boldsymbol{\eta},\boldsymbol{\beta}))$, which can be expressed as

$$G(\boldsymbol{\eta},\boldsymbol{\beta}) = \int_{y_{i1}}\int_{y_{i2}}\cdots\int_{y_{ip_1}}\int_0^\tau \mathrm{E}\Big[(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\left\{\boldsymbol{M}_n^{(1)}(\tau) + \int_0^\tau \boldsymbol{S}_n^{(1)}(t;\widetilde{y}_{i1},\boldsymbol{\beta}_0)d\Lambda_0(t)\right\}$$
$$-n^{-1}\int_0^\tau \log\frac{S_n^{(0)}(t;y_{i1},\boldsymbol{\beta})}{S_n^{(0)}(t;\widetilde{y}_{i1},\boldsymbol{\beta}_0)}d\bar{N}_n(t) + n^{-1}\sum_{i=1}^{n}\Delta_i\{\beta_2 + \boldsymbol{\beta}_3^T\boldsymbol{Z}_i(T_i)\}$$
$$\{I\left(y_{i1} > 0 \ge \widetilde{y}_{i1}\right) - I\left(\widetilde{y}_{i1} > 0 \ge y_{i1}\right)\}|\boldsymbol{X}_i\Big]\, f(y_{i1},y_{i2},\ldots,y_{ip_1})dy_{i1}dy_{i2}\ldots dy_{ip_1}d\Lambda_0(t)$$

Given $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, we define that $f(\epsilon) \equiv G\{\epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0\}$, where $\epsilon \in [0,1]$. Then, $G(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - G(\eta, \boldsymbol{\beta}_0) \equiv f(0) - f(1)$. If $\boldsymbol{\eta}$ is close to $\boldsymbol{\eta}_0$, then we obtain

$$
\begin{aligned}
\dot{f}(\epsilon) =& \lim_{\epsilon' \to \epsilon, \epsilon > \epsilon'} \frac{1}{\epsilon - \epsilon'} (G\{\epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0\} - G\{\epsilon'\boldsymbol{\eta} + (1-\epsilon')\boldsymbol{\eta}_0, \boldsymbol{\beta}_0\}) \\
=& \lim_{\epsilon' \to \epsilon, \epsilon > \epsilon'} \frac{1}{\epsilon - \epsilon'} \int_{y_{i1}<0} \int_{y_{i2}} \cdots \int_{y_{ip_1}} \int_0^\tau - \frac{s^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)}{s^{(0)}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} \\
& \{s^{(0)}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(0)}(t; \epsilon'\boldsymbol{\eta} + (1-\epsilon')\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} \\
& - \beta_{20}\{s^{(0)-}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(0)-}(t; \epsilon'\boldsymbol{\eta} + (1-\epsilon')\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} \\
& - \boldsymbol{\beta}_{30}^T\{s^{(1)-}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(1)-}(t; \epsilon'\boldsymbol{\eta} + (1-\epsilon')\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} \\
& f(y_{i1}, y_{i2}, ..., y_{ip_1}) dy_{i1} dy_{i2} ... dy_{ip_1} d\Lambda_0(t) \\
& + \lim_{\epsilon' \to \epsilon, \epsilon > \epsilon'} \frac{1}{\epsilon - \epsilon'} \int_{y_{i1}>0} \int_{y_{i2}} \cdots \int_{y_{ip_1}} \int_0^\tau - \frac{s^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)}{s^{(0)}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} \\
& \{s^{(0)}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(0)}(t; \epsilon'\boldsymbol{\eta} + (1-\epsilon')\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} \\
& - \beta_{20}\{s^{(0)+}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(0)+}(t; \epsilon'\boldsymbol{\eta} + (1-\epsilon')\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} \\
& - \boldsymbol{\beta}_{30}^T\{s^{(1)+}(t; \epsilon\boldsymbol{\eta} + (1-\epsilon)\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(1)+}(t; \epsilon'\boldsymbol{\eta} + (1-\epsilon')\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} \\
& f(y_{i1}, y_{i2}, ..., y_{ip_1}) dy_{i1} dy_{i2} ... dy_{ip_1} d\Lambda_0(t).
\end{aligned}
$$

Thus, we obtain that

$$
\begin{aligned}
\dot{f}(\epsilon) =& \int_{y_{i1}<0} \int_{y_{i2}} \cdots \int_{y_{ip_1}} \int_0^\tau y_{i1} \mathrm{E}\left[\exp(\boldsymbol{\beta}_{10}^T \boldsymbol{Z}_i(t)) \frac{\{\beta_{2*} + \boldsymbol{\beta}_{3*}^T \boldsymbol{Z}_i(t)\}^2}{2} | \widetilde{y}_{i1} = 0\right] f_{\widetilde{y}_{i1}}(0) \\
& f(y_{i1}, y_{i2}, ..., y_{ip_1}) dy_{i1} dy_{i2} ... dy_{ip_1} d\Lambda_0(t) \\
& - \int_{y_{i1}>0} \int_{y_{i2}} \cdots \int_{y_{ip_1}} \int_0^\tau y_{i1} \mathrm{E}\left[\exp^{\boldsymbol{\beta}_{10}^T \boldsymbol{Z}_i(t) + \beta_{20} + \boldsymbol{\beta}_{30} \boldsymbol{Z}_i(t)} \frac{\{\beta_{2*} + \boldsymbol{\beta}_{3*}^T \boldsymbol{Z}_i(t)\}^2}{2} | \widetilde{y}_{i1} = 0\right] \\
& f_{\widetilde{y}_{i1}}(0) f(y_{i1}, y_{i2}, ..., y_{ip_1}) dy_{i1} dy_{i2} ... dy_{ip_1} d\Lambda_0(t) \\
& < 0,
\end{aligned}
$$

where β_{2*} is between 0 and β_{20}, $\boldsymbol{\beta}_{3*}$ is between 0 and $\boldsymbol{\beta}_{30}$, $f_{\widetilde{y}_{i1}}(0)$ is the density function of $\widetilde{y}_{i1}$ at 0, and $f(Y_{i1}, Y_{i2}, ..., Y_{ip_1})$ is the density function of $(Y_{i1}, Y_{i2}, ..., Y_{ip_1})$.

Consequently, if $\widehat{\boldsymbol{\eta}} \in V_\delta(\boldsymbol{\eta}_0)$ and $\widehat{\boldsymbol{\beta}} \in V_\epsilon(\boldsymbol{\beta}_0)$, from the uniform convergence of $G_n(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}})$ to $G(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}})$, then it gives $\liminf G(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) \geq G(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ with probability one. Since $G(\boldsymbol{\eta}, \boldsymbol{\beta})$ has the unique maximum in $V_\epsilon(\boldsymbol{\beta}_0)$ and $V_\delta(\boldsymbol{\eta}_0)$, we conclude that $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}})$ should converge to $(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ in probability. Thus, Theorem 4.3.2 holds.

Proof of Theorem 4.3.3

First, we define $U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) = \{(\boldsymbol{\eta}, \boldsymbol{\beta}) : A < n^{1/2} \left(\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2\right)^{1/2} \leq n^{1/2}\epsilon\}$ and $V_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) = \{(\boldsymbol{\eta}, \boldsymbol{\beta}) : \left(\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2\right)^{1/2} < \epsilon\}$, for a given ϵ. From Theorem 4.3.2, $P_0 \{(\boldsymbol{\eta}, \boldsymbol{\beta}) \in V_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\} > 1 - \zeta$ for any $\zeta > 0$, when n is large enough. Hence,

$$P_0 \left\{ n^{1/2} \left(\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| + \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 \right)^{1/2} > A \right\} = P_0 \left\{ (\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) \in U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) \right\} + P_0 \left\{ (\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) \in V_\epsilon^C(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) \right\}$$
$$\leq P_0 \left\{ \sup_{\boldsymbol{\eta}, \boldsymbol{\beta} \in U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} L_n(\boldsymbol{\eta}, \boldsymbol{\beta}) \geq L_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) \right\} + \zeta = P_0 \left\{ \sup_{\boldsymbol{\eta}, \boldsymbol{\beta} \in U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) \geq 0 \right\} + \zeta.$$

The Taylor expression gives $G(\boldsymbol{\eta}, \boldsymbol{\beta}) = \dot{G}_{\boldsymbol{\eta}}(\boldsymbol{\eta}, \boldsymbol{\beta})(\boldsymbol{\eta} - \boldsymbol{\eta}_0)^T - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{I}(\boldsymbol{\eta}^*, \boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1)$, where $\boldsymbol{\beta}^*$ is between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. In the proof of Theorem 4.3.2, $\dot{G}_{\boldsymbol{\eta}}(\boldsymbol{\eta}, \boldsymbol{\beta})(\boldsymbol{\eta} - \boldsymbol{\eta}_0)^T$ is negative. In addition, the matrix $I(\boldsymbol{\eta}^*, \boldsymbol{\beta}^*)$ is positive definite by C.2. Therefore, there exists a positive constant k_0 which ensures $G(\boldsymbol{\eta}, \boldsymbol{\beta}) \leq -k_0(\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2)$. Additionally, we split $U_\epsilon(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ into subsets $H_{n,j} = \left\{ (\boldsymbol{\eta}, \boldsymbol{\beta}) : g(j) < n^{1/2} \left(\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 \right)^{1/2} < g(j+1) \right\}$, where $g(j) = 2^j$, and $j = 1, 2, ...$ Similar to Lemma A.1 in Chapter 3, there exists a constant $k > 0$ such that $\mathrm{E} \sup_{(\boldsymbol{\eta}, \boldsymbol{\beta}) \in V_\delta(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} |n^{1/2} \{G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - G(\boldsymbol{\eta}, \boldsymbol{\beta})\}| \leq k\epsilon$ as $n \to \infty$. Thus, we obtain

$$\limsup_n \sum_{j; g(j) > A} P_0 \left[\sup_{H_{n,j}} n^{1/2} \{G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - G(\boldsymbol{\eta}, \boldsymbol{\beta})\} \geq n^{-1/2} g^2(j) k_0 \right]$$
$$\leq \limsup_n \sum_{j; g(j) > A} \frac{\mathrm{E} \left[\sup_{H_{n,j}} n \{G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - G(\boldsymbol{\eta}, \boldsymbol{\beta})\} \right]^2}{g^4(j) k_0^2} \leq \sum_{j; g(j) > A} \frac{k^2 g^2(j+1)}{k_0^2 g^4(j)} \to 0,$$

as A goes to infinity. Hence, it gives $\lim_A \limsup_n P_0 \left\{ n^{1/2} \left(\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| + \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 \right)^{1/2} > A \right\} = 0$. Theorem 4.3.3 has been proved.

Proof of Theorem 4.3.4

We recall the definitions of $\boldsymbol{\eta}_{n,u_1} = \boldsymbol{\eta}_0 + n^{-1}\boldsymbol{u}_1$ and $\boldsymbol{\beta}_{n,u_2} = \boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{u}_2$, where $\boldsymbol{u}_1 = (a_1, a_2, ..., a_{p_1})^T$ and $\boldsymbol{u}_2 = (b_1, b_2, ..., b_{2p_2+1})^T$ in Section 4.3. The whole proof can be divided into the following steps. First, we partition and obtain the asymptotic expansions of the $l_n(\boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$. Second, we derive the asymptotic distributions based on the partitions. Finally, we obtain Theorem 4.3.4 using the argmax continuous mapping theorem.

In the first step, similar to the partition in Theorem 4.3.2, $l_n(\boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$ is partitioned as

$$
\begin{aligned}
& l_n(\boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0) \\
&= (\boldsymbol{\beta}_{n,u_2} - \boldsymbol{\beta}_0)^T \left\{ \sum_{i=1}^n \int_0^\tau \widetilde{\boldsymbol{Z}}_i(t; \boldsymbol{\eta}_{n,u_1}) dN_i \right\} - \int_0^\tau \log \left\{ \frac{S_n^{(0)}(t; \boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2})}{S_n^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} \right\} d\bar{N}_n(t) \\
&+ \sum_{i=1}^n \Delta_i \{\beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_i(T_i)\} \\
&\left[I\left\{0 \geq W_{i0} > n^{-1}\boldsymbol{X}_i^T \boldsymbol{u}_1\right\} I(n^{-1}\boldsymbol{X}_i^T \boldsymbol{u}_1 < 0) - I\left\{0 < W_{i0} \leq n^{-1}\boldsymbol{X}_i^T \boldsymbol{u}_1\right\} I(n^{-1}\boldsymbol{X}_i^T \boldsymbol{u}_1 \geq 0)\right],
\end{aligned}
$$

where $\boldsymbol{X}_i^T \boldsymbol{u}_1 = (-1, -a_1, -a_2, ..., a_{p_1})(X_{i1}, X_{i2}, ..., X_{ip_1}, 1)^T$ and $W_{i0} = \boldsymbol{\eta}_0^T \boldsymbol{X}_i$. By the Taylor expansion for $\boldsymbol{\beta}_{n,u_2}$ at $\boldsymbol{\beta}_0$,

$$
\begin{aligned}
\log \left\{ \frac{S_n^{(0)}(t; \boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_{n,u_2})}{S_n^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} \right\} =& \frac{S_n^{(0)}(t; \boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_0) - S_n^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)}{S_n^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} + n^{-1/2} \boldsymbol{u}_2^T \frac{\boldsymbol{S}_n^{(1)}(t; \boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_0)}{S_n^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} \\
&+ \frac{n^{-1}}{2} \boldsymbol{u}_2^T \boldsymbol{V}_n(t; \boldsymbol{\eta}_{n,u_1}, \boldsymbol{\beta}_0) \boldsymbol{u}_2 + o_p(n^{-1}),
\end{aligned}
$$

where $\boldsymbol{V}_n(t;\boldsymbol{\eta},\boldsymbol{\beta}) = \boldsymbol{S}_n^{(2)}(t;\boldsymbol{\eta},\boldsymbol{\beta})/S_n^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta}) - \left\{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta},\boldsymbol{\beta})/S_n^{(0)}(t;\boldsymbol{\eta},\boldsymbol{\beta})\right\}^{\otimes 2}$. Thus, we have

$$l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0) = Q_n(\boldsymbol{u}_1) + \boldsymbol{u}_2^T\boldsymbol{C}_n(\boldsymbol{u}_1) + \frac{1}{2}\boldsymbol{u}_2^T\boldsymbol{V}_n(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)\boldsymbol{u}_2 + o_p(n^{-1}),$$

where

$$\begin{aligned} Q_n(\boldsymbol{u}_1) = &\sum_{i=1}^n \Delta_i \Big[\left\{\beta_{20} + \boldsymbol{\beta}_{30}^T\boldsymbol{Z}_i(T_i)\right\} \\ &\left[I\left\{0 \ge W_{i0} > n^{-1}\boldsymbol{X}_i^T\boldsymbol{u}_1, \boldsymbol{X}_i^T\boldsymbol{u}_1 < 0\right\} - I\left\{0 < W_{i0} \le n^{-1}\boldsymbol{X}_i^T\boldsymbol{u}_1, \boldsymbol{X}_i^T\boldsymbol{u}_1 \ge 0\right\}\right] \\ &-\frac{S_n^{(0)}(T_i;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0) - S_n^{(0)}(T_i;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(T_i;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\Big], \end{aligned}$$

$$\begin{aligned} \boldsymbol{C}_n(\boldsymbol{u}_1) = & n^{-1/2}\sum_{i=1}^n \int_0^\tau \left\{\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_{n,u_1}) - \frac{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right\} dM_i(t) \\ &+ n^{-1/2}\int_0^\tau \sum_{i=1}^n \widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_{n,u_1})\left(\exp\left[r_{\boldsymbol{\eta}_0,\boldsymbol{\beta}_0}\left\{\boldsymbol{W}_i(t)\right\}\right] - \exp\left[r_{\eta_{n,u_1},\beta_0}\left\{\boldsymbol{W}_i(t)\right\}\right]\right) d\Lambda_0(t). \end{aligned}$$

We can verify that $\boldsymbol{C}_n(\boldsymbol{u}_1)$ uniformly converges to $\widetilde{\boldsymbol{l}}_n = n^{-1/2}\sum_{i=1}^n \int_0^\tau \left\{\widetilde{\boldsymbol{Z}}_i(t;\boldsymbol{\eta}_0) - \frac{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right\}$ $dM_i(t)$ in probability, as $\|\boldsymbol{u}_1\| \le A$. Then we have

$$l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2}) - l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0) = Q_n(\boldsymbol{u}_1) + \boldsymbol{u}_2^T\widetilde{\boldsymbol{l}}_n - \frac{1}{2}\boldsymbol{u}_2^T\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\boldsymbol{u}_2 + o_p(1). \tag{4.11}$$

In the second step, we derive the asymptotic distributions of $Q_n(\boldsymbol{u}_1)$ and $\widetilde{\boldsymbol{l}}_n$ in (4.11). The variable $\widetilde{\boldsymbol{l}}_n$ converges weakly to a Gaussian variable following the normal distribution $N(\boldsymbol{0},\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)^{-1})$. Clearly, since $\left\{\frac{1}{nS_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}d\bar{N}_n(t) - d\Lambda_0(t)\right\} = o_p(1)$ uniformly in t, we obtain that $Q_n = Q_n^+ - Q_n^- + o_p(1)$, where

$$Q_n^-(\boldsymbol{u}_1) = -\sum_{i=1}^n \Delta_i \left(\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_i(T_i) \right\} - \int_0^\tau \phi_i(t) d\Lambda_0(t) \right) I\left\{ \boldsymbol{X}_i^T \boldsymbol{u}_1 \geq 0, 0 < W_{i0} \leq n^{-1} \boldsymbol{X}_i^T \boldsymbol{u}_1 \right\},$$

$$Q_n^+(\boldsymbol{u}_1) = -\sum_{i=1}^n \Delta_i \left(\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_i(T_i) \right\} + \int_0^\tau \phi_i(t) d\Lambda_0(t) \right) I\left\{ \boldsymbol{X}_i^T \boldsymbol{u}_1 < 0, 0 \geq W_{i0} > n^{-1} \boldsymbol{X}_i^T \boldsymbol{u}_1 \right\},$$

$$\phi_i(t) = Y_i(t) \exp\left\{ \boldsymbol{\beta}_{10}^T \boldsymbol{Z}_i(t) \right\} \left[1 - \exp\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_i(t) \right\} \right].$$

Note that Q_n is defined as a random process on the space D of right-continuous functions with left-hand limits on R equipped with the Skorohod topology. We first show that the finite dimensional convergence holds for $Q_n^+(\boldsymbol{u}_1)$ (the same holds for $Q_n^-(\boldsymbol{u}_1)$), and we will identify its limit process based on this finite dimensional convergence. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_S$ be a sequence of vectors then we wish to obtain the limit distribution of any linear combination $\sum_{s=1}^S q_s Q_n^+(\boldsymbol{v}_s)$, where $q_1, q_2, ..., q_S$ are any fixed constants.

Let

$$\eta^- = -\Delta \left(\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}(T) \right\} - \int_0^\tau \phi(t) d\Lambda_0(t) \right)$$

and

$$\eta^+ = -\Delta \left(\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}(T) \right\} + \int_0^\tau \phi(t) d\Lambda_0(t) \right).$$

We let $H^{(1)}, ..., H^{(S)}$ be the order statistic of $\boldsymbol{X}^T \boldsymbol{v}_1, ..., \boldsymbol{X}^T \boldsymbol{v}_S$. Correspondingly, we let $q_{(1)}, ..., q_{(S)}$ be the ordered sequence of $q_1, ..., q_S$. We then define set $A_s = \{H^{(s-1)} < 0 < H^{(s)}\}$

for $1 \le s \le S$ and let A_0 be the set of $H^{(S)} \le 0$. We have

$$\begin{aligned}
&\mathrm{E}\left\{\exp\left(it\sum_{s=1}^{S} q_s Q_n^+(\boldsymbol{v}_s)\right)\right\}\\
&=\left\{P(A_0)+\sum_{s=1}^{S}P(A_s)\left(E\left[I(0<W<H^{(s)}/n)e^{(q_{(s)}+\ldots+q_{(S)})it\eta^+}\middle|A_s\right]\right.\right.\\
&\qquad +E\left[I(H^{(s)}/n\le W<H^{(s+1)}/n)e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta^+}\middle|A_s\right]+\ldots\\
&\qquad\qquad \left.\left.+E\left[I(H^{(S-1)}/n\le W<H^{(S)}/n)e^{q_{(S)}it\eta^+}\middle|A_s\right]\right)\right\}^n\\
&=\left\{1+\sum_{s=1}^{S}P(A_s)\left(E\left[I(0<W<H^{(s)}/n)\{e^{(q_{(s)}+\ldots+q_{(S)})it\eta^+}-1\}\middle|A_s\right]\right.\right.\\
&\qquad +E\left[I(H^{(s)}/n\le W<H^{(s+1)}/n)\{e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta^+}-1\}\middle|A_s\right]+\ldots\\
&\qquad\qquad \left.\left.+E\left[I(H^{(S-1)}/n\le W<H^{(S)}/n)\{e^{q_{(S)}it\eta^+}-1\}\middle|A_s\right]\right)\right\}^n.
\end{aligned}$$

Since

$$\begin{aligned}
&P(A_s)E\left[I(H^{(s)}/n\le W<H^{(s+1)}/n)\{e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta^+}-1\}\middle|A_s\right]\\
&=n^{-1}E\left[(H^{(s+1)}-H^{(s)})I(A_s)\{e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta^+}-1\}\middle|W=0\right]f_W(0+)+O(n^{-2}),
\end{aligned}$$

we conclude that

$$\begin{aligned}
&\mathrm{E}\left\{\exp\left(it\sum_{s=1}^{S} q_s Q_n^+(\boldsymbol{v}_s)\right)\right\}\\
&=\left\{1+n^{-1}f_W(0+)\sum_{s=1}^{S}\left(E\left[H^{(s)}I(A_s)\{e^{(q_{(s)}+\ldots+q_{(S)})it\eta^+}-1\}\middle|W=0\right]\right.\right.\\
&\qquad +E\left[(H^{(s+1)}-H^{(s)})I(A_s)\{e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta^+}-1\}\middle|W=0\right]+\ldots\\
&\qquad\qquad \left.\left.+E\left[(H^{(S)}-H^{(S-1)})I(A_s)\{e^{q_{(S)}it\eta^+}-1\}\middle|W=0\right]\right)+O(n^{-2})\right\}^n
\end{aligned}$$

so it converges to

$$\exp\left\{f_W(0+)\sum_{s=1}^{S}\sum_{k=s}^{S}\left(E\left[(H^{(k)}-H^{(k-1)})I(A_s)\{e^{(q_{(k)}+\ldots+q_{(S)})it\eta^+}-1\}\middle|W=0\right]\right)\right\}.$$

Recall

$$Q^+(\boldsymbol{u}_1) \equiv \sum_{\boldsymbol{x}} \sum_{0 \le l \le v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{u}_1)} V_l^+(\boldsymbol{x}),$$

where $V_l^+(\boldsymbol{x})$ is a random variable with the conditional distribution of η^+ given $W = 0$ and $\boldsymbol{X} = \boldsymbol{x}$ and it is assumed to be independent for any $\boldsymbol{x}$, and $v^+(\boldsymbol{x}, t)$ is a multivariate Poisson process with Poisson intensity $E[v^+(d\boldsymbol{x}, dt)] = f_{\boldsymbol{X}}(\boldsymbol{x}) f_W(0+) d\boldsymbol{x} dt$ for $t > 0$. We want to show that the limit distribution of $\sum_{s=1}^S q_s Q_n^+(\boldsymbol{v}_s)$ is the same as $\sum_{s=1}^S q_s Q^+(\boldsymbol{v}_s)$. To this end, we note

$$\begin{aligned}
& E\left[\exp\{it \sum_{s=1}^S q_s Q^+(\boldsymbol{v}_s)\}\right] \\
= \; & E\left[\exp\left\{it \sum_{s=1}^S q_s \sum_{\boldsymbol{x}} \left\{ I(\boldsymbol{x}^T \boldsymbol{v}_s > 0) \sum_{0 \le l \le v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_s)} V_l^+(\boldsymbol{x}) \right\}\right\}\right] \\
= \; & E\left[\exp\left\{it \sum_{s=1}^S \sum_{\boldsymbol{x}} I(\boldsymbol{x} \in A_s) \sum_{k=s}^S q_{(k)} \sum_{0 \le l \le v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_{(k)})} V_l^+(\boldsymbol{x}) \right\}\right] \\
= \; & \prod_{s=1}^S \prod_{k=s}^S E\left[\exp\left\{it \sum_{\boldsymbol{x}} I(\boldsymbol{x} \in A_s)(q_{(k)} + \ldots + q_{(S)}) \sum_{v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_{(k-1)}) < l \le v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_{(k)})} V_l^+(\boldsymbol{x}) \right\}\right] \\
= \; & \prod_{s=1}^S \prod_{k=s}^S \prod_{\boldsymbol{x} \in A_s} E\left[\exp\left\{it \sum_{0 < l \le v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_{(k)}) - v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_{(k-1)})} (q_{(k)} + \ldots + q_{(S)}) V_l^+(\boldsymbol{x}) \right\}\right],
\end{aligned}$$

where the last equality uses the fact that $v^+(d\boldsymbol{x}, [\boldsymbol{x}^T \boldsymbol{v}_{(k-1)}, \boldsymbol{x}^T v_{(k)}))$ is independent for each k for given $\boldsymbol{x}$ and $V_l^+(\boldsymbol{x})$'s are all i.i.d. We note that for a given $\boldsymbol{x}$,

$$\sum_{0 < l \le v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_{(k)}) - v^+(d\boldsymbol{x}, \boldsymbol{x}^T \boldsymbol{v}_{(k-1)})} (q_{(k)} + \ldots + q_{(S)}) V_l^+(\boldsymbol{x})$$

is a weighted summation of independent compound Poisson processes, where the Poisson rate is $(\boldsymbol{x}^T \boldsymbol{v}_{(k)} - \boldsymbol{x}^T \boldsymbol{v}_{(k-1)}) f_{\boldsymbol{X}}(\boldsymbol{x}) f_W(0+) d\boldsymbol{x}$, so its characteristic function is given as

$$\exp\left\{ f_W(0+) f_{\boldsymbol{X}}(\boldsymbol{x}) (\boldsymbol{x}^T \boldsymbol{v}_{(k)} - \boldsymbol{x}^T \boldsymbol{v}_{(k-1)}) E\left[e^{it(q_{(k)} + \ldots + q_{(S)})\eta^+} - 1 \middle| \boldsymbol{X} = \boldsymbol{x}, W = 0 \right] d\boldsymbol{x} \right\}.$$

Hence, we obtain

$$
\begin{aligned}
& \mathrm{E}\left[\exp\{it\sum_{s=1}^{S} q_s Q^+(\boldsymbol{v}_s)\}\right] \\
= & \left[\prod_{s=1}^{S}\prod_{k=s}^{S}\prod_{\boldsymbol{x}\in A_s} \exp\left\{ f_W(0+)(\boldsymbol{x}^T\boldsymbol{v}_{(k)} - \boldsymbol{x}^T\boldsymbol{v}_{(k-1)}) f_{\boldsymbol{X}}(\boldsymbol{x}) \right.\right. \\
& \left.\left. \times \mathrm{E}\left[e^{it(q_{(k)}+\ldots+q_{(S)})\eta^+} - 1 \middle| \boldsymbol{X} = \boldsymbol{x}, W = 0 \right] d\boldsymbol{x} \right\}\right] \\
= & \exp\left\{ \sum_{s=1}^{S}\sum_{k=s}^{S} \mathrm{E}_{\boldsymbol{x}}\left[I(A_s) f_W(0+)(\boldsymbol{x}^T\boldsymbol{v}_{(k)} - \boldsymbol{x}^T\boldsymbol{v}_{(k-1)}) \right.\right. \\
& \left.\left. \times E\left[e^{it(q_{(k)}+\ldots+q_{(S)})\eta^+} - 1 \middle| \boldsymbol{X} = \boldsymbol{x}, W = 0 \right]\right]\right\},
\end{aligned}
$$

which is the same as the limit distribution of $\sum_{s=1}^{S} q_s Q_n^+(\boldsymbol{v}_s)$.

Let $0 \le \boldsymbol{v}_1 \le \boldsymbol{v} \le \boldsymbol{v}_2 \le A$, $I_{n1} =]\boldsymbol{\eta}_0 + n^{-1}\boldsymbol{v}_1, \boldsymbol{\eta}_0 + n^{-1}\boldsymbol{v}]$ and $I_{n2} =]\boldsymbol{\eta}_0 + n^{-1}\boldsymbol{v}, \boldsymbol{\eta}_0 + n^{-1}\boldsymbol{v}_2]$. To prove Q_n^+ is tight, we have $E|Q_n^+(\boldsymbol{v}) - Q_n^+(\boldsymbol{v}_1)||Q_n^+(\boldsymbol{v}_2) - Q_n^+(\boldsymbol{v})|$ is bounded by $\|\boldsymbol{v}_2 - \boldsymbol{v}_1\|$ times a constant. Thus, the processes Q_n^+ converge weakly to Q^+, using the D-tightness criterion. (Billingsley 2009) Similarly, we can prove that Q_n^- converges weakly to Q^-. Therefore, the finite dimensional distribution of $Q_n(\boldsymbol{u}_1)$ converges the corresponding distribution of $Q(\boldsymbol{u}_1)$. Finally, because of argmax continuous mapping theorem, $n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = \arg\max_{u_1} Q_n(\boldsymbol{u}_1) + o_p(1)$ and it converges weakly to $\arg\max Q(s)$.

CHAPTER 5: CHANGE HYPERPLANE HAZARDS MODELS FOR CLUSTERED SURVIVAL DATA

5.1 Introduction

The Strong Heart Family Study is a longitudinal family-based cohort study for American Indians such that observations within the same families are correlated. For this study, Chapter 3 applied the Cox-type marginal model to address the correlation within the extended families and identified a change point in LTL for diabetic incidences. In the same study, however, the change point based on LTL has been observed to depend on other biomarkers, such as triglycerides, HDL, and etc. A good way to model the LTL change point depending on other biomarkers is to incorporate a change hyperplane, which is a linear function of LTL and other biomarkers as studied in Chapter 4 for univariate survival analysis. Therefore, in this chapter, we generalize the change hyperplane model in Chapter 4 to model clustered survival data as present in the Strong Heart Family study.

Specifically, our proposed method is a Cox-type marginal hazards model (Lee et al. 1992) with a change hyperplane based on multiple covariates for clustered survival data. The convergence rate of the proposed change hyperplane estimator is shown to be the rate of $1/n$, where n is the number of clusters. We prove that the asymptotic distribution of the proposed change hyperplane estimator follows a more complicated integrated composite Poisson process.

The structure of this paper is as follows. In Section 5.2, we adopt a two-step estimating method, the m out of n bootstrap, and a testing procedure of change hyperplane parameters for clustered survival data. In Section 5.3, we provide the asymptotic properties of the proposed estimators. Simulation studies evaluating the small sample performance of the method are presented in Section 5.4. In Section 5.5, we provide applications to the Strong Heart Family Study.

5.2 Methods

5.2.1 Model and Parameter Estimation

We consider n independent and identically distributed (i.i.d) clusters, and each cluster may have varying numbers of subjects K_i, where $i = 1, ..., n$ indicates the ith cluster. We define the observed failure time $T_{ij} = \min(\widetilde{T}_{ij}, C_{ij})$ and the failure indicator $\Delta_{ij} = I(\widetilde{T}_{ij} \leq C_{ij})$, where $j = 1, ..., K_i$ indicates the jth subject in the ith cluster, C_{ij} is the censoring time, and $\widetilde{T}_{ij}$ is the failure time. We assume that C_{ij} and $\widetilde{T}_{ij}$ are independent given the covariates $\boldsymbol{W}_{ij} = (\boldsymbol{X}_{ij}^T, \boldsymbol{Z}_{ij}(t)^T)^T$ in the model. The Cox-type marginal hazards model with a change hyperplane takes a form

$$\lambda(t|\boldsymbol{W}_{ij}) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_{ij}(t) + \beta_2 I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} > 0) + \boldsymbol{\beta}_3^T \boldsymbol{Z}_{ij}(t) I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} > 0)\right\},$$

where $\lambda_0(t)$ is an unknown baseline function, $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1^T, \beta_2, \boldsymbol{\beta}_3^T)^T$ is a vector of $2p_2 + 1$ unknown parameters, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_{p_1}, \eta_0)^T$ is a vector of $p_1 + 1$ unknown change hyperplane parameters, $\boldsymbol{X}_{ij} = (X_{ij1}, ..., X_{ijp_1}, -1)^T$ is a vector of the change hyperplane covariates, and $\boldsymbol{Z}_{ij}(t) = (Z_{ij1}(t), ..., Z_{ijp_2}(t))^T$ is a vector of other potentially time-dependent covariates.

We define $r_{\boldsymbol{\beta},\boldsymbol{\eta}}\{\boldsymbol{W}_{ij}(t)\} = \boldsymbol{\beta}_1^T \boldsymbol{Z}_{ij}(t) + \beta_2 I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} > 0) + \boldsymbol{\beta}_3^T \boldsymbol{Z}_{ij}(t) I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} > 0)$. Under the right censoring, the pseudo-partial likelihood function for n clusters with K_i subjects is expressed as

$$L(\boldsymbol{\eta}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{j=1}^{K_i} \left(\frac{\exp\left[r_{\boldsymbol{\beta},\boldsymbol{\eta}}\{\boldsymbol{W}_{ij}(T_{ij})\}\right]}{\sum_{l=1}^{n} \sum_{k=1}^{K_l} I(T_{lk} \geq T_{ij}) \exp\left[r_{\boldsymbol{\beta},\boldsymbol{\eta}}\{\boldsymbol{W}_{\boldsymbol{lk}}(T_{ij})\}\right]} \right)^{\Delta_{ij}}.$$

We extend the two-step estimating approach in Chapter 4 for univariate survival data to clustered events. The plausibility of such extension is because the Cox-type marginal model estimates the parameters of interest by assuming that the subjects are independent in the same clusters. Thus, the pseudo-partial likelihood for the Cox marginal model for clustered data results in the same form as the Cox proportional hazards model for univariate data. Consequently, the estimation of the change hyperplane parameters and the regression coefficients follows a similar two-step approach as in Chapter 4. First, given fixed change hyperplane parameters in the prespecified range, we maximized the pseudo-partial likelihood to obtain the estimates of regression coefficients. The profile function of the change hyperplane parameters over regression coefficients is formed. Second, we maximized this profile function by applying the genetic algorithm (Sekhon and Mebane 1998). Consequently, the maximum likelihood estimators are specified as $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) = \arg\max_{\boldsymbol{\eta} \in S_0, \boldsymbol{\beta}} l_n(\boldsymbol{\eta}, \boldsymbol{\beta})$, where $l_n(\boldsymbol{\eta}, \boldsymbol{\beta})$ is the log of the pseudo-partial likelihood function, and $S_0 = \{(\eta_1, ..., \eta_{p_1}, \eta_0) : \eta_1^2 + \eta_2^2 + ... + \eta_{p_1}^2 = 1, \eta_1 > 0, \eta_0 \in [\eta_a, \eta_b]\}$. In the prespecified set S_0, we impose two conditions, i.e., condition (1) $\eta_1^2 + \eta_2^2 + ... + \eta_{p_1}^2 = 1$ and condition (2) the coefficient of the first continuous covariate has positive η_1. These two conditions are needed to show the identifiability of the change hyperplane and regression parameters (Theorem 5.3.1).

5.2.2 Inference

The inference for parameters of interests consists of two parts: the inference of regression coefficients $\boldsymbol{\beta}$ and change hyperplane parameters $\boldsymbol{\eta}$. In Section 5.3, we have proved the asymptotic independence between the regression coefficients and the change hyperplane parameters. Thus, the inference of these two groups of parameters can be conducted separately. For the regression coefficient $\boldsymbol{\beta}$, we have proved that $\widehat{\boldsymbol{\beta}}$ converges weakly to the same normal distribution in Lee et al. (1992). Consequently, it is straightforward to calculate the covariance matrix and generate the 95% confidence interval based on the normal approximation for regression parameters. For change hyperplane parameters $\boldsymbol{\eta}$, it is hard to derive the closed form of the covariance matrix for the change hyperplane estimators based on the proved asymptotic distributions. In addition, the regular resampling approach, e.g. bootstrap, cannot provide consistent estimates for such non-continuous and non-$1/\sqrt{n}$ convergence rates. Here, we apply the m out of n bootstrap and adopt the algorithm in Bickel and Sakov (2008) to select m. Specifically, we select m_j, $j = 1, ..., p_1$, for each dimension of the change hyperplane parameters. We choose $\widehat{m}_j$ from a sequence of subset values of the total n which achieves the minimum distance between two adjacent empirical distributions. The final m equals to the minimum among all m_j. The obtained standard error based on the m out of n bootstrap are further adjusted by the factor n/m.

5.2.3 Hypothesis Testing for the Change Hyperplane

To verify the existence of the change hyperplane, we proposed the supremum tests with robust score statistics for clustered survival time. The supremum test is proposed by Davies (1977), Davies (1987), and Kosorok and Song (2007). Under the null hypothesis $H_0 : \beta_2 = 0$ and $\boldsymbol{\beta}_3 = \mathbf{0}$, the supremum test with robust score statistic is

$$\text{SUP}_{k^{p_1}} = \sup_{\eta_j \in [\eta_{j1}, ..., \eta_{jk}], j=2,...,p_1, 0} \boldsymbol{U}(\boldsymbol{\eta})^T \boldsymbol{\Sigma}(\boldsymbol{\eta})^{-1} \boldsymbol{U}(\boldsymbol{\eta}),$$

where $\boldsymbol{U}(\boldsymbol{\eta}) = \frac{\partial l_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}(\boldsymbol{\eta}) = \sum_{i=1}^n \sum_{j=1}^{K_i} \sum_{l=1}^{K_i} \boldsymbol{H_{ij}}(\boldsymbol{\eta}) \boldsymbol{H_{il}}(\boldsymbol{\eta})^T$,

$$\begin{aligned} \boldsymbol{H_{ij}}(\boldsymbol{\eta}) \quad = \quad & \left\{ \widetilde{\boldsymbol{Z}}_{ij}(T_{ij}) - \frac{\boldsymbol{S}_n^{(1)}(T_{ij}; \boldsymbol{\eta}, \boldsymbol{\beta})}{S_n^{(0)}(T_{ij}; \boldsymbol{\eta}, \boldsymbol{\beta})} \right\} \\ & - \sum_{s=1}^n \sum_{l=1}^{K_s} \frac{\Delta_{sl} \exp\{\boldsymbol{\beta}^T \widetilde{\boldsymbol{Z}}_{ij}(T_{sl})\}}{n S_n^{(0)}(T_{sl}; \boldsymbol{\eta}, \boldsymbol{\beta})} \left\{ \widetilde{\boldsymbol{Z}}_{ij}(T_{sl}) - \frac{\boldsymbol{S}_n^{(1)}(T_{sl}; \boldsymbol{\eta}, \boldsymbol{\beta})}{S_n^{(0)}(T_{sl}; \boldsymbol{\eta}, \boldsymbol{\beta})} \right\}, \end{aligned}$$

where $\boldsymbol{S}_n^{(r)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \frac{1}{n} \left(\sum_{i=1}^n \sum_{j=1}^{K_i} Y_{ij}(t) \widetilde{\boldsymbol{Z}}_{ij}^{\otimes r}(t; \boldsymbol{\eta}) \exp\left[r_{\boldsymbol{\eta}, \boldsymbol{\beta}} \{\boldsymbol{W}_{ij}(t)\} \right] \right)$, $Y_{ij}(t) = I(T_{ij} \geq t)$, $\widetilde{\boldsymbol{Z}}_{ij}(t; \boldsymbol{\eta}) = \left(\boldsymbol{Z}_{ij}(t)^T, I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} > 0), \boldsymbol{Z}_{ij}(t)^T I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} > 0) \right)^T$ for $r = 0, 1$, and k is the number of grids selected for each dimension of the change hyperplane parameters. We use permutation to generate the null distribution of the proposed test statistic. The null hypothesis is rejected at a significance level of α if $\text{SUP}_{k^{p_1}}$ is larger than the upper α-quantile of the permutation distribution.

5.3 Asymptotic Properties

The following conditions are needed to establish the asymptotic properties.

(C.1) The joint density of $(\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, ..., \boldsymbol{X}_{ik_i})$ is assumed to be strictly positive, bounded and continuous in a neighborhood $V_0 = \{\boldsymbol{x} : |\boldsymbol{\eta}_0^T \boldsymbol{x}| < \epsilon\}$. In addition, the joint density of $(\boldsymbol{Z}_{i1}(t), ..., \boldsymbol{Z}_{ik_i}(t))$ is assumed to be strictly positive and bounded.

(C.2) For any $V_\delta(\boldsymbol{\eta}_0) = \{\boldsymbol{\eta} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \delta\}$, the covariance matrix

$$I(\boldsymbol{\eta}, \boldsymbol{\beta}) = \int_0^\tau v(t; \boldsymbol{\eta}, \boldsymbol{\beta}) s^{(0)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) \lambda_0(t) dt$$

is positive definite, where $\boldsymbol{\eta}_0$ is the true value of $\boldsymbol{\eta}$, $\boldsymbol{v}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \boldsymbol{s}^{(2)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) / s^{(0)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) - [\boldsymbol{s}^{(1)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) / s^{(0)}(t; \boldsymbol{\eta}, \boldsymbol{\beta})]^{\otimes 2}$, $\boldsymbol{s}^{(r)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \mathrm{E}\left(\sum_{j=1}^{K_i} Y_{ij}(t) \widetilde{\boldsymbol{Z}}_{ij}^{\otimes r}(t; \boldsymbol{\eta}) \exp\left[r_{\boldsymbol{\eta}, \boldsymbol{\beta}}\{\boldsymbol{W}_{ij}(t)\}\right]\right)$, and $r = 0, 1, 2$. In addition,

$$\lambda_{\min}\left(\int_0^\tau \mathrm{E}\left[Y_{ij}(t)\left\{1, \boldsymbol{Z}_{ij}^T(t)\right\}^{\otimes 2} \middle| \boldsymbol{\eta}_0^T \boldsymbol{X}_{ij} = 0\right] d\Lambda_0(t)\right) > 0,$$

where $\lambda_{\min}(\boldsymbol{A})$ is the smallest eigenvalue of any square matrix $\boldsymbol{A}$.

(C.3) The random process $\sup_{t \in [0,\tau]} \sup_{\boldsymbol{\eta} \in V_\delta(\boldsymbol{\eta}_0), \boldsymbol{\beta} \in V_\epsilon(\boldsymbol{\beta}_0)} \left\| \boldsymbol{S}_n^{(r)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) - \boldsymbol{s}^{(r)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) \right\|$ converges almost surely to zero, where $\boldsymbol{s}^{(r)}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) < \infty$, and $r = 0, 1, 2$. When $r = 0$, $\boldsymbol{s}^{(0)}(t; \boldsymbol{\eta}, \boldsymbol{\beta})$ is bounded away from zero.

(C.4) $\sup_{t \in [0,\tau]} \lambda_0(t) < \infty$, and $P(Y_{ij}(t) = 1) > 0$ for all $t \in [0, \tau]$.

(C.5) $P(K_i \leq k_0) = 1$, where $1 \leq k_0 < \infty$.

Our first theorem establishes the identifiability of the change hyperplane parameters and regression coefficient parameters. Theorem 5.3.2 and Theorem 5.3.3 show the consistency and convergence rates of the change hyperplane estimators and regression coefficients estimators. Theorem 5.3.3 implies that the convergence rates for $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\beta}}$ are $1/n$ and $1/\sqrt{n}$, respectively. These rates will be applied to establish the asymptotic distributions of the estimators in Theorem 5.3.4. The proofs of these theorems follow the similar arguments as Chapter 4. We show the sketch of the proofs in Proof of Lemma and Theorems.

Theorem 5.3.1. *Under the condition that $\|\boldsymbol{\eta}\| = 1$ with the first continuous covariate coefficient $\eta_1 > 0$ and at least one of the elements in β_2 or $\boldsymbol{\beta}_3$ is nonzero, the change hyperplane parameters $\boldsymbol{\eta}$ and regression parameters $\boldsymbol{\beta}$ are identifiable.*

Theorem 5.3.2. *Under conditions (C.1)-(C.5), there exists a local maximizer $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}})$ in a neighborhood of $\boldsymbol{\eta}_0$ and $\boldsymbol{\beta}_0$ such that $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\beta}}) \to_p (\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)$.*

Theorem 5.3.3. *Under conditions (C.1)-(C.5),*

$$\lim_{A\to\infty} \lim_{n\to\infty} P_0 \left(n \left\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right\| > A\right) = 0,$$
$$\lim_{A\to\infty} \lim_{n\to\infty} P_0 \left(n^{1/2} \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| > A\right) = 0.$$

Thus, $\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| = o_p(1/n)$ and $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = o_p(1/\sqrt{n})$.

Let $\boldsymbol{\eta}_{n,u_1} = \boldsymbol{\eta}_0 + n^{-1}\boldsymbol{u}_1$, $\boldsymbol{\beta}_{n,u_2} = \boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{u}_2$, and $W_{ij0} = \boldsymbol{\eta}_0^T \boldsymbol{X}_{ij}$, where $\boldsymbol{u}_1 = (a_1, a_2, ..., a_{p_1})^T$ and $\boldsymbol{u}_2 = (b_1, b_2, ..., b_{2p_2+1})^T$. We define that $V^+_{mq,l}(\boldsymbol{x})$ and $V^-_{mq,l}(\boldsymbol{x})$ are random variables with the conditional distributions of η^+_{ij} and η^-_{ij} given $W_{iq0} = 0$, $K_i = m$, S^{q+}_{m1} and $\boldsymbol{X} = \boldsymbol{x}$, respectively, where

$$\eta^-_{ij} = -\Delta_{ij} \left(\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(T_{ij}) \right\} - \int_0^\tau \phi_{ij}(t) d\Lambda_0(t) \right),$$
$$\eta^+_{ij} = -\Delta_{ij} \left(\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(T_{ij}) \right\} + \int_0^\tau \phi_{ij}(t) d\Lambda_0(t) \right),$$
$$\phi_{ij}(t) = Y_{ij}(t) \exp\left\{ \boldsymbol{\beta}_{10}^T \boldsymbol{Z}_{ij}(t) \right\} \left[1 - \exp\left\{ \beta_{20} + \boldsymbol{\beta}_{30}^T \boldsymbol{Z}_{ij}(t) \right\} \right],$$

$S^{+}_{m1} = \{K_i = m, \text{only one } \boldsymbol{\eta}_0^T \boldsymbol{X}_{ij} > 0, \text{all the other } \boldsymbol{\eta}_0^T \boldsymbol{X}_{i1}, ..., \boldsymbol{\eta}_0^T \boldsymbol{X}_{im} \le 0\}$, and S^{j+}_{m1} is the jth element in the set S^{+}_{m1}, $j = 1, ..., m$. In addition, $V^{+}_{mq,l}(\boldsymbol{x})$ and $V^{-}_{mq,l}(\boldsymbol{x})$ are assumed to be independent for any $\boldsymbol{x}$. We further define that $v^{+}_{mq}(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{im}, t)$ and $v^{-}_{mq}(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{im}, t)$ are multivariate Poisson processes with Poisson intensity $E[v^{+}_{mq}(d\boldsymbol{x}_{i1}, d\boldsymbol{x}_{i2}, ..., d\boldsymbol{x}_{im}, dt)] = f_{\boldsymbol{X}_i}(\boldsymbol{x}_i) f_{W_{iq0}}(0+) p(m) P(S^{q+}_{m1} | K_i = m, S^{+}_{m1}) d\boldsymbol{x}_{i1} d\boldsymbol{x}_{i2} ... d\boldsymbol{x}_{im} dt$ for $t > 0$ and $E[v^{-}_{mq}(d\boldsymbol{x}_{i1}, d\boldsymbol{x}_{i2}, ..., d\boldsymbol{x}_{im}, dt)] = f_{\boldsymbol{X}_i}(\boldsymbol{x}_i) f_{W_{iq0}}(0-) p(m) P(S^{q-}_{m1} | K_i = m, S^{-}_{m1}) d\boldsymbol{x}_{i1} d\boldsymbol{x}_{i2} ... d\boldsymbol{x}_{im} dt$ for $t < 0$, respectively, where $f_{\boldsymbol{X}_i}(\boldsymbol{x}_i)$ is the joint density of $(\boldsymbol{X}_{i1}, ..., \boldsymbol{X}_{im})$, $f_{W_{iq0}}(0+)$ is the density function of $\boldsymbol{\eta}_0^T \boldsymbol{X}_{iq}$ at zero given $\boldsymbol{\eta}_0^T \boldsymbol{X}_{iq} \le 0$, and $f_{W_{iq0}}(0-)$ is the density function of $\boldsymbol{\eta}_0^T \boldsymbol{X}_{iq}$ at zero given $\boldsymbol{\eta}_0^T \boldsymbol{X}_{iq} > 0$. Thus, the asymptotic distribution is defined as $Q(\boldsymbol{u}_1) \equiv Q^{+}(\boldsymbol{u}_1) - Q^{-}(\boldsymbol{u}_1)$, where

$$Q^{+}(\boldsymbol{u}_1) \equiv \sum_{m=1}^{K} \sum_{q=1}^{m} \sum_{\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{im}} \sum_{0 \le l \le v^{+}_{mq}(d\boldsymbol{x}_{i1}, ..., d\boldsymbol{x}_{im}, \boldsymbol{x}_{iq}^T \boldsymbol{u}_1)} V^{+}_{mq,l}(\boldsymbol{x}_{iq}),$$

$$Q^{-}(\boldsymbol{u}_1) \equiv \sum_{m=1}^{K} \sum_{q=1}^{m} \sum_{\boldsymbol{x}_{i1}, ... \boldsymbol{x}_{im}} \sum_{0 \le l \le v^{-}_{mq}(d\boldsymbol{x}_{i1}, ..., d\boldsymbol{x}_{im}, \boldsymbol{x}_{iq}^T \boldsymbol{u}_1)} V^{-}_{mq,l}(\boldsymbol{x}_{iq}).$$

Theorem 5.3.4. *Under conditions (C.1)-(C.5), $n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ and $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ are asymptotically independent. Furthermore, $n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges weakly to $\inf\{\boldsymbol{u}_1 : \arg\max Q(\boldsymbol{u}_1)\}$, and $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges weakly to $N(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0) \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1})$.*

5.4 Simulation Studies

To verify the finite sample performance of our proposed methods, we assumed that the model included three covariates, i.e., the coverate Z, the change hyperplane, and their interaction. The covariate Z followed a Uniform distribution from -1 to 1. The change hyperplane contained two variables, which were $X_1 \sim N(2, 1.5^2)$ and $X_2 \sim N(0, 1)$. We generated the marginal survival times $\widetilde{T}_{ij}$ under the proportional hazards model $\Lambda(t | X_1, X_2, Z) = t \exp\{\beta_1 Z + \beta_2 I(\eta_1 X_1 + \eta_2 X_2 - \eta_0 > 0) + \beta_3 Z I(\eta_1 X_1 + \eta_2 X_2 - \eta_0 > 0)\}$, where $(\beta_1, \beta_2, \beta_3) = (-1, -1.8, 1.5)$, and $(\eta_1, \eta_2, \eta_0) = (0.8, -0.6, 1.7)$ with $\eta_1^2 + \eta_2^2 = 1$. The censoring time follows

Uniform(0, 35) to ensure the 10% censoring rate. Cai and Shen (2000) proposed the following conditional cumulative density function for the survival time of the jth subject in the ith cluster.

$$F_i(\widetilde{T}_{ij}|\widetilde{T}_{i1},...,\widetilde{T}_{i(j-1)}) = 1 - \left\{\sum_{h=1}^{j} S_{ih}(\widetilde{T}_{ih})^{-1/\gamma} - (j-1)\right\}\left\{\sum_{h=1}^{j} S_{ih}(\widetilde{T}_{ih})^{1/\gamma} - (j-2)\right\}^{\gamma+j-1},$$

where $S_{ij}(t) = P(\widetilde{T}_{ij} > t)$ is the marginal survival function, γ indicates the degree of dependence between $\widetilde{T}_{ij}$ and $\widetilde{T}_{ih}(h = 1,...,j-1)$. The Kendall's tau coefficient is a function of γ, which is expressed as $\tau_K = \frac{1}{2\gamma+1}$. We considered $\gamma = 0.25$, and the cluster sizes 2 or 20. The number of clusters is 100 or 200 for the cluster size 2, and 50 or 100 for the cluster size 20. We first assessed the finite sample performance for the change hyperplane and regression coefficient estimators. Then we examined the coverage rates of the confidence intervals based on the m out of n bootstrap and the normal approximation for the change hyperplane and regression parameters, respectively. All results are based on 500 replications and each m out of n bootstrap consists of 100 replicates.

Table 5.9 summarizes the results for the change hyperplane estimators and the m out of n bootstraps. The proposed estimator provides approximately unbiased estimates. The m out of n bootstrap generates proper coverage rates, which is close to 95%. Correspondingly, Table 5.10 shows that the estimates of the regression coefficients $\boldsymbol{\beta}$ are approximately unbiased. The standard error estimates (SEE) of the regression coefficients get close to their sample standard deviations (SSD) when the sample sizes increase. The confidence intervals using the normal approximation have proper coverage rates.

Our second set of simulation studies were aimed at comparing the type I error and power of the SUP_{5^2}, SUP_{10^2}, and SUP_{20^2} tests under different cluster sizes and sample sizes. The range for η_0 and η_2 is $[-10, 10]$ and $[-1, 1]$, respectively. We examine the performance of these tests with the sample sizes 100 and 200 for the cluster size 2, and the sample sizes 50 and 100 for the cluster size 20. The results for type I error and power are based on 10000

Table 5.9: Simulation Results for the Change Hyperplane Parameters.

Cluster Size	Number of Clusters	Parameters	Bias	SSD	95% CI	Length
2	100	$\widehat{\eta}_2$	-0.002	0.050	0.964	0.243
		$\widehat{\eta}_0$	-0.006	0.098	0.942	0.441
	200	$\widehat{\eta}_2$	-0.002	0.026	0.946	0.117
		$\widehat{\eta}_0$	0.000	0.046	0.950	0.220
20	50	$\widehat{\eta}_2$	-0.001	0.011	0.974	0.052
		$\widehat{\eta}_0$	0.001	0.022	0.974	0.102
	100	$\widehat{\eta}_2$	-0.001	0.006	0.958	0.027
		$\widehat{\eta}_0$	0.000	0.013	0.938	0.054

NOTE: SSD stands for sample standard deviation. Length is the length of the 95% CI.

Table 5.10: Simulation Results for the Regression Parameters.

Cluster Size	Number of Clusters	Parameters	Bias	SSD	SEE	95% CI
2	100	$\widehat{\beta}_1$	-0.020	0.186	0.195	0.932
		$\widehat{\beta}_2$	-0.085	0.206	0.222	0.936
		$\widehat{\beta}_3$	0.032	0.285	0.289	0.960
	200	$\widehat{\beta}_1$	0.005	0.132	0.138	0.928
		$\widehat{\beta}_2$	-0.036	0.146	0.156	0.934
		$\widehat{\beta}_3$	-0.007	0.199	0.195	0.946
20	50	$\widehat{\beta}_1$	-0.014	0.115	0.122	0.936
		$\widehat{\beta}_2$	-0.048	0.188	0.193	0.942
		$\widehat{\beta}_3$	0.039	0.180	0.199	0.920
	100	$\widehat{\beta}_1$	-0.015	0.083	0.089	0.930
		$\widehat{\beta}_2$	-0.031	0.136	0.140	0.944
		$\widehat{\beta}_3$	0.028	0.130	0.135	0.928

NOTE: SSD and SEE stand for sample standard deviation and average standard error estimate, respectively.

and 1000 replicates, respectively. All the other specifications are the same as the first set of simulations.

Table 5.11 shows that type I errors of all three tests get close to 0.05 when the sample sizes increase. For the large cluster size 20, the small number of grids has an inflated type I error. For power, the performance of the supremum tests is determined by the numbers of grids and sample sizes. Given the same sample size, the power gets stabilized after the number of grids reaches 10 for each parameter. Given the tests with the same number of grids, the power increases as the sample size increases.

Table 5.11: Type I Error and Power for SUP Tests for the Existence of the Change Hyperplane

			SUP Tests		
Cluster Size	(β_{20}, β_{30})	Number of Clusters	SUP_{5^2}	SUP_{10^2}	SUP_{20^2}
2	(0, 0)	100	0.051	0.052	0.051
		200	0.053	0.051	0.049
	(0.5, -0.5)	100	0.101	0.393	0.397
		200	0.209	0.724	0.874
20	(0, 0)	50	0.076	0.068	0.072
		100	0.068	0.050	0.055
	(0.25, -0.20)	50	0.179	0.519	0.499
		100	0.349	0.879	0.903

5.5 Application to the Strong Heart Family Study

We applied the proposed method to the Strong Heart Family Study. The SHFS followed 94 extended American Indian families from 2001 to 2009. There are 2315 participants free of diabetes at baseline, among whom 292 developed incident diabetes with a median survival time 5.4 years. Zhao et al. (2014) observed that those individuals with LTL less than the 25th percentile had a significantly higher risk of developing new diabetes than the other individuals. Later, Deng. et al (2016) estimated a more accurate change point of LTL at 0.870. Here, we applied the proposed change hyperplane model to identify the change point effect based on both LTL and the lipid profiles.

We included gender, age, body mass index (BMI) and the change hyperplane as predictors in the Cox-type marginal hazards model. The change hyperplane was assumed to be a linear combination of the standardized LTL, total triglycerides, HDL, and LDL. Based on the preliminary findings in Zhao et al. (2014) and Chapter 3, the individuals with the shorter LTL has a higher risk in obtaining diabetes. Hence, we specified the change hyperplane as $I(\eta_1\mathrm{LTL} + \eta_2\mathrm{Triglycerides} + \eta_3\mathrm{HDL} + \eta_4\mathrm{LDL} < \eta_0)$, where $\eta_1 > 0$.

We carried out our change hyperplane analysis as follows. First, we applied the proposed supremum test with the robust score statistics to verify the existence of the change hyperplane. We applied the SUP_{20^4} test with $\{(\eta_1, \eta_2, \eta_3, \eta_4) : \eta_1^2 + \eta_2^2 + \eta_3^2 + \eta_4^2 = 1, \eta_1 > 0\}$ and $\{\eta_0 : \eta_0 \in [-20, 20]\}$. Compared to the α-level of 0.05, the test is highly significant with p-value 0.01. This indicates the existence of a change hyperplane for diabetes incidence. Next, we applied the two-step procedure to estimate the change hyperplane and the m out of n bootstrap with 150 replicates to generate the 95% confidence interval of the change hyperplane parameters. The results was summarized in Model 1 of Table 5.12. The estimated change hyperplane parameters are $(0.497, -0.868, 0.005, 0.009)$. Both LTL and total triglycerides are statistically significant with p-values smaller than 0.05. After removing the non-significant change hyperplane covariates, we presented the results for the final change hyperplane parameters in Model 2 of Table 5.12. Based on the change hyperplane parameters in Model 2, we removed gender (p-value = 0.15) in the model. The regression coefficients for the final model were shown in Table 5.13. For this group of participants, the increase in the level of total triglycerides results in an increase in the risk of developing incident diabetes. The hazard ratio of diabetes for the participants with $I(0.465\text{LTL} - 0.885\text{Triglycerides} < 0.579)$ is 3.483 with 95% CI $[2.184, 5.555]$.

We checked the proportional hazard assumptions for all covariates in Figure 5.3. For the categorical variable $I(\boldsymbol{\eta}^T\boldsymbol{X} < 0)$, we generated a plot of log of negative log of survival functions versus time, which shows a parallel trend between the groups of $I(\boldsymbol{\eta}^T\boldsymbol{X} < 0)$ and $I(\boldsymbol{\eta}^T\boldsymbol{X} \geq 0)$. For continuous variables (Age, BMI), the scatter plots show that the Schoenfeld residuals are evenly distributed on both sides of the reference line, suggesting that the proportional hazards assumptions are reasonable.

Table 5.12: Change Hyperplane Estimates Based on the Strong Heart Family Study: Model 1 (full model) and Model 2 (final reduced model).

	Model 1			Model 2		
Parameters	Estimate ($\widehat{\boldsymbol{\eta}}$)	SE ($\widehat{\boldsymbol{\eta}}$)	95% CI	Estimate ($\widehat{\boldsymbol{\eta}}$)	SE ($\widehat{\boldsymbol{\eta}}$)	95% CI
LTL	0.497	0.132	[0.287, 0.707]*	0.465	0.213	[0.018, 0.912]
Triglycerides	-0.868	0.192	[-1.390, -0.346]*	-0.885	0.292	[-1.660, -0.11]
HDL	0.005	0.118	[-0.278, 0.288]	-	-	-
LDL	0.009	0.097	[-0.513, 0.531]	-	-	-
η_0	0.533	-	-	-0.579	-	-

NOTE: * stand for the significant results with p-value smaller than 0.05.

Table 5.13: Regression Coefficients Estimates Based on Model 2 (final reduced model) in the Strong Heart Family Study.

	Estimate	exp(Estimate)	SE	p-value
AGE	0.021	1.021	0.004	<0.001
BMI	0.840	2.316	0.115	<0.001
$I(\boldsymbol{\eta}^T \boldsymbol{X} < 0)$	1.248	3.483	0.223	<0.001

5.6 Conclusion

The clustered data are commonly seen in epidemiological studies due to its cost-efficiency in collecting data. Chapter 4 has developed a Cox proportional hazards model with a change hyperplane for univariate data. However, the proposed method and asymptotic properties cannot be directly extended to clustered data. In this paper, we extended the inference procedures based on the Cox-type marginal model. We proposed a two-step approach to maximize the pseudo-partial likelihood to obtain the change hyperplane and regression coefficient parameters. We adopted the m out of n approach and supremum tests to make inferences for the change hyperplane parameters. The simulations demonstrated that the proposed estimators are consistent, m out of n bootstrap generated proper coverage rates, and the supremum tests preserved the type I error.

Here, we applied the Cox-type marginal model to address the correlation within clusters. There exist other approaches to analyze clustered survival data, e.g. frailty model. The

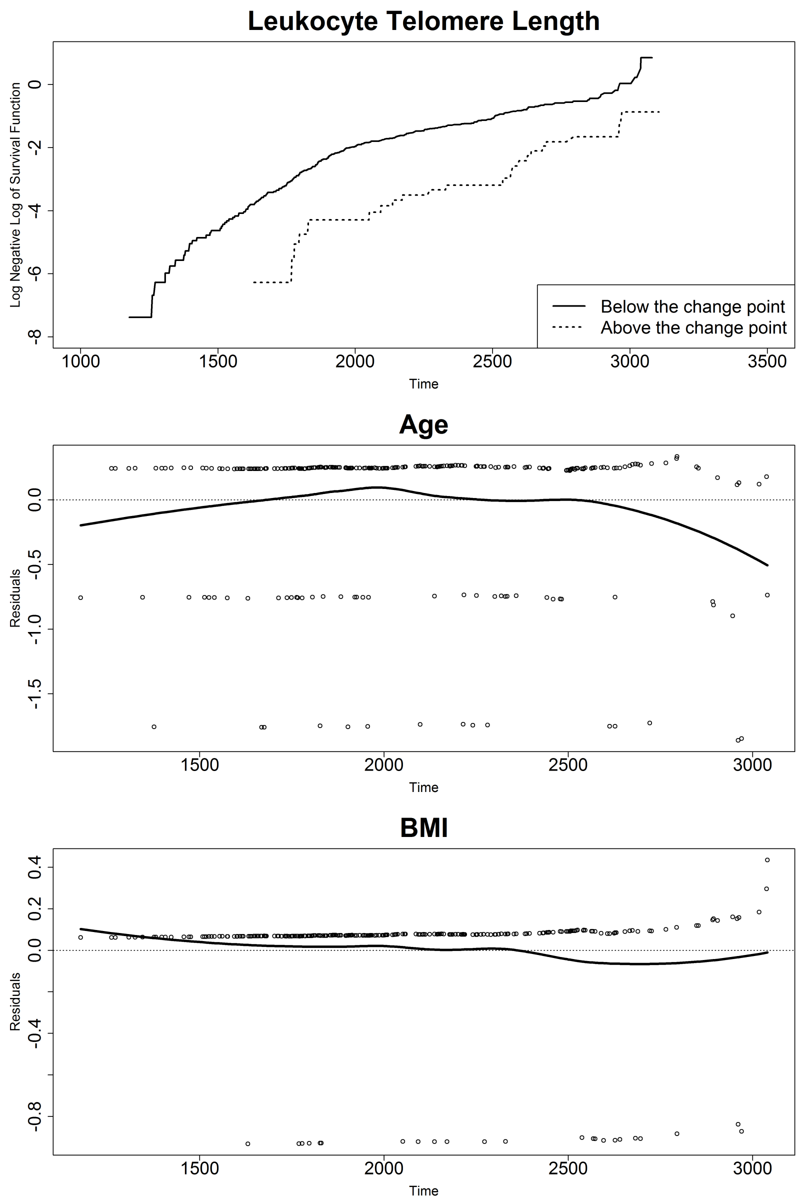

Figure 5.3: Diagnostic Plots. The log of negative log of survival functions versus time are plotted for leukocyte telomere length. Schoenfeld residuals are plotted for age, BMI.

frailty model is more desirable when the individual trajectory is of more interests than the population average effects. However, there is no literature about how to conduct the change point analysis in the frailty survival model for clustered data. The estimation and inference approaches for the Cox-type marginal model is no longer applicable to the frailty model. In future, we can develop a new inference method to conduct the change hyperplane analysis in the frailty model.

5.7 Proof of Lemma and Theorems

We redefine the change hyperplane $I(\eta_1^* X_{ij1} + \eta_2^* X_{ij2} + ... + \eta_{p_1}^* X_{ijp_1} - \eta_0^* > 0)$ by using an equivalent form $I(X_{ij1} + \eta_2 X_{ij2} + ... + \eta_{p_1} X_{ijp_1} - \eta_0 > 0)$, where $\eta_1^* > 0$ is the parameter for the first continuous covariate. These two equivalent forms are linked by $\eta_1^* = \frac{1}{\sqrt{1+\eta_2^2+...+\eta_{p_1}^2}}$, $\eta_j^* = \frac{\eta_j}{\sqrt{1+\eta_2^2+...+\eta_{p_1}^2}}$ for $j = 2, ..., p_1$, and $\eta_0^* = \frac{\eta_0}{\sqrt{1+\eta_2^2+...+\eta_{p_1}^2}}$. Thus, we define $\boldsymbol{\eta} = (1, \eta_2, ..., \eta_{p_1}, \eta_0)^T$ to be a vector of $p_1 + 1$ elements and $\boldsymbol{\eta}^T \boldsymbol{X}_{ij} = X_{ij1} + \eta_2 X_{ij2} + ... + \eta_{p_1} X_{ijp_1} - \eta_0$. The following proofs follow similar arguments in Chapter 4. Here, we only present the sketch of the key steps in the proofs.

We define

$$\boldsymbol{s}^{(r)+}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \mathrm{E}\left[\sum_{j=1}^{K_i} Y_{ij}(t) I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} > 0) \boldsymbol{Z}_{ij}^{\otimes r}(t) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_{ij}(t) + \beta_2 + \boldsymbol{\beta}_3^T \boldsymbol{Z}_{ij}(t)\right\}\right],$$
$$\boldsymbol{s}^{(r)-}(t; \boldsymbol{\eta}, \boldsymbol{\beta}) = \mathrm{E}\left[\sum_{j=1}^{K_i} Y_{ij}(t) I(\boldsymbol{\eta}^T \boldsymbol{X}_{ij} \le 0) \boldsymbol{Z}_{ij}^{\otimes r}(t) \exp\left\{\boldsymbol{\beta}_1^T \boldsymbol{Z}_{ij}(t)\right\}\right],$$

where $r = 0, 1$.

The proof of Theorem 5.3.1 is the same as Theorem 4.3.1 in Chapter 4. The proof of convergence rates follows the same proof as Theorem 4.3.3 in Chapter 4. So omitted.

Proof of Theorem 5.3.2

Consider $G_n(\boldsymbol{\eta}, \boldsymbol{\beta}) = n^{-1}\{l_n(\boldsymbol{\eta}, \boldsymbol{\beta}) - l_n(\boldsymbol{\eta}_0, \boldsymbol{\beta}_0)\}$ and let $G(\boldsymbol{\eta}, \boldsymbol{\beta})$ be

$$\int_0^\tau (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{s}^{(1)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0) \log \frac{s^{(0)}(t; \boldsymbol{\eta}, \boldsymbol{\beta})}{s^{(0)}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0)} d\Lambda_0(t) + g(\boldsymbol{\eta}, \boldsymbol{\beta})$$

with $g(\boldsymbol{\eta}, \boldsymbol{\beta}) = \int_0^\tau \beta_2 \{s^{(0)-}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(0)-}(t; \boldsymbol{\eta}, \boldsymbol{\beta}_0)\} + \boldsymbol{\beta}_3^T \{\boldsymbol{s}^{(1)-}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - \boldsymbol{s}^{(1)-}(t; \boldsymbol{\eta}, \boldsymbol{\beta}_0)\} - \beta_2 \{s^{(0)+}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - s^{(0)+}(t; \boldsymbol{\eta}, \boldsymbol{\beta}_0)\} - \boldsymbol{\beta}_3^T \{\boldsymbol{s}^{(1)+}(t; \boldsymbol{\eta}_0, \boldsymbol{\beta}_0) - \boldsymbol{s}^{(1)+}(t; \boldsymbol{\eta}, \boldsymbol{\beta}_0)\} d\Lambda_0(t)$. We then follow the same argument in the Theorem 3.3.1 in Chapter 3 to obtain that $\sup_{\boldsymbol{\eta},\boldsymbol{\beta}\in\Theta}|G_n(\boldsymbol{\eta},\boldsymbol{\beta})-G(\boldsymbol{\eta},\boldsymbol{\beta})|$ converges in probability to zero. To verify the strict concavity of $G(\boldsymbol{\eta},\boldsymbol{\beta})$ in a neighborhood of $\boldsymbol{\eta}_0$ and $\boldsymbol{\beta}_0$, we first define $\boldsymbol{Y}_{i1}=(\boldsymbol{\eta}^T\boldsymbol{X}_{i1},...,\boldsymbol{\eta}^T\boldsymbol{X}_{iK_i})^T$, $\boldsymbol{Y}_{i2}=(X_{i12},...,X_{iK_i2})^T$,..., $\boldsymbol{Y}_{ip_1}=(X_{i1p_1},...,X_{iK_ip_1})^T$, $\widetilde{\boldsymbol{Y}}_{i1}=(\boldsymbol{\eta}_0^T\boldsymbol{X}_{i1},...,\boldsymbol{\eta}_0^T\boldsymbol{X}_{iK_i})^T$, $\boldsymbol{Z}_i(t)=(\boldsymbol{Z}_{i1},...,\boldsymbol{Z}_{iK_i})$, and $f(\epsilon)\equiv G\{\epsilon\boldsymbol{\eta}+(1-\epsilon)\boldsymbol{\eta}_0,\boldsymbol{\beta}\}$, where $\epsilon\in[0,1]$. Then, $G(\boldsymbol{\eta}_0,\boldsymbol{\beta})-G(\eta,\boldsymbol{\beta})\equiv f(0)-f(1)$. If $\boldsymbol{\eta}$ is close to $\boldsymbol{\eta}_0$, then we obtain

$$\begin{aligned}
\dot{f}(\epsilon)=&\int_0^\tau\int_{\boldsymbol{y}_{i1}<0}\int_{\boldsymbol{y}_{i2}}\cdots\int_{\boldsymbol{y}_{ip_1}}\boldsymbol{y}_{i1}^T\mathrm{E}\left[\exp(\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_i(t))\frac{\{\beta_{2*}+\boldsymbol{\beta}_{3*}^T\boldsymbol{Z}_i(t)\}^2}{2}|\widetilde{\boldsymbol{y}}_{i1}=\boldsymbol{0}\right]\\
&f(\boldsymbol{y}_{i1},\boldsymbol{y}_{i2},...,\boldsymbol{y}_{ip_1})f_{\widetilde{\boldsymbol{y}}_{i1}}(\boldsymbol{0})d\boldsymbol{y}_{i1}d\boldsymbol{y}_{i2}...d\boldsymbol{y}_{ip_1}d\Lambda_0(t)\\
&-\int_0^\tau\int_{\boldsymbol{y}_{i1}>0}\int_{\boldsymbol{y}_{i2}}\cdots\int_{\boldsymbol{y}_{ip_1}}\boldsymbol{y}_{i1}^T\mathrm{E}\Big[\exp^{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_i(t)+\beta_{20}+\boldsymbol{\beta}_{30}\boldsymbol{Z}_i(t)}\\
&\frac{\{\beta_{2*}+\boldsymbol{\beta}_{3*}^T\boldsymbol{Z}_i(t)\}^2}{2}|\widetilde{\boldsymbol{y}}_{i1}=\boldsymbol{0}\Big]f(\boldsymbol{y}_{i1},\boldsymbol{y}_{i2},...,\boldsymbol{y}_{ip_1})f_{\widetilde{\boldsymbol{y}}_{i1}}(\boldsymbol{0})d\boldsymbol{y}_{i1}d\boldsymbol{y}_{i2}...d\boldsymbol{y}_{ip_1}d\Lambda_0(t)\\
&<0,
\end{aligned}$$

where β_{2*} is between 0 and β_{20}, $\boldsymbol{\beta}_{3*}$ is between 0 and $\boldsymbol{\beta}_{30}$, $f_{\widetilde{\boldsymbol{y}}_{i1}}(\boldsymbol{0})$ is the density function of $\widetilde{\boldsymbol{Y}}_{i1}$ at $\boldsymbol{0}$, and $f(\boldsymbol{Y}_{i1},\boldsymbol{Y}_{i2},...,\boldsymbol{Y}_{ip_1})$ is the density function of $(\boldsymbol{Y}_{i1},\boldsymbol{Y}_{i2},...,\boldsymbol{Y}_{ip_1})$. Thus, $G(\boldsymbol{\eta}_0,\boldsymbol{\beta})-G(\boldsymbol{\eta},\boldsymbol{\beta})\geq 0$ in a neighborhood of $\boldsymbol{\eta}$. Furthermore, $G(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)-G(\boldsymbol{\eta}_0,\boldsymbol{\beta})\geq 0$ when $\boldsymbol{\beta}$ is in a neighborhood of $\boldsymbol{\beta}_0$. Then, we conclude that $(\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}})$ should converge to $(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)$ in probability.

Proof of Theorem 5.3.4

Following the proof in the Theorem 4.3.4 of Chapter 4, we define that $\boldsymbol{\eta}_{n,u_1}=\boldsymbol{\eta}_0+n^{-1}\boldsymbol{u}_1$, $\boldsymbol{\beta}_{n,u_2}=\boldsymbol{\beta}_0+n^{-1/2}\boldsymbol{u}_2$, and $W_{ij0}=\boldsymbol{\eta}_0^T\boldsymbol{X}_{ij}$, where $\boldsymbol{u}_1=(a_1,a_2,...,a_{p_1})^T$ and $\boldsymbol{u}_2=(b_1,b_2,...,b_{2p_2+1})^T$. We first obtain

$$l_n(\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_{n,u_2})-l_n(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)=Q_n(\boldsymbol{u}_1)+\boldsymbol{u}_2^T\widetilde{\boldsymbol{l}}_n-\frac{1}{2}\boldsymbol{u}_2^T\boldsymbol{I}(\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)\boldsymbol{u}_2+o_p(1),$$

where

$$
\begin{aligned}
Q_n(\boldsymbol{u}_1) =& \sum_{i=1}^{n}\sum_{j=1}^{K_i}\Delta_{ij}\Big[\left\{\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(T_{ij})\right\} \\
&\left[I\left\{0 \geq W_{ij0} > n^{-1}\boldsymbol{X}_{ij}^T\boldsymbol{u}_1, \boldsymbol{X}_{ij}^T\boldsymbol{u}_1 < 0\right\} - I\left\{0 < W_{ij0} \leq n^{-1}\boldsymbol{X}_{ij}^T\boldsymbol{u}_1, \boldsymbol{X}_{ij}^T\boldsymbol{u}_1 \geq 0\right\}\right] \\
&-\frac{S_n^{(0)}(T_{ij};\boldsymbol{\eta}_{n,u_1},\boldsymbol{\beta}_0) - S_n^{(0)}(T_{ij};\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(T_{ij};\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\Big], \\
\widetilde{\boldsymbol{l}}_n =& n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\int_0^\tau\left\{\widetilde{\boldsymbol{Z}}_{ij}(t;\boldsymbol{\eta}_0) - \frac{\boldsymbol{S}_n^{(1)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}{S_n^{(0)}(t;\boldsymbol{\eta}_0,\boldsymbol{\beta}_0)}\right\}dM_{ij}(t).
\end{aligned}
$$

Obviously, the variable $\widetilde{\boldsymbol{l}}_n$ converges weakly to a Gaussian variable following the normal distribution $N(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1})$. Similarly, we further obtain $Q_n = \sum_{i=1}^n (Q_i^+ - Q_i^-) + o_p(1)$, where

$$
\begin{aligned}
Q_i^-(\boldsymbol{u}_1) =& -\sum_{j=1}^{K_i}\Delta_{ij}\left(\left\{\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(T_{ij})\right\} - \int_0^\tau \phi_{ij}(t)d\Lambda_0(t)\right) \\
& I\left\{\boldsymbol{X}_{ij}^T\boldsymbol{u}_1 \geq 0, 0 < W_{ij0} \leq n^{-1}\boldsymbol{X}_{ij}^T\boldsymbol{u}_1\right\}, \\
Q_i^+(\boldsymbol{u}_1) =& -\sum_{j=1}^{K_i}\Delta_{ij}\left(\left\{\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(T_{ij})\right\} + \int_0^\tau \phi_{ij}(t)d\Lambda_0(t)\right) \\
& I\left\{\boldsymbol{X}_{ij}^T\boldsymbol{u}_1 < 0, 0 \geq W_{ij0} > n^{-1}\boldsymbol{X}_{ij}^T\boldsymbol{u}_1\right\}, \\
\phi_{ij}(t) =& Y_{ij}(t)\exp\left\{\boldsymbol{\beta}_{10}^T\boldsymbol{Z}_{ij}(t)\right\}\left[1-\exp\left\{\beta_{20}+\boldsymbol{\beta}_{30}^T\boldsymbol{Z}_{ij}(t)\right\}\right].
\end{aligned}
$$

We first show that the finite dimensional convergence holds for $\sum_{i=1}^n Q_i^+$ (the same holds for $\sum_{i=1}^n Q_i^-$), and we will identify its limit process based on this finite dimensional convergence. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_S$ be a sequence of vectors, $H_i^{(1)}, ..., H_i^{(M)}$ be the order statistic of all subjects $(\boldsymbol{X}_{i1}^T\boldsymbol{v}_1, ..., \boldsymbol{X}_{i1}^T\boldsymbol{v}_S, ..., \boldsymbol{X}_{iK_i}^T\boldsymbol{v}_1, ..., \boldsymbol{X}_{iK_i}^T\boldsymbol{v}_S)$ in the cluster i, and $q_{(1)}, ..., q_{(M)}$ be the unique ordered sequence of $q_1, ..., q_M$, where $S \leq M \leq K_iS$. We then define set $A_{i,s} = \left\{H_i^{(s-1)} < 0 < H_i^{(s)}\right\}$ for $1 \leq s \leq M$ and let $A_{i,0}$ be the set of $H_i^{(M)} \leq 0$. The purpose is to obtain the limit distribution of any linear combination $\sum_{s=1}^M q_s Q_n^+(\boldsymbol{v}_s)$, where $q_1, ..., q_M$ are any fixed constants.

Let

$$\eta_{ij}^{-} = -\Delta_{ij}\left(\left\{\beta_{20} + \boldsymbol{\beta}_{30}^{T}\boldsymbol{Z}_{ij}(T_{ij})\right\} - \int_{0}^{\tau}\phi_{ij}(t)d\Lambda_{0}(t)\right),$$

and

$$\eta_{ij}^{+} = -\Delta_{ij}\left(\left\{\beta_{20} + \boldsymbol{\beta}_{30}^{T}\boldsymbol{Z}_{ij}(T_{ij})\right\} + \int_{0}^{\tau}\phi_{ij}(t)d\Lambda_{0}(t)\right).$$

Since the cluster sizes are not fixed, then the characteristic function for $Q_n^{+}(\boldsymbol{u}_1)$ is expressed as

$$\phi_n^{+}(t) = \left\{\mathrm{E}\left(\exp\sum_{s\le M} itq_sQ_i^{+}(\boldsymbol{v}_s)\right)\right\}^n = \left\{\sum_{m=1}^{K} P(K_i = m)\mathrm{E}\left(\exp\sum_{s\le M} itq_sQ_i^{+}(\boldsymbol{v}_s)\middle|K_i = m\right)\right\}^n,$$

where $K = \max(K_i)$ is the maximum cluster size. In addition, similar to the argument in Theorem 3 in Chapter 3, if there are more than two subjects in the same cluster satisfying $W_{ij0} > 0$, then the weighted summation of $\mathrm{E}\left\{\exp\left(it\sum_{s=1}^{M} q_sQ_i^{+}(\boldsymbol{v}_s)\middle|K_i = m\right)\right\}$ under those cases equals to $o(1/n)$. Thus, we define the set $S_{m1}^{+} = \{K_i = m, \text{only one } \boldsymbol{\eta}_0^T\boldsymbol{X}_{ij} > 0, \text{all the other } \boldsymbol{\eta}_0^T\boldsymbol{X}_{i1}, ..., \boldsymbol{\eta}_0^T\boldsymbol{X}_{im} \le 0\}$, and S_{m1}^{j+} is the jth element in the set S_{m1}^{+}, $j = 1, ..., m$. We have

$$\begin{aligned}
&\mathrm{E}\left\{\exp\left(it\sum_{s=1}^{M} q_sQ_i^{+}(\boldsymbol{v}_s)\middle|K_i = m\right)\right\}\\
=&1 + \sum_{s=1}^{M} P(A_{i,s}|K_i = m)\sum_{j=1}^{m} P(S_{m1}^{j+}|K_i = m, S_{m1}^{+})\\
&\left(E\left[I(0 < W_{ij0} < H_i^{(s)}/n)\{e^{(q_{(s)}+...+q_{(S)})it\eta_{ij}^{+}} - 1\}\middle|A_{i,s}, K_i = m, S_{m1}^{j+}\right]\right.\\
&\quad +E\left[I(H_i^{(s)}/n \le W_{ij0} < H_i^{(s+1)}/n)\{e^{(q_{(s+1)}+...+q_{(S)})it\eta_{ij}^{+}} - 1\}\middle|A_{i,s}, K_i = m, S_{m1}^{j+}\right] + ...\\
&\quad\quad \left. +E\left[I(H_i^{(S-1)}/n \le W_{ij0} < H_i^{(S)}/n)\{e^{q_{(S)}it\eta_{ij}^{+}} - 1\}\middle|A_{i,s}, K_i = m, S_{m1}^{j+}\right]\right).
\end{aligned}$$

Since

$$
\begin{aligned}
&P(A_{i,s}|K_i = m) \sum_{j=1}^{m} P(S_{m1}^{j+}|K_i = m, S_{m1}^{+}) \\
&\quad E\left[I(H_i^{(s)}/n \le W_{ij0} < H_i^{(s+1)}/n) e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta_{ij}^{+}} \middle| A_{i,s}, K_i = m, S_{m1}^{j+}\right] \\
=&n^{-1} \sum_{j=1}^{m} E\left[(H_i^{(s+1)} - H_i^{(s)}) I(A_{i,s})\{e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta_{ij}^{+}} - 1\} \middle| W_{ij0} = 0+, K_i = m, S_{m1}^{j+}\right] \\
&\quad f_{W_{ij0}}(0+) + O(n^{-2}),
\end{aligned}
$$

we conclude that

$$
\begin{aligned}
&\left[\mathrm{E}\left\{\sum_{m=1}^{K} P(K_i = m) \exp\left(it \sum_{s=1}^{M} q_s Q_i^{+}(\boldsymbol{v}_s) \middle| K_i = m\right)\right\}\right]^n \\
=&\left[\sum_{m=1}^{K} P(K_i = m)\right. \\
&\left\{1 + n^{-1} f_{W_{ij0}}(0+) \sum_{s=1}^{M} \sum_{j=1}^{m} \left(E\left[H_i^{(s)} I(A_{i,s})\{e^{(q_{(s)}+\ldots+q_{(S)})it\eta_{ij}^{+}} - 1\} \middle| W_{ij0} = 0+, K_i = m, S_{m1}^{j+}\right]\right.\right. \\
&+E\left[(H_i^{(s+1)} - H_i^{(s)}) I(A_{i,s})\{e^{(q_{(s+1)}+\ldots+q_{(S)})it\eta_{ij}^{+}} - 1\} \middle| W_{ij0} = 0+, K_i = m, S_{m1}^{j+}\right] + \ldots \\
&\left.\left.\left.+E\left[(H_i^{(S)} - H_i^{(S-1)}) I(A_{i,s})\{e^{q_{(S)}it\eta_{ij}^{+}} - 1\} \middle| W_{ij0} = 0+, K_i = m, S_{m1}^{j+}\right]\right) + O(n^{-2})\right\}\right]^n .
\end{aligned}
$$

So it converges to

$$
\begin{aligned}
&\exp\left\{f_{W_{ij0}}(0+) \sum_{m=1}^{K} P(K_i = m)\right. \\
&\left.\sum_{s=1}^{M} \sum_{k=s}^{M} \sum_{j=1}^{m} \left(E\left[(H_i^{(k)} - H_i^{(k-1)}) I(A_{i,s})\{e^{(q_{(k)}+\ldots+q_{(M)})it\eta_{ij}^{+}} - 1\} \middle| W_{ij0} = 0+, K_i = m, S_{m1}^{j+}\right]\right)\right\}.
\end{aligned}
$$

Recall

$$
Q^{+}(\boldsymbol{u}_1) \equiv \sum_{m=1}^{K} \sum_{q=1}^{m} \sum_{\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{im}} \sum_{0 \le l \le v_{mq}^{+}(d\boldsymbol{x}_{i1},\ldots,d\boldsymbol{x}_{im},\boldsymbol{x}_{iq}^{T}\boldsymbol{u}_1)} V_{mq,l}^{+}(\boldsymbol{x}_{iq}),
$$

where $V_{mq,l}^{+}(\boldsymbol{x}_{iq})$ is a random variable with the conditional distribution of η_{iq}^{+} given $W_{iq0} = 0$, $K_i = m$, and $\boldsymbol{X}_{iq} = \boldsymbol{x}_{iq}$ and it is assumed to be independent for any $\boldsymbol{x}_{iq}$, and $v_{mq}^{+}(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}, t)$

is a multivariate Poisson process with Poisson intensity

$$E[v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},dt)] = f_{\boldsymbol{X}_i}(\boldsymbol{x}_i) f_{W_{iq0}}(0+)p(m)P(S^{q+}_{m1}|K_i = m, S^+_{m1})d\boldsymbol{x}_{i1}d\boldsymbol{x}_{i2}...d\boldsymbol{x}_{im}dt,$$

where $f_{\boldsymbol{X}_i}(\boldsymbol{x}_i)$ is the joint density of $(\boldsymbol{X}_{i1},...,\boldsymbol{X}_{im})$ and $t > 0$. We want to show that the limit distribution of $\sum_{s=1}^M q_s Q_n(\boldsymbol{v}_s)$ is the same as $\sum_{s=1}^M q_s Q(\boldsymbol{v}_s)$. To this end, we note

$$\begin{aligned}
& E\left[\exp\{it\sum_{s=1}^{M} q_s Q(\boldsymbol{v}_s)\}\right] \\
= \; & E\Bigg[\exp\Bigg\{it\sum_{m=1}^{K} I(K_i = m)\sum_{q=1}^{m}\sum_{s=1}^{M} q_s \\
& \sum_{\boldsymbol{x}_{i1},...,\boldsymbol{x}_{im}}\Bigg\{I(\boldsymbol{x}_{iq}^T\boldsymbol{v}_s > 0)\sum_{0\le l\le v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},\boldsymbol{x}_{iq}^T\boldsymbol{v}_s)} V^+_{mq,l}(\boldsymbol{x}_{iq})\Bigg\}\Bigg\}\Bigg] \\
= \; & E\Bigg[\exp\Bigg\{it\sum_{m=1}^{K} I(K_i = m)\sum_{q=1}^{m}\sum_{s=1}^{M}\sum_{\boldsymbol{x}_{i1},...,\boldsymbol{x}_{im}} I(\boldsymbol{x}_{iq}\in A_{i,s}) \\
& \sum_{k=s}^{S} q_{(k)}\sum_{0\le l\le v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},H_i^{(k)})} V^+_{mq,l}(\boldsymbol{x}_{iq})\Bigg\}\Bigg] \\
= \; & \prod_{m=1}^{K}\prod_{q=1}^{m}\prod_{s=1}^{M}\prod_{k=s}^{M} E\Bigg[\exp\Bigg\{itI(K_i = m)\sum_{\boldsymbol{x}_{i1},...,\boldsymbol{x}_{im}} I(\boldsymbol{x}_{iq}\in A_{i,s})(q_{(k)} + ... + q_{(M)}) \\
& \sum_{v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},H_i^{(k-1)}) < l \le v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},H_i^{(k)})} V^+_{mq,l}(\boldsymbol{x}_{iq})\Bigg\}\Bigg] \\
= \; & \prod_{m=1}^{K}\prod_{q=1}^{m}\prod_{s=1}^{M}\prod_{k=s}^{M}\prod_{\boldsymbol{x}\in A_{i,s}} E\Big[\exp\Big\{itI(K_i = m)(q_{(k)} + ... + q_{(M)}) \\
& \sum_{0<l\le v^+(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},H_i^{(k)}) - v^+(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},H_i^{(k-1)})} V^+_{mq,l}(\boldsymbol{x}_{iq})\Big\}\Big],
\end{aligned}$$

where the last equality uses the fact that $v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},[H_i^{(k-1)},H_i^{(k)}))$ is independent for each k for given $\boldsymbol{x}_{iq}$ and $V^+_{mq,l}(\boldsymbol{x}_{iq})$'s are all i.i.d. We note that for a given $\boldsymbol{x}_{iq}$,

$$\sum_{0<l\le v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},H_i^{(k)}) - v^+_{mq}(d\boldsymbol{x}_{i1},...,d\boldsymbol{x}_{im},H_i^{(k-1)})} (q_{(k)} + ... + q_{(S)})V^+_{mq,l}(\boldsymbol{x}_{iq})$$

is a weighted summation of independent compound Poisson processes, where the Poisson rate is $(H_i^{(k)} - H_i^{(k-1)}) f_{\boldsymbol{X}_i}(\boldsymbol{x}_i) f_{W_{iq0}}(0+) p(m) P(S_{m1}^{q+} | K_i = m, S_{m1}^{+}) d\boldsymbol{x}_{i1} d\boldsymbol{x}_{i2} ... d\boldsymbol{x}_{im}$, so its characteristic function is given as

$$\exp\Big\{ f_{W_{iq0}}(0+) f_{\boldsymbol{X}_i}(\boldsymbol{x}_i) p(m) P(S_{m1}^{q+} | K_i = m, S_{m1}^{+}) (H_i^{(k)} - H_i^{(k-1)}) \\ E\Big[e^{it(q_{(k)} + ... + q_{(S)})\eta_{iq}^{+}} - 1 \Big| \boldsymbol{X}_i = \boldsymbol{x}_i, W_{iq0} = 0, K_i = m, S_{m1}^{q+} \Big] d\boldsymbol{x}_{i1} d\boldsymbol{x}_{i2} ... d\boldsymbol{x}_{im} \Big\}.$$

Hence, we obtain

$$\begin{aligned} & \mathrm{E}\Big[\exp\{ it \sum_{s=1}^{S} q_s Q(\boldsymbol{v}_s) \} \Big] \\ = \ & \exp\Big\{ \sum_{m=1}^{K} \sum_{q=1}^{m} \sum_{s=1}^{S} \sum_{k=s}^{S} \mathrm{E}_{\boldsymbol{x}_i} \Big[I(A_{i,s}) f_{W_{iq0}}(0+) (H_i^{(k)} - H_i^{(k-1)}) I(K_i = m) \\ & \times E\Big[e^{it(q_{(k)} + ... + q_{(S)})\eta_{iq}^{+}} - 1 \Big| \boldsymbol{X}_i = \boldsymbol{x}_i, W_{iq0} = 0, K_i = m, S_{m1}^{q+} \Big]\Big]\Big\}, \end{aligned}$$

which is the same as the limit distribution of $\sum_{s=1}^{S} q_s Q_n^{+}(\boldsymbol{v}_s)$. Finally, because of argmax continuous mapping theorem, $n(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = \arg\max_{u_1} Q_n(\boldsymbol{u}_1) + o_p(1)$ and it converges weakly to $\arg\max Q(s)$.

CHAPTER 6: SUMMARY AND FUTURE RESEARCH

In epidemiological studies and clinical trials, it is essential to identify the high-risk individuals for certain diseases based on prognostic biomarkers measured at baseline. The change point models become more and more popular in survival analysis for predicting the survival probability of individuals. In this dissertation, we have investigated the generalized change-point models for both univariate and clustered censored data.

Clustered survival data are commonly seen in large epidemiology studies because of its cost-efficiency in sampling. In Chapter 3, we developed a Cox marginal model with an unknown change point based on a continuous covariate for clustered data. Our focus was on how to make inferences of the unknown change point and derive the asymptotic distributions of the change point estimator. We established the consistency and asymptotic distributions, which is the composite Poisson process, for the proposed change point estimator. Additionally, the consistency and asymptotic normality of the regression coefficient estimators are proved. In simulation studies, the proposed estimators produced virtually unbiased estimates and m out of n bootstrap generated satisfactory coverage rates of the confidence intervals. Additionally, the proposed SUP tests preserved type I errors under various occasions. We implemented the proposed methods in the Strong Heart Family Study (SHFS) and compared the findings with the results in Zhao et al. (2014).

The change point analysis based on a single covariate has been extensively studied for the univariate case. Motivated by the growing popularity in classifying the high-risk individuals based on multiple covariates, we developed a Cox proportional hazards model with a change hyperplane for univariate survival data in Chapter 4. The consistency and asymptotic distribution of an integrated composite Poisson process were established for the proposed change hyperplane estimators. To generate the confidence intervals of each change hyperplane parameters, we generalized the algorithm to select m for multi-dimensional parameters in the m out of n bootstrap. In simulations, the proposed estimators performed well by yielding virtually unbiased estimates and the m out of n bootstraps produced proper coverage rates. The SUP test preserved the type I error in testing the existence of the change hyperplane effect. In addition, we applied our proposed method to the Cardiovascular Health Study. To gain a broader application of the change hyperplane model, we extended the proposed change hyperplane model to clustered event data in Chapter 5. Specifically, we proved the asymptotic properties of the proposed estimators. In simulation studies, the proposed inference approaches performed well for the cases of small and larger clusters. The Strong Heart Family Study was analyzed to verify the existence and evaluate the effect of a change hyperplane based on the LTL and lipid profiles.

The future work includes:

Firstly, we have considered the situation that there is only one "jump" effect in a single continuous covariate or a linear combination of the multiple risk factors. However, in reality, there may exist multiple change point or change hyperplane effects. For example, instead of categorizing the participants into low and high risk groups, we may further define a moderate risk group, which provides a more accurate evaluation of the risk categories for each individual. In this case, the inference procedures and the asymptotic properties proposed for the single change point or change hyperplane cannot be directly extended to the model with multiple change points or change hyperplanes. Muggeo (2003) proposed to apply the piecewise terms to address multiple change points in risk factors regression models. Goodman et al. (2011) proposed multiple change points in time in piecewise constant hazards function. Here, we can consider a spline-type estimation approach (Molinari et al. 2001) for the multiple change point or change hyperplane models in the future.

Secondly, we developed the change point or change hyperplane model based on the Cox proportional hazards model and Cox-type marginal hazards model for univariate and clustered survival data, respectively. However, O'neill (1986) demonstrated that fitting the Cox proportional hazards model could cause substantial bias if the underlying failure time distribution follows an additive hazard model. In this case, the additive risk models are desirable for the situation when the hazard function has a linear relationship with the risk factors. The change point analysis in additive hazards models has been studied for the change point based on an unknown time. Yao (1986) proposed a constrained maximum likelihood estimator to estimate the change point in time, and proved the consistency and asymptotic distribution of the proposed estimator. Chang et al. (1994) proposed an Nelson-Aalen type estimator for the cumulative hazard function and constructed a functional form of the proposed cumulative hazard function to estimate the change point. They further established the consistency, the convergence rates of $1/n$, and the asymptotic distribution of the proposed estimator. Later, Gijbels and Gürler (2003) proposed a least square estimator based on the slope of the cumulative hazard function. In their paper, they compared all three approaches in the simulation studies and found that the least square estimator has less bias and larger variances than the other two approaches. So another feasible extension is to generalize the change point or change hyperplane model to the additive hazards models.

Thirdly, to our best knowledge, there was a lack of methodology for the goodness of fit test to validate the reason to apply the change point model. So far most applications of the change point model have been explained by the underlying biological findings. It is of utmost interest to develop a goodness of fit test to detect the functional form of a covariate or a function of multiple covariates. Such methodology can help the researchers to decide whether to apply the change point analysis or not.

REFERENCES

Abrevaya, J. and Huang, J. (2005), "On the bootstrap of the maximum score estimator," *Econometrica*, 73, 1175–1204.

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

Battiti, R. (1990), "Optimization methods for back-propagation: Automatic parameter tuning and faster convergence," in *International Joint Conference on Neural Networks*, vol. 1, pp. 593–596.

Bickel, P. J., Götze, F., and van Zwet, W. R. (2012), *Resampling fewer than n observations: gains, losses, and remedies for losses*, Springer.

Bickel, P. J. and Sakov, A. (2005), "On the choice of m in the m out of n bootstrap and its application to confidence bounds for extreme percentiles," *Unpublished manuscript.*

— (2008), "On the choice of m in the m out of n bootstrap and confidence bounds for extrema," *Statistica Sinica*, 18, 967–985.

Billingsley, P. (2009), *Convergence of probability measures*, vol. 493, John Wiley & Sons.

Cai, J. and Prentice, R. L. (1995), "Estimating equations for hazard ratio parameters based on correlated failure time data," *Biometrika*, 82, 151–164.

— (1997), "Regression estimation using multivariate failure time data and a common baseline hazard function model," *Lifetime data analysis*, 3, 197–213.

Cai, J. and Shen, Y. (2000), "Permutation tests for comparing marginal survival functions with clustered failure time data." *Statistics in medicine*, 19, 2963–73.

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997), "Generalized partially linear single-index models," *Journal of the American Statistical Association*, 92, 477–489.

Chakraborty, B., Laber, E. B., and Zhao, Y. (2013), "Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme." *Biometrics*, 69, 714–23.

Chang, I.-S., Chen, C.-H., and Hsiung, C. A. (1994), "Estimation in change-point hazard rate models with random censorship," *Lecture Notes-Monograph Series*, 78–92.

Cheung, K., Lee, S. M., and Young, G. A. (2005), "Iterating the m out of n bootstrap in nonregular smooth function models," *Statistica Sinica*, 15, 945.

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

Clegg, L. X., Cai, J., and Sen, P. K. (1999), "A marginal mixed baseline hazards model for multivariate failure time data," *Biometrics*, 805–812.

Cook, R. D. and Li, B. (2002), "Dimension reduction for conditional mean in regression," *Annals of Statistics*, 455–474.

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

— (1975), "Partial likelihood," *Biometrika*, 62, 269–276.

Datta, S. and McCormick, W. P. (1995), "Bootstrap inference for a first-order autoregression with positive innovations," *Journal of the American Statistical Association*, 90, 1289–1300.

Davies, R. B. (1977), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247–254.

— (1987), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74, 33–43.

De Boor, C. (1978), "A practical guide to splines. Number 27 in Applied Mathematical Sciences," .

Dümbgen, L. (1993), "On nondifferentiable functions and the bootstrap," *Probability Theory and Related Fields*, 95, 125–140.

Efron, B. and Tibshirani, R. (1986), "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical science*, 54–75.

Efron, B. and Tibshirani, R. J. (1994), *An introduction to the bootstrap*, CRC press.

Fletcher, R. and Powell, M. J. (1963), "A rapidly convergent descent method for minimization," *Computer journal*, 6, 163–168.

Gandy, A. and Jensen, U. (2005), "On Goodness-of-Fit Tests for Aalen's Additive Risk Model," *Scandinavian Journal of Statistics*, 32, 425–445.

Gandy, A., Jensen, U., and Lütkebohmert, C. (2005), "A Cox model with a change-point applied to an actuarial problem," *Brazilian Journal of Probability and Statistics*, 19, 93–109.

Gastwirth, J. L. (1966), "On robust procedures," *Journal of the American Statistical Association*, 61, 929–948.

— (1985), "The use of maximin efficiency robust tests in combining contingency tables and survival analysis," *Journal of the American Statistical Association*, 80, 380–384.

Gijbels, I. and Gürler, Ü. (2003), "Estimation of a change point in a hazard function based on censored data," *Lifetime Data Analysis*, 9, 395–411.

Gill, P. E., Murray, W., and Wright, M. H. (1981), "Practical optimization," .

Goldberg, D. E. (1989), "Genetic Algorithm in Search, Optimization and Machine Learning, Addison," *W esley Publishing Company, R eading, MA*, 1, 9.

Goodman, M. S., Li, Y., and Tiwari, R. C. (2011), "Detecting multiple change points in piecewise constant hazard functions," *Journal of applied statistics*, 38, 2523–2532.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995), "On blocking rules for the bootstrap with dependent data," *Biometrika*, 82, 561–574.

Hardle, W., Hall, P., Ichimura, H., et al. (1993), "Optimal smoothing in single-index models," *The annals of Statistics*, 21, 157–178.

Härdle, W. and Stoker, T. M. (1989), "Investigating smooth multiple regression by the method of average derivatives," *Journal of the American Statistical Association*, 84, 986–995.

Holland, J. H. (1975), "Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence." .

Hougaard, P. (1986a), "A class of multivanate failure time distributions," *Biometrika*, 73, 671–678.

— (1986b), "Survival models for heterogeneous populations derived from stable distributions," *Biometrika*, 73, 387–396.

Huang, J., Sen, P., and Shao, J. (1996), "Bootstrapping a sample quantile when the density has a jump," *Statistica Sinica*, 6, 299–309.

Huang, J. Z. and Liu, L. (2006), "Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form," *Biometrics*, 62, 793–802.

Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58, 71–120.

Jensen, U. and Lütkebohmert, C. (2008), "A Cox-type regression model with change-points in the covariates," *Lifetime data analysis*, 14, 267–285.

Kantorovich, L. (1948), "On Newtonąŕs method for functional equations," in *Dokl. Akad. Nauk SSSR*, vol. 59, pp. 1237–1240.

Kantorovitch, L. (1939), "The method of successive approximation for functional equations," *Acta Mathematica*, 71, 63–97.

Klein, J. P. (1992), "Semiparametric estimation of random effects using the Cox model based on the EM algorithm," *Biometrics*, 795–806.

Kosorok, M. R. and Song, R. (2007), "Inference under right censoring for transformation models with a change-point based on a covariate threshold," *The Annals of Statistics*, 35, 957–989.

Lee, E., Wei, L., Amato, D., and Leurgans, S. (1992), "Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations," in *Survival Analysis: State of the Art*, eds. Klein, J. and Goel, P., Springer Netherlands, vol. 211 of *Nato Science*, pp. 237–247.

Lee, S. and Klein, J. (1988), "Bivariate models with a random environmental factor," *Industrial Journal of Productivity, Reliability, and Quality Control*, 13, 18.

Lee, S. M. (1999), "On a class of m out of n bootstrap confidence intervals," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 901–911.

Liang, K. Y., Self, S. G., and Liu, X. H. (1990), "The Cox proportional hazards model with change point: an epidemiologic application." *Biometrics*, 46, 783–93.

Luo, X. (1996), "The asymptotic distribution of MLE of treatment lag threshold," *Journal of statistical planning and inference*, 53, 33–61.

Luo, X. and Boyett, J. M. (1997), "Estimations of a threshold parameter in cox regression," *Communications in Statistics - Theory and Methods*, 26, 2329–2346.

Luo, X., Turnbull, B. W., and Clark, L. C. (1997), "Likelihood ratio tests for a changepoint with survival data," *Biometrika*, 84, 555–565.

Marquis, K., Debigaré, R., Lacasse, Y., LeBlanc, P., Jobin, J., Carrier, G., and Maltais, F. (2002), "Midthigh muscle cross-sectional area is a better predictor of mortality than body mass index in patients with chronic obstructive pulmonary disease," *American Journal of Respiratory and Critical Care Medicine*, 166, 809–813.

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

Mebane Jr, W. R., Sekhon, J. S., et al. (2011), "Genetic optimization using derivatives: the rgenoud package for R," *Journal of Statistical Software*, 42, 1–26.

Molinari, N., Daures, J.-P., and Durand, J.-F. (2001), "Regression splines for threshold selection in survival data analysis," *Statistics in medicine*, 20, 237–247.

Muggeo, V. M. (2003), "Estimating regression models with unknown break-points," *Statistics in medicine*, 22, 3055–3071.

Murphy, S. A. (1994), "Consistency in a proportional hazards model incorporating a random effect," *The Annals of Statistics*, 712–731.

— (1995), "Asymptotic theory for the frailty model," *The Annals of Statistics*, 182–198.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. (1992), "A counting process approach to maximum likelihood estimation in frailty models," *Scandinavian journal of statistics*, 25–43.

O'neill, T. J. (1986), "Inconsistency of the misspecified proportional hazards model," *Statistics & probability letters*, 4, 219–222.

Parner, E. et al. (1997), "Inference in semiparametric frailty models," Ph.D. thesis, Institute of Mathematical Sciences, Department of Theoretical Statistics, University of Aarhus.

Politis, D. and Romano, J. (1999), *Subsampling*, Springer, New York.

Pons, O. (2002), "Estimation in a Cox regression model with a change-point at an unknown time," *Statistics: A Journal of Theoretical and Applied Statistics*, 36, 101–124.

— (2003), "Estimation in a Cox regression model with a change-point according to a threshold in a covariate," *The Annals of Statistics*, 31, 442–463.

Sasieni, P. (1992a), "Information bounds for the conditional hazard ratio in a nested family of regression models," *Journal of the Royal Statistical Society. Series B (Methodological)*, 617–635.

— (1992b), "Non-orthogonal projections and their application to calculating the information in a partly linear Cox model," *Scandinavian Journal of Statistics*, 215–233.

Sekhon, J. S. and Mebane, W. R. (1998), "Genetic optimization using derivatives," *Political Analysis*, 7, 187–210.

Sen, B., Banerjee, M., Woodroofe, M., et al. (2010), "Inconsistency of bootstrap: The Grenander estimator," *The Annals of Statistics*, 38, 1953–1977.

Shao, J. (1994), "Bootstrap sample size in nonregular cases," *Proceedings of the American Mathematical Society*, 122, 1251–1262.

Spiekerman, C. F. and Lin, D. (1998), "Marginal regression models for multivariate failure time data," *Journal of the American Statistical Association*, 93, 1164–1175.

Tapp, R., Zimmet, P., Harper, C., de Courten, M., McCarty, D., Balkau, B., Taylor, H., Welborn, T., Shaw, J., Group, A. S., et al. (2006), "Diagnostic thresholds for diabetes: the association of retinopathy and albuminuria with glycaemia," *Diabetes research and clinical practice*, 73, 315–321.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990), "Martingale-based residuals for survival models," *Biometrika*, 77, 147–160.

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

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer Series in Statistics, New York, NY: Springer New York.

Wang, W. (2004), "Proportional hazards regression models with unknown link function and time-dependent covariates," *Statistica Sinica*, 14, 885–906.

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.

Wells, M. (1965), "Algorithm 251: function minimisation [E4]," *Communications of the ACM*, 8, 169–170.

Willeit, P., Raschenberger, J., Heydon, E. E., Tsimikas, S., Haun, M., Mayr, A., Weger, S., Witztum, J. L., Butterworth, A. S., Willeit, J., et al. (2014), "Leucocyte telomere length and risk of type 2 diabetes mellitus: new prospective cohort study and literature-based meta-analysis," *PloS one*, 9, e112483.

Xia, Y., Tong, H., Li, W., and Zhu, L.-X. (2002), "An adaptive estimation of dimension reduction space," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 363–410.

Xu, G., Sen, B., and Ying, Z. (2014), "Bootstrapping a change-point Cox model for survival data," *Electronic Journal of Statistics*, 8, 1345–1379.

Yao, X., Liu, Y., and Lin, G. (1999), "Evolutionary programming made faster," *Evolutionary Computation, IEEE Transactions on*, 3, 82–102.

Yao, Y.-C. (1986), "Maximum likelihood estimation in hazard rate models with a change-point," *Communications in Statistics-Theory and Methods*, 15, 2455–2466.

Yin, X. and Cook, R. D. (2002), "Dimension reduction for the conditional kth moment in regression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 159–175.

Zhao, J., Zhu, Y., Lin, J., Matsuguchi, T., Blackburn, E., Zhang, Y., Cole, S. A., Best, L. G., Lee, E. T., and Howard, B. V. (2014), "Short leukocyte telomere length predicts risk of diabetes in american indians: the strong heart family study." *Diabetes*, 63, 354–62.

Zhu, Y., Voruganti, V. S., Lin, J., Matsuguchi, T., Blackburn, E., Best, L. G., Lee, E. T., MacCluer, J. W., Cole, S. A., and Zhao, J. (2013), "QTL mapping of leukocyte telomere length in American Indians: the Strong Heart Family Study," *Aging (Albany NY)*, 5, 704.

Zucker, D. M., Agami, S., and Spiegelman, D. (2013), "Testing for a Changepoint in the Cox Survival Regression Model," *Journal of Statistical Theory and Practice*, 7, 360–380.